N78-22824

NASA Technical Memorandum 78703

ESTIMATION FROM INCOMPLETE MULTINOMIAL DATA

Karen Rackley Credeur

April 1978

NASA
National Aeronautics and Space Administration
Langley Research Center
Hampton, Virginia 23665

ESTIMATION FROM INCOMPLETE MULTINOMIAL DATA

A thesis presented

by

Karen Rackley Credeur

to

The Department of Statistics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Statistics

Harvard University

Cambridge, Massachusetts

January, 1978

ABSTRACT

We want to estimate the vector of multinomial cell probabilities $\underset{\sim}{p}$ from incomplete data, incomplete in that it contains partially classified observations. Each such partially classified observation is observed to fall in one of two or more selected categories but is not classified further into a single category. The data is assumed to be incomplete at random. The estimation criterion is minimization of risk for quadratic loss. The estimators are the classical maximum likelihood estimate, the Bayesian posterior mode, and the posterior mean. An approximation we develop is used for the posterior mean. The Dirichlet, the conjugate prior for the multinomial distribution, is assumed for the prior distribution.

We show these three estimators to be approximately equal in large samples. We then study risk in small- and medium-size samples through Monte-Carlo simulation studies for the trinomial distribution. Samples are of size 25 and 50, percentage of incomplete data varies around 15 and 40, and probabilities range from the center of the probability simplex P_2 to one of its corners. Probabilities equal the means of the prior distributions for varying prior parameters or are randomly generated from these distributions. Priors used in the Bayesian estimators are the correct prior, a uniform prior, and a perturbed prior. The EM iterative algorithm of Dempster, Laird, and Rubin (1977) is used to evaluate all three estimators.

Results indicated that the relationship between the probability $\underset{\sim}{p}$ being estimated and the prior parameters $\underset{\sim}{\beta}$ used in the Bayesian estimators was one of the most important factors in determining which estima-

tor was preferable. If the mean $\bar{\underset{\sim}{p}}$ of the Dirichlet distribution given the prior parameters $\underset{\sim}{\beta}$ was within a fairly wide range of $\underset{\sim}{p}$, then the posterior mean was the best estimator of $\underset{\sim}{p}$. If the mean was far from $\underset{\sim}{p}$, then the maximum likelihood estimate was best. Between these extremes was a region in which the posterior mode was often best when $\underset{\sim}{p}$ was toward a corner of P_2. The maximum likelihood estimate and posterior mode were equally best at a corner. When the best estimator was used, risk was usually reduced by one-fourth to one-third over that of the next best estimator and by one-third to one-half over that of the worst estimator. However, the reduction in risk was sometimes substantial. The largest reduction occurred at the corner $\underset{\sim}{p}=(0,0,1)$; the risk of the posterior mean was as much as 33,000 times larger than the risk of the posterior mode or maximum likelihood estimate.

As the percentage of incomplete data increased, the risk of the three estimators did not greatly increase and the relationship among the estimators changed little. As sample size increased, risk and the difference in risk between estimators usually decreased.

Because numerical evaluation of the exact posterior central moments is generally unfeasible, we also develop approximations for elements of the posterior mean and covariance matrices. The best of three approximations considered for the posterior mean is based on a first-order Taylor-series expansion of the exact posterior mean that has accuracy of order $O(n^{-1})$. Because terms in the expansion are then approximated, the final approximation, called the Taylor-series approximate posterior mean, is not necessarily accurate to order $O(n^{-1})$. However, we show that this

approximation asymptotically equals the exact posterior mean. Further, we give two conditions which guarantee that the error between the exact posterior mean and an iterative solution of the Taylor-series approximate posterior mean is of magnitude $O(n^{-1})$.

Approximations used for elements of the posterior covariance matrix are based on Taylor-series expansions accurate to order $O(n^{-3/2})$. When the iterative solution for the Taylor-series approximate posterior mean has accuracy of magnitude $O(n^{-1})$, then the Taylor-series approximate posterior variance and covariance can be evaluated noniteratively to have accuracy of magnitude $O(n^{-3/2})$. These approximations can also be evaluated iteratively. However, insurance of accuracy of magnitude $O(n^{-3/2})$ then depends on satisfaction of the two conditions discussed for iterative solution of the Taylor-series approximate posterior mean.

An important property of the Taylor-series approximations is that, as the percentage of incomplete data goes to zero, they go to the exact posterior moments. In addition, the relationship between the Taylor-series approximate posterior mean and the posterior mode parallels their complete-data relationship.

In the same Monte-Carlo simulation study used for the risk study, the Taylor-series approximation for the posterior mean was usually accurate to at least four significant figures; that for the posterior variance, to at least three significant figures; and that for the posterior covariance, to at least two significant figures. In practice, the Taylor-series approximations will generally be more accurate than numerical evaluation of the corresponding exact posterior moments.

ACKNOWLEDGEMENTS

I am indebted to Dr. Nan Laird for her patience, encouragement, and interest in my work during the past 16 months. Her suggestions greatly improved the composition and clarity of this thesis. I am also grateful to Professor Frederick Mosteller for his thorough review of an earlier draft of this work and his excellent criticism. I also thank Professor Arthur P. Dempster for suggesting the approximation part of this thesis and for many helpful discussions concerning the material in Chapter 3.

I am grateful to the Graduate Study Program of the National Aeronautics and Space Administration (NASA), Langley Research Center, and to my current NASA supervisors, especially Dr. James W. Ramsey, Jr., Head, Structural Engineering Section, Research Facilities Engineering Division, and Dr. Ross L. Goble, Chief, Research Facilities Engineering Division, and my former supervisors, especially Dr. Belinda Adams, Dr. James Ortega, Dr. John Shoosmith, and Robert E. Smith, for their encouragement and support of much of this work. For useful discussion at the beginning of this work, I also thank the following: Dr. David Hoaglin (random-number generators), Dr. George Canavos (analysis of variance and simulation), Dr. Carl Andersen (MACSYMA), Dr. Janet Campbell (simulation), Jim Schiess (analysis of variance), and Dr. Steve Park (computer programing of exact posterior central moments). Thanks also go to Mrs. Anne Leatherman who reformated data from my computer programs for use in canned analysis-of-variance computer programs.

Finally, I am deeply grateful to my husband Leonard whose constant encouragement, support, and personal sacrifice were invaluable in my finishing this thesis.

TABLE OF CONTENTS

LISTING OF TABLES

LISTING OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Overview:

This thesis is concerned with simultaneous estimation of the vector $\underset{\sim}{p}$ of cell probabilities from incomplete multinomial data where the criterion of goodness is minimization of risk for quadratic loss. As is well known, the posterior mean will minimize expected risk. However, complete-data results indicate that for at least boundary probabilities, the maximum likelihood estimate might be a better estimator. Hence, we study both estimators for specified values of $\underset{\sim}{p}$. In addition, we investigate a third estimator, the posterior mode, which has some advantages of each of the other two estimators.

Because numerical evaluation is generally unfeasible, we also develop approximations for the posterior mean and covariance matrices. Therefore, part of this thesis concerns derivation of the approximations and proof of their accuracy.

In the next section, we define the risk problem and detail reasons for choosing the posterior mean, maximum likelihood estimator, and posterior mode. We begin by defining special notation for the incomplete-data problem. We also outline a robustness study concerning use of the correct prior in the Bayesian estimators. In the third section, we review the literature of estimation from incomplete multinomial data.

Chapter 2 describes the estimators. First we derive the exact posterior mean and central moments and illustrate the problems in their numerical computation. Then we give derivations for the mode estimators, the maximum likelihood estimate and posterior mode. In Chapter 3, we develop truncated

Taylor-series approximations for the exact posterior mean and covariance matrices. In Chapter 4 we prove the asymptotic, large-sample, accuracy of these approximations. For these large samples, the posterior mean, maximum likelihood estimate, and posterior mode are all approximately equal; hence, there will be little difference in their risks.

We then turn to small-sample behavior of the estimators. For small- and medium-size samples, we investigate (1) the accuracy of the Taylor-series approximations for the posterior mean and covariance matrices, (2) which of the Taylor-series approximation, maximum likelihood estimate, and posterior mode best approximates the posterior mean, (3) which estimator best minimizes risk for quadratic loss at specified values of $\underset{\sim}{p}$, and (4) how robust results in (3) are to use of the correct prior in the Bayesian estimators. Because we could not answer these questions analytically, we performed Monte-Carlo simulation studies for the trinomial distribution. In Chapter 5 we discuss the design and relevant computational procedures for two such studies. Chapters 6 and 7 give results of these two studies and guidelines for practical implementation of the results.

In Chapter 8 we summarize the main research of the thesis, draw conclusions, and recommend areas for future study.

1.2 Problem Statement:

Assume that we have a k-dimensional Dirichlet prior

$$g(\underset{\sim}{p}|\underset{\sim}{\nu}) = [\Gamma(\sum_{i=1}^{k+1} \nu_i)/\prod_{i=1}^{k+1} \Gamma(\nu_i)] \prod_{i=1}^{k+1} p_i^{\nu_i-1}, \tag{1.1}$$

where $\nu_i>0$ and $\underset{\sim}{p}$ takes values in the k-dimensional probability simplex $P_k=\{(p_1,\ldots,p_{k+1}):p_i\geq 0, \sum_{i=1}^{k+1} p_i=1\}$. The Dirichlet density is the conjugate prior for the multinomial distribution. Assume also that we have complete data $\underset{\sim}{x}=(x_1,\ldots,x_{k+1})$, $n=\sum_{i=1}^{k+1} x_i$, denoting nonnegative integer sample values of the random vector $\underset{\sim}{X}=(X_1,\ldots,X_{k+1})$ having the k-dimensional multinomial distribution $M(n;\underset{\sim}{p})$ with density

$$h(\underset{\sim}{x}|\underset{\sim}{p}) = [n!/\prod_{i=1}^{k+1} x_i!] \prod_{i=1}^{k+1} p_i^{x_i}. \tag{1.2}$$

Thus, the k+1 components of $\underset{\sim}{x}$ respectively denote the number of the n observations that fall in k+1 mutually exclusive categories $C_1,\ldots,C_{k+1}$.

Suppose, however, that n observations are made on k+1 mutually exclusive categories but that some of these observations are only partially observed in that each of these observations falls in one of two or more of the k+1 categories but cannot further be classified into a single category. That is, for some of the n observations one knows only that the observation falls in one of l particular categories for $1\leq l\leq k+1$ but not which one of these l categories. This set of categories among which an observation is shared is called a pattern of incomplete data.

We denote such a set of categories as C suffixed by the indices of the sharing categories. For example, if an observation is known to fall in one

of categories C_i, C_j, or C_l, for $1 \leq i,j,l \leq k+1$, but cannot be specified further, we write that the observation falls in C_{ijl}. More commonly, we write the total of all such observations falling in C_{ijl} as z_{ijl} or $z_{\{i,j,l\}}$ where, following a few more comments, we elaborate on these two z subscript notations. Corresponding to the use of $\underset{\sim}{x}=(x_1,x_2,\ldots,x_{k+1})$, we write $\underset{\sim}{z}=(z_1,z_2,\ldots,z_{12},z_{13},\ldots,z_{12\ldots k})$ [or $(z_{\{1\}},z_{\{2\}},\ldots,z_{\{1,2\}},z_{\{1,3\}},\ldots,z_{\{1,2,\ldots,k\}})$] to denote the vector of incomplete data. Thus, $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12},z_{13},z_{23})$ represents the vector of incomplete trinomial data having, for example, z_2 completely specified observations falling in C_2 and z_{13} incompletely specified observations such that each observation is known to fall in one of C_1 or C_3 (C_{13}) but is not specified further.

However, we need some way to abbreviate notation for summing and multiplying over all collections containing a particular integer in forthcoming equations. The least cumbersome approach is to adopt set notation and then, for convenience and to parallel complete-data notation (i.e., complete-data notation is x_1, not $x_{\{1\}}$), drop braces and commas where possible. Therefore, in the next few paragraphs, we formally define the set notation used.

We first note that we want the notation to allow for dividing the data into separate multinomial groups in the Hocking and Oxspring manner to be described in the next section. Although we observe data in the general, unrestricted, form $z_1, z_2,\ldots,z_{12},\ldots,z_{12\ldots k}$, where the completely specified data $z_1, z_2,\ldots,z_{k+1}$ need not be subdivided, we use the Hocking and Oxspring restrictive form in writing the likelihood for the exact posterior central moments in Chapter 2 and for some of the asymptotic proofs in Chapter 4. Thus, for each incomplete-data pattern, we create notation to allow for enough artificial completely specified observations to complete a multinomial group.

For example, if we observe z_1, z_2, z_3, z_{12}, and z_{13}, we can treat the data in the Hocking and Oxspring manner as coming from three independent distributions, one trinomial and two binomials as follows: $v_1=z_1$, v_2, v_3; $y_{12}=z_{12}$,y_3; and $w_{13}=z_{13}$, w_2; where $v_2+w_2=z_2$ and $v_3+y_3=z_3$. Here, v_1, v_2, and v_3 have a trinomial distribution with probabilities p_1, p_2, and p_3; y_{12} and y_3 have a binomial distribution with probabilities (p_1+p_2) and p_3; and w_{13} and w_2 have a binomial distribution with probabilities (p_1+p_3) and p_2.

Therefore, for k the dimension of a multinomial distribution, let $\mathcal{S}$ be a nonempty subset of $\{1,2,\ldots,k+1\}$ and let P be the set of mutually exclusive and exhaustive subsets $\mathcal{S}$. For example, for the trinomial distribution we could have the following P and $\mathcal{S}$:

$P_1 = \{\{1\},\{2\},\{3\}\}$ containing $\mathcal{S}_{1,1}=\{1\}$, $\mathcal{S}_{2,1}=\{2\}$, and $\mathcal{S}_{3,1}=\{3\}$;

$P_2 = \{\{1,2\},\{3\}\}$ containing $\mathcal{S}_{1,2}=\{1,2\}$ and $\mathcal{S}_{2,2}=\{3\}$;

$P_3 = \{\{1,3\}\ \{2\}\}$ containing $\mathcal{S}_{1,3}=\{1,3\}$ and $\mathcal{S}_{2,3}=\{2\}$; and

$P_4 = \{\{1\},\{2,3\}\}$ containing $\mathcal{S}_{1,4}=\{1\}$ and $\mathcal{S}_{2,4}=\{2,3\}$.

Define $\mathcal{S},P$ to be the set element $\mathcal{S}$ in the set P. Suppose that there are $\beta_{\mathcal{S},P}$ elements in $\mathcal{S},P$. Let $z_{\mathcal{S},P}$ be the number of observations such that each observation falls in one of the $\beta_{\mathcal{S},P}$ categories C_i for $i\in\mathcal{S}$, but is not further classified into a particular one of these $\beta_{\mathcal{S},P}$ categories if $\beta_{\mathcal{S},P}>1$. Incomplete multinomial data is data of the form $z_{\mathcal{S},P}$ for $\mathcal{S}$ containing more than one element; i.e., $\beta_{\mathcal{S},P}>1$.

Thus, for the example given in the third preceding paragraph, we have that $\underset{\sim}{z}_{\mathcal{S},P_1}=(z_{\{1\},\{\{1\},\{2\},\{3\}\}},z_{\{2\},\{\{1\},\{2\},\{3\}\}},z_{\{3\},\{\{1\},\{2\},\{3\}\}})$ $=(v_1,v_2,v_3)$, $\underset{\sim}{z}_{\mathcal{S},P_2}=(z_{\{1,2\},\{\{1,2\},\{3\}\}},z_{\{3\},\{\{1,2\},\{3\}\}}) = (y_{12},y_3)$, and

$\underset{\sim}{z}_{\mathcal{S}},P_3=(z_{\{1,3\},\{\{1,3\},\{2\}\}},z_{\{2\},\{\{1,3\},\{2\}\}}) = (w_{13},w_2)$.

We note that while we are deriving the posterior distribution of $\underset{\sim}{p}$ in Chapter 2 or calculating its limit in Chapter 4, we will use the P subscript. For all other purposes, however, we discard the P subscript and work with only the sufficient statistics of the Hocking and Oxspring observed data, defined by

$$z_{\mathcal{S}} = \sum_{P\supseteq\mathcal{S}} z_{\mathcal{S},P}. \tag{1.3}$$

Thus, in our trinomial example the sufficient statistics are

$z_{\{1\}}=z_{\{1\},\{\{1\},\{2\},\{3\}\}}$, $z_{\{2\}}=z_{\{2\},\{\{1\},\{2\},\{3\}\}}+z_{\{2\},\{\{1,3\},\{2\}\}}$, $z_{\{3\}}=z_{\{3\},\{\{1\},\{2\},\{3\}\}}+z_{\{3\},\{\{1,2\},\{3\}\}}$, $z_{\{1,2\}}=z_{\{1,2\},\{\{1,2\},\{3\}\}}$, and $z_{\{1,3\}}=z_{\{1,3\},\{\{1,3\},\{2\}\}}$

We let $\underset{\sim}{z}$ denote the vector of all $z_{\mathcal{S}}$. Therefore, as in our earlier discussion, $\underset{\sim}{z}$ is our vector of observed data. Similarly, $n=\sum_{\mathcal{S}} z_{\mathcal{S}}$ denotes the sum of all the observed data. Finally, we define $p_{\mathcal{S}}$ as the sum of probabilities p_i for i in $\mathcal{S}$. Thus, $p_{\{3\}}=p_3$ and $p_{\{3,5,6\}}=p_3+p_5+p_6$.

In summary, we use set notation because it is the least cumbersome mechanism for writing sums and products over all sets (or collections) containing a particular integer. The use of set notation also aids derivations of exact posterior central moments in Chapter 2 and calculation of limits in Chapter 4. On the other hand, where possible we delete the braces and commas to simplify equations and to parallel complete-data notation (i.e., complete-data notation is x_i, not $x_{\{i\}}$). For example, we usually write p_{12} instead of $p_{\{1,2\}}$. We also mix the simplified and full notations. For example, we usually write

$z_i + \sum_{D \ni i} z_D p_i / p_D$ rather than $z_{\{i\}} + \sum_{D \ni i} z_D p_{\{i\}} / p_D$ where $\sum_{D \ni i}$ means the sum over all those multiple-integer sets that contain i. Thus, in the trinomial example, $\sum_{D \ni 1}$ means the sum over all sets {1,2} and {1,3} that contain the integer 1. Note that we define D as a set containing more than one integer unless otherwise specified. That is, D can not denote the set {i} for any i.

Finally, we assume that the incompleteness of the data is random. That is [see Rubin (1976)], incomplete data is not a function of the values that would have been observed.

In this thesis we are interested in minimizing risk. Risk is defined as expected loss with respect to, in this work, the distribution of $\underset{\sim}{z}$ given $\underset{\sim}{p}$; that is, for some estimator $\dot{\underset{\sim}{p}}$ of $\underset{\sim}{p}$,

$$r(\underset{\sim}{p},\dot{\underset{\sim}{p}}) = E[L(\underset{\sim}{p},\dot{\underset{\sim}{p}})] = \sum_{Z_k} L(\underset{\sim}{p},\dot{\underset{\sim}{p}})\, h(\underset{\sim}{z}|\underset{\sim}{p}), \qquad (1.4)$$

where $r(\underset{\sim}{p},\dot{\underset{\sim}{p}})$ is the risk of $\dot{\underset{\sim}{p}}$, $L(\underset{\sim}{p},\dot{\underset{\sim}{p}})$ is the loss function for $\underset{\sim}{p}$, $Z_k=\{(z_1,\ldots,z_{k+1},z_{12},z_{13},\ldots,z_{12\ldots k})$: each $\underset{\sim}{z}$ component is a nonnegative integer and the $\underset{\sim}{z}$ components sum to n}, and $h(\underset{\sim}{z}|\underset{\sim}{p})$ is the density of $\underset{\sim}{z}$ given $\underset{\sim}{p}$.

In (1.4), the risk function depends on the value of the generally unknown probability $\underset{\sim}{p}$. As Zellner (1971,p25) points out, it is impossible to find an estimator $\dot{\underset{\sim}{p}}$ that minimizes risk $r(\underset{\sim}{p},\dot{\underset{\sim}{p}})$ for all possible values of $\underset{\sim}{p}$. He gives as an example that the vector $\dot{\underset{\sim}{p}}=\underset{\sim}{b}$ of constants will have minimum risk when $\underset{\sim}{p}=\underset{\sim}{b}$; hence, as $\underset{\sim}{p}$ varies over P_k, the minimizing estimator varies.

Therefore, a common practice is to choose as an estimator that one that minimizes the average risk $E[r(\underset{\sim}{p},\dot{\underset{\sim}{p}})]$, where

$$E[r(\underset{\sim}{p},\dot{\underset{\sim}{p}})] = \int_{P_k} r(\underset{\sim}{p},\dot{\underset{\sim}{p}})\, g(\underset{\sim}{p})\, d\underset{\sim}{p}$$

$$= \int_{P_k} \left[\sum_{Z_k} L(\underset{\sim}{p},\dot{\underset{\sim}{p}})\, h(\underset{\sim}{z}|\underset{\sim}{p})\right]\, g(\underset{\sim}{p})\, d\underset{\sim}{p} \qquad (1.5)$$

$$= \sum_{Z_k} \left[\int_{P_k} L(\underset{\sim}{p},\dot{\underset{\sim}{p}})\, f(\underset{\sim}{p}|\underset{\sim}{z})\, d\underset{\sim}{p}\right] q(\underset{\sim}{z})$$

for $g(\underset{\sim}{p})$ the prior density of $\underset{\sim}{p}$, $f(\underset{\sim}{p}|\underset{\sim}{z})$ the posterior density of $\underset{\sim}{p}$ given $\underset{\sim}{z}$, and $q(\underset{\sim}{z})$ the marginal density of $\underset{\sim}{z}$.

Now, the estimator minimizing the term in brackets in the last line of (1.5) also minimizes expected risk. For quadratic loss

$$L(\underset{\sim}{p},\dot{\underset{\sim}{p}}) = (\underset{\sim}{p}-\dot{\underset{\sim}{p}})'(\underset{\sim}{p}-\dot{\underset{\sim}{p}})$$

$$= \sum_{i=1}^{k+1} (p_i-\dot{p}_i)^2, \qquad (1.6)$$

this Bayes estimator is the posterior mean. We use quadratic loss (also called mean squared error) for the loss function because of its mathematical tractability, frequent past usage, accuracy in approximating other loss functions [see Mood and Graybill (1963,p165) and DeGroot (1970,p227)], and physical interpretation. The emphasis in quadratic loss is on minimization of the overall scatter of the estimates from the true value rather than concentration on a few extreme departures. In particular, the quadratic-loss criterion allows bias in an estimator if the variance is compensatingly small.

As noted just before (1.5), however, the posterior mean will not minimize risk in (1.4) for all values of $\underset{\sim}{p}$. Hence, there might be ranges

of $\underset{\sim}{p}$ for which other commonly used, and easily calculated, estimators improve on the posterior mean. Further, as Zellner (1971,p26) notes, many sampling theorists object to use of the prior density $g(\underset{\sim}{p})$ (because it is never known in practice). Thus, they do not consider the minimal average risk property of the posterior mean to be important.

Therefore, besides the posterior mean $\tilde{\underset{\sim}{p}}$, we also investigate two other estimators to minimize risk for at least some values of $\underset{\sim}{p}$. The first estimator is the maximum likelihood estimate $\hat{\underset{\sim}{p}}$. We include it because it is a classical estimator that is often used. In particular, it is frequently used when one has no prior knowledge. For complete data, the maximum likelihood estimator $\ddot{\underset{\sim}{p}}=\underset{\sim}{x}/n$ is the unique, minimum variance unbiased estimate of $\underset{\sim}{p}$. Hence, any estimator having smaller risk than $\ddot{\underset{\sim}{p}}$ must be biased. However, Johnson (1971) has shown that $\ddot{\underset{\sim}{p}}$ is admissible. That is, there does not exist any other estimator $\dot{\underset{\sim}{p}}$ having at least as small a risk for all values of $\underset{\sim}{p}$ and strictly smaller risk for at least one value of $\underset{\sim}{p}$.

The maximum likelihood estimate $\ddot{\underset{\sim}{p}}$ is admissible because no other estimators have smaller risk when all but one of the $\underset{\sim}{p}$ components are near zero. Since the risk of $\ddot{\underset{\sim}{p}}$ equals $1-\sum_{i=1}^{k+1} p_i^2$, the risk is close to zero when $\underset{\sim}{p}$ is near a corner of the P_k simplex. Hence, if the incomplete-data case parallels the complete-data case, we would expect the maximum likelihood estimate $\hat{\underset{\sim}{p}}$ to have smallest risk when all but one of the $\underset{\sim}{p}$ components are near zero and the posterior mean to have smallest risk furthest from the boundary; i.e., at the center of P_k.

We also include the posterior mode $\hat{\tilde{\underset{\sim}{p}}}$. It is an in-between estimator in that, like the maximum likelihood estimate, it is a mode and, like the

posterior mean, it is a Bayesian estimate and utilizes prior knowledge. Unlike the posterior mean, however, the posterior mode can have zero components for a nonzero prior. Hence, it is a strong competitor for the maximum likelihood estimate for extreme values of $\underset{\sim}{p}$, those values near a boundary of the P_k simplex.

Finally, we note that the posterior mean minimizing expected risk depends on knowledge of the prior $g(\underset{\sim}{p})$. In practice, we would not know the true prior $g(\underset{\sim}{p})$. At best we would have some estimate of $g(\underset{\sim}{p})$ that has, in general, undeterminable error. To investigate how robust our results are to use of the correct prior, we compare the three estimators by using two wrong priors, as well as the correct prior, in their calculations in the small-sample trinomial simulations. Note that the maximum likelihood estimate, not being a Bayesian estimate, is the same for all three studies.

For the first wrong prior, we choose the uniform prior with vector of parameters (1,1,1) because of its common use when one is uncertain of prior knowledge. The uniform prior gives equal weight to all components of $\underset{\sim}{p}$. For this prior, the posterior mode equals the maximum likelihood estimate. For the second wrong prior, we choose the vector of parameters $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$, where $\underset{\sim}{\nu}$ is the correct prior. This prior perturbs the three components of $\underset{\sim}{p}$ by .09, .05, and -.14, respectively. Hence, we call it the perturbed prior.

1.3 Literature Review:

To date, most of the published work on estimation from incomplete multinomial data has concerned maximum likelihood estimation. In 1958 Hartley presented an iterative method for calculating maximum likelihood estimates from those sets of discrete data for which a maximum-likelihood procedure is available for the corresponding complete-data sample. Because his method was later generalized and clarified by Dempster, Laird, and Rubin (1977) in a paper described at the end of this section, we do not further discuss Hartley's method now. Hartley gave examples for the Poisson, negative binomial, and binomial distributions. Hartley also proposed calculating the large-sample covariance matrix of the maximum likelihood estimates by using the calculus of finite differences. He used the iterates from the maximum-likelihood-estimate algorithm to estimate the second derivative of the log likelihood function via the standard finite difference formula.

Blumenthal (1968) considered maximum-likelihood estimation from incomplete multinomial data for the special case in which a category does not share data with more than one group of categories. That is, for the k-dimensional multinomial population, if category C_i shares data with category C_j for j in some subset $\emptyset$ of the k+1 indices of $\underset{\sim}{p}$, then C_i does not share data with any category C_h for which h is not an element of $\emptyset$. For the binomial case, Blumenthal also investigated the problem of non-random missingness.

Hocking and Oxspring (1971) considered the case in which data comes from populations all related to the same "parent" population. In a related

population, at least one parameter is the sum of two or more probabilities from the parent population. Those parameters for the related population that are not such sums, exhaust those probabilities of the parent population that are not elements of these sums. Hocking and Oxspring derived the maximum likelihood estimates and their large-sample covariance matrix in the usual manner (e.g., the large-sample inverse covariance matrix is the Fisher Information for $\underset{\sim}{p}$). They developed an iterative algorithm for solution of the resulting nonlinear equations.

A simple case of the Hocking and Oxspring situation is that of a parent population having probabilities p_1, p_2, and p_3 and a related population having probabilities p_1+p_2 and p_3. In general, however, we do not have sample information given twice on category C_3. That is, we have sample data given for p_1, p_2, p_3, and p_1+p_2 and do not have data on C_3 broken into two groups to help estimation.

Sundberg (1974) developed maximum-likelihood theory for the general problem of incomplete data from an exponential family, of which the multinomial distribution is a member. He proved that the derivatives of the log likelihood with respect to the natural (exponential) parameters can be written as the difference of an unconditional and conditional expectation of the complete-data sufficient statistics. He noted that this form for the first and second partial derivatives was first discovered in unpublished work by Martin-Löf. [However, Efron (1977) noted that this form was implicit in Fisher's 1925 paper.]

Dempster, Laird, and Rubin (1977) extended Sundberg's work to the general case where the problem need not involve an exponential family. They called their algorithm the EM algorithm because it consists of an

expectation step followed by a maximization step. Although this is the same algorithm proposed by Hartley (1958), Dempster, Laird, and Rubin generalized the algorithm, clarified the techniques, improved the mathematics, and extended the history and usage of the algorithm. They proved that the EM algorithm converges to a local maximum or a saddle point when the likelihood is bounded and the matrix of second partial derivatives of the complete-data likelihood is negative definite with nonzero bounded eigenvalues. They also gave a formula for the rate of convergence close to a stationary point. Finally, they showed how the EM algorithm can be used to calculate a posterior mode.

We describe the EM algorithm in the next chapter where we use it to calculate the mode estimators, the maximum likelihood estimate and the posterior mode. We also use the EM algorithm for solution of the approximation we develop in Chapter 3 for the exact posterior mean.

CHAPTER 2

THE ESTIMATORS

2.1 Introduction:

In this chapter we give formulas for the estimators. In the next section we derive the posterior central moments. We begin with known formulas for the complete-data case and then, utilizing notation defined at the beginning of Section 1.2, derive elements of the posterior mean and covariance matrices for the incomplete-data case. We then illustrate these derivations with an example and discuss difficulties in the numerical computation of these exact moments.

In the last section, we give derivations for the mode estimators based on theory from Sundberg (1974). We then show how values of these estimators are calculated with the EM algorithm of Dempster, Laird, and Rubin (1977). The first part of the section discusses the maximum likelihood estimate. The second part details results for the posterior mode.

2.2 Posterior Central Moments:

2.2.1 Complete Data:

For the k-dimensional Dirichlet prior $g(\underset{\sim}{p})$ in (1.1) and complete-data sample $\underset{\sim}{x}=(x_1,\ldots,x_{k+1})$ from the multinomial distribution with density (1.2), the posterior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{x}$ has the k-dimensional density

$$f(\underset{\sim}{p}|\underset{\sim}{x}) = \frac{[\Gamma(\sum_{i=1}^{k+1}\nu_i)/\prod_{i=1}^{k+1}\Gamma(\nu_i)]\prod_{i=1}^{k+1}p_i^{\nu_i-1}\;[n!/\prod_{i=1}^{k+1}x_i!]\prod_{i=1}^{k+1}p_i^{x_i}}{\int\cdots\int[\Gamma(\Sigma\nu_i)/\Pi\Gamma(\nu_i)]\Pi p_i^{\nu_i-1}\;[n!/\Pi x_i!]\Pi p_i^{x_i}d\underset{\sim}{p}} \tag{2.1}$$

$$= [\Gamma(n+\sum_{i=1}^{k+1}\nu_i)/\prod_{i=1}^{k+1}\Gamma(x_i+\nu_i)]\prod_{i=1}^{k+1}p_i^{x_i+\nu_i-1}.$$

Thus, the posterior distribution is again k-dimensional Dirichlet, this time with parameters $x_i+\nu_i$ for $1\leq i\leq k+1$.

As is well known, the posterior mean of p_i given $\underset{\sim}{x}$ is

$$E(p_i|\underset{\sim}{x}) = (x_i+\nu_i)/(n+\sum_{j=1}^{k+1}\nu_j). \tag{2.2}$$

Similarly, the posterior covariance matrix has elements

$$\mathrm{var}(p_i|\underset{\sim}{x}) = (n+\sum_{h=1}^{k+1}\nu_h+1)^{-1}\,E(p_i|\underset{\sim}{x})\,[1-E(p_i|\underset{\sim}{x})] \tag{2.3}$$

and

$$\mathrm{cov}(p_i,p_j|\underset{\sim}{x}) = -(n+\sum_{h=1}^{k+1}\nu_h+1)^{-1}\,E(p_i|\underset{\sim}{x})\,E(p_j|\underset{\sim}{x}). \tag{2.4}$$

The vector of posterior means (2.2) is the Bayes estimator for quadratic loss defined in (1.6).

In general, for l a positive integer,

$$E(p_i^l|\underset{\sim}{x}) = \prod_{q=0}^{l-1} (x_i+\nu_i+q) / \prod_{q=0}^{l-1} (n+\sum_{h=1}^{k+1} \nu_h+q) \tag{2.5}$$

so that, from multinomial expansion and substitution of (2.2) and (2.5), the l^{th} moment of $p_i|\underset{\sim}{x}$ about $E(p_i|\underset{\sim}{x})$ is

$$E\{[p_i-E(p_i|\underset{\sim}{x})]^l|\underset{\sim}{x}\} = \sum_{j=0}^{l} \left[(-1)^j \binom{l}{j} \left(\frac{x_i+\nu_i}{n+\Sigma\nu_h}\right)^j \prod_{q=0}^{l-j-1} \frac{x_i+\nu_i+q}{n+\Sigma\nu_h+q} \right]$$

$$= \sum_{j=0}^{l} \left[(-1)^j \binom{l}{j} \left(\frac{x_i/n+\nu_i/n}{1+\Sigma\nu_h/n}\right)^j \prod_{q=0}^{l-j-1} \frac{x_i/n+(\nu_i+q)/n}{1+(\Sigma\nu_h+q)/n} \right] \tag{2.6}$$

where we use the convention that $\prod_{q=0}^{-1} f(q)=1$ for any function of q.

2.2.2 Incomplete Data:

Recall the notation defined at the beginning of Section 1.2. Let $\underset{\sim}{p}$ again have the Dirichlet prior density $g(\underset{\sim}{p})$ of (1.1). Further, assume that given $\underset{\sim}{p}$, and thus all p_g, each $\underset{\sim}{z}_{g,P}$ has the multinomial distribution

$$h_P(\underset{\sim}{z}_{g,P}|\underset{\sim}{p}) = [(\sum_{g\in P} z_{g,P})! / \prod_{g\in P} z_{g,P}!] \prod_{g\in P} p_g^{z_{g,P}}. \tag{2.7}$$

Then, the likelihood of the total incomplete data $\underset{\sim}{z}$ given $\underset{\sim}{p}$ is

$$h(\underset{\sim}{z}|\underset{\sim}{p}) = \prod_{P} h_P(\underset{\sim}{z}_{g,P}|\underset{\sim}{p}). \tag{2.8}$$

The posterior density of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is therefore

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = g(\underset{\sim}{p})\, h(\underset{\sim}{z}|\underset{\sim}{p}) / \int_{p_k} g(\underset{\sim}{p})\, h(\underset{\sim}{z}|\underset{\sim}{p})\, d\underset{\sim}{p}. \tag{2.9}$$

To evaluate $f(\underset{\sim}{p}|\underset{\sim}{z})$, recall that $p_{\mathcal{S}}^{\Sigma z_{\mathcal{S}}, P}$ is $p_{\mathcal{S}}^{z_{\mathcal{S}}}$ and that $p_{\mathcal{S}}$ is a sum of probabilities; i.e., $p_{\mathcal{S}} = \sum_{j\in\mathcal{S}} p_j$. Thus, we can rewrite $p_{\mathcal{S}}^{z_{\mathcal{S}}}$ as a multinomial expansion. For example, if $p_{\mathcal{S}}=p_1+p_3+p_5$, then we can write $p_{\mathcal{S}}^{z_{\mathcal{S}}}$ as

$$(p_1+p_3+p_5)^{z_{\mathcal{S}}} = \sum_{j=0}^{z_{\mathcal{S}}} \sum_{i=0}^{j} \binom{z_{\mathcal{S}}}{j}\binom{j}{i} p_1^{\,i}\, p_3^{\,j-i}\, p_5^{\,z_{\mathcal{S}}-j}. \quad (2.10)$$

Rewriting the posterior density (2.9) in this manner, multiplying resulting terms times each other and the prior, and collecting terms yields the numerator as a sum of ω terms of the form

$$c_l p_1^{\gamma_{11}-1} p_2^{\gamma_{21}-1} \cdots p_{k+1}^{\gamma_{(k+1)1}-1} \quad (2.11)$$

where $1\le l\le\omega$, $\omega=\prod_{D}(z_D+1)$ for D containing more than one integer, c_l is a function of the incomplete data only (hence, not a function of $\underset{\sim}{p}$), and $\sum_{j=1}^{k+1}\gamma_{jl}=n+\sum_{i=1}^{k+1}\nu_i\equiv m$. That is, $\sum_{j=1}^{k+1}\gamma_{jl}\equiv m$ is the sum of the prior parameters ν_i plus the total number of observations and thus is independent of l. [See following Section 2.2.3 for an example.]

Hence, each term (2.11) of the numerator can be written as a Dirichlet density times a coefficient that is not a function of $\underset{\sim}{p}$ Therefore, integrating the numerator with respect to $\underset{\sim}{p}$ to evaluate the denominator yields that the posterior density of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = \sum_{l=1}^{\omega} c_l \prod_{j=1}^{k+1} p_j^{\gamma_{jl}-1} \Big/ \Big\{ \sum_{l=1}^{\omega} \Big[c_l \prod_{j=1}^{k+1} \Gamma(\gamma_{jl})/\Gamma(m)\Big]\Big\}. \quad (2.12)$$

Let $B=\sum_{l=1}^{\omega} c_l \prod_{j=1}^{k+1}\Gamma(\gamma_{jl})$. Then the posterior mean of p_i given $\underset{\sim}{z}$ is

$$E(p_i|\underset{\sim}{z}) = m^{-1} \sum_{l=1}^{\omega} c_l \Gamma(\gamma_{il}+1) \prod_{j\neq i}^{k+1} \Gamma(\gamma_{jl})/B. \tag{2.13}$$

Similarly,

$$E(p_i^2|\underset{\sim}{z}) = [m(m+1)]^{-1} \sum_{l=1}^{\omega} c_l \Gamma(\gamma_{il}+2) \prod_{j\neq i}^{k+1} \Gamma(\gamma_{jl})/B \tag{2.14}$$

and

$$E(p_i p_h|\underset{\sim}{z}) = [m(m+1)]^{-1} \sum_{l=1}^{\omega} c_l \Gamma(\gamma_{il}+1)\Gamma(\gamma_{hl}+1) \prod_{j\neq i,h}^{k+1} \Gamma(\gamma_{jl})/B, \tag{2.15}$$

for variance and covariance calculations

$$\text{var}(p_i|\underset{\sim}{z}) = E(p_i^2|\underset{\sim}{z}) - [E(p_i|\underset{\sim}{z})]^2 \tag{2.16}$$

and

$$\text{cov}(p_i,p_h|\underset{\sim}{z}) = E(p_i p_h|\underset{\sim}{z}) - E(p_i|\underset{\sim}{z})\, E(p_h|\underset{\sim}{z}), \tag{2.17}$$

respectively.

2.2.3 Example:

We now give an example for a small artificial data set to illustrate derivations given in Section 2.2.2. We also want to indicate difficulties that would be encountered in numerically evaluating these elements of the exact posterior mean and covariance matrices for larger or more complex data sets unless one has unusual computing equipment.

We created the data in the more restrictive form of Hocking and Oxspring to show how their form relates to ours. Suppose that we have observed the following data on three categories C_i, $1 \leq i \leq 3$,

$$\begin{array}{ccc} C_1 & C_2 & C_3 \end{array} \qquad n_i$$

$$\begin{pmatrix} 2 & 3 & 3 \\ \leftarrow 4 \rightarrow & & 3 \\ 2 & 2 & \end{pmatrix} \qquad \begin{array}{l} 8 \\ 7 \\ \underline{4} \\ 19 = n \end{array} \tag{2.18}$$

where the arrows denote the two categories between which the incompletely specified observations fall. The amount of incomplete data is 32% of the total sample size. In the notation of Section 1.2, we have that,

$$P_1 = \{\{1\},\{2\},\{3\}\}, \quad \mathcal{S}_{1,1}=\{1\}, \quad \mathcal{S}_{2,1}=\{2\}, \quad \mathcal{S}_{3,1}\{3\};$$

$$P_2 = \{\{1,2\}\ \{3\}\}, \quad \mathcal{S}_{1,2}=\{1,2\}, \quad \mathcal{S}_{2,2}=\{3\};$$

$$P_3 = \{\{1,3\}\ \{2\}\}, \quad \mathcal{S}_{1,3}=\{1,3\}, \quad \mathcal{S}_{2,3}=\{2\};$$

$z_{\{1\}}=2$, $z_{\{2\}}=3+2=5$, $z_{\{3\}}=3+3=6$, $z_{\{1,2\}}=4$, $z_{\{1,3\}}=2$, $z_{\{2,3\}}=0$, $\underset{\sim}{z}=(2,5,6,4,2,0)$, and $n=\sum_{\mathcal{S}} z_{\mathcal{S}} = 2+5+6+4+2=19$.

From (2.7) and (2.8), the likelihood of $\underset{\sim}{z}$ given $\underset{\sim}{p}$ is

$$\begin{aligned} L(\underset{\sim}{p};\underset{\sim}{z}) &= \frac{(2+3+3)!}{2!3!3!} p_1^2 p_2^3 p_3^3 \frac{(4+3)!}{4!3!} p_{\{1,2\}}^4 p_3^2 \frac{(2+2)!}{2!2!} p_{\{1,3\}}^2 p_2^2 \\ &= \frac{8!7!4!}{2!3!3!4!3!2!2!} p_1^2 p_2^5 p_3^6 (p_1+p_2)^4 (p_1+p_3)^2. \end{aligned} \tag{2.19}$$

Suppose that we have a uniform prior $g(p_1,p_2)=2$; that is, $\nu_i=1$ for $1 \leq i \leq 3$ in the Dirichlet prior (1.1). Then, the posterior density of $\underset{\sim}{p}$ given the incomplete data $\underset{\sim}{z}$ is

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = \frac{p_1^2 p_2^5 p_3^6 (p_1+p_2)^4 (p_1+p_3)^2}{\int_0^1 \int_0^{1-p_2} p_1^2 p_2^5 p_3^6 (p_1+p_2)^4 (p_1+p_3)^2 \, dp_1 \, dp_2}. \qquad (2.20)$$

Through expansion, multiplication, and collection of terms, the numerator of the posterior density (2.20) can be written as

$$\begin{aligned} p_1^2 p_2^5 p_3^6 \Big(\sum_{a=0}^{4} \binom{4}{a} p_1^a p_2^{4-a} \Big) \Big(\sum_{b=0}^{2} \binom{2}{b} p_1^b p_3^{2-b} \Big) &= p_1^2 p_2^9 p_3^8 + 4 p_1^3 p_2^8 p_3^8 \\ &+ 6p_1^4 p_2^7 p_3^8 + 4p_1^5 p_2^6 p_3^8 + p_1^6 p_2^5 p_3^8 + 2(p_1^3 p_2^9 p_3^7 + 4p_1^4 p_2^8 p_3^7 + \\ &+ 6p_1^5 p_2^7 p_3^7 + 4p_1^6 p_2^6 p_3^7 + p_1^7 p_2^5 p_3^7) + p_1^4 p_2^9 p_3^6 + 4p_1^5 p_2^8 p_3^6 \\ &+ 6p_1^6 p_2^7 p_3^6 + 4p_1^7 p_2^6 p_3^6 + p_1^8 p_2^5 p_3^6 . \end{aligned} \qquad (2.21)$$

Adding $\nu_i - 1 = 1 - 1 = 0$ to each exponent in (2.21), we have that the numerator is a sum of $\omega = \prod_D (z_D + 1) = 5 \times 3 = 15$ terms of the form

$$c_l \, p_1^{\gamma_{1l}-1} \, p_2^{\gamma_{2l}-1} \, p_3^{\gamma_{3l}-1}$$

with $\sum_{i=1}^{3} \gamma_{il} = n + \sum_{i=1}^{3} \nu_i = 19+3 = 22$ for all $1 \le l \le 15$. Integrating the numerator (2.21) with respect to $\underset{\sim}{p}$ to evaluate the denominator yields the posterior density (2.12) of $\underset{\sim}{p}$ given $\underset{\sim}{z}$.

The smaller the variance of a distribution, the better a point estimate, such as the mean, is as a descriptor of the distribution. Therefore, as a rough indication of how large the variance is, we define a sample coefficient of variation

$$C.V.(p_i|\underset{\sim}{z}) = [var(p_i|\underset{\sim}{z})]^{\frac{1}{2}} / E(p_i|\underset{\sim}{z}). \qquad (2.22)$$

[Note that the coefficient of variation is usually defined as a standard

deviation of an estimator (not a distribution) divided by the estimator.]

Calculating the mean (2.13), variance (2.16), covariance (2.17), and sample coefficient of variation (2.22) yields results shown in the following Table 2.1.

TABLE 2.1

EXAMPLE 2.2.3 RESULTS

moment \ i	1	2	3
$E(p_i\|\underset{\sim}{z})$	0.241202	0.384927	0.373871
$var(p_i\|\underset{\sim}{z})$	.011921	.012725	.011203
$C.V.(p_i\|\underset{\sim}{z})$	.4527	.2931	.2831
$cov(p_1,p_2\|\underset{\sim}{z})$=-0.006721,		$cov(p_1,p_3\|\underset{\sim}{z})$=-0.005199	
$cov(p_2,p_3\|\underset{\sim}{z})$=-0.006004			

As expected, the sample coefficient of variation is highest for p_1 because category 1 has the highest proportion of shared data. [Compare $(z_{12}+z_{13})/(z_1+z_{12}+z_{13})=.75$ with $z_{12}/(z_2+z_{12})=.44$ and $z_{13}/(z_3+z_{13})=.25$.] The posterior variance of p_1 is larger, in proportion to the posterior mean of p_1, than is that of p_2 or p_3 to their respective posterior means.

2.2.4 Evaluation Problems:

In general, we have the following problems in evaluating the exact posterior central moments:

(1) large number of terms - hence, pocket calculators and many desk calculators cannot be used;

(2) rounding errors (in the large number of terms to sum, the products of gamma functions and factorial-like constants c_1, approximations for the gamma functions, and final divisions, sums, and subtractions) - hence, computers must carry many figures of precision; and

(3) large magnitude of terms (each term is a product of generally large gamma functions and factorial-like constants c_1) - hence, computers must have an unusually large range of values unless much extra computer programing and execution cost, time, and storage are used.

In the next few paragraphs, we discuss these problems and give several illustrative examples. An example of an unusual electronic computer that can be straightforwardly used to calculate these moments in small enough samples is discussed in Sections 5.4 and 5.10.

The example given in the last section is among the smallest data sets one could have. Yet, even for it there are 15 terms in each of the numerators for $E(p_1|\underset{\sim}{z})$, $E(p_2|\underset{\sim}{z})$, $E(p_1^2|\underset{\sim}{z})$, $E(p_2^2|\underset{\sim}{z})$, $E(p_3^2|\underset{\sim}{z})$, $E(p_1p_2|\underset{\sim}{z})$, $E(p_1p_3|\underset{\sim}{z})$, and $E(p_2p_3|\underset{\sim}{z})$. The denominator, the same for all calculations, also had 15 terms. Hence, there were 135 terms plus all the multiplications within terms, additions, divisions, and subtractions to evaluate the final moments. For a trinomial sample having incompletely specified observations $z_{12}=z_{13}=z_{23}=9$, the number of terms in each numerator (and the one denominator) is 1000. Hence, there are a total of 9,000 terms to evaluate, not including any multiplication within terms, addition of

the 1000 terms, and subtractions and divisions for the final moments.

Finally, for a trinomial sample having incompletely specified observations z_{12}=16, z_{13}=17, and z_{23}=17 (corresponding to 50% incomplete data in a sample size of 100 and 15% incomplete data in a sample size of 330), there would be 5,202 terms in each numerator (and the one denominator) for evaluating the posterior mean and covariance matrix. Thus, the total number of terms, excluding the multiplications within terms, addition of the 5,202 terms, etc. would be 46,818.

To evaluate these moments even on a large electronic computer can be difficult. Because of the gamma function in the terms, we need a computer having an unusually large range. In the second example, a term of $\binom{9}{5}\binom{9}{5}\binom{9}{4}\,\Gamma(35)\,\Gamma(40)\,\Gamma(32) \doteq 10^{134}$ would exceed the range of most electronic computers. Most have ranges smaller than 10^{-100} - 10^{100}. Yet, depending on the prior, this is a term for a sample size of only 100, and this is only one of 1,000 terms. We can circumvent the range problem by dividing each term of the numerator and denominator by a large value; hence, scaling down the terms. However, doing so takes more computer programing and execution time, cost, and storage. Further, it also creates problems with roundoff error. We might also have to scale down more than once, depending on the values involved. Each successive such scaling involves increasing cost and roundoff error.

The cost and time involved in evaluating these moments is important. The loss in precision, however, is critical. For the third example, a computer carrying even eight significant-figure accuracy will yield an answer for the exact solution that can be counted on for only one or two significant figures. [The large loss in precision owes to rounding errors

in approximations for the gamma functions, in the several multiplications within each term, in the additions of the 5,202 terms (roundoff error is approximately $\sqrt{5202}$ or 2 to 3 significant figures), in the final divisions, additions, and subtractions, along with roundoff error from any divisions necessary to scale down the magnitude of terms to fit within the range of the computer.]

If a computer carries six significant-figure accuracy, which is common, one might not get any accurate evaluation. Hence, any canned computer program would be particularly susceptible to wrong usage and interpretation. Someone not understanding the numerical problems or heeding any package warnings might use it on a six significant-figure single-precision accurate computer and think his answers were correct.

On many large electronic computers, one can use double-precision significant-figure calculations. However, doing so would usually at least quadruple the cost. Further, on those large electronic computers, as well as those numerous kinds of desk and pocket calculators, not allowing double-precision calculations, or enough single-precision accuracy, there is no way to obtain an accurate evaluation of the exact posterior mean and covariance elements.

One driving factor in these problems is the large magnitude of the terms. The other driving factor is the number $\omega=\prod_{D}(z_D+1)$ of these terms in each numerator of $E(p_i^j|\underset{\sim}{z})$. As either sample size or percentage of incomplete data increases, ω increases. For a sample size of 200 and percentage of incomplete data of 50% with $z_{12}=z_{13}=33$ and $z_{23}=34$, the number of terms in each numerator for the moments is 40,460. Hence, the

total number of terms for just the numerators (excluding multiplications and gamma approximations involved in each term) is 364,140. However, if we consider the same sample size and percentage of incomplete data for a 5-nomial having, say $z_{12}=15$, $z_{23}=34$, $z_{25}=16$, $z_{35}=3$, $z_{123}=15$, and $z_{1234}=17$, the number of terms in each numerator is 645,120 and the total number of terms for just the numerators (with same preceding exclusions) is 5,806,080. Hence, the problems illustrated for the trinomial data samples, as well as the cost, increase in somewhat factorial manner as the number of multinomial dimensions increase.

Finally, it would be nice to have a short, easily remembered and easily evaluated, formula for at least the posterior mean. As Hoaglin (1977) notes, such a formula is valuable. It can be evaluated by pocket calculators anywhere. The maximum likelihood estimate and posterior mode, to be given in the next section, both have short, easily remembered formulas. Although these formulas can often be evaluated by pocket calculator, they are not simple to evaluate in general. However, they are very easy and inexpensive to program for computer evaluation. In particular, they do not have the three computational problems just outlined for the exact posterior mean. We find in Chapter 3 that we can derive a similar, although approximate, formula for the posterior mean.

2.3 Mode Estimators:

2.3.1 Background:

In this section we show how the maximum likelihood estimate and posterior mode are derived. First consider the complete-data equivalent $\underset{\sim}{x}$ of $\underset{\sim}{z}$. Let $z_D^{(i)}$ denote the (unknown) number of the z_D observations that fall in category C_i. Then, for $1 \le i \le k+1$,

$$x_i = z_i + \sum_{D \ni i} z_D^{(i)}. \tag{2.23}$$

For the theory of this section, we want to express the complete-data density

$$h(\underset{\sim}{x}|\underset{\sim}{p}) = (n!/\sum_{i=1}^{k+1} x_i!) \prod_{i=1}^{k+1} p_i^{x_i} \tag{2.24}$$

in terms of exponential-family parameters. Therefore, for $1 \le i \le k$, define

$$\phi_i = \ln(p_i/p_{k+1}). \tag{2.25}$$

Definition (2.25) and $\sum_{i=1}^{k+1} p_i = 1$ yield that

$$p_{k+1} = 1/(1+\sum_{i=1}^{k} e^{\phi_i}) \tag{2.26}$$

and

$$p_i = e^{\phi_i}/(1+\sum_{j=1}^{k} e^{\phi_j}). \tag{2.27}$$

For $1 \le i \le k$, define the sufficient statistics for $\underset{\sim}{p}$ as

$$t_i(\underset{\sim}{x}) = x_i. \tag{2.28}$$

Then $h(\underset{\sim}{x}|\underset{\sim}{p})$ can be written in exponential-family form as

$$h(\underset{\sim}{x}|\underset{\sim}{\phi}) = b(\underset{\sim}{x}) \exp[\underset{\sim}{\phi}\, \underset{\sim}{t}(\underset{\sim}{x})']/a(\underset{\sim}{\phi}) \tag{2.29}$$

for $b(\underset{\sim}{x}) = n!/\prod_{i=1}^{k+1} x_i!$ and $a(\underset{\sim}{\phi}) = (1+\sum_{i=1}^{k} e^{\phi_i})^{-n}$ since $\sum_{i=1}^{k+1} x_i = n$.

2.3.2 Maximum Likelihood Estimate:

For the multinomial distribution, the likelihood is the density. Thus, we seek to maximize $h(\underset{\sim}{z}|\underset{\sim}{\phi})$, the incomplete-data density (2.8) rewritten in terms of the exponential parameters $\underset{\sim}{\phi}$. From Sundberg (1974), the first and second partial derivatives of the log likelihood are

$$\partial \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/\partial\underset{\sim}{\phi} = -E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}] + E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}] \tag{2.30}$$

and

$$\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial\underset{\sim}{\phi}\partial\underset{\sim}{\phi}') = -\mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}] + \mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}]. \tag{2.31}$$

At the maximum of the likelihood, the vector(2.30) of first partial derivatives is zero, so that

$$E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}] = E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}]. \tag{2.32}$$

Since

$$E[t_i(\underset{\sim}{x})|\underset{\sim}{\phi}] = np_i, \tag{2.33}$$

and, from (2.23),

$$E(z_i|\underset{\sim}{z},\underset{\sim}{\phi}) = z_i \tag{2.34}$$

and

$$E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{\phi}) = z_D\, p_i/p_D \tag{2.35}$$

where, again, $p_D = \sum_{j\in D} p_j$, evaluation of (2.32) yields that the maximum likelihood estimate $\hat{p}_i$ of p_i is

$$\hat{p}_i = [z_i + \sum_{D\ni i} z_D\, \hat{p}_i/\hat{p}_D]/n. \tag{2.36}$$

To solve the nonlinear system of equations arising from (2.36), we use the EM algorithm of Dempster, Laird, and Rubin (1977). The algorithm is divided into two steps. In the expectation step (their E-step), the complete-data sufficient statistics $\underset{\sim}{t}(\underset{\sim}{x})$ are estimated by finding a solution to

$$[\underset{\sim}{t}(\underset{\sim}{x})]^{(l)}=E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}^{(l)}]. \tag{2.37}$$

In the maximization step (their M-step), $\underset{\sim}{\phi}^{(l+1)}$ is determined as the solution of the equations

$$E[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}] = [\underset{\sim}{t}(\underset{\sim}{x})]^{(l)}. \tag{2.38}$$

Thus, translating back from $\underset{\sim}{\phi}$ to $\underset{\sim}{p}$, we estimate an initial value $\hat{p}_i^{(0)}$ of $\hat{p}_i$ for $1\leq i\leq k$. We then substitute $\hat{\underset{\sim}{p}}^{(0)}$, together with $\underset{\sim}{z}$, into the right-hand side $z_i+\sum_{D\ni i}\hat{p}_i^{(0)}/\hat{p}_D^{(0)}$ of (2.37) and evaluate for $[\underset{\sim}{t}(\underset{\sim}{x})]^{(0)}$. Given $[\underset{\sim}{t}(\underset{\sim}{x})]^{(0)}$, we then solve (2.38); i.e., we solve

$$n\hat{\underset{\sim}{p}} = [\underset{\sim}{t}(\underset{\sim}{x})]^{(0)} \tag{2.39}$$

for $\hat{\underset{\sim}{p}}^{(1)}$; hence, $\hat{\underset{\sim}{p}}^{(1)}=[\underset{\sim}{t}(\underset{\sim}{x})]^{(0)}/n$. We then successively repeat the E and M steps until convergence; that is, until successive values of $\hat{\underset{\sim}{p}}^{(l)}$ agree to the desired number of significant figures.

Since we are concerned only with finite values of $\underset{\sim}{z}$, the likelihood $h(\underset{\sim}{z}|\underset{\sim}{\phi})$ is bounded. Hence, the first condition of Dempster, Laird, and Rubin (1977) for guaranteeing convergence of the EM algorithm to a local maximum or saddle point is satisfied. Further, the complete-data multinomial distribution is a member of the regular exponential family. Hence, the last convergence condition is simply that the eigenvalues of $\mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}]$ be bounded above zero on some path joining all $\underset{\sim}{\phi}^{(l)}$. From Graybill (1969,

p187), the eigenvalues λ are the solution to the characteristic equation

$$[1-\sum_{i=1}^{k} p_i^2/(p_i-\lambda)] \prod_{i=1}^{k} (p_i-\lambda) = 0. \tag{2.40}$$

In general, we want (2.31) to be negative semidefinite.

Dempster, Laird, and Rubin (1977) give the rate of convergence of the EM algorithm. For the multinomial distribution the rate of convergence is the largest eigenvalue of

$$\text{cov}[t(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}^{(t)}] \{\text{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}^{(t)}]\}^{-1}, \tag{2.41}$$

for $\underset{\sim}{\phi}^{(t)}$ the converged estimate of $\underset{\sim}{\phi}^{(1)}$, provided that this eigenvalue is less than 1. As expected, when the percentage of incomplete data is small, the algorithm converges rapidly. As the percentage of incomplete data increases, the number of iterations increases. Dempster, Laird, and Rubin also note that, since the allocation of incompletely specified observations often varies across different components of $\underset{\sim}{p}$, certain components of $\underset{\sim}{p}$ may converge rapidly while others may converge slowly.

2.3.3 Posterior Mode:

The derivation for the posterior mode of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is similar to that for the maximum likelihood estimate. For the posterior mode, however, the prior must be included in the maximization.

Recall from (2.9) that the posterior density of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = g(\underset{\sim}{p})\, h(\underset{\sim}{z}|\underset{\sim}{p})/\int_{P_k} g(\underset{\sim}{p})\, h(\underset{\sim}{z}|\underset{\sim}{p})\, d\underset{\sim}{p}. \tag{2.42}$$

From definition (1.1) of the prior $g(\underset{\sim}{p})$, that piece of $\log[f(\underset{\sim}{p}|\underset{\sim}{z})]$ from (2.42) that depends on $\underset{\sim}{p}$ is the same as that piece of $\log[h(\underset{\sim}{z}|\underset{\sim}{p})]$ that

depends on $\underset{\sim}{p}$ except that, for $1 \leq i \leq k+1$, z_i is replaced by $(z_i+\nu_i-1)$ and, hence, n is replaced by $n+\sum_{j=1}^{k+1} \nu_j-(k+1)$. Therefore, from (2.36), the posterior mode $\hat{\tilde{p}}$ of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is given by

$$\hat{\tilde{p}}_i = (z_i+\nu_i-1+\sum_{D \ni i} z_D \hat{\tilde{p}}_i/\hat{\tilde{p}}_D)/[n+\sum_{j=1}^{k+1} \nu_j-(k+1)] \tag{2.43}$$

for $1 \leq i \leq k+1$. As for the maximum likelihood estimate, we evaluate the nonlinear system of equations arising from (2.43) by the EM algorithm. The comments in Section 2.3.2 concerning convergence also hold for the posterior mode. In general, the prior should reduce the effect of incomplete data so that convergence should be somewhat faster for the posterior mode than for the maximum likelihood estimate. The numerator for the convergence matrix in (2.41) is given in Appendix 4D.2 for the maximum likelihood estimate. Derivation for the posterior mode is similar. Calculating second partial derivatives of the two log likelihoods for the complete-data case yields for elements of $\{\text{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}^{(t)}]\}^{-1}$ in the denominator of (2.41):

for the maximum likelihood estimate -

$$\begin{aligned} \hat{\sigma}^{ii} &= n[-1/\hat{p}_i^{(t)}+1/\hat{p}_{k+1}^{(t)}] \\ \text{and} \qquad & \qquad (2.44) \\ \hat{\sigma}^{ij} &= n\hat{p}_{k+1}^{(t)} \end{aligned}$$

and for the posterior mode -

$$\hat{\tilde{\sigma}}^{ii} = [n+\sum_{j=1}^{k+1} \nu_j-(k+1)]\ \{-[\hat{\tilde{p}}_i^{(t)}+\nu_i-1]/[\hat{\tilde{p}}_i^{(t)}]^2+[\hat{\tilde{p}}_{k+1}^{(t)}+\nu_{k+1}-1]/[\hat{\tilde{p}}_{k+1}^{(t)}]^2\}$$

and (2.45)

$$\hat{\tilde{\sigma}}^{ij} = [n+\sum_{j=1}^{j+1} \nu_j-(k+1)]\ [\hat{\tilde{p}}_{k+1}^{(t)}+\nu_{k+1}-1]/[\hat{\tilde{p}}_{k+1}^{(t)}]^2.$$

In most cases, the prior parameters ν_i are greater than 1; hence, the denominator in (2.41) is usually larger for the posterior mode than for the maximum likelihood estimate. Which of the posterior mode and the maximum likelihood estimate actually has the faster rate of convergence, of course, depends also on the relative sizes of the numerators in (2.41).

Note from the "-1" term that it is possible for (2.43) to be negative. If so, the mode occurs at a boundary point; i.e., the posterior mode is zero. Also observe that if $\nu_i=1$, for $1\leq i\leq k+1$, then the posterior mode and the maximum likelihood estimate are identical.

CHAPTER 3

APPROXIMATIONS FOR POSTERIOR MEAN AND COVARIANCE MATRICES

3.1 Introduction:

As discussed in Section 2.2.4, numerical evaluation of elements of the mean and covariance matrices of the posterior distribution of $\underset{\sim}{p}$ given incomplete data $\underset{\sim}{z}$ is unfeasible for all but those cases having only a small number of incompletely specified observations. Therefore, we seek approximations for these posterior moments.

In the next chapter, we prove that the limiting central moments of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ are corresponding moments of the limiting distribution. In particular, the limit of the posterior mean is the mean of the limiting posterior distribution. We also prove that the mean of the limiting posterior distribution is the maximum likelihood estimate (2.36). Finally, from equations (2.36) and (2.43), the posterior mode equals the maximum likelihood estimate in the limit and, hence, equals the limiting posterior mean. Therefore, two natural candidates to approximate the exact posterior mean are the maximum likelihood estimate and the posterior mode. However, there are also problems in using these estimates as approximations.

The maximum likelihood estimate is best known for being good in large samples; it is not necessarily good in small samples. In particular, if a value of z_i has been observed that has very small probability for given p_i, then the maximum likelihood estimate will be poor if the sample size is small. For example, if $p_i = .20$ and we observe $z_i = 10$ in a sample of size 25, then the maximum likelihood

estimate $\hat{p}_i = .40$ is a poor estimate of p_i. Further, the maximum likelihood estimate is the correct estimate for an estimation criterion of choosing that value of $\underset{\sim}{p}$ that maximizes the likelihood (2.8) and not for an estimation criterion of minimizing expected risk (1.5). Finally, the maximum likelihood estimate has no place for a prior, which is important in all but those cases in which the current data is of large enough sample size, or significantly greater relevance, to drown out past information.

The posterior mode (2.43) does incorporate the prior. However, the posterior mode is the correct estimate for an estimation criterion of choosing that value of $\underset{\sim}{p}$ that maximizes the posterior density given the prior density $g(\underset{\sim}{p})$ and observed data $\underset{\sim}{z}$ and not for an estimation criterion of minimizing expected risk. Finally, from equation (2.43) we observe that, for small enough prior ν_i, a component of the posterior mode $\hat{\tilde{p}}_i$ can be approximately zero even though an observation ($z_i=1$) has been observed.

A different approach for approximating the exact posterior mean $\tilde{\underset{\sim}{p}}$ is to note that the posterior mean of the complete-data Dirichlet density with prior parameters $(\nu_1,\ldots,\nu_k;\nu_{k+1})$ equals the posterior mode of the complete-data Dirichlet density with prior parameters $(\nu_1+1,\ldots,\nu_k+1;\nu_{k+1}+1)$; that is, from (2.2)

$$(x_i+\nu_i)/(n+\sum_{j=1}^{k+1}\nu_j) = [x_i+(\nu_i+1)-1]/[n+\sum_{j=1}^{k+1}(\nu_j+1)-(k+1)]. \qquad (3.1)$$

Therefore, paralleling the incomplete-data posterior mode (2.46), we could estimate the incomplete-data exact posterior mean (2.13) by

$$\dot{\tilde{p}}_i = (z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/(n+\sum_{j=1}^{k+1}\nu_j). \qquad (3.2)$$

A very important property of approximation (3.2) is that as the proportion $\sum_{D\ni i} z_D/(n+\Sigma\nu_j)$ of incomplete data goes to zero, approximation (3.2) equals the exact posterior mean (3.1).

However, there are problems with this approach to obtain (3.2). We find in this chapter that the relationship between the posterior mean and posterior mode for complete data does not hold for incomplete data. Thus, (3.2) is an approximation and this approach does not enable us to assess its accuracy. Finally, from consideration of the definition and from small-sample examples (one given at the end of this chapter), we do not expect the large-sample covariance matrix of the posterior mode or maximum likelihood estimate to be a good approximation for the exact posterior covariance matrix. Therefore, we seek another type of approach for estimating the exact posterior central moments.

As noted, both the posterior mode and maximum likelihood estimate are derived from consideration of an estimation criterion other than minimization of expected risk (1.5). Therefore, one way to seek another approximation is to start with the desired estimation criterion; that is begin with the exact solutions for the posterior mean and covariance matrices. However, approximating exact solutions (2.13) - (2.17) for the posterior moments given incomplete data is difficult because of the number and structure of terms. An alternative method starts with exact solutions (2.2) - (2.6) for the posterior moments given complete data and then transforms these solutions via conditional probability to the

incomplete-data case, making any necessary approximations along the way.

In this chapter, we follow the above approach to derive approximations for the posterior central moments by making extended use of conditional probability and first-order Taylor-series approximations. Section 2.2 gives the posterior moments given complete data. Therefore, for incomplete data $\underset{\sim}{z}$ we substitute fictitious complete data consistent with $\underset{\sim}{z}$ and write the results of Section 2.2.1. Then, twice applying known lemmas on conditioning, we average results from the complete-data step over the posterior distribution of the unknown, substituted, complete data. At this point we still have unknown terms in the expressions. For these, we use Taylor-series approximations. The resulting approximation for the posterior mean is equation (3.2); hence, as the percentage of incomplete data goes to zero, the approximation goes to the exact posterior mean. From (2.36) and (2.43), neither the maximum likelihood estimate nor the posterior mode has this important property. Also, since asymptotically (3.2) equals the maximum likelihood estimate (2.36), it equals the limiting exact posterior mean. Further, since Taylor-series expansions are used, we can assess the accuracy of the approximations. Finally, we can use the same approach to approximate elements of the posterior covariance matrix. Doing so, we find the same important property in the resulting approximations that they go to the exact posterior variances and covariances as the percentage of incomplete data goes to zero. Note that, since the Taylor-series approximation (3.2) for the posterior mean is also a posterior mode

[for the prior ν_i+1], it can be evaluated by the EM algorithm discussed in Section 2.3.2.

In the next section, we derive the Taylor-series approximations for elements of the posterior mean and covariance matrices. Intermediate calculations are given in Appendices 3A, 3B, and 3C. Section 3.3 algebraically illustrates the resulting approximations for the trinomial distribution. Section 3.4 concludes the chapter with a comparison of the Taylor-series approximations, maximum likelihood estimate, and the posterior mode on the small-sample data set given in Section 2.2.3.

3.2 Derivation of Taylor-Series Approximations:

3.2.1 Posterior Mean Vector:

Again let D denote the set $\mathcal{S}$ from Section 1.2 containing more than one element and, for $i \in D$, define $z_D^{(i)}$ as the number of the z_D observations that fall in category i. If $z_D^{(i)}$ were known for all i and all D, then the data would be complete and the posterior central moments would directly follow from Section 2.2.1. Therefore, assume that we know all $z_D^{(i)}$ and denote the vector of this unknown information by $\underset{\sim}{u}$. Thus, $\underset{\sim}{u}$ is the vector of all $z_D^{(i)}$ for all D and all $1 \leq i \leq k$. For example, in Section 2.2.3 we would have that $\underset{\sim}{u}=(z_{12}^{(1)}, z_{13}^{(1)})$ and $\underset{\sim}{z}=(z_1, z_2, z_3, z_{12}, z_{13}, z_{23})$. Given $\underset{\sim}{z}$ and $\underset{\sim}{u}$, then, for $1 \leq i \leq k$, we have complete data

$$x_i = \sum_{\mathcal{S} \ni i} z_{\mathcal{S}}^{(i)} = z_i + \sum_{D \ni i} z_D^{(i)}. \tag{3.3}$$

Thus, from Section 2.2.1, recalling that $m = \sum_{j=1}^{k+1} \nu_j + n$, we have from (2.2) the posterior mean

$$E(p_i | \underset{\sim}{z}, \underset{\sim}{u}) = (x_i + \nu_i)/m = (z_i + \sum_{D \ni i} z_D^{(i)} + \nu_i)/m. \tag{3.4}$$

To obtain moments of $\underset{\sim}{p}$ given only the observed data $\underset{\sim}{z}$, then, we average result (3.4) over the distribution of $\underset{\sim}{u} | \underset{\sim}{z}$. To do so, write the posterior density $f(\underset{\sim}{p} | \underset{\sim}{z})$ as

$$\begin{aligned} f(\underset{\sim}{p} | \underset{\sim}{z}) &= \int \ell(\underset{\sim}{p}, \underset{\sim}{u} | \underset{\sim}{z}) \, d\underset{\sim}{u} \\ &= \int g(\underset{\sim}{p} | \underset{\sim}{z}, \underset{\sim}{u}) \, h(\underset{\sim}{u} | \underset{\sim}{z}) \, d\underset{\sim}{u} \end{aligned} \tag{3.5}$$

for $\ell(\underset{\sim}{p}, \underset{\sim}{u} | \underset{\sim}{z})$ the joint posterior density of $\underset{\sim}{p}$ and $\underset{\sim}{u}$ given $\underset{\sim}{z}$ and $g(\underset{\sim}{p} | \underset{\sim}{z}, \underset{\sim}{u})$ and $h(\underset{\sim}{u} | \underset{\sim}{z})$, conditional densities.

From (3.5) we obtain the following standard lemma [see Parzen (1962, p55) or Rao (1968,p79)] on conditioning, which we write in terms of general random variables V and W because we apply the lemma to one other density besides $f(\underset{\sim}{p}|\underset{\sim}{z})$:

Lemma 3.1: For random variables V and W, and where the variable under the expected-value sign E is the variable with respect to which the expectation is to be taken: $E(V)=\underset{W}{E}[E(V|W)]$, $var(V)=\underset{W}{E}[var(V|W)]+\underset{W}{var}[E(V|W)]$, and $cov(V_1,V_2)=\underset{W}{E}[cov(V_1,V_2|W)]+\underset{W}{cov}[E(V_1|W),E(V_2|W)]$.

By using Lemma 3.1 and (3.4) we have, defining $r_{iD}=p_i/p_D$, that

$$
\begin{aligned}
E(p_i|\underset{\sim}{z}) &= \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}[E(p_i|\underset{\sim}{z},\underset{\sim}{u})] \\
&= \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}\{[\sum_{g\ni i} z_g^{(i)}+\nu_i]/m\} \\
&= [z_i+\sum_{D\ni i} E(z_D^{(i)}|\underset{\sim}{z})+\nu_i]/m \qquad (3.6) \\
&= \{z_i+\nu_i+\sum_{D\ni i} \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})]\}/m \\
&= [z_i+\nu_i+\sum_{D\ni i} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z})]/m.
\end{aligned}
$$

The first line of (3.6) follows from applying Lemma 3.1 to $E(p_i|\underset{\sim}{z})$; the second line, from complete-data posterior mean (2.2); and the third line, from separating out that part of $\sum_{g\ni i} z_g^{(i)}$ that is already known. The fourth line of (3.6) follows from applying Lemma 3.1 to $E(z_D^{(i)}|\underset{\sim}{z})$; and the last line, from the complete-data multinomial specification.

In Appendix 3B we show through Taylor-series expansions that

$$\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z}) = E(p_i|\underset{\sim}{z})/E(p_D|\underset{\sim}{z}) + O(n^{-1}), \tag{3.7}$$

where the symbol O giving the order of magnitude of the error is defined in Appendix 3B. Details are given in Appendix 3B. Therefore, substituting (3.7) into (3.6) and collecting terms yields, for $\tilde{p}_i \equiv E(p_i|\underset{\sim}{z})$ and the error ε_i to be determined in Chapter 4, that

$$\tilde{p}_i = (z_i+\nu_i+\sum_{D\ni i} z_D\tilde{p}_i/\tilde{p}_D)/m + \varepsilon_i. \tag{3.8}$$

Dropping the error term in (3.8) yields, for $1\leq i\leq k$, the Taylor-series approximate posterior mean vector $\dot{\underset{\sim}{p}}$; i.e.,

$$\dot{\tilde{p}}_i = (z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/m. \tag{3.9}$$

Observe that (3.9) is the same approximation (3.2) obtained by paralleling the complete-data relationship between the exact posterior mean and the posterior mode.

Calculations for Taylor-Series Approximate Posterior Mean: For those categories i that have only complete data, the Taylor-series approximation is the exact posterior mean (2.2). For those categories i that have incomplete data, we use the EM iterative algorithm of Dempster, Laird, and Rubin (1977) described in Section 2.3.2 since (3.9) is a posterior mode for the prior $\beta_i=\nu_i+1$. Thus, for those categories i that have incomplete data, s denoting the number of iteration, and

$$\dot{\tilde{r}}_{iD}^{(s)} = \dot{\tilde{p}}_i^{(s)}/\dot{\tilde{p}}_D^{(s)}, \tag{3.10}$$

we approximate the exact posterior mean from (3.9) by the iterative algorithm

$$\dot{\tilde{p}}_i^{(s+1)} = [z_i+\nu_i+\sum_{D\ni i} z_D \dot{\tilde{r}}_{iD}^{(s)}]/m. \qquad (3.11)$$

To begin (3.11), we use the data $\underset{\sim}{z}$, prior parameters ν_j, and any other available information to choose an initial estimate $\dot{\tilde{p}}_i^{(0)}$, $1 \leq i \leq k$, and thus an initial estimate of $\dot{\tilde{r}}_{iD}^{(0)}$. Substituting $\dot{\tilde{r}}_{iD}^{(0)}$ into the right-hand side of (3.11), we evaluate (3.11) to obtain $\dot{\tilde{p}}_i^{(1)}$ and $\dot{\tilde{r}}_{iD}^{(1)}$ for all i referring to categories having incomplete data. Using $\dot{\tilde{r}}_{iD}^{(1)}$, we then reevaluate (3.11) to calculate $\dot{\tilde{p}}_i^{(2)}$. We continue in this cyclic fashion until results from successive iterations agree to the desired number of significant figures.

Note that the system of k equations arising from (3.9) for the Taylor-series approximate posterior mean is nonlinear. Thus, as for the maximum likelihood estimate (2.36) and the posterior mode (2.43), the number of solutions to this system can range from zero to infinity. [See Ortega and Rheinboldt (1970,p2).] If there are solutions, none need be in P_k. If a solution is in P_k, it need not be close to the exact posterior mean. However, since (3.9) is a posterior mode for the prior $\underset{\sim}{\beta}=\underset{\sim}{\nu}+\underset{\sim}{1}$, Dempster, Laird, and Rubin (1977) give conditions (discussed in Sections 2.3, 4.3.2, and 5.8.3 and Appendix 4E) under which an iterative solution for (3.9) converges to a local maximum in P_k. Hence, when these conditions are met, there is at least one solution in P_k. In Chapter 4 we give conditions under which an iterative solution converges to within a small error of the exact posterior mean. We also speculate that this solution, when it exists, is given by the global

maximum in P_k, which is found by choosing that one of the local maximum in P_k that maximizes the likelihood.

When there are only a few patterns of incomplete data, the non-linear system of equations arising from (3.9) for the posterior mean vector can sometimes be solved analytically. Several solutions will be obtained but usually all but one will fail to satisfy the constraints $0 \leq \tilde{p}_i \leq 1$ and $\sum_{i=1}^{k+1} \tilde{p}_i = 1$. Examples of analytic solutions for the asymptotic posterior mean and covariance matrices are given in Appendix 4D.5.

3.2.2 Posterior Covariance Matrix:

For approximating elements of the posterior covariance matrix, we follow the same procedure given in the last section. For the complete-data step that lead to (3.4), we obtain

$$\mathrm{var}(p_i|\underset{\sim}{z},\underset{\sim}{u}) = \{E(p_i|\underset{\sim}{z},\underset{\sim}{u})\,[1-E(p_i|\underset{\sim}{z},\underset{\sim}{u})]\}/(m+1) \tag{3.12}$$

and

$$\mathrm{cov}(p_i,p_h|\underset{\sim}{z},\underset{\sim}{u}) = -\,[E(p_i|\underset{\sim}{z},\underset{\sim}{u})\,E(p_h|\underset{\sim}{z},\underset{\sim}{u})]/(m+1). \tag{3.13}$$

For the conditioning step that lead to (3.6), we obtain

$$\begin{aligned}\mathrm{var}(p_i|\underset{\sim}{z}) = {} & \sum_{D\ni i} [z_D{}^2 \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{var}}(r_{iD}|\underset{\sim}{z}) + \sum_{\substack{Q\ni i \\ Q\neq D}} z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{cov}}(r_{iD},r_{iQ}|\underset{\sim}{z})]/[m(m+1)] \\ & + \{\sum_{D\ni i} (z_D/m) \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[r_{iD}(1-r_{iD})|\underset{\sim}{z}] + E(p_i|\underset{\sim}{z})[1-E(p_i|\underset{\sim}{z})]\}/(m+1),\end{aligned} \tag{3.14}$$

and, for $h \neq i$,

$$\operatorname{cov}(p_i,p_h|\underset{\sim}{z}) = \sum_{D\ni i}\sum_{Q\ni h} z_D z_Q/[m(m+1)]\ \operatorname{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD},r_{hQ}|\underset{\sim}{z})$$

$$-[\sum_{D\ni i,h}(z_D/m)\ E_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}r_{hD}|\underset{\sim}{z})+E(p_i|\underset{\sim}{z})E(p_h|\underset{\sim}{z})]/(m+1). \tag{3.15}$$

Derivations for (3.14) and (3.15) are given in Appendix 3A.

Finally, for the ratio-approximation step that lead to (3.8), we have, with ratio moments given in Appendix 3B and substitution details for (3.14) given in Appendix 3C, that

$$\tilde{\sigma}_{ii} = \sum_{D\ni i}(z_D/m)[(z_D-1)/(m+1)]/\tilde{p}_D^{\,4}\{[\tilde{p}_{\emptyset}^{\,2}\tilde{\sigma}_{ii}+\tilde{p}_i\sum_{j\in\emptyset}[\tilde{p}_i\tilde{\sigma}_{jj}-2\tilde{p}_{\emptyset}\tilde{\sigma}_{ji}+2\tilde{p}_i\sum_{\substack{l\in\emptyset\\ l>j}}\tilde{\sigma}_{jl}]]\}$$

$$+\sum_{D\ni i}\sum_{\substack{Q\ni i\\ Q\neq D}}(z_D/m)[z_Q/(m+1)]/(\tilde{p}_D\tilde{p}_Q)^2\{\tilde{p}_{\emptyset}[\tilde{p}_{\emptyset}\tilde{\sigma}_{ii}-\tilde{p}_i\sum_{l\in\emptyset}\tilde{\sigma}_{il}]+\tilde{p}_i\sum_{j\in\emptyset}[\tilde{p}_i\sum_{l\in\emptyset}\tilde{\sigma}_{jl}-\tilde{p}_{\emptyset}\tilde{\sigma}_{jl}]\}$$

$$+[\sum_{D\ni i}(z_D/m)\tilde{p}_i\tilde{p}_{\emptyset}/\tilde{p}_D^{\,2}+\tilde{p}_i(1-\tilde{p}_i)]/(m+1)+\delta_{ii}, \tag{3.16}$$

and, for $h>i$ and $\mathbb{D}$ denoting D minus the integer h,

$$\tilde{\sigma}_{ih} = \sum_{D\ni i}\sum_{Q\ni h}(z_D/m)[z_Q/(m+1)]/(\tilde{p}_D^{\,2}\tilde{p}_Q^{\,2})\{\tilde{p}_{\emptyset}[\tilde{p}_{\emptyset}\tilde{\sigma}_{ih}-\tilde{p}_h\sum_{l\in\emptyset}\tilde{\sigma}_{il}]+\tilde{p}_i\sum_{j\in\emptyset}[\tilde{p}_h\sum_{l\in Q}\tilde{\sigma}_{jl}-\tilde{p}_Q\tilde{\sigma}_{jh}]\}$$

$$-[\sum_{D\ni i,h}(z_D/m)/\tilde{p}_D^{\,4}(\tilde{p}_i\tilde{p}_h\,\tilde{p}_D^{\,2}+\tilde{p}_{\emptyset}\tilde{p}_{\mathbb{D}}\tilde{\sigma}_{ih}-\tilde{p}_h\tilde{p}_{\emptyset}\sum_{j\in\mathbb{D}}\tilde{\sigma}_{ij}-\tilde{p}_i\tilde{p}_{\mathbb{D}}\sum_{j\in\emptyset}\tilde{\sigma}_{jh}+\tilde{p}_h\tilde{p}_i\sum_{j\in\emptyset}\sum_{l\in\mathbb{D}}\tilde{\sigma}_{jl})$$

$$+\tilde{p}_i\tilde{p}_h]/(m+1)+\delta_{ih}, \tag{3.17}$$

where $\tilde{\sigma}_{ii}\equiv\operatorname{var}(p_i|\underset{\sim}{z})$, $\tilde{\sigma}_{ih}\equiv\operatorname{cov}(p_i,p_h|\underset{\sim}{z})$, $\emptyset$ and $\not\emptyset$ denote D and Q, respectively, minus the integer over which they are summed (so that $\emptyset$ is D minus i and $\not\emptyset$ is Q minus h or Q minus i depending on the definition of Q given under the summation sign), and, again, $\tilde{p}_i\equiv E(p_i|\underset{\sim}{z})$ so that $\tilde{p}_{ij}=\tilde{p}_i+\tilde{p}_j$. The terms δ_{ii} and δ_{ih} represent the error made by

approximating posterior moments of the ratios r_{iD} in equations (3.14) and (3.15), respectively.

Dropping the error terms in (3.16) and (3.17) and then solving the resulting nonlinear system of equations for $1 \leq i \leq k$ and $i < h \leq k$ yields the Taylor-series approximate posterior covariance matrix with elements $\dot{\tilde{\sigma}}_{ii}$ and $\dot{\tilde{\sigma}}_{ih}$. Note that, as for the Taylor-series approximation for the posterior mean, the Taylor-series approximate posterior covariance matrix goes to the exact posterior covariance matrix as the percentage of incomplete data goes to zero.

Calculations for Taylor-Series Approximate Covariances: Thus, to solve the nonlinear system of equations for the Taylor-series approximate posterior covariance matrix, first note that for those categories that have only complete data,

$$\dot{\tilde{\sigma}}_{ii} = \tilde{p}_i(1-\tilde{p}_i)/(m+1) \tag{3.18}$$

and, for category h also having only complete data,

$$\dot{\tilde{\sigma}}_{ih} = -\tilde{p}_i\tilde{p}_h/(m+1), \tag{3.19}$$

in agreement with (2.3) and (2.4), respectively. Recall that $\dot{\tilde{p}}_i=\tilde{p}_i$ and $\dot{\tilde{p}}_h=\tilde{p}_h$ in this case of complete data.

For those categories i that have incomplete data, results are a noniterative estimate of $\tilde{\sigma}_{ih}$ for category h having only complete data and a choice of iterative and noniterative estimates for elements $\tilde{\sigma}_{ij}$ for category j, as well as i, having incomplete data.

For category h having only complete data and category i having incomplete data, we approximate $\text{cov}(p_i, p_h|\underset{\sim}{z})$ by

$$\dot{\tilde{\sigma}}_{ih} = -\dot{\tilde{p}}_i^{(t)} \tilde{p}_h/(m+1) \tag{3.20}$$

for $\dot{\tilde{p}}_i^{(t)}$ denoting the converged estimate $\dot{\tilde{p}}_i^{(s)}$ from (3.11). Approximation (3.20) is noniterative in $\dot{\tilde{\sigma}}_{ih}$.

For i and h referring to categories that have incomplete data and for s again denoting the number of iterations, we can write (3.16) and (3.17) as iterative algorithms. To do so, we drop the error terms δ_{ih} and write $\tilde{\sigma}_{ih}$ on the left-hand side of (3.16) and (3.17) as $\dot{\tilde{\sigma}}_{ih}^{(s+1)}$ and $\tilde{p}_i$ and $\tilde{\sigma}_{ih}$ on the right-hand side of these equations as $\dot{\tilde{p}}_i^{(t)}$ and $\dot{\tilde{\sigma}}_{ih}^{(s)}$, respectively, for $\dot{\tilde{p}}_i^{(t)}$ denoting the converged estimate from (3.11). These equations are given for the trinomial distribution in the next section.

To obtain initial estimates $\dot{\tilde{\sigma}}_{ii}^{(0)}$ and $\dot{\tilde{\sigma}}_{ih}^{(0)}$, we assume, for the first iteration only, that the ratio $\dot{\tilde{r}}_{iD} \equiv \dot{\tilde{p}}_i/\dot{\tilde{p}}_D$ is nonrandom. With this assumption, we have from (3.11), (3.14), and (3.15) that

$$\dot{\tilde{\sigma}}_{ii}^{(0)} = [\dot{\tilde{p}}_i^{(t)}(1-\dot{\tilde{p}}_i^{(t)}) + \sum_{D \ni i} (z_D/m)\, \dot{\tilde{r}}_{iD}^{(t)}(1-\dot{\tilde{r}}_{iD}^{(t)})]/(m+1) \tag{3.21}$$

and

$$\dot{\tilde{\sigma}}_{ih}^{(0)} = [\dot{\tilde{p}}_i^{(t)}\dot{\tilde{p}}_h^{(t)} + \sum_{D \ni i,h} (z_D/m)\, \dot{\tilde{r}}_{iD}^{(t)}\dot{\tilde{r}}_{hD}^{(t)}]/(m+1). \tag{3.22}$$

The second procedure for estimating elements of the posterior covariance matrix for those q categories that have incomplete data is noniterative in $\tilde{\sigma}_{ij}$. For both i and j referring to categories having incomplete data, $\tilde{a}_{lh}$ coefficients of $\tilde{\sigma}_{lh}$, and $\tilde{b}_{ij}$ a term that is not a function of $\tilde{\sigma}_{lh}$ for any l or h, we can write equations (3.16) and (3.17) as

$$\tilde{\sigma}_{ij} = \sum_{D \ni i} \sum_{Q \ni j} \left[\sum_{l \in D} \sum_{h \in Q} \tilde{a}_{lh} \tilde{\sigma}_{lh} \right] + \tilde{b}_{ij} + \delta_{ij}, \tag{3.23}$$

where we note that δ_{ij} also contains terms in $\tilde{\sigma}_{ij}$ [for example, second-order terms in the approximation for $E(r_{iD}|\underset{\sim}{z})$ are terms in $\tilde{\sigma}_{ii}$].

Thus, we can write (3.23) as a linear system of $q(q+1)/2$ equations in the $q(q+1)/2$ unknowns $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{ij}$:

$$[\underset{\sim}{\tilde{A}} + \underset{\sim}{\delta}_A] \, \underset{\sim}{\tilde{\sigma}} = \underset{\sim}{\tilde{B}} \, [\underset{\sim}{I} + \underset{\sim}{\delta}_B] \, \underset{\sim}{1} \tag{3.24}$$

where $\underset{\sim}{\tilde{\sigma}}$ is the $q(q+1)/2 \times 1$ vector of $\tilde{\sigma}_{ij}$ for both i and j referring to categories having incomplete data, $\underset{\sim}{\tilde{A}}$ is the $q(q+1)/2 \times q(q+1)/2$ matrix of the $\tilde{a}_{lh}$, $\underset{\sim}{\tilde{B}}$ is the $q(q+1)/2 \times q(q+1)/2$ matrix with $\tilde{b}_{ij}$ on the diagonal and 0's elsewhere, $\underset{\sim}{I}$ is the $q(q+1)/2 \times q(q+1)/2$ identity matrix, $\underset{\sim}{\delta}_A$ is the $q(q+1)/2 \times q(q+1)/2$ matrix containing those terms in δ_{ij} that are terms in $\underset{\sim}{\tilde{\sigma}}$, $\underset{\sim}{\delta}_B$ is the $q(q+1)/2 \times q(q+1)/2$ matrix containing zeros on the off-diagonal and the remaining terms of δ_{ij} divided by $\tilde{b}_{ij}$ on the diagonal, and $\underset{\sim}{1}$ is the $q(q+1)/2 \times 1$ vector containing all 1's.

The Taylor-series approximation $\dot{\tilde{\sigma}}_{ij}$ for these terms $\tilde{\sigma}_{ij}$ of the covariance matrix is then given from (3.24) by dropping the error terms $\underset{\sim}{\delta}_A$ and $\underset{\sim}{\delta}_B$; substituting the converged approximation $\dot{\tilde{p}}_i^{(t)}$ from (3.11) for $\tilde{p}_i$ in $\underset{\sim}{\tilde{A}}$ and $\underset{\sim}{\tilde{B}}$, yielding the matrices $\underset{\sim}{\dot{\tilde{A}}}$ and $\underset{\sim}{\dot{\tilde{B}}}$, respectively; and computing $\underset{\sim}{\dot{\tilde{\sigma}}}$ as

$$\underset{\sim}{\dot{\tilde{\sigma}}} = \underset{\sim}{\dot{\tilde{A}}}^{-1} \, \underset{\sim}{\dot{\tilde{B}}} \, \underset{\sim}{1}. \tag{3.25}$$

The tradeoff between the two procedures to approximate elements of the posterior covariance matrix for those categories that have incomplete data is the cost of the one-time expense of the larger-

dimensional operation (3.25) in the noniterative procedure versus the cost of iteratively evaluating the smaller-dimensional $[q(q+1)/2]\times 1$ covariance vector written directly from (3.16) and (3.17). In the next section we illustrate these Taylor-series approximations by writing them for the general case for the trinomial distribution. We conclude the chapter by giving a numerical example.

3.3 Algebraic Trinomial Illustration:

Suppose that, having taken minimization of expected risk as the criterion for choosing a point estimate of the posterior distribution of $\underset{\sim}{p}=(p_1,p_2,p_3)$ given incomplete trinomial data $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12},z_{13},z_{23})$, we want to calculate elements $\tilde{p}_1$, $\tilde{p}_2$, and $\tilde{p}_3=1-\tilde{p}_1-\tilde{p}_2$ of the posterior mean vector. Suppose also that, for the same estimation criterion, we have past estimates $\bar{p}_i$ calculated from a recent data sample of size $\bar{n}$ and prior parameters $\bar{\nu}_i$, whence we calculate new prior parameters

$$\nu_i=(\bar{n}+\sum_{j=1}^{k+1}\bar{\nu}_j)\bar{p}_i. \tag{3.26}$$

If we had no information other than $\underset{\sim}{z}$, we could set $\nu_i=1$ to obtain a uniform prior.

Recall that $\overset{*}{\tilde{p}}_{ij}^{(s)}=\overset{*}{\tilde{p}}_i^{(s)}+\overset{*}{\tilde{p}}_j^{(s)}$ and that $\overset{*}{\tilde{r}}_{ij}^{(s)}=\overset{*}{\tilde{p}}_i^{(s)}/\overset{*}{\tilde{p}}_{ij}^{(s)}$. Then, from (3.11) iterative estimates of elements of $\underset{\sim}{\tilde{p}}$ are given by

$$\overset{*}{\tilde{p}}_1^{(s+1)}=(z_1+\nu_1+z_{12}\overset{*}{\tilde{r}}_{12}^{(s)}+z_{13}\overset{*}{\tilde{r}}_{13}^{(s)})/m$$

and (3.27)

$$\overset{*}{\tilde{p}}_2^{(s+1)}=(z_2+\nu_2+z_{12}\overset{*}{\tilde{r}}_{21}^{(s)}+z_{23}\overset{*}{\tilde{r}}_{23}^{(s)})/m.$$

To choose an initial estimate $\overset{*}{\tilde{p}}_i^{(0)}$ to calculate $\overset{*}{\tilde{r}}_{ij}^{(0)}$ for (3.27), we use the previous estimate $\bar{p}_i$, theoretical results (such as from genetic or engineering laws), and/or current data. Then, calculating $\overset{*}{\tilde{r}}_{ij}^{(0)}$ for $1\leq i,j\leq 3$ and substituting results into the right-hand side of (3.27), we iterate on (3.27) until results converge.

To estimate the posterior covariance matrix, we have from (3.16) and (3.17) that

$$\tilde{\sigma}_{11} = (z_{12}/m)[(z_{12}-1)/(m+1)]\{\tilde{p}_2^2\tilde{\sigma}_{11}+\tilde{p}_1[\tilde{p}_1\tilde{\sigma}_{22}-2\tilde{p}_2\tilde{\sigma}_{12}]\}/\tilde{p}_{12}^4$$

$$+(z_{13}/m)[(z_{13}-1)/(m+1)]\{\tilde{p}_3^2\tilde{\sigma}_{11}+\tilde{p}_1[\tilde{p}_1\tilde{\sigma}_{33}-2\tilde{p}_3\tilde{\sigma}_{13}]\}/\tilde{p}_{13}^4$$

$$+2(z_{12}/m)[z_{13}/(m+1)]\{\tilde{p}_2[\tilde{p}_3\tilde{\sigma}_{11}-\tilde{p}_1\tilde{\sigma}_{13}]+\tilde{p}_1[\tilde{p}_1\tilde{\sigma}_{23}-\tilde{p}_3\tilde{\sigma}_{12}]\}/(\tilde{p}_{12}\tilde{p}_{13})^2$$

$$+\{(z_{12}/m)\tilde{p}_1\tilde{p}_2/(\tilde{p}_{12})^2+(z_{13}/m)\tilde{p}_1\tilde{p}_3/(\tilde{p}_{13})^2 + \tilde{p}_1(1-\tilde{p}_1)\}/(m+1) + \delta_{11},$$

$$\tilde{\sigma}_{12} = z_{12}^2/[m(m+1)]\{-\tilde{p}_2^2\tilde{\sigma}_{11}+2\tilde{p}_1\tilde{p}_2\tilde{\sigma}_{12}-\tilde{p}_1^2\tilde{\sigma}_{22}\}/\tilde{p}_{12}^4$$

$$+z_{12}z_{23}/[m(m+1)]\{\tilde{p}_2[\tilde{p}_3\tilde{\sigma}_{12}-\tilde{p}_2\tilde{\sigma}_{13}]+\tilde{p}_1[\tilde{p}_2\tilde{\sigma}_{23}-\tilde{p}_3\tilde{\sigma}_{22}]\}/(\tilde{p}_{12}\tilde{p}_{23})^2$$

$$+z_{12}z_{13}/[m(m+1)]\{\tilde{p}_3[\tilde{p}_1\tilde{\sigma}_{12}-\tilde{p}_2\tilde{\sigma}_{11}]+\tilde{p}_1[\tilde{p}_2\tilde{\sigma}_{13}-\tilde{p}_1\tilde{\sigma}_{23}]\}/(\tilde{p}_{12}\tilde{p}_{13})^2 \qquad (3.28)$$

$$+z_{13}z_{23}/[m(m+1)]\{\tilde{p}_3[\tilde{p}_3\tilde{\sigma}_{12}-\tilde{p}_2\tilde{\sigma}_{13}]+\tilde{p}_1[\tilde{p}_2\tilde{\sigma}_{33}-\tilde{p}_3\tilde{\sigma}_{23}]\}/(\tilde{p}_{13}\tilde{p}_{23})^2$$

$$-\{(z_{12}/m)[\tilde{p}_1\tilde{p}_2\tilde{p}_{12}^2-\tilde{p}_2^2\tilde{\sigma}_{11}+2\tilde{p}_1\tilde{p}_2\tilde{\sigma}_{12}-\tilde{p}_1^2\tilde{\sigma}_{22}]/\tilde{p}_{12}^4 + \tilde{p}_1\tilde{p}_2\}/(m+1) + \delta_{12},$$

$$\tilde{\sigma}_{22} = (z_{12}/m)[(z_{12}-1)/(m+1)]\{\tilde{p}_1^2\tilde{\sigma}_{22}+\tilde{p}_2[\tilde{p}_2\tilde{\sigma}_{11}-2\tilde{p}_1\tilde{\sigma}_{12}]\}/\tilde{p}_{12}^4$$

$$+(z_{23}/m)[(z_{23}-1)/(m+1)]\{\tilde{p}_3^2\tilde{\sigma}_{22}+\tilde{p}_2[\tilde{p}_2\tilde{\sigma}_{33}-2\tilde{p}_3\tilde{\sigma}_{23}]\}/\tilde{p}_{23}^4$$

$$+2(z_{12}/m)[z_{23}/(m+1)]\{\tilde{p}_1[\tilde{p}_3\tilde{\sigma}_{22}-\tilde{p}_2\tilde{\sigma}_{23}]+\tilde{p}_2[\tilde{p}_2\tilde{\sigma}_{23}-\tilde{p}_3\tilde{\sigma}_{12}]\}/(\tilde{p}_{12}\tilde{p}_{23})^2$$

$$+\{(z_{12}/m)\tilde{p}_1\tilde{p}_2/\tilde{p}_{12}^2+(z_{23}/m)\tilde{p}_2\tilde{p}_3/\tilde{p}_{23}^2 + \tilde{p}_2(1-\tilde{p}_2)\}/(m+1) + \delta_{22}.$$

To estimate the posterior covariance matrix by the iterative procedure, we iterate on (3.27) until the convergence condition is met on, say, the t^{th} iteration. Then, for $f_i=\dot{\tilde{p}}_i^{(t)}$, $f_{ij}=\dot{\tilde{p}}_{ij}^{(t)}$, and $g_{ij}=\dot{\tilde{r}}_{ij}^{(t)}$, we rewrite (3.28) as

$$\dot{\tilde{\sigma}}_{11}^{(s+1)} = z_{12}^2/[m(m+1)]\{f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} - 2f_{12}\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/f_{12}^4$$

$$+z_{13}^2/[m(m+1)]\{f_{13}^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_1f_{13}\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/f_{13}^4$$

$$+2z_{12}z_{13}/[m(m+1)]\{f_2f_{13}\dot{\tilde{\sigma}}_{11}^{(s)} + f_1[f_2 - f_{13}]\dot{\tilde{\sigma}}_{12}^{(s)} - f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/(f_{12}f_{13})^2$$

$$-\{(z_{12}/m)[f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} - 2f_1f_2\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)} - f_1f_2f_{12}^2]/f_{12}^4$$

$$+(z_{13}/m)[f_{13}^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_1f_{13}\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)} - f_1f_3f_{13}^2]/f_{13}^4 - f_1(1-f_1)\}/(m+1),$$

$$\dot{\tilde{\sigma}}_{12}^{(s+1)} = z_{12}^2/[m(m+1)]\{-f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_1f_2\dot{\tilde{\sigma}}_{12}^{(s)} - f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/f_{12}^4$$

$$+z_{12}z_{13}/[m(m+1)]\{-f_2f_{13}\dot{\tilde{\sigma}}_{11}^{(s)} + f_1[f_{13} - f_2]\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/(f_{12}f_{13})^2$$

$$+z_{12}z_{23}/[m(m+1)]\{-f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + f_2[f_{23} - f_1]\dot{\tilde{\sigma}}_{12}^{(s)} - f_1f_{23}\dot{\tilde{\sigma}}_{22}\}/(f_{12}f_{23})^2 \qquad (3.29)$$

$$+z_{13}z_{23}/[m(m+1)]\{f_2f_{13}\dot{\tilde{\sigma}}_{11}^{(s)} + [f_3 + 2f_1f_2]\dot{\tilde{\sigma}}_{12}^{(s)} + f_1f_{23}\dot{\tilde{\sigma}}_{22}^{(s)}\}/(f_{13}f_{23})^2$$

$$-\{(z_{12}/m)[-f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_1f_2\dot{\tilde{\sigma}}_{12}^{(s)} - f_1^2\dot{\tilde{\sigma}}_{22}^{(s)} + f_1f_2f_{12}^2]/f_{12}^4 + f_1f_2\}/(m+1),$$

$$\dot{\tilde{\sigma}}_{22}^{(s+1)} = z_{12}^2/[m(m+1)]\{f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} - 2f_1f_2\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/f_{12}^4$$

$$+z_{23}^2/[m(m+1)]\{f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_2f_{23}\dot{\tilde{\sigma}}_{12}^{(s)} + f_{23}^2\dot{\tilde{\sigma}}_{22}^{(s)}\}/f_{23}^4$$

$$+2z_{12}z_{23}/[m(m+1)]\{-f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + f_2[f_1 - f_{23}]\dot{\tilde{\sigma}}_{12}^{(s)} + f_1f_{23}\dot{\tilde{\sigma}}_{22}^{(s)}\}/(f_{12}f_{23})^2$$

$$-\{(z_{12}/m)[f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} - 2f_1f_2\dot{\tilde{\sigma}}_{12}^{(s)} + f_1^2\dot{\tilde{\sigma}}_{22}^{(s)} - f_1f_2f_{12}^2]/f_{12}^4$$

$$+(z_{23}/m)[f_2^2\dot{\tilde{\sigma}}_{11}^{(s)} + 2f_2f_{23}\dot{\tilde{\sigma}}_{12}^{(s)} + f_{23}^2\dot{\tilde{\sigma}}_{22}^{(s)} - f_2f_3f_{23}^2]/f_{23}^4 - f_2(1-f_2)\}/(m+1).$$

where we calculate initial estimates $\dot{\tilde{\sigma}}_{ii}^{(0)}$ and $\dot{\tilde{\sigma}}_{ij}^{(0)}$ from (3.21) and (3.22) as

$$\overset{*}{\sigma}_{11}^{(0)} = [f_1(1-f_1)+(z_{12}/m)g_{12}(1-g_{12})+(z_{13}/m)g_{13}(1-g_{13})]/(m+1),$$

$$\overset{*}{\sigma}_{12}^{(0)} = -[f_1f_2+(z_{12}/m)g_{12}g_{21}]/(m+1) \qquad (3.30)$$

$$\overset{*}{\sigma}_{22}^{(0)} = [f_2(1-f_2)+(z_{12}/m)g_{21}(1-g_{21})+(z_{23}/m)g_{23}(1-g_{23})]/(m+1).$$

After evaluating (3.30) we iterate on (3.29) until results converge.

To estimate the posterior covariance matrix by the noniterative procedure, we substitute in equations (3.28) for $\tilde{\sigma}_{13}$, $\tilde{\sigma}_{23}$, and $\tilde{\sigma}_{33}$ in terms of $\tilde{\sigma}_{11}$, $\tilde{\sigma}_{12}$, and $\tilde{\sigma}_{22}$, collect terms, and rewrite equations (3.28) as

$$\begin{aligned}
&-[f_1(1-f_1)+(z_{12}/m)g_{12}g_{21}+(z_{13}/m)g_{13}g_{31}]/(m+1) + \tau_{11} \\
&\quad = \tilde{\sigma}_{11}\{-1+[(z_{12}g_{21}/f_{12}+z_{13}/f_{13})^2-z_{12}(g_{21}/f_{12})^2-z_{13}/f_{13}^2]/[m(m+1)]\} \\
&\qquad +2\tilde{\sigma}_{12}[(z_{13}g_{13}/f_{13}-z_{12}g_{12}/f_{12})(z_{12}g_{21}/f_{12}+z_{13}/f_{13})+z_{12}g_{12}g_{21}/f_{12}^2 \\
&\qquad\quad -z_{13}g_{13}/f_{13}^2]/[m(m+1)] \\
&\qquad +\tilde{\sigma}_{22}[(z_{12}g_{12}/f_{12}-z_{13}g_{13}/f_{13})^2- z_{12}(g_{12}/f_{12})^2+z_{13}(g_{13}/f_{13})^2] /[m(m+1)],
\end{aligned}$$

$$\begin{aligned}
&f_1f_2[1+z_{12}/(mf_{12}^2)]/(m+1) + \tau_{12} \\
&\quad = \tilde{\sigma}_{11}[(z_{12}g_{21}/f_{12}+z_{13}/f_{13})(z_{23}g_{23}/f_{23}-z_{12}g_{21}/f_{12})+z_{12}(g_{21}/f_{12})^2]/[m(m+1)] \\
&\qquad +\tilde{\sigma}_{12}\{-1+[(z_{12}f_1/f_{12}^2+z_{23}(1-2f_1)/f_{23}^2)(z_{12}f_2/f_{12}^2+z_{13}(1-2f_2)/f_{13}^2) \\
&\qquad\quad + z_{12}(z_{12}-2)g_{12}g_{21}/f_{12}^2+z_{13}z_{23}(f_{12}-2f_1f_2)/(f_{13}f_{23})^2]/[m(m+1)]\} \\
&\qquad +\tilde{\sigma}_{22}[(z_{12}g_{12}/f_{12}+z_{23}/f_{23})(z_{13}g_{13}/f_{13}-z_{12}g_{12}/f_{12})+z_{12}(g_{12}/f_{12})^2]/[m(m+1)],
\end{aligned} \qquad (3.31)$$

and

$$-[f_2(1-f_2)+(z_{12}/m)g_{12}g_{21}+(z_{23}/m)g_{23}g_{32}]/(m+1) + \tau_{22}$$

$$= \tilde{\sigma}_{11}[(z_{12}g_{21}/f_{12}-z_{23}g_{23}/f_{23})^2-z_{12}(g_{21}/f_{12})^2-z_{23}(g_{23}/f_{23})^2]/[m(m+1)]$$

$$+2\tilde{\sigma}_{12}[(-z_{12}g_{12}/f_{12}-z_{23}/f_{23})(z_{12}g_{21}/f_{12}-z_{23}g_{23}/f_{23})+z_{12}g_{12}g_{21}/f_{12}^2$$

$$-z_{23}g_{23}/f_{23}^2]/[m(m+1)]$$

$$+\tilde{\sigma}_{22}\{-1+[(z_{12}g_{12}/f_{12}+z_{23}/f_{23})^2-z_{12}(g_{12}/f_{12})^2-z_{23}/f_{23}^2]/[m(m+1)]\},$$

for $\tau_{ij}=\delta_{ij}$ plus the error made from approximating $\tilde{p}_i$ by $f_i \equiv \dot{\tilde{p}}_i^{(t)}$.

Dropping the error terms τ_{ij}, we have that equations (3.31) are three equations linear in the approximations $\dot{\tilde{\sigma}}_{11}$, $\dot{\tilde{\sigma}}_{12}$, and $\dot{\tilde{\sigma}}_{22}$ of the posterior covariances $\tilde{\sigma}_{11}$, $\tilde{\sigma}_{12}$, and $\tilde{\sigma}_{22}$, respectively. That is, we approximate elements of the posterior covariance matrix by

$$\dot{\underset{\sim}{\tilde{\sigma}}} = \dot{\underset{\sim}{A}}^{-1} \dot{\underset{\sim}{B}} \tag{3.32}$$

where $\dot{\underset{\sim}{\tilde{\sigma}}}=(\dot{\tilde{\sigma}}_{11},\dot{\tilde{\sigma}}_{12},\dot{\tilde{\sigma}}_{22})$, $\dot{\underset{\sim}{A}}$ is the 3×3 coefficient matrix of $\dot{\underset{\sim}{\tilde{\sigma}}}$ from the right-hand side of (3.31), and $\dot{\underset{\sim}{B}}$ is the 3×1 column vector of constants given on the left-hand side of (3.31).

3.4 Numerical Example:

We now compare the Taylor-series approximations for elements of the posterior mean and covariance matrices with approximations given by the maximum likelihood estimate and by the posterior mode on a small sample. We use the example given in Section 2.2.3. Initial estimates for the iterative algorithms were $\dot{p}_1^{(0)}=1/4$ and $\dot{p}_2^{(0)}=\dot{p}_3^{(0)}=3/8$. The condition for convergence was that the absolute relative difference $|\dot{p}_i^{(s+1)}-\dot{p}_i^{(s)}|/\dot{p}_i^{(s)}$ be less than 0.001 where $\dot{p}_i^{(s)}$ denotes the s^{th} iteration of approximation $\dot{p}_i$. Because a uniform prior was used, the posterior mode equals the maximum likelihood estimate.

Results from these approximations are given in the following Table 3.1. The Taylor-series approximations are by far the better approximations for elements of the posterior mean and covariance matrices. Further, they are excellent approximations for such a small sample. For example, values of the Taylor-series approximate posterior mean differ from the three corresponding elements of the exact posterior mean by only 0.3%, 0.1%, and 0.1% in percentage absolute relative difference $100\times|\dot{\tilde{p}}_i-\tilde{p}_i|/\tilde{p}_i$. Corresponding percentage absolute relative differences for the maximum likelihood esimate ($\equiv$ posterior mode) are 9.7%, 3.8%, and 2.4%, respectively.

TABLE 3.1

COMPARISON OF APPROXIMATIONS FOR ELEMENTS OF POSTERIOR MEAN AND COVARIANCE MATRICES FOR EXAMPLE 2.2.3

Approximation / Moment	Exact Posterior Moments	Taylor-Series Approximated Posterior Moments	Maximum Likelihood Solution*
$E(p_1\|\underset{\sim}{z})$	0.241202	0.242103	0.217718
$E(p_2\|\underset{\sim}{z})$	.384927	.384484	.399394
$E(p_3\|\underset{\sim}{z})$	.373871	.373273	.382888
$var(p_1\|\underset{\sim}{z})$	.011921	.011956	.015703
$var(p_2\|\underset{\sim}{z})$	.012725	.012772	.016648
$var(p_3\|\underset{\sim}{z})$	.011204	.011182	.014062
$100\times C.V.(p_1\|\underset{\sim}{z})$	45.27	45.16	57.56
$100\times C.V.(p_2\|\underset{\sim}{z})$	29.31	29.39	32.31
$100\times C.V.(p_3\|\underset{\sim}{z})$	28.31	28.36	30.97
$cov(p_1,p_2\|\underset{\sim}{z})$	-.006721	-.006773	-.009144

*Since a uniform prior is used, also equals posterior mode

3.5 Summary:

In this chapter we considered three different approximations for elements of the posterior mean vector and one approximation for elements of the posterior covariance matrix. The maximum likelihood estimate (2.36) and posterior mode (2.43) were considered because they asymptotically equal the limiting posterior mean. However, as discussed in the first section, there are problems with using these two estimates to approximate the posterior mean. We then derived an approximation by conditioning twice from the complete-data posterior mean and using Taylor-series expansions for the unknown terms. An important property of the resulting Taylor-series approximation is that as the percentage of incomplete data goes to zero, the approximation goes to the exact posterior mean. Neither the maximum likelihood estimate nor the posterior mode has this property. The Taylor-series approximation also relates to the posterior mode (2.43) in the same manner that the complete-data posterior mean relates to the complete-data posterior mode. Because the Taylor-series approximation is thus a posterior mode (for $\beta_i=\nu_i+1$), we were able to solve its nonlinear system of equations by the EM algorithm discussed in Section 2.3.2. Approximations for the posterior mean and their complete-data counterparts are given in the following Table 3.2.

The same approach of conditioning and using Taylor-series expansions was also used to derive approximations for elements of the posterior covariance matrix. The resulting approximations also have the important property that as the percentage of incomplete data goes to zero, the approximations go to the exact elements of the posterior covariance matrix. We showed how to solve the system of equations from the approximations either

iteratively or noniteratively, once the posterior mean has been approximated.

We illustrated the Taylor-series approximations algebraically for the trinomial distribution and then compared them numerically with the maximum likelihood estimate and the posterior mode for a uniform prior on a small sample.

TABLE 3.2

APPROXIMATIONS FOR EXACT POSTERIOR MEAN

	complete data	incomplete data
exact posterior mean	$(x_i+\nu_i)/(n+\sum_{j=1}^{k+1}\nu_j)$	$\sum_{l=1}^{\omega} c_l\Gamma(\gamma_{il}+1)\prod_{j\neq i}^{k+1}\Gamma(\gamma_{jl})/[(n+\Sigma\nu_j)\sum_{l=1}^{\omega} c_l\prod_{j=1}^{k+1}\Gamma(\gamma_{jl})]$
Taylor-series approximation	$(x_i+\nu_i)/(n+\sum_{j=1}^{k+1}\nu_j)$	$(z_i+\nu_i+\sum_{D\ni i} z_D\mathring{p}_i/\mathring{p}_D)/(n+\Sigma\nu_j)$
posterior mode	$(x_i+\nu_i-1)/[n+\sum_{j=1}^{k+1}\nu_j-(k+1)]$	$(z_i+\nu_i-1+\sum_{D\ni i} z_D\hat{\tilde{p}}_i/\hat{\tilde{p}}_D)/[n+\Sigma\nu_j-(k+1)]$
maximum likelihood estimate	x_i/n	$(z_i+\sum_{D\ni i} z_D\hat{p}_i/\hat{p}_D)/n$

APPENDIX 3A

CONDITIONAL DERIVATIONS FOR TAYLOR-SERIES APPROXIMATIONS FOR POSTERIOR COVARIANCE MATRIX

3A.1 Posterior Variance:

For $1 \leq i \leq k$ and Q, like D, denoting $\not\ni$ containing more than one element,

$$
\begin{aligned}
\operatorname{var}(p_i|\underset{\sim}{z}) &= \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}[\operatorname{var}(p_i|\underset{\sim}{z},\underset{\sim}{u})] + \underset{\underset{\sim}{u}|\underset{\sim}{z}}{\operatorname{var}}[E(p_i|\underset{\sim}{z},\underset{\sim}{u})] \\
&= \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}\{[m^2(m+1)]^{-1}[z_{\{i\}}+\sum_{D\ni i} z_D^{(i)}+\nu_i][m-(z_{\{i\}} \\
&\quad + \sum_{D\ni i} z_D^{(i)}+\nu_i)]\}+m^{-2}\underset{\underset{\sim}{u}|\underset{\sim}{z}}{\operatorname{var}}[z_{\{i\}}+\sum_{D\ni i} z_D^{(i)}+\nu_i] \\
&= \{[m-(z_{\{i\}}+\nu_i)][z_{\{i\}}+\nu_i+\sum_{D\ni i} E(z_D^{(i)}|\underset{\sim}{z})]-(z_{\{i\}}+\nu_i) \\
&\quad \times \sum_{D\ni i} E(z_D^{(i)}|\underset{\sim}{z})-E[(\sum_{D\ni i} z_D^{(i)}|\underset{\sim}{z})^2] \\
&\quad +(m+1)\operatorname{var}[\sum_{D\ni i} z_D^{(i)}|\underset{\sim}{z}]\}/[m^2(m+1)] \qquad (3A.1) \\
&= \Big(\{m-[z_{\{i\}}+\nu_i+\sum_{D\ni i} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z})]\}\{z_{\{i\}}+\nu_i+ \\
&\quad \sum_{D\ni i} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z})\} - \sum_{D\ni i}\{z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[r_{iD}(1-r_{iD})|\underset{\sim}{z}] \\
&\quad +z_D^2 \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{var}}(r_{iD}|\underset{\sim}{z}) + \sum_{\substack{Q\ni i \\ Q\neq D}} z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{cov}}(r_{iD},r_{iQ}|\underset{\sim}{z})\} \\
&\quad +(m+1)\sum_{D\ni i} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[r_{iD}(1-r_{iD})|\underset{\sim}{z}]+z_D^2 \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{var}}(r_{iD}|\underset{\sim}{z}) \\
&\quad + \sum_{\substack{Q\ni i \\ Q\neq D}} z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{cov}}(r_{iD},r_{iQ}|\underset{\sim}{z})\}\Big)/[m^2(m+1)],
\end{aligned}
$$

because

$$E(z_D^{(i)}|\underset{\sim}{z}) = \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_D^{(i)}|\underset{\sim}{p},\underset{\sim}{z})] = z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z}) \tag{3A.2}$$

and

$$\begin{aligned}
E[(\sum_{D\ni i} z_D^{(i)}|\underset{\sim}{z})^2] &= \sum_{D\ni i}\{E[(z_D^{(i)}|\underset{\sim}{z})^2] + \sum_{\substack{Q\ni i\\ Q\neq D}} E(z_D^{(i)} z_Q^{(i)}|\underset{\sim}{z})\} \\
&= \sum_{D\ni i}\Big(\mathrm{var}(z_D^{(i)}|\underset{\sim}{z}) + [E(z_D^{(i)}|\underset{\sim}{z})]^2 \\
&\quad + \sum_{\substack{Q\ni i\\ Q\neq D}} \{\mathrm{cov}(z_D^{(i)}, z_Q^{(i)}|\underset{\sim}{z}) + [E(z_D^{(i)}|\underset{\sim}{z})][E(z_Q^{(i)}|\underset{\sim}{z})]\}\Big) \\
&= \sum_{D\ni i}\Big[\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[\mathrm{var}(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] + \mathrm{var}_{\underset{\sim}{p}|\underset{\sim}{z}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\quad + \{\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})]\}^2 \\
&\quad + \sum_{\substack{Q\ni i\\ Q\neq D}} \Big(\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[\mathrm{cov}(z_D^{(i)}, z_Q^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\qquad + \mathrm{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p}), E(z_Q^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\qquad + \{\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})]\}\{\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_Q^{(i)}|\underset{\sim}{z},\underset{\sim}{p})]\}\Big)\Big] \qquad (3A.3) \\
&= \sum_{D\ni i}\{z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[r_{iD}(1-r_{iD})|\underset{\sim}{z}] + z_D^2 \mathrm{var}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z}) + z_D^2[\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z})]^2 \\
&\quad + \sum_{\substack{Q\ni i\\ Q\neq D}} [z_D z_Q \mathrm{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}, r_{iQ}|\underset{\sim}{z}) + z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z}) \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iQ}|\underset{\sim}{z})]\} \\
&= \sum_{D\ni i}\{z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[r_{iD}(1-r_{iD})|\underset{\sim}{z}] + z_D^2 \mathrm{var}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z})
\end{aligned}$$

$$+ \sum_{\substack{Q \ni i \\ Q \neq D}} z_D z_Q \operatorname*{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}, r_{iQ}|\underset{\sim}{z})\} + [\sum_{D \ni i} z_D \operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z})]^2$$

and

$$\begin{aligned}
\operatorname{var}(\sum_{D \ni i} z_D^{(i)}|\underset{\sim}{z}) &= \sum_{D \ni i}[\operatorname{var}(z_D^{(i)}|\underset{\sim}{z}) + \sum_{\substack{Q \ni i \\ Q \neq D}} \operatorname{cov}(z_D^{(i)}, z_Q^{(i)}|\underset{\sim}{z})] \\
&= \sum_{D \ni i}\Big(\operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}[\operatorname{var}(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] + \operatorname*{var}_{\underset{\sim}{p}|\underset{\sim}{z}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\qquad + \sum_{\substack{Q \ni i \\ Q \neq D}} \{ \operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}[\operatorname{cov}(z_D^{(i)}, z_Q^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\qquad + \operatorname*{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p}), E(z_Q^{(i)}|\underset{\sim}{z},\underset{\sim}{p})]\}\Big) \\
&\quad + \sum_{D \ni i}\{z_D \operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}[r_{iD}(1-r_{iD})|\underset{\sim}{z}] + z_D^2 \operatorname*{var}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z}) \\
&\qquad + \sum_{\substack{Q \ni i \\ Q \neq D}} z_D z_Q \operatorname*{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}, r_{iQ}|\underset{\sim}{z})\}.
\end{aligned} \tag{3A.4}$$

since $\operatorname{cov}(z_D^{(i)}, z_Q^{(i)}) = 0$ for $Q \neq D$.

Therefore, combining terms in (3A.1) and recalling (3.6),

$$\begin{aligned}
\operatorname{var}(p_i|\underset{\sim}{z}) &= \sum_{D \ni i}\{z_D^2/[m(m+1)] \operatorname*{var}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z}) + \sum_{\substack{Q \ni i \\ Q \neq D}} z_D z_Q/[m(m+1)] \operatorname*{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}, r_{iQ}|\underset{\sim}{z})\} \\
&\quad + \{E(p_i|\underset{\sim}{z})[1-E(p_i|\underset{\sim}{z})] + \sum_{D \ni i}(z_D/m) \operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}[r_{iD}(1-r_{iD})|\underset{\sim}{z}]\}/(m+1).
\end{aligned} \tag{3A.5}$$

3A.2 Posterior Covariance

Finally, for $1 \leq i,h \leq k, i \neq h$, and Q defined as in 4A.1, we have that

$$\begin{aligned}
\operatorname{cov}(p_i,p_h|\underset{\sim}{z}) &= \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}[\operatorname{cov}(p_i,p_h|\underset{\sim}{z},\underset{\sim}{u})] + \underset{\underset{\sim}{u}|\underset{\sim}{z}}{\operatorname{cov}}[E(p_i|\underset{\sim}{z},\underset{\sim}{u}),E(p_h|\underset{\sim}{z},\underset{\sim}{u})] \\
&= -[m^2(m+1)]^{-1} \underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}\{[z_{\{i\}} + \sum_{D \ni i} z_D^{(i)} + \nu_i][z_{\{h\}} + \sum_{Q \ni h} z_Q^{(h)} + \nu_h] \\
&\quad + m^{-2} \underset{\underset{\sim}{u}|\underset{\sim}{z}}{\operatorname{cov}}[z_{\{i\}} + \sum_{D \ni i} z_D^{(i)} + \nu_i, z_{\{h\}} + \sum_{Q \ni h} z_Q^{(h)} + \nu_h] \qquad (3A.6) \\
&= \{(z_{\{h\}}+\nu_h) \sum_{D \ni i} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z}) + (z_{\{i\}}+\nu_i)[z_{\{h\}}+\nu_h + \sum_{Q \ni h} z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{hQ}|\underset{\sim}{z})] \\
&\quad - \sum_{D \ni i,h} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}r_{ih}|\underset{\sim}{z}) + \sum_{D \ni i} \sum_{Q \ni h} z_D z_Q [\underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{cov}}(r_{iD},r_{hQ}|\underset{\sim}{z}) \\
&\quad + \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z}) \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{hQ}|\underset{\sim}{z})] - (m+1)[\sum_{D \ni i} \sum_{Q \ni h} z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{cov}}(r_{iD},r_{hQ}|\underset{\sim}{z}) \\
&\quad - \sum_{D \ni i,h} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}r_{hD}|\underset{\sim}{z})]\}
\end{aligned}$$

because

$$\begin{aligned}
\underset{\underset{\sim}{u}|\underset{\sim}{z}}{E}[(\sum_{D \ni i} z_D^{(i)})(\sum_{Q \ni h} z_Q^{(h)})] &= \sum_{D \ni i} \sum_{Q \ni h} E(z_D^{(i)} z_Q^{(h)}|\underset{\sim}{z}) \\
&= \sum_{D \ni i} \sum_{Q \ni h} \{ \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[\operatorname{cov}(z_D^{(i)},z_Q^{(h)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&\qquad + \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\operatorname{cov}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p}),E(z_Q^{(h)}|\underset{\sim}{z},\underset{\sim}{p})] \qquad (3A.7) \\
&\qquad + \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p})] \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[E(z_Q^{(h)}|\underset{\sim}{z},\underset{\sim}{p})] \\
&= -\sum_{D \ni i,h} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}r_{hD}|\underset{\sim}{z}) + \sum_{D \ni i} \sum_{Q \ni h} z_D z_Q
\end{aligned}$$

$$\times \{\underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{cov}}(r_{iD},r_{hQ}|\underset{\sim}{z}) + \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}|\underset{\sim}{z})\,\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{hQ}|\underset{\sim}{z})\}$$

and

$$\underset{\underset{\sim}{u}|\underset{\sim}{z}}{\mathrm{cov}}\Big(\sum_{D\ni i} z_D^{(i)}, \sum_{Q\ni h} z_Q^{(h)}\Big) = \sum_{D\ni i}\sum_{Q\ni h}\{\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}[\mathrm{cov}(z_D^{(i)},z_Q^{(h)}|\underset{\sim}{z},\underset{\sim}{p})] + \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{cov}}[E(z_D^{(i)}|\underset{\sim}{z},\underset{\sim}{p}),E(z_Q^{(h)}|\underset{\sim}{z},\underset{\sim}{p})]\} \quad (3A.8)$$

$$= -\sum_{D\ni i,h} z_D \underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}r_{hD}|\underset{\sim}{z}) + \sum_{D\ni i}\sum_{Q\ni h} z_D z_Q \underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{cov}}(r_{iD},r_{hQ}|\underset{\sim}{z}).$$

Therefore, combining terms in (3A.6) and recalling (3.6),

$$\mathrm{cov}(p_i,p_h|\underset{\sim}{z}) = \sum_{D\ni i}\sum_{Q\ni h}(z_D/m)[z_Q/(m+1)]\underset{\underset{\sim}{p}|\underset{\sim}{z}}{\mathrm{cov}}(r_{iD},r_{hQ}|\underset{\sim}{z}) - [\sum_{D\ni i,h}(z_D/m)\underset{\underset{\sim}{p}|\underset{\sim}{z}}{E}(r_{iD}r_{hD}|\underset{\sim}{z}) + E(p_i|\underset{\sim}{z})E(p_h|\underset{\sim}{z})]/(m+1). \quad (3A.9)$$

APPENDIX 3B

APPROXIMATIONS OF RATIOS AND THEIR MOMENTS

3B.1 Introduction

For $i\in D$, consider the ratio $r_{iD}=p_i/p_D$ for $p_D=\sum_{j\in D} p_j$. Define $e_j=[p_j-E(p_j|\underset{\sim}{z})]|\underset{\sim}{z}$ and let $\underset{\sim}{e}_D$ be the vector of e_j for $j\in D$. Let $E(\underset{\sim}{p}|\underset{\sim}{z}) = (E(p_1|\underset{\sim}{z}), \ldots, E(p_k|\underset{\sim}{z}))$. Define $\emptyset$ to be D minus the integer i. Let $d_D(w_i, \prod_{j\in\emptyset} w_j,)$ be a vector function of dimension equal to the number of integers in D and be defined by

$$d_D(w_i, \prod_{j\in\emptyset} w_j,) = w_i/\sum_{j\in D} w_j \tag{3B.1}$$

[Thus, for $D=\{1,2,3\}$ and $i=1$, $\emptyset=\{2,3\}$ and $d_D(w_1, \prod_{j\in\emptyset} w_j,) = d_D(w_1,w_2,w_3) = w_1/(w_1+w_2+w_3)$]. Then, for $\ell\in D$,

$$\partial d_D(w_i, \prod_{j\in\emptyset} w_j,)/\partial w_\ell = \begin{cases} \sum_{j\in\emptyset} w_j/(\sum_{j\in D} w_j)^2 & \text{for } \ell=i \\ -w_i/(\sum_{j\in D} w_j)^2 & \text{for } \ell\neq i \end{cases} \tag{3B.2}$$

To characterize errors in the ratio approximations, we define the Landau symbols O and o and their stochastic parallels O_p and o_p. [See Bishop, Fienberg, and Holland (1975,chpt.14), Cox and Hinkley (1974, chpt.9), Cramer (1951,chpt.12), and Schmetterer (1974,p17).] Let $\|\underset{\sim}{y}\|$ denote the length $(\sum_{i=1}^{k+1} y_i^2)^{\frac{1}{2}}$ of the k-dimensional vector $\underset{\sim}{y}$.

Definition 3B.1: For $\{a_n\}$ a sequence of real numbers or vectors and $\{b_n\}$ a sequence of positive real numbers

a. $a_n=O(b_n)$ if there exists a number K and an integer n(K) such that if n exceeds n(K) then $\|a_n\| < Kb_n$;

b. $a_n=o(b_n)$ if for every $\varepsilon>0$, there exists an integer $n(\varepsilon)$ such that if n exceeds $n(\varepsilon)$ then $\|a_n\| < \varepsilon b_n$.

Definition 3B.2: For a(x) and b(x) continuous functions of the real number or vector x.

a. $a(x)=O(b(x))$ as $x\longrightarrow y$ if, for any sequence $\{x_n\}$ such that $x_n\longrightarrow y$, $a(x_n)=O(b(x_n))$;

b. $a(x)=o(b(x))$ as $x\longrightarrow y$ if, for any sequence $\{x_n\}$ such that $x_n\longrightarrow y$, $a(x_n)=o(b(x_n))$.

Definition 3B.3: For random variable, or vector of random variables, V_n and sequence $\{a_n\}$ of positive real numbers

a. $V_n=O_p(a_n)$ if for every $\eta>0$ there exists a constant $K(\eta)$ and an integer $n(\eta)$ such that if $n\geq n(\eta)$, then $P\{\|V_n\|/a_n\leq K(\eta)\}\geq 1-\eta$;

b. $V_n=o_p(a_n)$ if for every $\varepsilon>0$, $\lim_{n\to\infty} P\{\|V_n\|/a_n\leq\varepsilon\} = 1$.

Lemma 3B.1: For O, o, O_p, and o_p as just defined:

a. For the nonzero constant c, $O(cx_n)=O(x_n)$ and $o(cx_n)=o(x_n)$.

b. $O(o(x_n))=o(x_n)$; $o(O(x_n))=o(x_n)$; $O(O(x_n))=O(x_n)$; and $o(o(x_n))=o(x_n)$;

c. $o(x_n)+O(y_n)=O(\|x_n\|+\|y_n\|)$; $o(x_n)O(y_n)=o(x_n\cdot y_n)$; and $O(x_n)O(y_n)=O(x_n\cdot y_n)$;

d. $x_n=O(a_n^{-j})$ implies that $x_n=o(a_n^{-j+\frac{1}{2}})$ but $x_n=o(a_n^{-j+\frac{1}{2}})$ does not imply that $x_n=O(a_n^{-j})$. [For example, let $x_n=c/n^{3/4}$ for c a constant.]; and

e. a. through d. hold if O is replaced by O_p and/or o by o_p with the exception that if a subscript p appears anywhere on the left-hand

side of an equality in a. through d., then a subscript p must also appear on the right-hand side.

To justify results from calculating the expected value of the error terms, we have that

Lemma 3B.2: for j an even integer,

$$|E\{\prod_{g=1}^{j+1}(p_{h_g}-\tilde{p}_{h_g})|\underset{\sim}{z}\}| \leq |E\{\prod_{g=1}^{j}(p_{h_g}-\tilde{p}_{h_g})|\underset{\sim}{z}\}| = O(n^{-j/2})$$

where again $\tilde{p}_{h_g}=E(p_{h_g}|\underset{\sim}{z})$.

A proof of Lemma 3B.2 is given in Chapter 4. Thus, Lemma 3B.2 gives the magnitude of elements of the posterior covariance matrix and proves that posterior central cross-product moments significantly decrease as their order increases. Therefore, Taylor-series approximations in this appendix are valid.

From definition 3B.2, we can write the first-order Taylor-series expansion of $d_D(p_i|\underset{\sim}{z}, \prod_{j\in\emptyset} p_j|\underset{\sim}{z},) = r_{iD}|\underset{\sim}{z}$ about the value

$$d_D(E(p_i|\underset{\sim}{z}), \prod_{j\in\emptyset} E(p_i|\underset{\sim}{z}),) = E(p_i|\underset{\sim}{z})/E(p_D|\underset{\sim}{z}) = E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})$$

as

$$r_{iD}|\underset{\sim}{z} = E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z}) + \underset{\sim}{e}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + o(\|\underset{\sim}{e}_D\|) \qquad (3B.3)$$

as $\underset{\sim}{p}|\underset{\sim}{z}\longrightarrow E(\underset{\sim}{p}|\underset{\sim}{z})$, for $[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'$ denoting the transposed vector of $\partial d_D(w_i, \prod_{j\in\emptyset} w_j)/\partial w_l$, for $l\in D$, evaluated at $\underset{\sim}{w}=E(\underset{\sim}{p}|\underset{\sim}{z})$. That is, for $l\in D$,

$$\partial d_D/\partial E(p_\ell|\underset{\sim}{z}) = \begin{cases} \sum_{j\in D} E(p_j|\underset{\sim}{z})/[\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 & \text{for } \ell=i \\ -E(p_i|\underset{\sim}{z})/[\sum_{j\in D} E(p_j|\underset{\sim}{z})^2] & \text{for } \ell\neq i. \end{cases} \tag{3B.4}$$

By Tchebychev's inequality and the definition of O_p,

$$p_i|\underset{\sim}{z} - E(p_i|\underset{\sim}{z}) = O_p\left([\mathrm{var}(p_i|\underset{\sim}{z})]^{1/2}\right). \tag{3B.5}$$

From Lemma 3B.2 we have that $\mathrm{var}(p_i|\underset{\sim}{z}) = O(n^{-1})$. Therefore, by Lemma 3B.1, the error term in the first-order Taylor-series approximation (3B.3) of $r_{iD}|\underset{\sim}{z}$ is

$$\begin{aligned} o(\|\underset{\sim}{e}_D\|) &= o([\sum_{j\in D} e_j^2]^{1/2}) \\ &= o[O_p(n^{-1/2})] \\ &= o_p(n^{-1/2}). \end{aligned} \tag{3B.6}$$

Because we know the magnitude of $\underset{\sim}{e}_D$, we can also write (3B.6) as

$$o(\|\underset{\sim}{e}_D\|) = O_p(n^{-1}). \tag{3B.7}$$

Recalling from Lemma 3B.2 that the expected value of the error term with respect to the posterior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{z}$ is small relative to the first-order terms, we approximate moments of each ratio $r_{iD}|\underset{\sim}{z}$ by calculating expected values of the left- and right-hand sides of each $r_{iD}|\underset{\sim}{z}$ Taylor-series approximation.

Recall that $\tilde{\underset{\sim}{\Sigma}} = (\tilde{\sigma}_{ij})$ is the posterior covariance matrix of $\underset{\sim}{p}$ given $\underset{\sim}{z}$. Let $\tilde{\underset{\sim}{\Sigma}}_D$ denote that portion of $\tilde{\underset{\sim}{\Sigma}}$ that pertains to $j\in D$. That

is, $\tilde{\underset{\sim}{\Sigma}}_D$ is the matrix that results from deleting from $\tilde{\underset{\sim}{\Sigma}}$ all ℓ^{th} rows and columns for $\ell \notin D$.

3B.2 Posterior Mean:

Then, since $E(\underset{\sim}{e}_D)=\underset{\sim}{0}$ and $E[\underset{\sim}{e}_D' \underset{\sim}{e}_D] = \tilde{\underset{\sim}{\Sigma}}_D$, we have from (3B.3), (3B.6), and (3B.7) that

$$r_{iD}|\underset{\sim}{z} = E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z}) + \underset{\sim}{e}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + o_p(n^{-1/2}) \qquad (3B.8)$$

and

$$\begin{aligned} E(r_{iD}|\underset{\sim}{z}) &= E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z}) + o(n^{-1/2}) \\ &= E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z}) + O(n^{-1}). \end{aligned} \qquad (3B.9)$$

Note that we can write $O(n^{-1})$ in the last line of (3B.9) because $o(n^{-1/2})$ in the first line comes from an n^{-1} term. [Recall (3B.5) - (3B.7).]

3B.3 Posterior Variance:

Similarly,

$$r_{iD}^2|\underset{\sim}{z} = [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 + [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\underset{\sim}{e}_D'\underset{\sim}{e}_D[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'$$
$$+ 2\{[E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z}) + o_p(n^{-1/2})]\,\underset{\sim}{e}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'$$
$$+ [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(\underset{\sim}{p}|\underset{\sim}{z})]\,o_p(n^{-1/2})\} + o_p(n^{-1}), \tag{3B.10}$$

so that

$$E(r_{iD}^2|\underset{\sim}{z}) = [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 + [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\tilde{\Sigma}_D[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'$$
$$+ 2[E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(\underset{\sim}{p}|\underset{\sim}{z})]o(n^{-1/2}) + o(n^{-1}) \tag{3B.11}$$

or

$$E(r_{iD}^2|\underset{\sim}{z}) = [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 + \sum_{j\in D}[\partial d_D/\partial E(p_j|\underset{\sim}{z})]^2\tilde{\sigma}_{jj}$$
$$+ 2\sum_{j\in D}\sum_{\substack{\ell\in D\\ \ell>j}}[\partial d_D/\partial E(p_j|\underset{\sim}{z})]\,[\partial d_D/\partial E(p_\ell|\underset{\sim}{z})]\tilde{\sigma}_{j\ell} \tag{3B.12}$$
$$+ O(n^{-1}).$$

For use in Chapter 4, substitution from (3B.4) into (3B.12) yields that

$$E(r_{iD}^2|\underset{\sim}{z}) = \{2[E(p_\emptyset|\underset{\sim}{z})\sum_{j\in\emptyset}\tilde{\sigma}_{ij} + E(p_i|\underset{\sim}{z})\sum_{j\in\emptyset}\sum_{\substack{\ell\in\emptyset\\ \ell>j}}\tilde{\sigma}_{j\ell}] - [E(p_i|z)]^2\sum_{j\in\emptyset}\tilde{\sigma}_{jj}$$
$$+ [E(p_\emptyset|\underset{\sim}{z})]^2\tilde{\sigma}_{ii}\}/[E(p_D|\underset{\sim}{z})]^4 + [E(p_i|\underset{\sim}{z})/E(p_D|\underset{\sim}{z})]^2 + O(n^{-1}). \tag{3B.13}$$

From (3B.9) we have that

$$[E(r_{iD}|\underset{\sim}{z})]^2 = [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 + 2\,[E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]\,o(n^{-1/2}) + o(n^{-1}) \tag{3B.14}$$

or

$$[E(r_{iD}|\underset{\sim}{z})]^2 = [E(p_i|\underset{\sim}{z})/\sum_{j\in D} E(p_j|\underset{\sim}{z})]^2 + O(n^{-1}). \tag{3B.15}$$

Therefore, from equations (3B.11) and (3B.14) we have that

$$\begin{aligned} \mathrm{var}(r_{iD}|\underset{\sim}{z}) &= [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\,\tilde{\underset{\sim}{\Sigma}}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + o(n^{-1}) \\ &= [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\,\tilde{\underset{\sim}{\Sigma}}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + O(n^{-3/2}) \\ &= \sum_{j\in D}[\partial d_D/\partial E(p_j|\underset{\sim}{z})]^2\tilde{\sigma}_{jj} + 2\sum_{j\in D}\sum_{\substack{\ell\in D\\ \ell>j}}[\partial d_D/\partial E(p_j|\underset{\sim}{z})] \\ &\quad \times [\partial d_D/\partial E(p_\ell|\underset{\sim}{z})]\tilde{\sigma}_{j\ell} + O(n^{-3/2}). \end{aligned} \tag{3B.16}$$

Substituting from (3B.4) into (3B.16) yields that

$$\begin{aligned} \mathrm{var}(r_{iD}|\underset{\sim}{z}) = \{&2E(p_i|\underset{\sim}{z})\,[-E(p_\emptyset|\underset{\sim}{z})\sum_{j\in\emptyset}\tilde{\sigma}_{ji} + E(p_i|\underset{\sim}{z})\sum_{j\in\emptyset}\sum_{\substack{\ell\in\emptyset\\ \ell>j}}\tilde{\sigma}_{j\ell}] \\ &+ [E(p_\emptyset|\underset{\sim}{z})]^2\tilde{\sigma}_{ii} + [E(p_i|\underset{\sim}{z})]^2\sum_{j\in\emptyset}\tilde{\sigma}_{jj}\}/[E(p_D|\underset{\sim}{z})]^4 \\ &+ O(n^{-3/2}). \end{aligned} \tag{3B.17}$$

3B.4 Posterior Covariance:

Similarly, for all cases except those for which i=h at the same time that D=Q,

$$
\begin{aligned}
r_{iD}r_{hQ}|\underset{\sim}{z} = {} & [E(p_i|\underset{\sim}{z})/\sum_{j\in D}E(p_j|\underset{\sim}{z})]\,[E(p_h|\underset{\sim}{z})/\sum_{\ell\in Q}E(p_\ell|\underset{\sim}{z})] \\
& + \underset{\sim}{e}_D\{[E(p_h|\underset{\sim}{z})/\sum_{\ell\in Q}E(p_\ell|\underset{\sim}{z})][\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'\} \\
& + \underset{\sim}{e}_Q\{[E(p_i|\underset{\sim}{z})/\sum_{j\in D}E(p_j|\underset{\sim}{z})][\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'\} \\
& + [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\,\underset{\sim}{e}_D'\,\underset{\sim}{e}_Q\,[\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' \\
& + o_p(n^{-1/2})\,\{E(p_i|\underset{\sim}{z})/\sum_{j\in D}E(p_j|\underset{\sim}{z}) + E(p_h|\underset{\sim}{z})/\sum_{\ell\in Q}E(p_\ell|\underset{\sim}{z}) \\
& \qquad + \underset{\sim}{e}_D\,[\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + \underset{\sim}{e}_Q\,[\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]'\} + o_p(n^{-1}).
\end{aligned}
\tag{3B.18}
$$

Therefore,

$$
\begin{aligned}
E(r_{iD}r_{hQ}|\underset{\sim}{z}) = {} & [E(p_i|\underset{\sim}{z})/\sum_{j\in D}E(p_j|\underset{\sim}{z})]\,[E(p_h|\underset{\sim}{z})/\sum_{\ell\in Q}E(p_\ell|\underset{\sim}{z})] \\
& + [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\,\tilde{\underset{\sim}{\Sigma}}_{DQ}\,[\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' \\
& + o(n^{-1/2})\,\{E(p_i|\underset{\sim}{z})/\sum_{j\in D}E(p_j|\underset{\sim}{z}) + E(p_h|\underset{\sim}{z})/\sum_{\ell\in Q}E(p_\ell|\underset{\sim}{z})\} \\
& + o(n^{-1})
\end{aligned}
\tag{3B.19}
$$

for $\tilde{\underset{\sim}{\Sigma}}_{DQ}$ being the matrix whose elements are $\tilde{\sigma}_{j\ell}$ for all $j\in D$ and all $\ell\in Q$. That is, $\tilde{\sigma}_{j\ell}\in\tilde{\underset{\sim}{\Sigma}}_{DQ}$ if and only if $j\in D$ and $\ell\in Q$. If k_D is the number of integers in D and k_Q is the number of integers in Q, then the dimension of $\tilde{\underset{\sim}{\Sigma}}_{DQ}$ is $k_D\times k_Q$.

Recall that $\emptyset$ denotes D minus the integer i and let $\not{\emptyset}$ denote Q minus the integer h. Then, substitution from (3B.4) into (3B.19) yields that

$$\begin{aligned} E(r_{iD}r_{hQ}|\underset{\sim}{z}) = {} & \{E(p_{\emptyset}|\underset{\sim}{z})[E(p_{\not{\emptyset}}|\underset{\sim}{z})\tilde{\sigma}_{ih} - E(p_h|\underset{\sim}{z}) \sum_{\ell\in\not{\emptyset}} \tilde{\sigma}_{i\ell}] \\ & + E(p_i|\underset{\sim}{z}) \sum_{j\in\emptyset} [E(p_h|\underset{\sim}{z}) \sum_{\ell\in\not{\emptyset}} \tilde{\sigma}_{j\ell} - E(p_{\not{\emptyset}}|\underset{\sim}{z})\tilde{\sigma}_{jh}]\}/[E(p_D|\underset{\sim}{z})E(p_Q|\underset{\sim}{z})]^2 \\ & + E(p_i|\underset{\sim}{z})E(p_h|\underset{\sim}{z})/[E(p_D|\underset{\sim}{z})E(p_Q|\underset{\sim}{z})] + O(n^{-1}). \end{aligned} \tag{3B.20}$$

From (3B.9) and (3B.19) we have that

$$\begin{aligned} E(r_{iD}|\underset{\sim}{z})E(r_{hQ}|\underset{\sim}{z}) = {} & E(r_{iD}r_{hQ}|\underset{\sim}{z}) - [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\, \tilde{\underset{\sim}{\Sigma}}_{DQ}\, [\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' \\ & + o(n^{-1}). \end{aligned} \tag{3B.21}$$

Therefore, from (3B.19) and (3B.21),

$$\begin{aligned} \operatorname{cov}(r_{iD},r_{hQ}|\underset{\sim}{z}) &= [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\, \tilde{\underset{\sim}{\Sigma}}_{DQ}\, [\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + o(n^{-1}) \\ &= [\partial d_D/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]\, \tilde{\underset{\sim}{\Sigma}}_{DQ}\, [\partial d_Q/\partial E(\underset{\sim}{p}|\underset{\sim}{z})]' + O(n^{-3/2}) \\ &= \sum_{j\in D} \sum_{\ell\in Q} [\partial d_D/\partial E(p_j|\underset{\sim}{z})]\, [\partial d_Q/\partial E(p_\ell|\underset{\sim}{z})]\, \tilde{\sigma}_{j\ell} + O(n^{-3/2}). \end{aligned} \tag{3B.22}$$

Substituting from (3B.4) yields that

$$\begin{aligned} \operatorname{cov}(r_{iD},r_{hQ}|\underset{\sim}{z}) = {} & \{E(p_{\emptyset}|\underset{\sim}{z})[E(p_{\not{\emptyset}}|\underset{\sim}{z})\tilde{\sigma}_{ih} - E(p_h|\underset{\sim}{z}) \sum_{\ell\in Q} \tilde{\sigma}_{i\ell}] \\ & + E(p_i|\underset{\sim}{z}) \sum_{j\in\emptyset} [E(p_h|\underset{\sim}{z}) \sum_{\ell\in\not{\emptyset}} \tilde{\sigma}_{j\ell} - E(p_{\not{\emptyset}}|\underset{\sim}{z})\tilde{\sigma}_{jh}]\} \\ & /[E(p_D|\underset{\sim}{z})E(p_Q|\underset{\sim}{z})]^2 + O(n^{-3/2}). \end{aligned} \tag{3B.23}$$

APPENDIX 3C

INTERMEDIATE CALCULATION FOR VARIANCE

From (3.14)

$$\operatorname{var}(p_i|\underset{\sim}{z}) = \sum_{D\ni i}[z_D^2 \operatorname*{var}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}|\underset{\sim}{z}) + \sum_{\substack{Q\ni i\\ Q\neq D}} z_D z_Q \operatorname*{cov}_{\underset{\sim}{p}|\underset{\sim}{z}}(r_{iD}, r_{iQ}|\underset{\sim}{z})]/[m(m+1)] \tag{3C.1}$$

$$+\{\sum_{D\ni i}(z_D/m) \operatorname*{E}_{\underset{\sim}{p}|\underset{\sim}{z}}[r_{iD}(1-r_{iD}|\underset{\sim}{z})] + E(p_i|\underset{\sim}{z})[1-E(p_i|\underset{\sim}{z})]\}/(m+1).$$

Substituting from (3B.17), (3B.13), and (3B.23), we have, for $\emptyset$ denoting D minus the integer i and $\emptyset$ denoting Q minus the integer i, that

$$\begin{aligned}\operatorname{var}(p_i|\underset{\sim}{z}) = {} & \sum_{D\ni i}(z_D/m)[z_D/(m+1)]/[E(p_D|\underset{\sim}{z})]^4\{2E(p_i|\underset{\sim}{z})[E(p_i|\underset{\sim}{z})\sum_{j\in\emptyset}\sum_{\substack{\ell\in\emptyset\\ \ell>j}}\tilde{\sigma}_{j\ell} \\ & - E(p_\emptyset|\underset{\sim}{z})\sum_{j\in\emptyset}\tilde{\sigma}_{ji}] + [E(p_\emptyset|\underset{\sim}{z})]^2\tilde{\sigma}_{ii} + [E(p_i|\underset{\sim}{z})]^2\sum_{j\in\emptyset}\tilde{\sigma}_{jj}\} \qquad (3C.2)\\ & + \sum_{D\ni i}\sum_{\substack{Q\ni i\\ Q\neq D}}(z_D/m)[z_Q/(m+1)]/[E(p_D|\underset{\sim}{z})E(p_Q|\underset{\sim}{z})]^2 \\ & \times\{E(p_\emptyset|\underset{\sim}{z})[E(p_\emptyset|\underset{\sim}{z})\tilde{\sigma}_{ii} - E(p_i|\underset{\sim}{z})\sum_{\ell\in Q}\tilde{\sigma}_{i\ell}] \\ & + E(p_i|\underset{\sim}{z})\sum_{j\in D}[E(p_i|\underset{\sim}{z})\sum_{\ell\in Q}\tilde{\sigma}_{j\ell} - E(p_\emptyset|\underset{\sim}{z})\tilde{\sigma}_{ji}]\} \\ & + \Big\{\sum_{Q\ni i}(z_D/m)\Big(E(p_i|\underset{\sim}{z})E(p_\emptyset|\underset{\sim}{z})/[E(p_D|\underset{\sim}{z})]^2 \\ & - \{2E(p_i|\underset{\sim}{z})E(p_\emptyset|\underset{\sim}{z})\sum_{j\in\emptyset}\tilde{\sigma}_{ij} + 2[E(p_i|\underset{\sim}{z})]^2\sum_{j\in\emptyset}\sum_{\substack{\ell\in\emptyset\\ \ell>j}}\tilde{\sigma}_{j\ell} \\ & + [E(p_\emptyset|\underset{\sim}{z})]^2\tilde{\sigma}_{ii} - [E(p_i|\underset{\sim}{z})]^2\sum_{j\in\emptyset}\tilde{\sigma}_{jj}\}/[E(p_D|\underset{\sim}{z})]^4\Big)\end{aligned}$$

$$+ E(p_i|\underset{\sim}{z})[1-E(p_i|\underset{\sim}{z})]\Big\}/(m+1) + \delta_{ii}$$

for δ_{ii} denoting the error in (3C.2) made by approximating posterior moments of the ratios r_{iD} in (3C.1).

Therefore,

$$\begin{aligned}\tilde{\sigma}_{ii} = {} & \sum_{D\ni i}(z_D/m)[(z_D-1)/(m+1)]/[E(p_D|\underset{\sim}{z})]^4\{[E(p_\emptyset|\underset{\sim}{z})]^2\tilde{\sigma}_{ii} \\ & +E(p_i|\underset{\sim}{z})\sum_{j\in\emptyset}[E(p_i|\underset{\sim}{z})\tilde{\sigma}_{jj}-2E(p_\emptyset|\underset{\sim}{z})\tilde{\sigma}_{ji}+2E(p_i|\underset{\sim}{z})\sum_{\substack{\ell\in\emptyset\\ \ell>j}}\tilde{\sigma}_{j\ell}]\} \\ & + \sum_{D\ni i}\sum_{\substack{Q\ni i\\ Q\neq D}}(z_D/m)[z_Q/(m+1)]/[E(p_D|\underset{\sim}{z})E(p_Q|\underset{\sim}{z})]^2\{E(p_\emptyset|\underset{\sim}{z})[E(p_\emptyset|\underset{\sim}{z})\tilde{\sigma}_{ii} \\ & -E(p_i|\underset{\sim}{z})\sum_{\ell\in Q}\tilde{\sigma}_{i\ell}]+E(p_i|\underset{\sim}{z})\sum_{j\in\emptyset}[E(p_i|\underset{\sim}{z})\sum_{\ell\in Q}\tilde{\sigma}_{j\ell}-E(p_\emptyset|\underset{\sim}{z})\tilde{\sigma}_{ji}]\} \\ & +\{\sum_{D\ni i}(z_D/m)E(p_i|\underset{\sim}{z})E(p_\emptyset|\underset{\sim}{z})/[E(p_D|\underset{\sim}{z})]^2 \\ & + E(p_i|\underset{\sim}{z})[1-E(p_i|\underset{\sim}{z})]\}/(m+1) + \delta_{ii}.\end{aligned} \tag{3C.3}$$

CHAPTER 4

ASYMPTOTICS FOR TAYLOR-SERIES APPROXIMATIONS

4.1 Introduction:

In Chapter 3, we used low-order Taylor-series expansions for unknown terms in deriving Taylor-series approximations for the posterior mean and covariance matrices. For these Taylor-series expansions to allow accurate approximations, higher-order central cross-product posterior moments of $\underset{\sim}{p}$ must be substantially smaller than lower-order central cross-product moments. In this chapter we prove this condition. We then assess the accuracy of the Taylor-series approximations. Because results are in terms of orders of magnitude or otherwise involve limiting distributions, we call this chapter the asymptotics for Taylor-series approximations. For the asymptotics we use the sampling-theory approach. We fix the probability $\underset{\sim}{p}$ and then study the limiting distribution of the data as the sample size n goes to infinity.

In the next section we determine the magnitude of the central cross-product moments and show that this magnitude substantially decreases as the order of the moment increases. The first part of the section gives results for complete data; the last part, results for incomplete data. In the third section we assess the accuracy of the Taylor-series approximations for the posterior mean and covariance matrices. We begin by giving the accuracy for the ratio approximations of Appendix 3B. A summary concludes the chapter.

Five appendices give derivations used in the chapter. The first appendix calculates the posterior central moments given complete

multinomial data. The second appendix derives the limiting posterior distribution given complete data. The third appendix calculates central moments of the k-dimensional multivariate normal distribution, giving results more general than found in the literature. The fourth appendix derives the limiting posterior distribution given incomplete multinomial data. Finally, the fifth appendix gives the error in evaluating a function by an iterative solution of an approximation to the function. Note that techniques developed in the appendices are applicable to distributions other than the Dirichlet or multinomial.

4.2 Central Cross-Product Moments:

4.2.1 Complete Data:

In this section we obtain the order of magnitude of central cross-product moments given complete multinomial data $\underset{\sim}{x}$. We begin by obtaining the order of magnitude of the l^{th} posterior central moment (2.6). To do so, in Appendix 4A we write (2.6) in a Taylor series in $(n+\Sigma\nu_h)^{-1}$ about 0 for enough values of l to detect a pattern for the low-order term in n^{-1}. We then extend moment results from Kendall and Stuart (1969,v1,p148) for Pearson distributions to prove by induction that for

$$l!! = l(l-2)(l-4)(l-6)\cdots 1 \quad \text{for } l \text{ odd}^*,$$

$$\mu_i \equiv E(p_i|\underset{\sim}{x}),$$

and (4.1)

$$\sigma_{ii} \equiv \text{var}(p_i|\underset{\sim}{x}) = \mu_i(1-\mu_i)/(n+\sum_{j=1}^{k+1}\nu_j),$$

we have that

$$E[(p_i-\mu_i)^l|\underset{\sim}{x}] \doteq \begin{cases} (l-1)!!\sigma_{ii}^{l/2} & \text{for } l \text{ even} \\ (l-1)l!!\sigma_{ii}^{(l+1)/2}(1-2\mu_i)/[3\mu_i(1-\mu_i)] & \text{for } l \text{ odd,} \end{cases} \quad (4.2)$$

where the approximation in (4.2) means that we have given the lowest-order term in n^{-1}. [Recall from (2.6) that $E[(p_i-\mu_i)^l|\underset{\sim}{x}]$ is a function of n.]

Hence, noting the n in the denominator of σ_{ii} in (4.1), we have that

$$\left.\begin{array}{l} \text{for } l \text{ even} \\ \lim_{n\to\infty} n^{l/2} E[(p_i-\mu_i)^l|\underset{\sim}{x}] = (l-1)!![\mu_i(1-\mu_i)]^{l/2}, \\ \text{and for } l \text{ odd} \\ \lim_{n\to\infty} n^{(l+1)/2} E[(p_i-\mu_i)^l|\underset{\sim}{x}] = (l-1)l!!(1-2\mu_i)[\mu_i(1-\mu_i)]^{(l-1)/2}/3. \end{array}\right\} \quad (4.3)$$

*Standard mathematical notation; for example, see Gradshtevy and Ryzhik (1967,pxliii); $l!!$ is not defined for l even.

Therefore,

$$E[(p_i-\mu_i)^l|\underset{\sim}{x}] = \begin{cases} O(n^{-l/2}) & \text{for } l \text{ even} \\ O(n^{-(l+1)/2}) & \text{for } l \text{ odd.} \end{cases} \tag{4.4}$$

In Appendix 4A we also found that

$$E[(p_i-\mu_i)^l(p_j-\mu_j)^h|\underset{\sim}{x}] = \begin{cases} O(n^{-(l+h)/2}) & \text{for } l+h \text{ even} \\ O(n^{-(l+h+1)/2}) & \text{for } l+h \text{ odd.} \end{cases}$$

for $2 \leq l,h \leq 8$.

However, the methods of Appendix 4A were unfeasible for evaluating the general l^{th} posterior central cross-product moment $E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}]$ for $1 \leq h_g \leq k$ and $h_g \neq h_1$ for at least one g. Therefore, to obtain general results similar to those given in (4.4), we use the Helly-Bray Theorem [Rao (1968,p97)]:

Theorem 4.1 (Helly-Bray Theorem): If the distribution function F_n converges to the distribution function F, then

$$\int g \, dF_n \to \int g \, dF$$

for every bounded continuous function g.

Since $\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})$ is bounded and is continuous in $\underset{\sim}{p}$, by Theorem 4.1 limits of posterior central cross-product moments equal corresponding

moments of the limiting distribution. The latter moments are usually referred to as the asymptotic moments. [See Bishop, Fienberg, and Holland (1975,p485).] Hence, we calculate the limiting distribution of the posterior distribution of $\underset{\sim}{p}$ given complete data $\underset{\sim}{x}$ and then calculate the central cross-product moments of this limiting distribution.

By using Stirling's approximation [Cramer (1951,p130)] for the logarithm of the gamma function, theorems from Graybill (1969,p8,170,184) to calculate the determinant and inverse of the covariance matrix, series approximations [CRC Tables (1962,p373)] for $\log(1+\varepsilon)$ for $|\varepsilon|<1$, and Tchebychev's inequality [Bishop, Fienberg, and Holland (1975,p476)] to determine the magnitude of error in approximations, we prove in Appendix 4B that the k-dimensional Dirichlet density with mean $\underset{\sim}{\mu}$ and covariance matrix $\underset{\sim}{\Sigma}$ differs from a k-dimensional multivariate normal density with mean $\underset{\sim}{\mu}$ and covariance matrix $\underset{\sim}{\Sigma}$ by order of magnitude $O_p(n^{-1/2})$. [Recall definition 3B.3 of O_p.] Rao (1968,p104) gives the following convergence theorem involving densities:

Theorem 4.2: If the density $f_n(\underset{\sim}{x})$ converges to the density $f(\underset{\sim}{x})$ as $n\to\infty$, then the distribution function $F_n(\underset{\sim}{x})$ converges to the distribution function $F(\underset{\sim}{x})$ as $n\to\infty$.

Therefore, from Theorem 4.2 the limiting posterior distribution of $\underset{\sim}{p}$ given complete data $\underset{\sim}{x}$ is $N_k(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$.

To obtain central cross-product moments of this limiting distribution, in Appendix 4C we multiply the multivariate-normal moment-generating function [Wilks (1963,p168)] by $\exp(-\underset{\sim}{t}\underset{\sim}{\mu}')$, continuously differentiate the

results with respect to $\underset{\sim}{t}$, and then set $\underset{\sim}{t}$ to $\underset{\sim}{0}$ in the differentiated results. Doing so yields that the central cross-product moment for the multivariate normal distribution is zero for l odd and, for l even, is a sum of $l-1$ terms, each of which is a product of $l/2$ elements of the covariance matrix and thus, from (2.3) and (2.4), is of magnitude $O(n^{-l/2})$. Therefore, application of these results with the Helly-Bray Theorem yields for the l^{th} posterior central cross-product moment that

$$\text{for } l \text{ even}, \quad E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}] = O(n^{-l/2}) \tag{4.5}$$

for $1 \leq h_g \leq k$. For l odd, however, these results yield only that

$$\text{for } l \text{ odd}, \quad \lim_{n\to\infty} E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}] = 0. \tag{4.6}$$

Therefore, to calculate the order of magnitude for odd posterior central cross-product moments, we have the following lemma:

Lemma 4.1: for l a positive integer,

$$|E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}]| \leq |E[\prod_{g=1}^{l-1}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}]|. \tag{4.7}$$

Proof:

First note that, since h_a can equal h_b for any $1 \leq a,b \leq l$ one of the l values of g, the density function f_l for the l^{th} central cross-product moment will be of dimension $1 \leq \alpha \leq k$.

In going from the $(l-1)^{st}$ to the l^{th} central cross-product moment, the density function will remain the same if the additional variable p_{h_l} for $p_{h_l}-\mu_{h_l}$ is a variable of f_{l-1}. In such case, the proof follows from the fact that, for all g,

$$-1 \leq p_{h_g} - \mu_{h_g} \leq 1. \tag{4.8}$$

Hence, the integrand for the l^{th} posterior central cross-product moment is a fraction of that for the $(l-1)^{st}$ central cross-product moment and (4.7) therefore follows.

If the additional variable p_{h_l} in going from the $(l-1)^{st}$ to the l^{th} central cross-product moment is not a variable of f_{l-1}, then for P_α the α-dimensional simplex of the vector of those distinct probabilities p_{h_g} in $\prod_{g=1}^{l} p_{h_g}$ and $P_{\alpha-1}$ the $(\alpha-1)$-dimensional subspace of P_α obtained by deleting variable p_{h_l}, we have that

$$\begin{aligned} |E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}]| &= |\int_{P_\alpha} \prod_{g=1}^{l}(p_{h_g}-\mu_{h_g}) f_l \, d\underset{\sim}{p}| \\ &= |\int_{P_{\alpha-1}} \prod_{g=1}^{l-1}(p_{h_g}-\mu_{h_g})[\int(p_{h_l}-\mu_{h_l}) f_l \, dp_{h_l}] d\underset{\sim}{p}| \qquad (4.9) \\ &\leq |\int_{P_{\alpha-1}} \prod_{g=1}^{l-1}(p_{h_g}-\mu_{h_g}) f_{l-1} \, d\underset{\sim}{p}| = |E[\prod_{g=1}^{l-1}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}]| \end{aligned}$$

since (4.8) yields that $\int(p_{h_l}-\mu_{h_l}) f_l \, dp_{h_l}$ is bounded by $\pm\int f_l \, dp_{h_l} = \pm f_{l-1}$.

From bound (4.7), magnitude (4.5), low-order terms for cross-product moments $|E[(p_i-\mu_i)^a(p_j-\mu_j)^b|\underset{\sim}{x}]|$ for $2 \leq a,b \leq 8$ from Appendix 4A, and results (4.4) for $E[(p_i-\mu_i)^l|\underset{\sim}{x}]$ for l odd, we would expect that, in general, $E[\prod_{g=1}^{l}(p_{h_g}-\mu_{h_g})|\underset{\sim}{x}] = O(n^{-(l+1)/2})$ for l odd.

Note that for incomplete data, we can duplicate all complete-data results but one. We can not parallel proof from Appendix 4A that for odd l of 3, 5, and 7 the cross-product moment is $O(n^{-(l+1)/2})$. Although we expect this result based on all complete-data results and on incomplete-

density as a product of complete-data Dirichlet densities, each having, from Appendix 4B, a limiting multivariate normal distribution. Because these densities are of differing dimensions and on differing combinations of variables, we do not immediately have that the resultant product of these multivariate normal densities is a k-dimensional multivariate normal density on the k components of $\underset{\sim}{p}$. However, by equating coefficients and solving for unknowns, we then prove that, owing to the special relationship between the first and each remaining product, the sum of exponents from each Dirichlet in the product does form the exponent of such a density. Following derivations in Appendix 4D, we have as final results for elements u_i, S^{ii}, and S^{ij}, respectively, of the asymptotic mean and inverse covariance matrices

$$u_i = (z_i + \sum_{D \ni i} z_D u_i / u_D)/n, \tag{4.10}$$

$$\begin{aligned} S^{ii} = {} & n(u_i+u_{k+1})/(u_i u_{k+1}) - \sum_{\substack{D \not\ni i \\ D \ni k+1}} (z_D/u_D)(u_D-u_{k+1})/(u_D u_{k+1}) \\ & - \sum_{\substack{D \ni i \\ D \not\ni k+1}} (z_D/u_D)(u_D-u_i)/(u_i u_D) - \sum_{\substack{D \ni i \\ D \ni k+1}} (z_D/u_D)(u_i+u_{k+1})/(u_i u_{k+1}), \end{aligned} \tag{4.11}$$

and

$$S^{ij} = n/u_{k+1} + \sum_{\substack{D \not\ni k+1 \\ D \ni i,j}} (z_D/u_D)/u_D + \sum_{\substack{D \ni k+1 \\ D \not\ni i, \not\ni j}} (z_D/u_D)/u_D - \sum_{D \ni k+1} (z_D/u_D)/u_{k+1}, \tag{4.12}$$

for D a set $\not\ni$ containing more than one element, "$D \ni i,j$" meaning the set D containing <u>both</u> i and j, and all conditions given under a summation sign to be met simultaneously [for example, the first summation sign in (4.12) means the sum over all sets D such that $D \not\ni k+1$ at the same time that $D \ni i,j$].

Note that expressions (4.11) and (4.12) for elements of the asymptotic inverse covariance matrix are simple [especially relative to expressions (4D.12) and (4D.13) given by the traditional derivation]. Furthermore, they parallel complete-data results given by the first term in each of expressions (4.11) and (4.12). [Note that once we have expressions (4.11) and (4.12) and thus know what to work toward, we show in lengthy reexpressions in Appendix 4D that results given by the traditional approach can be simplified to (4.11) and (4.12). Thus, the second approach might be useful in other kinds of problems to clarify and simplify any unwieldy results given by the traditional approach.]

From (4.11) and (4.12) we have that elements of the asymptotic covariance matrix are $O(n^{-1})$. Thus, paralleling (4.5) we have that

$$\text{for } l \text{ even}, \quad E\Big[\prod_{g=1}^{l}(p_{h_g}-\tilde{p}_{h_g})\,\Big|\,\underset{\sim}{z}\Big] = O(n^{-l/2}). \tag{4.13}$$

Further, Lemma 4.1 holds for the case of incomplete data $\underset{\sim}{z}$ as well as for that of complete data $\underset{\sim}{x}$. Therefore, again paralleling the case for complete data, since Lemma 4.1 gives that the odd l^{th} posterior central cross-product moment is bounded in magnitude by the even $(l-1)^{st}$ moment, from (4.13) the odd l^{th} moment is of magnitude no greater than $O(n^{-l/2})$. Therefore, conditions for using Taylor-series expansions in Chapter 3 are satisfied.

Note, from comparing (3.9) with (4.10), that asymptotically the Taylor-series approximate posterior mean equals the exact posterior mean.

4.3 Accuracy of the Taylor-Series Approximations:

4.3.1 Introduction:

To determine the accuracy of the Taylor-series approximations of Chapter 3, we note that the only terms we approximated in the derivations were moments of the ratios. Therefore, we calculate the error made in these approximations and then calculate the overall error made by substituting these approximations into equations (3.6), (3.14), and (3.15) for the posterior mean, variance, and covariance, respectively. We also apply results from Isaacson and Keller (1966) to determine the error made by iteratively solving the resulting equations and then using the solution to approximate the exact posterior central moments.

4.3.2 Accuracy of the Taylor-Series Approximation for Posterior Mean:

The approximation for the exact posterior mean

$$\tilde{p}_i = (z_i+\nu_i+\sum_{D\ni i} z_D\tilde{p}_i/\tilde{p}_D)/m + \varepsilon_i \tag{4.14}$$

obtained by dropping the error term ε_i is

$$\tilde{p}_i \doteq (z_i+\nu_i+\sum_{D\ni i} z_D\tilde{p}_i/\tilde{p}_D)/m. \tag{4.15}$$

Rewriting (4.15) as a nonlinear system of equations yields the Taylor-series approximate posterior mean

$$\dot{\tilde{p}}_i = (z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/m \tag{4.16}$$

given in (3.9).

We now give the asymptotic error in using (4.16) to approximate (4.14). We do so by first determining the error in approximating (4.14) by (4.15) and then determining the error in solving (4.15) by the EM iterative algorithm of Dempster, Laird, and Rubin (1977). We look for conditions under which an iterative solution to (4.15) agrees with (4.14) within some error bound. In the formulation of the iterative process, we rewrite (4.15) as (4.16).

To determine the error in approximating (4.14) by (4.15), we must determine the accuracy of the approximation of each ratio r_{iD} and its first two moments. These accuracies are given in Appendix 3B in terms of the O and o notations. Then, from (3.6) and (3.7), the exact posterior mean $\tilde{p}_i$ can be written as

$$\begin{aligned}\tilde{p}_i &= [z_i+\nu_i+\sum_{D\ni i} z_D E(r_{iD}|\underset{\sim}{z})]/m \\ &= \{z_i+\nu_i+\sum_{D\ni i} z_D[\tilde{p}_i/\tilde{p}_D+O(n^{-1})]\}/m \\ &= [z_i+\nu_i+\sum_{D\ni i} z_D\tilde{p}_i/\tilde{p}_D]/m + O(n^{-1}).\end{aligned} \tag{4.17}$$

Hence, the error ε_i in approximating (4.14) by (4.15) is $O(n^{-1})$.

We next investigate how this error is affected by solving (4.15) by the EM iterative algorithm. To do so, we find two conditions in Appendix 4E whose satisfaction guarantees that

$$|\dot{\tilde{p}}_i^{(s)}-\tilde{p}_i| \leq \delta/(1-\lambda) + \lambda^s[\rho_o-\delta/(1-\lambda)]. \tag{4.18}$$

In (4.18), δ is a bound on the error made by approximating (4.14) by (4.15) and, hence, from (4.17) is of magnitude $O(n^{-1})$. The term λ is a

positive proportion less than 1; λ differs from a constant by $O(n^{-1})$. Therefore, $\delta/(1-\lambda)=O(n^{-1})$. The term ρ_0 is a constant. Since s can be made as large as desired, the right-hand term can be considered to be 0; in particular, it can be made at least as small as $O(n^{-1})$. Thus, from (4.18), when the two conditions given in Appendix 4E are satisfied, the error in approximating (4.14) by (4.16) is $O(n^{-1})$; i.e.,

$$\dot{\tilde{p}}_i = \tilde{p}_i + O(n^{-1}). \tag{4.19}$$

The two conditions in Appendix 4E concern the region in which the initial iterative estimate is chosen and a bound on the partial derivatives of the right-hand side of (4.16) with respect to $\dot{\tilde{p}}_j$. If there exists a neighborhood $\|\dot{\tilde{p}}-\tilde{p}\|_\infty<\rho$, for $\rho>0$, of $\tilde{p}$ such that for all probabilities in this neighborhood

$$\max_{1\leq i\leq k} \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial \dot{\tilde{p}}_j| \leq \lambda < 1,$$

for

$$g_i(\dot{\tilde{p}})=(z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/(n+\sum_{h=1}^{k+1}\nu_h),$$

and if an initial iterative estimate $\dot{\tilde{p}}_i^{(0)}$ is chosen within the inner neighborhood $\|\dot{\tilde{p}}-\tilde{p}\|_\infty<\rho_0\leq\rho-\delta/(1-\lambda)$, where δ is a bound on the error in approximating the exact posterior mean by a first-order Taylor-series expansion, then the iterative solution to the defining equation of the Taylor-series approximate posterior mean $\dot{\tilde{p}}$ will converge to within $\delta/(1-\lambda)$ $=O(n^{-1})$ of the exact posterior mean.

If a neighborhood of the exact posterior mean can be found in which the second condition is satisfied, then, for large enough sample sizes,

the first condition can be satisfied by choosing an initial iterative estimate in a neighborhood within the first neighborhood. For moderate percentages of incomplete data, the inner neighborhood is almost as large as the outer neighborhood. In Appendix 4E, we show how to determine, in practice, whether the second condition can be expected to hold; hence, we show how to approximate the size of the outer neighborhood. Further, for incomplete trinomial data, Appendix 4E shows that a root of the defining equations of the Taylor-series approximate posterior mean that differs from the exact posterior mean by magnitude $O(n^{-1})$ exists in P_2.

However, this root need not be unique in P_2. Moreover, as Ortega and Rheinbolt (1970,p2) illustrate with a simple case, a nonlinear system of k equations in k unknowns may have no solution or may have arbitrarily many solutions. Therefore, we now consider when the Taylor-series approximate posterior mean for incomplete data from the general k-dimensional multinomial distribution not only has a solution but also has a solution that is in P_k and that differs from the exact posterior mean by magnitude $O(n^{-1})$.

Because the Taylor-series approximate posterior mean can be written as a posterior mode, it will always have at least one solution in P_k when certain conditions, soon to be discussed, are met. However, none of these solutions may be in the epm convergence region, which we define as the region in which an initial iterative estimate can be picked so that successive iterates are guaranteed to converge to within a small error of the exact posterior mean. In particular, for $k>2$, there may not exist an epm convergence region. That is, there may not exist a neighborhood

of $\tilde{p}$ such that for all probabilities $\mathring{p}$ in the neighborhood

$$\max_i \sum_{j=1}^{k} |\partial g_i(\mathring{p})/\partial \mathring{p}_j| \doteq \max_i \sum_{D \ni i} z_D/m\{1/\mathring{p}_D + [\beta(D)-2]\mathring{p}_i/\mathring{p}_D^2\} < 1. \qquad (4.20)$$

As the number k of dimensions, the percentage $100\times\Pi z_D/n$ of incomplete data, or the number $\beta(D)$ of variables sharing incomplete data increases, inequality (4.20) shows that this possibility increases.

The most likely values of $\tilde{p}$ not to have an epm convergence region are those in higher dimensions that have one or more components near zero and/or a component near 1 when the percentage of incomplete data is high. For example, consider incomplete multinomial data z_1, z_2, $\cdots$, z_{10}, z_{11}, and $z_{1\cdots10}$ where the percentage of incomplete data is $100\times(z_{1\cdots10}/n)=50$. Suppose that $\tilde{p}_{10}=.89$ and $\tilde{p}_i=.01$ otherwise. Further, suppose that the sample size n is large enough, or the sum Σv_j of prior parameters is small enough, that $z_{1\cdots10}/n \doteq z_{1\cdots10}/m \doteq .5$. Then, for probabilities $\mathring{p}_i^{(s)}$ near $\tilde{p}_i$ and $D=\{1,\cdots,10\}$, one term in (4.20) is

$$(z_{1\cdots10}/m)\{1/\mathring{p}_D^{(s)} + [\beta(D)-2]\mathring{p}_{10}^{(s)}/[\mathring{p}_D^{(s)}]^2\}$$

$$\doteq 0.5\{1/.99 + 8\times.89/(.99)^2\} = 4.14 > 1.$$

However, for probabilities having such small values for some components, results of Chapters 6 and 7 indicate that the posterior mean is a relatively poor estimator to minimize risk for quadratic loss; the posterior mode is much better. Hence, for this particular case, we do not have to be concerned with not being able to find an epm convergence region. This example illustrates, however, that the Taylor-series approximate posterior mean needs more study in higher dimensions,

since the largest factor, 8, in the last inequality is a function of the dimension of the P_k simplex.

When there does exist an epm convergence region, there can be a problem finding it because there may be multiple roots in P_k of the Taylor-series approximation. In particular, there may be multiple roots in P_k for which inequality (4.20) is satisfied. The problem then is choosing among these roots. Since the Taylor-series approximation can be written as

$$\dot{\tilde{p}}_i = [z_i+\beta_i-1+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D]/[n+\sum_{j=1}^{k+1}\beta_j-(k+1)], \qquad (4.21)$$

where $\beta_i=\nu_i+1$, the Taylor-series approximation is a posterior mode; i.e., (4.21) is in form (2.43). Thus, the Taylor-series approximate posterior mean enjoys the convergence properties of the EM algorithm. That is, define $t_i(\underset{\sim}{x})=z_i+\nu_i+\sum_{D\ni i} z_D^{(i)}$, $\phi_i=\ln(p_i/p_{k+1})$, and t as the number of iterations required to meet convergence conditions. Then, since the multinomial distribution is a member of the regular exponential family, $\dot{\underset{\sim}{p}}^{(s)}$ converges in P_k to at least a local maximum if the eigenvalues of $\mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}^{(s)}]$, $1\leq s\leq t$, are bounded above zero. [See Section 2.3.] To find a global maximum, choose that root that maximizes the likelihood function

$$\dot{\tilde{p}}_1^{\,z_1+\nu_1-1}\dot{\tilde{p}}_2^{\,z_2+\nu_2-1}\cdots\dot{\tilde{p}}_{k+1}^{\,z_{k+1}+\nu_{k+1}-1}\prod_D \dot{\tilde{p}}_D^{\,z_D}.$$

From the complete-data relationship between the posterior mode and posterior mean, we intuitively expect the global maximum to be in the epm convergence region, or at least be the closest root to $\tilde{\underset{\sim}{p}}$. However,

this hypothesis has not been proved and needs study. As for the two-dimensional case, however, Appendix 4E proves that if a root is in the epm convergence region, then the error in using it to approximate the exact posterior mean is of magnitude $O(n^{-1})$. Note again that the $O(n^{-1})$ error mainly comes from using a first-order Taylor-series expansion to approximate the exact posterior mean.

Observe, as we illustrate with examples in Appendix 4E, that the two guaranteed-convergence conditions are sufficient, not necessary. That is, an initial iterative estimate can fall far outside the epm convergence region and convergence to the exact posterior mean still occur to within the same small error incurred when an initial iterative estimate is chosen inside the epm convergence region. Moreover, as also exampled, the error bound given by Theorem 4E.1 when these two conditions are satisfied is extremely conservative.

Finally, one should not pick as an initial iterative estimate a probability containing zero components because $\dot{\tilde{p}}_i$ corresponding to those components will be the same for all iterations. Further, any initial iterative estimate that has components near zero may cause the convergence process to be extremely slow for those components; see Section 5.8.4 for an example.

4.3.3 Accuracy of Taylor-Series Approximations for Posterior Covariances:

For those categories i and h that have only complete data, there is no error in writing $\dot{\tilde{\sigma}}_{ii}$ of (3.18) and $\dot{\tilde{\sigma}}_{ih}$ of (3.19) for $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{ih}$, respectively. For those categories h having only complete data and those categories i having incomplete data, we have from equation (3.15),

Section 4.3.2, and Lemma 3B.1 that the error in writing $\dot{\tilde{\sigma}}_{ih}$ of (3.20) for $\tilde{\sigma}_{ih}$ is $O(n^{-2})$. For both categories i and h having incomplete data, we have a choice of approximating the variance and covariance by procedures that are iterative or noniterative in $\tilde{\sigma}_{ih}$.

For the iterative procedure, we first evaluate initial estimates $\dot{\tilde{\sigma}}_{ii}^{(0)}$ of (3.21) and $\dot{\tilde{\sigma}}_{ih}^{(0)}$ of (3.22). Equations (3.14), (3.15), (3B.9), and Lemma 3B.1 yield that the error in these approximations $\dot{\tilde{\sigma}}_{ii}^{(0)}$ and $\dot{\tilde{\sigma}}_{ih}^{(0)}$ is $O(n^{-2})$, provided that parallel conditions from Appendix 4E for $\dot{\tilde{\sigma}}_{ii}$ and $\dot{\tilde{\sigma}}_{ih}$ can be satisfied.

To calculate the error in making (3.16) and (3.17) iterative algorithms, we note from the form of (3.14) and (3.15) and from approximations (3B.9), (3B.12), (3B.15), (3B.16), (3B.19), (3B.21), and (3B.22) that the largest error for $\dot{\tilde{\sigma}}_{ih}^{(s)}$ will come from approximating $\text{var}(r_{iD}|\underset{\sim}{z})$ and $\text{cov}(r_{iD},r_{hQ}|\underset{\sim}{z})$. [That is, the error in approximating terms multiplied by $1/(m+1)$ in (3.14) and (3.15) is $1/(m+1)$ times the error for those terms and in total is less than the error made in approximating $\text{var}(r_{iD}|\underset{\sim}{z})$ and $\text{cov}(r_{iD},r_{hQ}|\underset{\sim}{z})$.] At the same time, note from (3B.16) and (3B.22) that these errors are $O(n^{-3/2})$. Thus, if parallel conditions from Appendix 4E are satisfied for $\dot{\tilde{\sigma}}_{ii}$ and $\dot{\tilde{\sigma}}_{ih}$, then, recalling Lemma 3B.1, we have that the errors in approximating $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{ih}$ by $\dot{\tilde{\sigma}}_{ii}^{(s)}$ and $\dot{\tilde{\sigma}}_{ih}^{(s)}$, respectively, are $O(n^{-3/2})$.

The second procedure to approximate the variance and covariance for those q variables referring to categories that have incomplete data is a method that is noniterative in $\tilde{\sigma}_{ih}$. Recall from (3.23) that, for both i and j referring to categories having incomplete data, $\tilde{a}_{1h}$

coefficients of $\tilde{\sigma}_{lh}$, and $\tilde{b}_{ij}$ a term that is not a function of $\tilde{\sigma}_{lh}$ for any l or h, in this procedure we write each $\tilde{\sigma}_{ij}$ as

$$\tilde{\sigma}_{ij} = \sum_{D \ni i} \sum_{Q \ni j} \left[\sum_{l \in D} \sum_{h \in Q} \tilde{a}_{lh} \tilde{\sigma}_{lh} \right] + \tilde{b}_{ij} + \delta_{ij}. \tag{4.22}$$

In (4.22) δ_{ij} is an infinite series containing terms in $E(e_i^h|\underset{\sim}{z})$ and $E(e_i^l e_j^h|\underset{\sim}{z})$ for $l,h \geq 2$ and $e_i = p_i - \tilde{p}_i$. Thus, some of these terms are in $\tilde{\sigma}_{lh}$ [for example, second-order terms in the approximation for $E(r_{iD}|\underset{\sim}{z})$ are terms in $\tilde{\sigma}_{ii}$]. Therefore, we can divide δ_{ij} into a component δ_A containing terms in $\tilde{\sigma}_{lh}$ and a component δ_B containing the remaining terms.

Doing so, we can write (4.22) as a linear system of $q(q+1)/2$ equations in the $q(q+1)/2$ unknowns $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{ij}$:

$$[\underset{\sim}{\tilde{A}} + \underset{\sim}{\delta}_A] \underset{\sim}{\tilde{\sigma}} = \underset{\sim}{\tilde{B}} [\underset{\sim}{I} + \underset{\sim}{\delta}_B] \underset{\sim}{1} \tag{4.23}$$

where $\underset{\sim}{\tilde{\sigma}}$ is the $q(q+1)/2 \times 1$ vector of $\tilde{\sigma}_{ij}$ for i and j both referring to categories having incomplete data, $\underset{\sim}{\tilde{A}}$ is the $q(q+1)/2 \times q(q+1)/2$ matrix of the $\tilde{a}_{lh}$, $\underset{\sim}{\tilde{B}}$ is the $q(q+1)/2 \times q(q+1)/2$ matrix with $\tilde{b}_{ij}$ on the diagonal and 0's elsewhere, $\underset{\sim}{I}$ is the $q(q+1)/2 \times q(q+1)/2$ identity matrix, δ_A is the $q(q+1)/2 \times q(q+1)/2$ matrix containing those terms in δ_{ij} that are terms in $\tilde{\sigma}$, $\underset{\sim}{\delta}_B$ is the $q(q+1)/2 \times q(q+1)/2$ matrix containing zeros on the off-diagonal and the remaining terms of δ_{ij} divided by $\tilde{b}_{ij}$ on the diagonal, and $\underset{\sim}{1}$ is the $q(q+1)/2 \times 1$ vector containing all 1's.

Now, from (3B.16) and (3B.22), the terms $var(r_{iD}|\underset{\sim}{z})$ and $cov(r_{iD}, r_{hQ}|\underset{\sim}{z})$ in (3.14) and (3.15) contain no terms in $\tilde{\sigma}_{ii}$ and $\tilde{\sigma}_{ij}$ that are not already included in $\underset{\sim}{\tilde{A}}$. The terms $E(r_{iD}|\underset{\sim}{z})$ and $E(r_{iD}, r_{hQ}|\underset{\sim}{z})$

do, however; in particular, the first terms dropped from their Taylor-series expansions. Since these terms have coefficients that are constant with respect to the sample size n and since $E(r_{iD}|\underset{\sim}{z})$ and $E(r_{iD},r_{hQ}|\underset{\sim}{z})$ in (3.14) and (3.15) are multiplied by $(m+1)^{-1}=O(n^{-1})$, by Lemma 3B.1 the component of δ_{ij} that goes on the left-hand side of (4.23) is $O(n^{-1})\times\tilde{\sigma}_{ij}$. Thus, all terms in $\underset{\sim}{\delta}_A$ are $O(n^{-1})$.

To determine $\underset{\sim}{\delta}_B$ we first note from (3.14) and (3.15) that we can write (4.23) as

$$[\tilde{\underset{\sim}{A}} + \underset{\sim}{\delta}_A]\,\tilde{\underset{\sim}{\sigma}} = (m+1)^{-1}\,\tilde{\underset{\sim}{F}}\,[\underset{\sim}{I} + \underset{\sim}{\delta}_B]\,\underset{\sim}{1} \tag{4.24}$$

for $\tilde{\underset{\sim}{F}}=(m+1)\tilde{\underset{\sim}{B}}$, because all terms in $\tilde{\underset{\sim}{B}}$ come from those terms in (3.14) and (3.15) that are multiplied by $1/(m+1)$. As discussed for the iterative estimate, the largest error in approximating terms in (3.14) and (3.15) comes from the $O(n^{-3/2})$ error in approximating $\text{var}(r_{iD}|\underset{\sim}{z})$ and $\text{cov}(r_{iD},r_{hQ}|\underset{\sim}{z})$. As discussed following (4.23), this error contains no terms in $\tilde{\sigma}_{ij}$ and thus is in that part of δ_{ij} that belongs to $\underset{\sim}{\delta}_B$. Since the diagonal terms of $\underset{\sim}{\delta}_B$ are terms from δ_{ij} divided by corresponding diagonal terms of $\tilde{\underset{\sim}{B}}=(m+1)^{-1}\,\tilde{\underset{\sim}{F}}=O(n^{-1})\,\tilde{\underset{\sim}{F}}$, we have that $\delta_{B,ii}$ $= O(n^{-3/2})\,[O(n)\,\tilde{F}_{ii}^{-1}] = O(n^{-1/2})\,\tilde{F}_{ii}^{-1} = O(n^{-1/2})$. Recall that off-diagonal elements of $\underset{\sim}{\delta}_B$ are 0.

Therefore, we can write (4.24) as

$$\begin{bmatrix} \left(\tilde{A}_{ij} + O(n^{-1})\right) \end{bmatrix} \tilde{\underset{\sim}{\sigma}} = (m+1)^{-1}\,\tilde{\underset{\sim}{F}} \begin{bmatrix} & & 0 \\ & 1+O(n^{-1/2}) & \\ 0 & & \end{bmatrix} \underset{\sim}{1}. \tag{4.25}$$

At this point, recalling from equations (3.16) and (3.17) the form of $\tilde{\underset{\sim}{A}}$ and $\tilde{\underset{\sim}{F}}$, we substitute $\dot{\tilde{p}}_i^{(t)}$ for $\tilde{p}_i$ in $\tilde{\underset{\sim}{A}}$ and $\tilde{\underset{\sim}{F}}$, where $\dot{\tilde{p}}_i^{(t)}$ again denotes the converged estimate $\dot{\tilde{p}}_i^{(s)}$ from (3.11). We denote the resulting matrices as $\dot{\tilde{\underset{\sim}{A}}}$ and $\dot{\tilde{\underset{\sim}{F}}}$, respectively. If the two conditions of Appendix 4E are satisfied, so that $\dot{\tilde{p}}_i^{(t)}=\tilde{p}_i+0(n^{-1})$ for $1\leq i\leq k$, then, from Lemma 3B.1, the error in approximating $\tilde{\underset{\sim}{A}}$ and $\tilde{\underset{\sim}{F}}$ by $\dot{\tilde{\underset{\sim}{A}}}$ and $\dot{\tilde{\underset{\sim}{F}}}$, respectively, is $0(n^{-1})$.

In this case, (4.25) can be rewritten as

$$\begin{bmatrix} \left(\dot{\tilde{A}}_{ij}+0(n^{-1})\right) \end{bmatrix} \tilde{\underset{\sim}{\sigma}} = (m+1)^{-1} \begin{bmatrix} \ddots & & 0 \\ & \dot{\tilde{F}}_{ii}+0(n^{-1}) & \\ 0 & & \ddots \end{bmatrix} \begin{bmatrix} \ddots & & 0 \\ & 1+0(n^{-1/2}) & \\ 0 & & \ddots \end{bmatrix} \underset{\sim}{1}. \tag{4.26}$$

To solve for $\tilde{\underset{\sim}{\sigma}}$, we must invert the coefficient matrix of $\tilde{\underset{\sim}{\sigma}}$ in (4.26), which matrix we assume to be nonsingular. To determine the error made by approximating the result by $\dot{\tilde{\underset{\sim}{A}}}^{-1}$, we use the following lemma:

Lemma 4.2: If $\hat{\underset{\sim}{A}}$ and $\underset{\sim}{A}$ are h-dimensional square matrices such that $\hat{\underset{\sim}{A}}=\underset{\sim}{A}+0(q)$ and $\hat{\underset{\sim}{A}}^{-1}$ and $\underset{\sim}{A}^{-1}$ exist, then $\hat{\underset{\sim}{A}}^{-1}=\underset{\sim}{A}^{-1}+0(q)$.

proof:

Define $\hat{A}_{ij}$ and A_{ij} to be the cofactor of $\hat{A}_{ij}$ and A_{ij}, respectively. Then, $\hat{A}_{ij}-A_{ij}=0(q)$ for all i and j implies that $\hat{A}_{ij}-A_{ij}\leq 0(q)$. Thus, from Lemma 3.2, $\hat{A}_{ij}\hat{A}_{ij}-A_{ij}A_{ij}=0(q)$ so that $\det(\hat{\underset{\sim}{A}})-\det(\underset{\sim}{A})=\sum_{j=1}^{h}(\hat{A}_{1j}\hat{A}_{1j}-A_{1j}A_{1j})$ $=0(q)$. Therefore, since a matrix inverse is the transposed matrix of

cofactors divided by the determinant, we have that, for $\tilde{q} \leq q$,

$$\hat{\underset{\sim}{A}}^{-1} = [\det(\hat{\underset{\sim}{A}})]^{-1} (\hat{A}_{ij})' = [\det(\underset{\sim}{A})+O(q)]^{-1} \left(A_{ij}+O(\tilde{q})\right)'$$

$$= \det(\underset{\sim}{A}) (A_{ij})' + O(q)$$

$$= \underset{\sim}{A}^{-1} + O(q).$$

Thus, assuming that $\tilde{A}$ and $\dot{\tilde{A}}$ exist (i.e., their determinants are not zero), solving for $\tilde{\underset{\sim}{\sigma}}$ in (4.26) and applying Lemmas 3B.1 and 4.2, we have that

$$\begin{aligned}
\tilde{\underset{\sim}{\sigma}} &= (m+1)^{-1} \left[\left(\dot{\tilde{A}}_{ij}^{-1}+O(n^{-1})\right)\right] \begin{bmatrix} \ddots & \\ & \dot{\tilde{F}}_{ii}+O(n^{-1}) \end{bmatrix} \begin{bmatrix} \ddots & \\ & 1+O(n^{-1/2}) \end{bmatrix} \underset{\sim}{1} \\
&= (m+1)^{-1} \left[\left(\dot{\tilde{A}}_{ij}^{-1}+O(n^{-1})\right)\right] \begin{bmatrix} \ddots & \\ & \dot{\tilde{F}}_{ii}+O(n^{-1/2}) \end{bmatrix} \underset{\sim}{1} \\
&= (m+1)^{-1} \left[\dot{\underset{\sim}{A}}^{-1} \dot{\underset{\sim}{F}} + \left(O(n^{-1/2})\right)\right] \underset{\sim}{1} \\
&= \dot{\tilde{\underset{\sim}{\sigma}}} + \left(O(n^{-3/2})\right) .
\end{aligned} \tag{4.27}$$

Therefore, for both i and j referring to categories having incomplete data, the errors in approximating the vector of $\tilde{\sigma}_{ij}$ by the procedure that is noniterative in $\tilde{\sigma}_{ij}$ are, like those of the iterative procedure, of order $O(n^{-3/2})$.

4.4 Summary:

In the first part of this chapter, we proved that the posterior central cross-product moments satisfy the conditions for a first-order Taylor-series expansion to be an accurate approximation of the exact posterior mean. We also proved that asymptotically the Taylor-series approximate posterior mean equals the exact posterior mean.

In the second part of the chapter, we studied how fast the Taylor-series approximate posterior mean approaches the limit, the exact posterior mean, and then investigated the accuracy of the Taylor-series approximate posterior variance and covariance. We began by showing that the Taylor-series expansions for elements of the exact posterior mean and covariance matrices are accurate to order $O(n^{-1})$ and $O(n^{-3/2})$, respectively. However, because the exact posterior moments in these expansions are then approximated, the errors in the final approximations, which we called the Taylor-series approximations, are not necessarily of magnitude $O(n^{-1})$ and $O(n^{-3/2})$, respectively.

Nearly always, the Taylor-series approximate posterior mean will be evaluated iteratively. For this type of evaluation, we gave two sufficient conditions guaranteeing accuracy of the Taylor-series approximate posterior mean to the exact posterior mean within order of magnitude $O(n^{-1})$. If there exists a neighborhood $\|\dot{\tilde{p}}-\tilde{p}\|_\infty<\rho$, for $\rho>0$, of $\tilde{p}$ such that for all probabilities $\dot{\tilde{p}}$ in this neighborhood

$$\max_{1\leq i\leq k} \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial \dot{\tilde{p}}_j| \leq \lambda < 1,$$

for

$$g_i(\dot{\tilde{p}})=(z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/(n+\sum_{h=1}^{k+1}\nu_h),$$

and if an initial iterative estimate $\dot{\tilde{p}}_i^{(0)}$ is chosen within the inner neighborhood $\|\dot{\tilde{p}}-\tilde{p}\|_\infty<\rho_0\leq\rho-\delta/(1-\lambda)$ where δ is a bound on the error in approximating the exact posterior mean by a first-order Taylor-series expansion, then the iterative solution to the defining equation of the Taylor-series approximate posterior mean $\dot{\tilde{p}}$ will converge to within $O(n^{-1})$ of the exact posterior mean. We also showed how to determine, in practice, whether these conditions can be expected to hold.

Further, for incomplete trinomial data, we showed that there does exist a root in P_2 of the defining equations for the Taylor-series approximate posterior mean that differs from the exact posterior mean by magnitude $O(n^{-1})$. We then investigated when the Taylor-series approximation for incomplete data from the general k-dimensional multinomial distribution has a solution that differs from the exact posterior mean by magnitude $O(n^{-1})$. Because the Taylor-series approximate posterior mean can be written as a posterior mode, it always has at least one solution in P_k if the eigenvalues of the covariance matrix of the complete-data sufficient statistics are bounded above zero. However, none of these solutions may be in the convergence region for the exact posterior mean ("epm convergence region"). In particular, for k>2, there may not exist an epm convergence region and we gave an example of such a case. In this example, many components of $\tilde{p}$ were very small. Since results of Chapters 6 and 7 indicate that the posterior mean is a poor estimator to use to minimize risk for quadratic loss when components of $\tilde{p}$ are very small, the posterior mode being

much better, absence of an epm convergence region was considered unimportant for this particular cause (because the posterior mean would not be calculated).

When there does exist an epm convergence region, there can be trouble finding it, because there may be multiple roots in P_k of the defining equations for the Taylor-series approximate posterior mean. The problem then is choosing among these roots. We showed how to find one choice, the global maximum. Although it was not proved, from the complete-data relationship between the posterior mode and posterior mean, we intuitively expect the global maximum to be in the epm convergence region, or at least be the closest root to $\tilde{p}$.

We also noted that the two guaranteed-convergence conditions, conditions given by Lemma 4E.1 on the initial iterative estimate and on the partial derivatives of the posterior mean, are sufficient but not necessary. We gave two illustrations in Appendix 4E where these conditions were not met but the iterates correctly converged. Further, as also illustrated, the error bound given by Lemma 4E.1 is extremely conservative.

Finally, for those categories having only complete data, there is no error in using the Taylor-series approximation for the exact posterior mean.

Recall that elements of the Taylor-series approximate posterior covariance matrix can be evaluated by procedures that are noniterative or iterative in elements of the posterior covariance matrix. The Taylor-series approximate posterior mean is used in both procedures. When the error in the Taylor-series approximate posterior mean is $O(n^{-1})$,

then the noniterative procedure yields approximations for elements of the posterior covariance matrix that are accurate to order $O(n^{-3/2})$. If, in addition to the $O(n^{-1})$ accuracy in the Taylor-series approximate posterior mean, parallel conditions given in Appendix 4E are met for the Taylor-series approximate covariances, then the iterative procedure also gives approximations for elements of the posterior covariance matrix that are accurate to order $O(n^{-3/2})$. Under these same conditions, when one of categories i and j has no incomplete data, then the error in the Taylor-series approximate variance and covariance is $O(n^{-2})$. For both i and j having only complete data, there is no error in approximating the exact posterior variance and covariance by the Taylor-series approximate posterior variance and covariance, respectively.

APPENDIX 4A

POSTERIOR CENTRAL MOMENTS GIVEN COMPLETE DATA

4A.1 Introduction:

In this appendix, we determine orders of magnitude for the posterior central moments. To do so, we prove by induction an expression for the lowest-order term in $(n+\sum_{h=1}^{k+1}\nu_h)^{-1}$ of the l^{th} posterior central moment $E[(p_i-\mu_i)^l|\underset{\sim}{x}]$. We first determine, in Section 4A.3, the expression for the first twenty-one central moments, enough moments to determine an algebraic pattern. Then, in Section 4A.4, we extend moment results from Kendall and Stuart (1969,v1,p148-150) for Pearson distributions, proving that if the expression is true for any two successive values of l, it must also be true for the next higher value of l.

We conclude Appendix 4A in Section 4A.5 by generalizing this method to cross-product moments. Order-of-magnitude results are given for forty-nine cases. Because of the variety and complexity of possible results for the lowest-order term in $(n+\Sigma\nu_h)^{-1}$, we do not further use this method. Instead, we describe a different approach in Section 4.2.1 of the main text. Although the different approach gives orders of magnitude for even cross-product moments, it gives only bounds for odd cross-product moments. Hence, results of Section 4A.5 are especially important for odd cross-product moments.

In Section 4A.2 we describe a symbolic computer system used to facilitate algebraic operations in the last three sections.

Remark: The usual procedure to calculate moments is through the

characteristic (moment-generating) function, cumulants, or factorial moments. However, in this case, calculation of the posterior central moments (2.6) was easiest done directly. As might be unsurprising in such case, none of the three usual procedures aided in obtaining the limit of these moments. The moment-generating function led directly to expression (2.6) for the lth posterior central moments; that is, differentiating $\exp(-\underset{\sim}{t}'\underset{\sim}{\mu})\phi(\underset{\sim}{t})$ with respect to $\underset{\sim}{t}$, for $\phi(\underset{\sim}{t})$ the moment-generating function, and setting results to 0 gives (2.6). Thus, use of the moment-generating function was not helpful in reexpressing (2.6) to obtain its limit. Calculation of the logarithm of the moment-generating function to obtain the cumulants (for purpose of translation back to the central moments) also did not aid in obtaining the limit of the lth central moment (2.6). Consideration of factorial moments, often useful for discrete distributions, was unfruitful for this continuous distribution.

4A.2 Symbolic Computer System:

In this section we describe a computer system used to facilitate algebraic operations in the remaining three sections. In Section 4A.3 we use this system to expand the first twenty-one central moments, $E[(p_i-\mu_i)^l|\underset{\sim}{x}]$ for $1\leq l\leq 21$, in a Taylor series in $(n+\Sigma\nu_h)^{-1}$ about the point $(n+\Sigma\nu_h)^{-1}=0$. In Section 4A.4 we use the computer system to algebraically solve in terms of $(n+\Sigma\nu_h)^{-1}$ and $\mu_i\equiv E(p_i|\underset{\sim}{x})$ a system of four equations in four unknowns to enable, for all l, the $(l+1)^{st}$ central moment to be written in terms of the two preceding, l^{th} and $(l-1)^{st}$, moments. In Section 4A.5, the computer system facilitates evaluation of cross-product moments $E[(p_i-\mu_i)^l(p_j-\mu_j)^h|\underset{\sim}{x}]$ for $2\leq l,h\leq 8$.

The computer system used is MACSYMA* (Project MAC's SYmbolic MAnipulation System), developed by the Mathlab Group, Project MAC at M.I.T. (Massachusetts Institute of Technology). MACSYMA is a versatile interactive computer system for manipulating algebraic or symbolic expressions as well as for performing high-precision numerical calculations. MACSYMA is written in LISP (a list procession programing language used for non-numerical applications) for a Digital Equipment Corporation PDP-10 computer with a KL10 processor and 500k 36-bit words of memory. The PDP-10 computer is located at the Laboratory for Computer Science at M.I.T. and is known as the MC (MACSYMA CONSORTIUM) computer. A large variety of computer terminals at NASA, Langley Research Center, allow access to MACSYMA.

---

*This work is supported by the Defense Advanced Research Projects Agency work order 2095, under Office of Naval Research Contract #N00014-75-C-0661.

MACSYMA can algebraically differentiate and integrate analytic expressions, take limits, solve systems of linear or polynomial equations, expand functions in Taylor series, manipulate matrices and tensors, factor complicated polynomials in many variables, plot functions, and calculate Laplace transforms. The system has "built-in knowledge" of many commonly used mathematical functions. Operations are done in rational, not floating-point, arithmetic. Thus, round-off error does not exist. Additional information can be found in MACSYMA manuals (1975a,1975b,1976) by the Mathlab Group at M.I.T.

4A.3 Derivation of General Expression:

In this section we determine an expression for the lowest-order term in $(n+\sum_{h=1}^{k+1}\nu_h)^{-1}$ of the first twenty-one posterior central moments. We do so by writing the l^{th} central moment [recall (2.6)]

$$E[(p_i-\mu_i)^l|\underset{\sim}{x}] = \left[\sum_{j=0}^{l}(-1)^{l-j}\binom{l}{j}\left(\frac{x_i+\nu_i}{n+\Sigma\nu_h}\right)^{l-j}\prod_{q=0}^{j-1}\frac{x_i+\nu_i+q}{n+\Sigma\nu_h+q}\right] \tag{4A.1}$$

in a Taylor series in $(n+\sum_{h=1}^{k+1}\nu_h)^{-1}$ about the point $(n+\sum_{h=1}^{k+1}\nu_h)^{-1}=0$.

Recall from (2.2) that

$$\mu_i = (x_i+\nu_i)/(n+\sum_{h=1}^{k+1}\nu_h) \tag{4A.2}$$

and let

$$l!! = l(l-2)(l-4)(l-6)\ldots 1 \quad \text{for } l \text{ odd}^*, \tag{4A.3}$$

$$r = (n+\sum_{h=1}^{k+1}\nu_h)^{-1}, \tag{4A.4}$$

$$s_i = [n+\sum_{h=1}^{k+1}\nu_h-(x_i+\nu_i)]/(x_i+\nu_i)$$

$$= (1-\mu_i)/\mu_i, \tag{4A.5}$$

and

$$y_i = p_i-\mu_i. \tag{4A.6}$$

Then

*Standard mathematical notation. For example, see Gradshteyn and Ryzhik (1967,pxliii). Note that $l!!$ is not defined for l even.

$$s_i+1 = \mu_i^{-1}, \tag{4A.7}$$

$$s_i^2-1 = (1-2\mu_i)/\mu_i^2, \tag{4A.8}$$

$$s_i/(s_i+1)^2 = \mu_i(1-\mu_i), \tag{4A.9}$$

and the variance is given by

$$\sigma_{ii} = rs_i/[(1+r)(s_i+1)^2] = r\mu_i(1-\mu_i) \tag{4A.10}$$

since

$$r/(1+r) = 1/(n+\Sigma\nu_h+1). \tag{4A.11}$$

Rewrite the l^{th} central moment (4A.1) as

$$\begin{aligned} E(y_i^l|\underset{\sim}{x}) &= \mu_i^l \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} \prod_{q=0}^{j-1} \frac{1+q/(x_i+\nu_i)}{1+q/(n+\Sigma\nu_h)} \\ &= \mu_i^l \sum_{j=0}^{l} (-1)^{l-j} \binom{l}{j} \prod_{q=0}^{j-1} [1+s_i q \frac{r}{1+qr}] \end{aligned} \tag{4A.12}$$

since

$$\begin{aligned} \frac{1+q/(x_i+\nu_i)}{1+q/(n+\Sigma\nu_h)} &= 1+ \frac{q[-1/(n+\Sigma\nu_h)+1/(x_i+\nu_i)]}{1+q/(n+\Sigma\nu_h)} \\ &= 1 + \frac{q}{n+\Sigma\nu_h} \left[\frac{(n+\Sigma\nu_h)-(x_i+\nu_i)}{x_i+\nu_i}\right] /[1+(n+\Sigma\nu_h)] \\ &= 1+s_i q \frac{r}{1+qr}. \end{aligned} \tag{4A.13}$$

Define

$$f(r) = r/(1+qr). \tag{4A.14}$$

Expanding $f(r)$ in a Taylor series in r about the point $r=0$ yields that

$$r/(1+qr) = \sum_{j=1}^{\infty} (-1)^{j-1} q^{j-1} r^{j}. \tag{4A.15}$$

Substituting (4A.15) into (4A.12) and, by using MACSYMA, expanding $\mu_i^{\,l} E(y_i^{\,l} | \underset{\sim}{x})$ in a Taylor series in r about the point $r=0$ yields for low-order terms for the first twenty-one central moments results given in Table 4A.1. Note that all results must be multiplied by $\mu_i^{\,l}$. To get the lowest-order term in r, we discard all those terms in the inner-most set of parenthesis, except for cases $l=2$ and $l=3$. The following pattern is detected for $1 \leq l \leq 21$:

$$E(y_i^{\,l} | \underset{\sim}{x}) \doteq \begin{cases} (l-1)!! s_i^{\,l/2} r^{l/2} \mu_i^{\,l} & \text{for } l \text{ even} \\ (l-1)l!!(s_i-1) s_i^{\,(l-1)/2} r^{(l+1)/2} \mu_i^{\,l}/3 & \text{for } l \text{ odd,} \end{cases} \tag{4A.16}$$

where the approximation "$\doteq$" in (4A.16) means that only the lowest-order term in r is given. As a check on these formulas, note that for $l=20$ and $l=21$, (4A.16) yields $19!! s_i^{\,10} r_i^{\,10} \mu_i^{\,20}$ and $(20\times 21!!/3)(s_i-1) s_i^{\,10} \times r_i^{\,11} \mu_i^{\,21}$, respectively, both of which agree with results in Table 4A.1.

To simplify results we multiply numerator and denominator of (4A.16) by $(1+r)^{l/2}$ for l even and by $(1+r)^{(l+1)/2}$ for l odd. Then, substituting into (4A.16) from (4A.2) - (4A.10) and again giving only the lowest-order term in $r=1/(n+\Sigma\nu_h)$ yields that, for $1 \leq l \leq 21$,

$$E(y_i^{\,l} | \underset{\sim}{x}) \doteq \begin{cases} (l-1)!! \sigma_{ii}^{\,l/2} & \text{for } l \text{ even} \\ (l-1)l!!(1-2\mu_i) \sigma_{ii}^{\,(l+1)/2} / [3\mu_i(1-\mu_i)] & \text{for } l \text{ odd.} \end{cases} \tag{4A.17}$$

TABLE 4A.1

LOW-ORDER TERMS* FOR FIRST 21 CENTRAL MOMENTS $E[(p_i-\mu_i)^{\ell}|\underset{\sim}{x}]$

ℓ	low-order term
2	$r\ s_i\ (r-1)$
4	$3\ r^2\ s_i\ (r(2s_i^2 - 8s_i + 2) + s_i)$
6	$5\ r^3\ s_i^2\ (r(26s_i^2 - 79s_i + 26) + 3s_i)$
8	$35\ r^4\ s_i^3\ (r(68s_i^2 - 184s_i + 68) + 3s_i)$
10	$315\ r^5\ s_i^4\ (r(140s_i^2 - 355s_i + 140) + 3s_i)$
12	$3465\ r^6\ s_i^5\ (r(250s_i^2 - 608s_i + 250) + 3s_i)$
14	$45045\ r^7\ s_i^6\ (r(406s_i^2 - 959s_i + 406) + 3s_i)$
16	$675675\ r^8\ s_i^7\ (r(616s_i^2 - 1424s_i + 616) + 3s_i)$
18	$34459425\ r^9\ s_i^8\ (r(296s_i^2 - 673s_i + 296) + s_i)$
20	$654729075\ r^{10}\ s_i^9\ (r(410s_i^2 - 920s_i + 410) + s_i)$
3	$-2\ r^2\ (s_i-1)\ s_i\ (3r-1)$
5	$4\ r^3\ (s_i-1)\ s_i\ (r(6s_i^2 - 50s_i + 6) + 5s_i)$
7	$42\ r^4\ (s_i-1)\ s_i^2\ (r(22s_i^2 - 115s_i + 22) + 5s_i)$
9	$56\ r^5\ (s_i-1)\ s_i^3\ (r(472s_i^2 - 1970s_i + 472) + 45s_i)$
11	$770\ r^6\ (s_i-1)\ s_i^4\ (r(916s_i^2 - 3335s_i + 916) + 45s_i)$
13	$60060\ r^7\ (s_i-1)\ s_i^5\ (r(314s_i^2 - 1042s_i + 314) + 9s_i)$
15	$210210\ r^8\ (s_i-1)\ s_i^6\ (r(2474s_i^2 - 7675s_i + 2474) + 45s_i)$
17	$4084080\ r^9\ (s_i-1)\ s_i^7\ (r(3668s_i^2 - 10810s_i + 3668) + 45s_i)$
19	$87297210\ r^{10}\ (s_i-1)\ s_i^8\ (r(5192s_i^2 - 14695s_i + 5192) + 45s_i)$
21	$6110804700\ r^{11}\ (s_i-1)\ s_i^9\ (r(2362s_i^2 - 6470s_i + 2362) + 15s_i)$

* all results must be multiplied by μ_i^{ℓ}

4A.4 Validity of General Expression:

In the last section we derived expression (A.17) for the l^{th} posterior moment for $1 \leq l \leq 21$, hence proving the expression true for these values of l. In this section, we prove that if the expression is true for any two successive values of l, it must also be true for the next higher value of l. Having done so, we will have proved that expression (A.17) holds for all positive integer values of l.

In (4A.1) we are calculating the l^{th} posterior central moment of p_i. Since the posterior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{x}$ is the k-dimensional Dirichlet $D(x_1+\nu_1,\ldots,x_k+\nu_k;x_{k+1}+\nu_{k+1})$, then the marginal posterior distribution of p_i for $1 \leq i \leq k$ is the one-dimensional Dirichlet $D(x_i+\nu_i; \sum_{j\neq i}^{k+1}(x_j+\nu_j))$ or beta $Be(x_i+\nu_i, \sum_{j\neq i}^{k+1}(x_j+\nu_j))$. [See Wilks (1963, p173-179).] That is, the posterior density of p_i given $\underset{\sim}{x}$ is

$$f(p_i|\underset{\sim}{x}) = \Gamma[\sum_{j=1}^{k+1}(x_j+\nu_j)]/\{\Gamma(x_i+\nu_i)\Gamma[\sum_{j\neq i}^{k+1}(x_j+\nu_j)]\}p_i^{x_i+\nu_i-1}(1-p_i)^{\sum_{j\neq i}^{k+1}(x_j+\nu_j)-1}. \quad (4A.18)$$

Now, the beta distribution is known as one of the Pearson distributions [Kendall and Stuart (1969,v1,p148)]. A Pearson distribution is defined as any frequency function f(w) for which

$$df(w)/dw = (w-a)f(w)/(b_0+b_1w+b_2w^2) \quad (4A.19)$$

for some a, b_0, b_1, and b_2. Kendall and Stuart derive the general moment for a Pearson distribution in terms of lower-order moments. We now generalize their method to the case of central moments. Note that we treat the most common case, f(0)=f(1)=0. However, results also hold

when one or both of f(0) and f(1) are not zero. Thus, results also hold for J-shaped, U-shaped, and flat beta distributions. [See also Kendall and Stuart (1969,v1,p151).]

Therefore, cross multiplying (4A.19), adding and subtracting powers of $\eta=E(w)$, and multiplying both resulting sides by $(w-\eta)^l$ yields that

$$(w-\eta)^l\ [(b_0+\eta b_1+\eta^2 b_2)+(b_1+2\eta b_2)(w-\eta)+b_2(w-\eta)^2]\ df(w)/dw\ dw = (w-\eta)^l\ [(w-\eta)+(\eta-a)]\ f(w)\ dw. \tag{4A.20}$$

Integrating the left-hand side of (4A.20) by parts over the range of the distribution, we find, assuming that the integrals exist, that

$$(w-\eta)^l\ \{[(b_0+\eta b_1+\eta^2 b_2)+(b_1+2\eta b_2)(w-\eta)+b_2(w-\eta)^2]\ f(w)\ \}_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(w)\ \{l \times(b_0+\eta b_1+\eta^2 b_2)(w-\eta)^{l-1}+(l+1)(b_1-2\eta b_2)(w-\eta)^l+(l+2)b_2(w-\eta)^{l+1}\}\ dx \tag{4A.21}$$
$$= \int_{-\infty}^{\infty} (w-\eta)^{l+1}\ f(w)\ dw + (\eta-a) \int_{-\infty}^{\infty} (w-\eta)^l\ f(w)\ dw.$$

For the beta density (4A.18), $f(p_i|\underset{\sim}{x})$ is positive for $0\leq p_i\leq 1$. Thus, in equation (4A.21) we replace endpoints $-\infty$ and $+\infty$ by 0 and 1, respectively, and w, η, and f(w) by $p_i|\underset{\sim}{x}$, μ_i, and $f(p_i|\underset{\sim}{x})$, respectively. We then note that, since $f(1|\underset{\sim}{x})=f(0|\underset{\sim}{x})=0$ and, for any positive integer j,

$$\lim_{p_i\to+1} p_i^j=1 \text{ and } \lim_{p_i\to+0} p_i^j=0,$$

$$\lim_{p_i\to\begin{cases}+1\\+0\end{cases}} p_i^j\ f(p_i|\underset{\sim}{x}) = 0. \tag{4A.22}$$

Therefore, the first term in equation (4A.21) is 0.

Hence, recalling from Section 4A.3 the definition $y_i=p_i-\mu_i$, we can write (4A.21) as

$$l[b_0+b_1\mu_i+\mu_i^2 b_2]\ E(y_i^{l-1}|\underset{\sim}{x}) + [(l+1)(b_1-2b_2\mu_i)+\mu_i-a]\ E(y_i^{l}|\underset{\sim}{x}) + [(l+2)b_2+1]\ E(y_i^{l+1}|\underset{\sim}{x}) = 0. \tag{4A.23}$$

Thus, if we knew b_0, b_1, b_2, and a we could use (4A.23) to calculate any $(l+1)^{st}$ central moment in terms of the l^{th} and $(l-1)^{st}$ central moments. To calculate b_0, b_1, b_2, and a, we successively let $l=0$, 1, 2, and 3 in (4A.23), substitute results from Section 4A.3 for $E(y_i^j|\underset{\sim}{x})$ for $2\leq j\leq 4$, and set $E(y_i^{-1}|\underset{\sim}{x})=0$, $E(y_i^0|\underset{\sim}{x})=1$, and $E(y_i^1|\underset{\sim}{x})=0$ to obtain four equations in four unknowns b_0, b_1, b_2, and a. Solving these four equations with MACSYMA yields

$$\begin{aligned} b_0 &= 4r/[(2r-1)(s_i+1)^2], \\ b_1 &= r(s_i-3)/[(2r-1)(s_i+1)], \\ b_2 &= -\,r/(2r-1), \end{aligned} \tag{4A.24}$$

and

$$a = (rs_i+r-1)/[(2r-1)(s_i+1)].$$

Substituting results (4A.24) into equation (4A.23) and collecting terms in $E(y_i^{l-1}|\underset{\sim}{x})$, $E(y_i^{l}|\underset{\sim}{x})$, and $E(y_i^{l+1}|\underset{\sim}{x})$ yields that

$$E(y_i^{l+1}|\underset{\sim}{x}) = lr\,[(s_i^2-1)\ E(y_i^{l}|\underset{\sim}{x}) + s_i E(y_i^{l-1}|\underset{\sim}{x})]/[(1+lr)(s_i+1)^2]. \tag{4A.25}$$

Therefore, we can use (4A.25) to show that if expression (4A.17) holds for the l^{th} and $(l-1)^{st}$ central moments, it must also hold for

the $(l+1)^{st}$ moment. Because we have different expressions for l even and l odd we have two cases. By using (4A.17), (4A.25), and (4A.2) - (4A.10), we have that, to the lowest-order term in r:

for $l+1$ even:

$$\begin{aligned} E(y_i^{l+1}|\underset{\sim}{x}) &\doteq l r[(s_i^2-1)(l-1)l!!(1-2\mu_i)/3\ r\sigma_{ii}^{(l-1)/2}(1+r)^{(l-1)/2} \\ &\quad +s_i(l-2)!!\sigma_{ii}^{(l-1)/2}(1+r)^{(l-1)/2}]/[(lr+1)(s_i+1)^2] \\ &\doteq l!!\sigma_{ii}^{(l-1)/2}(1+r)^{(l-1)/2}\{l(l-1)r^2(1-2\mu_i)^2/3+r\mu_i)\}/(1+lr) \\ &\doteq l!!\sigma_{ii}^{(l+1)/2}(1+r)^{(l+1)/2}\left(1+\{l(l-1)(1-2\mu_i)^2/[3\mu_i(1-\mu_i)]\}r\right)/(1+lr) \\ &\doteq l!!\sigma_{ii}^{(l+1)/2}{}^{*} \end{aligned} \tag{4A.26}$$

for $l+1$ odd:

$$\begin{aligned} E(y_i^{l+1}|\underset{\sim}{x}) &\doteq l r[(s_i^2-1)(l-1)!!\sigma_{ii}^{l/2}(1+r)^{l/2}+s_i(l-2)(l-1)!!(1-2\mu_i)r/3 \\ &\quad \times\sigma_{ii}^{(l-2)/2}(1+r)^{(l-2)/2}]/[(lr+1)(s_i+1)^2] \\ &\doteq l(l-1)!!\sigma_{ii}^{l/2}(1+r)^{l/2}[r(1-2\mu_i)+(l-2)(1-2\mu_i)r/3]/(1+lr) \\ &\doteq l(l-1)!!\sigma_{ii}^{l/2}(1+r)^{l/2}(1-2\mu_i)r/3\ (3+l-2)/(1+lr) \\ &\doteq l(l+1)!!\sigma_{ii}^{(l+2)/2}(1-2\mu_i)/[3\mu_i(1-\mu_i)]^{*}. \end{aligned} \tag{4A.27}$$

Therefore, from results of Sections 4A.3 and 4A.4, expression (4A.17) is true for all positive integer values of l.

---

*These expressions are actually divided by a finite constant $c(l)$ where

$$c(l) = \begin{cases} 1 & \text{if } l<n+\Sigma\nu_h \\ 2 & \text{if } l=n+\Sigma\nu_h \\ >2 & \text{if } l>n+\Sigma\nu_h. \end{cases}$$

The constants $c(l)$ arise from evaluation of the term $\xi=(1+r)^{(l+1)/2}/(1+lr)$. Since $r=1/(n+\Sigma\nu_h)$, then $r<1$. Hence, the numerator of ξ can be accurately approximated by the first two terms of the series expansion

$$(1+r)^{(l+1)/2} = \sum_{j=0}^{\infty} \binom{(l+1)/2}{j} r^j/j!$$

$$= 1+[(l+1)/2]r+[(l+1)/2][(l+1)/2-1]r^2/2!+\cdots.$$

When $lr<1$ (i.e., $l<n+\Sigma\nu_h$), then the term $1/(1+lr)$ in ξ can also be accurately approximated by the first two terms $1-lr$ of a series expansion $1-lr+(lr)^2-(lr)^3+\cdots$. In this case, ξ can be accurately approximated by $1+[(l+1)/2-l]r$, the low-order term in r resulting from the multiplication of the two series. Therefore, in this case of $lr<1$, expressions (4A.26) and (4A.27) are correct as given. When $lr=1$ (i.e., $l=n+\Sigma\nu_h$), however, then $\xi=(1+r)^{(l+1)/2}=\{1+[(l+1)/2]r\}/2$ and expressions (4A.26) and (4A.27) must be divided by 2. When $lr>1$ (i.e., $l>n+\Sigma\nu_h$), then $\xi>(1+r)^{(l+1)/2}/2$ so that expressions (4A.26) and (4A.27) must be divided by some constant larger than 2.

However, the interest in Chapter 4 is in very large n; in particular, the limiting case $n\to\infty$. For these cases, $l<n<n+\Sigma\nu_h$ and $c(l)=1$. Therefore, to avoid carrying around a term that is 1 in the cases in which we are interested, we do not include it. Further, the limit taken in (4.3) in the main text is not affected by $c(l)$.

4A.5 Cross-product Moments

The method of the preceding sections readily extends to cross-product central moments $E(\prod_{g=1}^{l} y_{i_g} | \underset{\sim}{x})$. We can write these cross-product moments as a nest of expressions where each expression is similar in form to the l^{th} central moment (4A.12) with the exception that each term of the sum is multiplied by results of inner nests.

For example, for $1 \leq i,j \leq k$; $j \neq i$, l, h positive integers; and, again $\mu_i = (x_i+\nu_i)/(n+\Sigma\nu_h)$ and $y_i^{\,l} = [p_i - \mu_i]^l$; we have that

$$E(y_i^{\,l} y_j^{\,h} | \underset{\sim}{x}) = E\left\{\left[\sum_{a=0}^{l} (-1)^{l-a} \binom{l}{a} \mu_i^{\,l-a} p_i^{\,a}\right] \times \left[\sum_{b=0}^{h} (-1)^{h-b} \binom{h}{b} \mu_j^{\,h-b} p_j^{\,b}\right] | \underset{\sim}{x}\right\}$$

$$= \mu_i^{\,l} \mu_j^{\,h} \sum_{a=0}^{l} \sum_{b=0}^{h} (-1)^{l-a}(-1)^{h-b} \binom{l}{a} \binom{h}{b} \tag{4A.28}$$

$$\times \left(\frac{x_i+\nu_i}{n+\Sigma\nu_h}\right)^{-a} \left(\frac{x_j+\nu_j}{n+\Sigma\nu_h}\right)^{-b} \times \frac{\prod_{q_i=0}^{a-1} (x_i+\nu_i+q_i) \prod_{q_j=0}^{b-1} (x_j+\nu_j+q_j)}{\prod_{q_b=0}^{a+b-1} (n+\Sigma\nu_h+q_b)}$$

since

$$E(p_i^{\,a} p_j^{\,b} | \underset{\sim}{x}) = \frac{\Gamma(x_i+\nu_i+a)\Gamma(x_j+\nu_j+b)\Gamma(n+\Sigma\nu_h)}{\Gamma(x_i+\nu_i)\Gamma(x_j+\nu_j)\Gamma(n+\Sigma\nu_h+a+b)}. \tag{4A.29}$$

In (4A.28) we again use the convention that $\prod_{q=0}^{-1} f(q) = 1$ for any function f of q. Now, we can write the last two lines of (4A.28) as

$$\frac{\prod_{q_i=0}^{a-1}\left[1+\frac{q_i}{x_i+\nu_i}\right]\prod_{q_j=0}^{b-1}\left[1+\frac{q_j}{x_j+\nu_j}\right]}{\prod_{q_i=0}^{a-1}\left[1+\frac{q_i}{n+\Sigma\nu_h}\right]\prod_{q_b=a}^{a+b-1}\left[1+\frac{q_b}{n+\Sigma\nu_h}\right]}$$

$$= \left(\prod_{q_i=0}^{a-1}\frac{1+q_i/(x_i+\nu_i)}{1+q_i/(n+\Sigma\nu_h)}\right)\left(\prod_{q_j=0}^{b-1}\frac{1+q_j/(x_j+\nu_j)}{1+(a+q_j)/(n+\Sigma\nu_h)}\right) \tag{4A.30}$$

$$= \left(\prod_{q_i=0}^{a-1}\left[1+s_i q_i\,\frac{r}{1+q_i r}\right]\right)\left(\prod_{q_j=0}^{b-1}\left[1+(s_j q_j-a)\,\frac{r}{1+(q_j+a)r}\right]\right)$$

where, again,

$$r = (n+\Sigma\nu_h)^{-1}.$$

and

$$s_j = [(n+\Sigma\nu_h)-(x_j+\nu_j)]/(x_j+\nu_j) = (1-\mu_i)/\mu_i.$$

The first term of the last line of (4A.30) was derived in (4A.13) and the second term has a similar derivation.

Therefore, we can write (4A.28) as

$$E(y_i^{\,l}y_j^{\,h}|\underset{\sim}{x}) = \mu_i^{\,l}\mu_j^{\,h}\sum_{a=0}^{l}(-1)^{l-a}\binom{l}{a}\prod_{q_i=0}^{a-1}\left[1+s_i q_i\,\frac{r}{1+q_i r}\right]$$

$$\times\left\{\sum_{b=0}^{h}(-1)^{h-b}\binom{h}{b}\prod_{q_j=0}^{b-1}\left[1+(s_j q_j-a)\,\frac{r}{1+(q_j+a)r}\right]\right\} \tag{4A.31}$$

Using MACSYMA, we can evaluate the (4A.31) factor in braces for enough values of h to establish a pattern for the low-order term in r for h even and odd. We can then use the method of Section 4A.4 to show that this pattern is valid for all values of h. The procedure can then

be repeated for the remaining factor of (4A.31).

Because we will have cross multiplication between the two factors of (4A.31), however, we must know not only the lowest-order term in r but also the next lowest-order term in r. In general, for each additional variable y_m in (4A.31) we must know an additional low-order term in r.

Further, the two lowest-order terms in r for the (4A.31) factor in braces will be a function of "a" from the first factor, so the final result for (4A.31) in terms of the low-order term in r will be more complex than that of (4A.17).

Therefore, the variety of possible results and greater complexity of intermediate evaluations, especially those of pattern recognition and algebraic manipulations, make this method generally unfeasible for cross-product moments. Hence, we adopt another approach, to be discussed in the main text, to evaluate the magnitude of cross-product moments.

We conclude this section by noting that evaluation of (4A.31) for $2 \leq 1$, $h \leq 8$ yields that

$$E\{[p_i-\mu_i]^1[p_j-\mu_j]^h|\underset{\sim}{x}\} = \begin{cases} O(n^{-(1+h)/2}) & \text{for } 1+h \text{ even} \\ O(n^{-(1+h+1)/2}) & \text{for } 1+h \text{ odd.} \end{cases} \quad (4A.32)$$

APPENDIX 4B

LIMITING POSTERIOR DISTRIBUTION GIVEN COMPLETE DATA

Cox and Hinkley (1974,p399) prove in general that when the data has an exponential-family distribution and the conjugate prior is used, the limiting posterior distribution is multivariate normal with the vector of maximum likelihood estimates for the mean. The inverse covariance matrix of this limiting distribution is the negative matrix of second partial derivatives of the log likelihood evaluated at the maximum likelihood estimates. In this appendix we prove this theorem in detail for our complete-data case where the data has a multinomial distribution and the conjugate prior, the Dirichlet, is used.

From (2.1) the posterior distribution of $\underset{\sim}{p}$ given complete data is k-dimensional Dirichlet. Therefore, we prove that the limiting Dirichlet is k-dimensional multivariate normal with mean and covariance matrices those of the Dirichlet. We proceed by proving that the log Dirichlet converges to the log multivariate normal as the sample size indefinitely increases. Important aids will be Stirling's approximation for the logarithm of the gamma function and theorems from Graybill (1969) on patterned matrices.

From (2.1) the posterior density $f(\underset{\sim}{p}|\underset{\sim}{x})$ of the k-dimensional variable $\underset{\sim}{p}$ given complete data $\underset{\sim}{x}$ is, in the notation of Wilks (1963,p178), that for the Dirichlet distribution $D(x_1+\nu_1,\ldots,x_k+\nu_k;\ x_{k+1}+\nu_{k+1})$; i.e.,

$$f(\underset{\sim}{p}|\underset{\sim}{x}) = \{\Gamma[\sum_{h=1}^{k+1}(x_h+\nu_h)]/\prod_{h=1}^{k+1}\Gamma(x_h+\nu_h)\}\prod_{h=1}^{k+1}p_h^{x_h+\nu_h-1}, \qquad (4B.1)$$

where x_h is the number of observations falling in category C_h, ν_h is the k real, positive parameter for the prior density (1.1) of $\underset{\sim}{p}$, and $p_{k+1}=1-\sum_{h=1}^{k}p_h$.

As the sample size $n=\sum_{h=1}^{k+1} x_h$ increases, x_i/n approaches a constant and ν_i/n approaches zero.

For the Dirichlet $D(x_1+\nu_1,\ldots,x_k+\nu_k;x_{k+1}+\nu_{k+1})$, recall (2.2) - (2.4) that the mean vector $\underset{\sim}{\mu}$ of $\underset{\sim}{p}|\underset{\sim}{x}$ has elements, for $1\leq i\leq k$,

$$\mu_i = (x_i+\nu_i)/(n+\sum_{h=1}^{k+1}\nu_h) \tag{4B.2}$$

and that the covariance matrix $\underset{\sim}{\Sigma}=(\sigma_{ij})$ of $\underset{\sim}{p}|\underset{\sim}{x}$ has elements, for $1\leq i\leq k$,

$$\sigma_{ii} = \mu_i(1-\mu_i)/(n+\sum_{h=1}^{k+1}\nu_h+1) \tag{4B.3}$$

and, for $i<j\leq k$,

$$\sigma_{ij} = \mu_i\mu_j/(n+\sum_{h=1}^{k+1}\nu_h+1). \tag{4B.4}$$

Now, $(n+\Sigma\nu_h+1)\ \underset{\sim}{\Sigma}$ is of a matrix pattern treated by Graybill (1969), who gives its determinant and inverse. Applying Theorems 1.5.4 and 8.4.3 of Graybill (1969,p8,184) to $(n+\Sigma\nu_h+1)\ \underset{\sim}{\Sigma}$ yields that

$$\det(\underset{\sim}{\Sigma}) = \prod_{i=1}^{k+1}\mu_i/(n+\Sigma\nu_h+1)^k \tag{4B.5}$$

for "det" denoting determinant. Applying Theorem 8.3.3 of Graybill (1969, p170) yields that the inverse $\underset{\sim}{\Sigma}^{-1} = (\sigma^{ij})$ of $\underset{\sim}{\Sigma}$ has, for $1\leq i\leq k$, elements

$$\sigma^{ii} = (n+\Sigma\nu_h+1)(\mu_i+\mu_{k+1})/(\mu_i\mu_{k+1}) \tag{4B.6}$$

and, for $i<j\leq k$,

$$\sigma^{ij} = (n+\Sigma\nu_h+1)/\mu_{k+1}. \tag{4B.7}$$

By dropping the term $O(q^{-1})$ in Stirling's formula [Cramer (1951,p130)]

$$\log[\Gamma(q)] = (q-\tfrac{1}{2})\log(q) - q + \tfrac{1}{2}\log(2\pi) + O(q^{-1}) \tag{4B.8}$$

for the logarithm of the gamma function for q positive and real, we can approximate the logarithm of the Dirichlet density (4B.1) as

$$\begin{aligned}\log[f(\underset{\sim}{p}|\underset{\sim}{x})] &= (n+\Sigma\nu_h-\tfrac{1}{2})\log(n+\Sigma\nu_h)+\tfrac{1}{2}\log(2\pi)-\sum_{i=1}^{k+1}[(x_i+\nu_i-\tfrac{1}{2})\log(x_i+\nu_i)\\ &\quad +\tfrac{1}{2}\log(2\pi)]+\sum_{i=1}^{k+1}(x_i+\nu_i-1)\log(p_i)+O(n^{-1})\\ &= -\sum_{i=1}^{k+1}(x_i+\nu_i)\log\{(x_i+\nu_i)/[(n+\Sigma\nu_h)p_i]\}+\tfrac{1}{2}\log\{\prod_{i=1}^{k+1}(x_i+\nu_i)\\ &\quad /[(n+\Sigma\nu_h)(2\pi)^k]\}-\log(\prod_{i=1}^{k+1}p_i)+O(n^{-1}).\end{aligned} \tag{4B.9}$$

Now, for $1\leq i\leq k+1$, let

$$z_i = (p_i-\mu_i)/\sqrt{\sigma_{ii}} \tag{4B.10}$$

where we define μ_{k+1} and $\sigma_{k+1,k+1}$ by (4B.2) and (4B.3), respectively. Then, for $1\leq i\leq k+1$,

$$E(z_i) = 0, \tag{4B.11}$$

$$\text{var}(z_i) = 1, \tag{4B.12}$$

and, for $1\leq i\leq k$, $i\leq j\leq k$,

$$\text{cov}(z_i,z_j) = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}. \tag{4B.13}$$

Thus, from (4B.10),

$$\begin{aligned}p_i &= \mu_i+z_i\sqrt{\sigma_{ii}}\\ &= \mu_i\left(1+z_i\{(1-\mu_i)/[\mu_i(n+\Sigma\nu_h+1)]\}^{\frac{1}{2}}\right).\end{aligned} \tag{4B.14}$$

From Tchebychev's inequality [Bishop, Fienberg, and Holland (1975,p476)] and (4B.3),

$$\begin{aligned} p_i - \mu_i &= O_p(\sqrt{\sigma_{ii}}) \\ &= O_p(n^{-\frac{1}{2}}). \end{aligned} \tag{4B.15}$$

Thus, the term

$$\begin{aligned} \varepsilon_i &= z_i \sqrt{(1-\mu_i)/[\mu_i(n+\Sigma\nu_h+1)]} \\ &= (p_i-\mu_i)/\mu_i \end{aligned} \tag{4B.16}$$

in the second line of (4B.14) is $O_p(n^{-\frac{1}{2}})$. Therefore, for large enough n, (4B.16) is bounded in absolute value by 1, so that [CRC Tables (1962, p373)],

$$\log [1+\varepsilon_i] = \varepsilon_i - \varepsilon_i^2/2 + o_p(n^{-1}). \tag{4B.17}$$

Hence, from (4B.2), (4B.14), and (4B.17), we have that, for $1 \leq i \leq k$,

$$\begin{aligned} \log \mu_i/p_i &= -\log [1+\varepsilon_i] \\ &= -\varepsilon_i + \varepsilon_i^2/2 + o_p(n^{-1}). \end{aligned} \tag{4B.18}$$

Therefore, substituting (4B.18) into (4B.9), we have that,

$$\begin{aligned} \log f(\underset{\sim}{p}|\underset{\sim}{x}) = &\sum_{i=1}^{k+1} (x_i+\nu_i)[\varepsilon_i - \varepsilon_i^2/2] + \tfrac{1}{2} \log\{\prod_{i=1}^{k+1} (x_i+\nu_i) \\ &/[(n+\Sigma\nu_h)(2\pi)^k]\} - \log\left(\prod_{i=1}^{k+1} p_i\right) + o_p(n^{-1}). \end{aligned} \tag{4B.19}$$

For the first of the four terms in (4B.19) we have, using (4B.3) and (4B.10), that

$$\sum_{i=1}^{k+1}(x_i+\nu_i)\varepsilon_i = (n+\Sigma\nu_h)\sum_{i=1}^{k+1}(p_i-\mu_i) = 0. \tag{4B.20}$$

For the second term in (4B.19) we have that

$$-\sum_{i=1}^{k+1}(x_i+\nu_i)\varepsilon_i^2/2 = [-\tfrac{1}{2}\sum_{i=1}^{k+1} z_i^2(1-\mu_i)][(n+\Sigma\nu_h)/(n+\Sigma\nu_h+1)] = -\tfrac{1}{2}\sum_{i=1}^{k+1} z_i^2(1-\mu_i) + O_p(n^{-1}) \tag{4B.21}$$

since, from (4B.15) (or meaning of standardized variable), $z_i^2=O_p(1)$ so that $O(n^{-1})\ \Sigma z_i^2(1-\mu) = O_p(n^{-1})$.

In (4B.21) we can write $z_{k+1}^2(1-\mu_{k+1})$ in terms of z_i^2 and μ_i for $1\leq i\leq k$ as

$$\begin{aligned} z_{k+1}^2(1-\mu_{k+1}) &= (n+\Sigma\nu_h+1)(p_{k+1}-\mu_{k+1})^2/\mu_{k+1} \\ &= (n+\Sigma\nu_h+1)[\sum_{i=1}^{k}(\mu_i-p_i)]^2/\mu_{k+1} \\ &= (n+\Sigma\nu_h+1)[\sum_{i=1}^{k}(\mu_i-p_i)^2+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}(\mu_i-p_i)(\mu_j-p_j)]/\mu_{k+1} \\ &= \sum_{i=1}^{k} z_i^2(1-\mu_i)\mu_i/\mu_{k+1} + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} z_i z_j \\ &\qquad \times[\mu_i(1-\mu_i)\mu_j(1-\mu_j)]^{\frac{1}{2}}/\mu_{k+1}\ , \end{aligned} \tag{4B.22}$$

since

$$\begin{aligned} (n+\Sigma\nu_h+1)(\mu_i-p_i)^2 &= (\mu_i-p_i)^2\{(n+\Sigma\nu_h+1)/[\mu_i(1-\mu_i)]\}\mu_i(1-\mu_i) \\ &= z_i^2\mu_i(1-\mu_i) \end{aligned} \tag{4B.23}$$

and, similarly, for $j \neq i$

$$(n+\Sigma\nu_h+1)(\mu_i-p_i)(\mu_j-p_j) = z_i z_j[\mu_i(1-\mu_i)\mu_j(1-\mu_j)]^{\frac{1}{2}}. \qquad (4B.24)$$

Therefore, substituting (4B.22) into (4B.21) and recalling (4B.6) and (4B.7), we have that

$$\begin{aligned}
-\sum_{i=1}^{k+1}(x_i+\nu_i)\epsilon_i^2/2 &= -\tfrac{1}{2}\Big[\sum_{i=1}^{k} z_i^2(1-\mu_i)(1+\mu_i/\mu_{k+1}) \\
&\quad +2\sum_{i=1}^{k-1}\sum_{j>i}^{k} z_i z_j[\mu_i(1-\mu_i)\mu_j(1-\mu_j)]^{\frac{1}{2}}/\mu_{k+1}\Big] + O_p(n^{-1}) \\
&= -\tfrac{1}{2}(n+\Sigma\nu_h+1)\Big[\sum_{i=1}^{k}(p_i-\mu_i)^2(\mu_i+\mu_{k+1})/(\mu_i\mu_{k+1}) \qquad (4B.25) \\
&\quad +2\sum_{i=1}^{k-1}\sum_{j>i}^{k}(p_i-\mu_i)(p_j-\mu_j)/\mu_{k+1}\Big] + O_p(n^{-1}) \\
&= -\tfrac{1}{2}\,(\underset{\sim}{p}-\underset{\sim}{\mu})\,\underset{\sim}{\Sigma}^{-1}\,(\underset{\sim}{p}-\underset{\sim}{\mu})' + O_p(n^{-1}).
\end{aligned}$$

Now, from (4B.15) we have that

$$\prod_{i=1}^{k+1} p_i = \prod_{i=1}^{k+1}\mu_i + O_p(n^{-\frac{1}{2}}). \qquad (4B.26)$$

Therefore, by using (4B.2), (4B.5), and (4B.26), we can write the last two terms of the log Dirichlet (4B.19) as

$$\begin{aligned}
&\log\{[\prod_{i=1}^{k+1}(x_i+\nu_i)]^{\frac{1}{2}}/[(n+\Sigma\nu_h)^{\frac{1}{2}}(2\pi)^{k/2}\prod_{i=1}^{k+1}p_i]\} \\
&= \log\{(2\pi)^{k/2}[(n+\Sigma\nu_h+1)/(n+\Sigma\nu_h)]^{k/2}[(\prod_{i=1}^{k+1}\mu_i)^{\frac{1}{2}}+O_p(n^{-\frac{1}{2}})]/(n+\Sigma\nu_h+1)^{k/2}\}^{-1} \qquad (4B.27) \\
&= \log\left((2\pi)^{k/2}[1+O(n^{-k/2})]\{[\det(\underset{\sim}{\Sigma})]^{\frac{1}{2}}+O_p(n^{-(k+1)/2})\}\right)^{-1} \\
&= \log\{(2\pi)^{k/2}[\det(\underset{\sim}{\Sigma})]^{\frac{1}{2}}[1+O_p(n^{-\frac{1}{2}})]\}^{-1}.
\end{aligned}$$

Therefore, substituting (4B.20), (4B.25), and (4B.27) into (4B.19), taking the antilogarithm of the result, and noting that

$$\exp[O_p(n^{-1})] = 1+O_p(n^{-1}), \tag{4B.28}$$

we have that

$$\begin{aligned} f(\underset{\sim}{p}|\underset{\sim}{x}) &= \{(2\pi)^{k/2}[\det(\underset{\sim}{\Sigma})]^{\frac{1}{2}}[1+O_p(n^{-\frac{1}{2}})]\}^{-1}\{\exp[-\tfrac{1}{2}(\underset{\sim}{p}-\underset{\sim}{\mu})\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{p}-\underset{\sim}{\mu})']\}[1+O_p(n^{-1})] \\ &= \{(2\pi)^{k/2}[\det(\underset{\sim}{\Sigma})]^{\frac{1}{2}}\}^{-1}\exp[-\tfrac{1}{2}(\underset{\sim}{p}-\underset{\sim}{\mu})\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{p}-\underset{\sim}{\mu})]\ [1+O_p(n^{-\frac{1}{2}})]. \end{aligned} \tag{4B.29}$$

Rao [1968,(xv)p104] proves that if the density of a random variable converges to some density, then the distribution of the random variable converges to the distribution for the limiting density. Therefore,

$$\lim_{n\to\infty} D(x_1+\nu_1,\ldots,x_k+\nu_k;x_{k+1}+\nu_{k+1}) = N_k(\underset{\sim}{\mu},\underset{\sim}{\Sigma}); \tag{4B.30}$$

that is, the limiting k-dimensional posterior distribution of $\underset{\sim}{p}$ given complete data $\underset{\sim}{x}$ is k-dimensional multivariate normal with mean and covariance matrices those of the Dirichlet.

APPENDIX 4C

CENTRAL MOMENTS OF k-DIMENSIONAL MULTIVARIATE NORMAL DISTRIBUTION

4C.1 Introduction:

Let $x_1,\ldots,x_k$ have the k-dimensional multivariate normal distribution $N_k(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$ with the 1×k mean vector $\underset{\sim}{\mu}$ and k×k covariance matrix $\underset{\sim}{\Sigma}=(\sigma_{ij})$. Anderson (1958,p39) gives the second and fourth central moments of this distribution. Lindley (1965,v1,p95) and Schmetterer (1974,p76) give the lth central moment for the one-dimensional distribution. In this appendix we derive the general central moment of the k-dimensional distribution. We conclude the appendix by illustrating the formula for the first six central cross-product moments and by showing it equals formulas from Anderson, Lindley, and Schmetterer for their specialized cases.

Because the moment-generating function of the multivariate normal distribution exists, we work with it rather than the characteristic function to avoid using the extra, complex, variable $\sqrt{-1}$. To obtain central moments, we multiply the moment-generating function $\phi(t_1,\ldots,t_k)$ by $\exp(-\underset{\sim}{t}\underset{\sim}{\mu}')$, continuously differentiate the results with respect to $\underset{\sim}{t}$, and then set $\underset{\sim}{t}$ to $\underset{\sim}{0}$ in the differentiated results. [See Lindley (1965,v1, p92) or Jeffreys (1939,p74).]

An alternative approach is to calculate cumulants $\kappa_{i_1\ldots i_k}$ and then, from the cumulants, central moments. Straightforward calculation yields results of Anderson (1958,p39) that $\kappa_{0\ldots0i_j0\ldots0}=\mu_j$ for $1\leq j\leq k$, $\kappa_{0\ldots0i_j0\ldots0i_l0\ldots0}=\sigma_{jl}$ for $1\leq j,l\leq k$, and $\kappa_{i_1,\ldots,i_k}=0$ for $\sum_{j=1}^{k} i_j>2$.

From Kendall and Stuart (1969,v1,p70) we can therefore write the first ten central moments $E[(p_i-\mu_i)^l|\underset{\sim}{x}]$ for $1 \leq l \leq 10$. The method to extend these results to the general central (cross-product) moment, however, is no briefer than the method using moment-generating functions that is given in the next section.

4C.2 Theory

From the characteristic function given by Anderson (1958,p36) and Wilks (1963,p168), we can write the moment-generating function $\phi(t_1,\ldots,t_k)$ of the k-dimensional multivariate normal distribution as

$$\phi(t_1,\ldots,t_k) = e^{\underset{\sim}{t}\,\underset{\sim}{\mu}' + \frac{1}{2}\,\underset{\sim}{t}\,\underset{\sim}{\Sigma}\,\underset{\sim}{t}'}. \qquad (4C.1)$$

Defining

$$f(t_1,\ldots,t_k) = e^{-\underset{\sim}{t}\,\underset{\sim}{\mu}'}\,\phi(t_1,\ldots,t_k), \qquad (4C.2)$$

we have that

$$f(t_1,\ldots,t_k) = e^{\frac{1}{2}\,\underset{\sim}{t}\,\underset{\sim}{\Sigma}\,\underset{\sim}{t}'} = e^{\frac{1}{2}\left(\sum_{i=1}^{k} t_i^2\,\sigma_{ii} + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} t_i\,t_j\,\sigma_{ij}\right)}. \qquad (4C.3)$$

Hence,

$$\partial f(t_1,\ldots,t_k)/\partial t_i = \left(t_i\,\sigma_{ii} + \sum_{j\neq i}^{k} t_j\,\sigma_{ij}\right)\; f(t_1,\ldots,t_k). \qquad (4C.4)$$

Define

$$C_i = t_i\,\sigma_{ii} + \sum_{j\neq i}^{k} t_j\,\sigma_{ij} \qquad (4C.5)$$

and rewrite (4C.4) as

$$\partial f(t_1,\ldots,t_k)/\partial t_i = C_i\; f(t_1,\ldots,t_k). \qquad (4C.6)$$

Now, for all $1\leq i,j\leq k$,

$$\partial C_i/\partial t_j = \sigma_{ij}. \qquad (4C.7)$$

Then,

$$\partial^2 f(t_1,\ldots,t_k)/(\partial t_i \partial t_j) = \sigma_{ij}\, f(t_1,\ldots,t_k) + C_i\, \partial f(t_1,\ldots,t_k)/\partial t_j, \tag{4C.8}$$

$$\partial^3 f(t_1,\ldots,t_k)/(\partial t_i \partial t_j \partial t_k) = \sigma_{ij}\, \partial f(t_1,\ldots,t_k)/\partial t_k + \sigma_{ik} \partial f(t_1,\ldots,t_k)/\partial t_j \tag{4C.9}$$
$$+ C_i\, \partial^2 f(t_1,\ldots,t_k)/(\partial t_j \partial t_k),$$

and

$$\begin{aligned}\partial^4 f(t_1,\ldots,t_k)/(\partial t_i \partial t_j \partial t_k \partial t_\ell) = {} & \sigma_{ij}\partial^2 f(t_1,\ldots,t_k)/(\partial t_k \partial t_\ell) \\ & + \sigma_{ik}\partial^2 f(t_1,\ldots,t_k)/(\partial t_j \partial t_\ell) \\ & + \sigma_{i\ell}\partial^2 f(t_1,\ldots,t_k)/(\partial t_j \partial t_k) \\ & + C_i \partial^3 f(t_1,\ldots,t_k)/(\partial t_j \partial t_k \partial t_\ell).\end{aligned} \tag{4C.10}$$

Continuing in this fashion, we have in general that, for $1 \leq h_g \leq k$ and m a positive integer,

$$\begin{aligned}\partial^\ell f(t_1,\ldots,t_k)/\Big(\prod_{g=1}^{m} \partial t_{h_g}\Big) = {} & \sum_{i=2}^{\ell} \sigma_{h_1 h_i}\, \partial^{\ell-2} f(t_1,\ldots,t_k)/\Big(\prod_{\substack{m=2 \\ m\neq i}}^{\ell} \partial t_{h_m}\Big) \\ & + C_{h_1}\, \partial^{\ell-1} f(t_1,\ldots,t_k)/\Big(\prod_{m=2}^{\ell} \partial t_{h_m}\Big)\end{aligned} \tag{4C.11}$$

We use the double subscript h_g rather than a single subscript h because t_h is meaningless for $k<h\leq\ell$ and we want a convenient way of allowing all possible permutations of the k integers and their powers j for $1\leq j\leq \ell$.

Now, odd central moments are 0 because the multivariate normal

distribution is symmetric about the mean.[1] For ℓ even and

$$y_{h_g} = x_{h_g} - \mu_{h_g}, \tag{4C.12}$$

we therefore have from (4C.11) that the ℓ^{th} central moment $E(\prod_{g=1}^{\ell} y_{h_g})$ is

$$E(\prod_{g=1}^{\ell} y_{h_g}) = \sum_{i=2}^{\ell} \sigma_{h_1 h_i} \partial^{\ell-2} f(t_1,\ldots,t_k)/(\prod_{\substack{m=2 \\ m\neq i}}^{\ell} \partial t_{h_m}) \Bigg]_{\underset{\sim}{t}=\underset{\sim}{0}} \tag{4C.13}$$

Therefore, each of the ℓ-1 terms in the ℓ^{th} central moment is a variance or covariance times an $(\ell-2)^{nd}$ central moment. Evaluating the second central moment (4C.8) at $\underset{\sim}{t}=\underset{\sim}{0}$ yields that

$$E(\prod_{g=1}^{2} y_{h_g}) = \sigma_{h_1 h_2}. \tag{4C.14}$$

Hence, by induction the ℓ^{th} central moment $E(\prod_{g=1}^{\ell} y_{h_g})$ is a sum of ℓ-1 terms, each of which is a product of those $\ell/2$ elements of the covariance matrix that are indexed by the subscripts h_g. That is, for

$$\text{(i)} \quad i_1 = 1, \tag{4C.15}$$

and

$$\text{(ii)} \quad i_{2j-1} = \min\left(2\times\prod_{b=2}^{2j-1} \bar{\delta}_{i_b,2}, 3\times\prod_{b=2}^{2j-1} \bar{\delta}_{i_b,3}, \ldots, (2j-2)\times\prod_{b=2}^{2j-2} \bar{\delta}_{i_b,2j-2}, 2j-1\right), \tag{4C.16}$$

where $2\leq j\leq \ell/2$ and

---

[1] also because $C_{\ell_i}(0) = 0$; $f'(0) = 0$, and by induction all odd moments are zero.

$$\bar{\delta}_{i_b,q} = 1-\delta_{i_b,q} = \begin{cases} 0 & \text{if } i_b=q \\ 1 & \text{if } i_b\neq q \end{cases} \qquad (4C.17)$$

is defined to be the one complement of the Kronecker Delta symbol $\delta_{i_b,q}$ [see Feller (1968,v1,p428), Korn and Korn (1968,p544), or CRC Standard Math Tables (1962,p501)], we have that

$$E\left(\prod_{g=1}^{\ell} y_{h_g}\right) = \sum_{i_2=2}^{\ell} \sigma_{h_{i_1}h_{i_2}} \sum_{\substack{i_4>i_3 \\ i_4\neq i_j \\ \text{for } j<4}} \sigma_{h_{i_3}h_{i_4}} \cdots \sum_{\substack{i_{\ell-4}>i_{\ell-3} \\ i_{\ell-4}\neq i_j \\ \text{for } j<\ell-3}} \sigma_{h_{i_{\ell-3}}h_{i_{\ell-4}}}$$

$$\times \sum_{\substack{i_\ell>i_{\ell-1} \\ i_\ell\neq i_j \\ \text{for } j<\ell-1}} \sigma_{h_{i_{\ell-1}}h_{i_\ell}}$$

$$= \sum_{i_1=1}^{1} \sum_{i_2=2}^{\ell} \sum_{\substack{i_3=2 \\ i_3\neq i_j \\ \text{for } j<3}}^{Q(3)} \sum_{\substack{i_4>i_3 \\ i_4\neq i_j \\ \text{for } j<4}}^{\ell} \sum_{\substack{i_5=3 \\ i_5\neq i_j \\ \text{for } j<5}}^{Q(5)} \sum_{\substack{i_6>i_5 \\ i_6\neq i_j \\ \text{for } j<6}}^{\ell} \qquad (4C.18)$$

$$\cdots \sum_{\substack{i_{2s}-1=s \\ i_{2s-1}\neq i_j \\ \text{for } j<2s-1}}^{Q(2s-1)} \sum_{\substack{i_{2s}>i_{2s-1} \\ i_{2s}\neq i_j \\ \text{for } j<2s}}^{\ell} \cdots$$

$$\cdots \sum_{\substack{i_{\ell-1}=\ell/2 \\ i_{\ell-1}\neq i_j \\ \text{for } j<\ell-1}}^{Q(\ell-1)} \sum_{\substack{i_\ell>i_{\ell-1} \\ i_\ell\neq i_j \\ \text{for } j<\ell}}^{\ell} \sigma_{h_1h_{i_2}} \sigma_{h_{i_3}h_{i_4}} \cdots \sigma_{h_{i_{2s-1}}h_{i_{2s}}} \cdots \sigma_{h_{i_{\ell-1}}h_{i_m}},$$

where

$$Q(2s-1) = s + \sum_{j=2}^{2s-2} \delta_{i_j,s} + \sum_{a=2}^{2s-2} \sum_{\substack{b=2 \\ b\neq a}}^{2s-2} \delta_{i_a,s}\,\delta_{i_b,s+1} \qquad (4C.19)$$

$$+\ldots+ \sum_{j_s=2}^{2s-2} \sum_{\substack{j_{s+1}=2 \\ j_{s+1}\neq j_s}}^{2s-2} \cdots \sum_{\substack{j_{2s-2}=2 \\ j_{2s-2}\neq a \\ \text{for } a<2s-2}}^{2s-2} \prod_{b=3}^{2s-2} \delta_{i_{j_b},b},$$

for $\delta_{i_j,s}$ the Kronecker Delta symbol. For example,

$$Q(3) = 2 + \delta_{i_2,2}$$

and

$$Q(5) = 3 + \sum_{j=2}^{4} \delta_{i_j,3} + \sum_{j_3=2}^{4} \sum_{\substack{j_4=2 \\ j_4\neq j_3}}^{4} \delta_{i_{j_3},3}\,\delta_{i_{j_4},4}.$$

4C.3 Illustrations

From (4C.18) the first six central moments of the k-dimensional multivariate normal distribution $N_k(\underset{\sim}{\mu},\underset{\sim}{\Sigma}=(\sigma_{ij}))$ are, for $1\leq a,b,c,d,e,f\leq k$,

$$E(y_a) = E(y_a y_b y_c) = E(\prod_{g=1}^{5} y_{h_g}) = 0, \tag{4C.20}$$

$$E(y_a y_b) = \sigma_{ab}, \tag{4C.21}$$

$$E(y_a y_b y_c y_d) = \sigma_{ab}\sigma_{cd} + \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}, \tag{4C.22}$$

and

$$\begin{aligned} E(y_a y_b y_c y_d y_e y_f) &= \sigma_{ab}[\sigma_{cd}\sigma_{ef}+\sigma_{ce}\sigma_{df}+\sigma_{cf}\sigma_{de}] \\ &+ \sigma_{ac}[\sigma_{bd}\sigma_{ef}+\sigma_{be}\sigma_{df}+\sigma_{bf}\sigma_{de}] + \sigma_{ad}[\sigma_{bc}\sigma_{ef}+\sigma_{be}\sigma_{cf}+\sigma_{bf}\sigma_{ce}] \\ &+ \sigma_{ae}[\sigma_{bc}\sigma_{df}+\sigma_{bd}\sigma_{cf}+\sigma_{bf}\sigma_{cd}] + \sigma_{af}[\sigma_{bc}\sigma_{de}+\sigma_{bd}\sigma_{ce}+\sigma_{be}\sigma_{cd}]. \end{aligned} \tag{4C.23}$$

Thus, for $N_3(\underset{\sim}{\mu},\underset{\sim}{\Sigma})$,

$$E(y_1^2 y_2 y_3) = \sigma_{11}\sigma_{23} + 2\sigma_{12}\sigma_{13} \tag{4C.24}$$

and

$$E(y_1^6) = 15\sigma_{11}^3. \tag{4C.25}$$

We note that the second and fourth moments agree with Anderson (1958,p39) and that the ℓth central moment for ℓ even is

$$E(y_i^{\ell}) = [\prod_{j=1}^{\ell/2}(2j-1)]\,\sigma_{ii}^{\ell/2} = \sigma_{ii}^{\ell/2}\ell!/[2^{\ell/2}(\ell/2)!], \tag{4C.26}$$

paralleling results for the one-dimensional normal distribution from Lindley (1965,v1,p95) and Schmetterer (1974,p76).

APPENDIX 4D

LIMITING POSTERIOR DISTRIBUTION GIVEN INCOMPLETE DATA

4D.1 Introduction:

In this appendix we calculate the limiting k-dimensional posterior distribution of $\underset{\sim}{p}$ given incomplete data $\underset{\sim}{z}$. We calculate this limiting distribution in two ways. In the traditional approach, given in Section 4D.2, we note that the prior density is continuous in $\underset{\sim}{p}$ and the likelihood is regular. Therefore, the limiting distribution for the posterior density $f(\underset{\sim}{p}|\underset{\sim}{z})$ is multivariate normal with the vector of maximum likelihood estimates for the mean. The negative matrix of second partial derivatives of the log likelihood, evaluated at the maximum likelihood estimate, is the large-sample inverse covariance matrix. Therefore, rewriting the log likelihood in terms of exponential parameters, using theory from Sundberg (1974) to calculate the first and second partial derivatives of the log likelihood with respect to these exponential parameters, and transforming results back to $\underset{\sim}{p}$ gives elements of the asymptotic posterior mean and covariance matrices. However, results for the asymptotic inverse covariance matrix are very long and complicated expressions that do not easily simplify.

Therefore, to obtain simpler expressions and, moreover, equations paralleling those for complete data, in Section 4D.3 we also derive the limiting posterior distribution another way. We rewrite the posterior density as a product of complete-data Dirichlet densities, each having, from Appendix 3B, a limiting multivariate normal distribution. Because these densities are of differing dimensions and on differing combinations of variables, we do not immediately have that the resultant product of these multivariate normal densities is a k-dimensional multivariate normal

density on the k components of $\underset{\sim}{p}$. However, by equating coefficients and solving for unknowns, we then prove that, owing to the special relationship between the first and each remaining product, the sum of exponents from each Dirichlet in the product does form the exponent of such a density.

As part of this proof, we check that the k-dimensional inverse matrix in the exponent is positive definite and symmetric; hence, a covariance matrix. We also obtain the nonexponential term for the limiting multivariate normal distribution and prove that the limit of the denominator of the posterior density (the marginal distribution) is 1. The essential step for the latter is that the limit of the integral that is this denominator can be taken inside the integral.

We begin this nontraditional approach by first considering, in Section 4D.3.1 the case having at least one category, say C_{k+1}, for which all data is complete. For this case we derive elements of the asymptotic mean and inverse covariance matrices as functions of a number of unknown ratios and as a large nonlinear system of equations. In Subsection 4D.3.2 we rewrite the moments to eliminate these ratios and reduce the nonlinear system of equations. As results, we get the maximum likelihood estimate for the asymptotic mean and very simple expressions for the asymptotic covariance matrix that parallel expressions from the complete-data case. In Subsection 4D.3.3 we extend proofs from the preceding two subsections to the general case allowing incomplete data on all categories. The expression for elements of the asymptotic mean vector is identical to that of Subsection 4D.3.2. Expressions for elements of the asymptotic inverse covariance matrix are those of 4D.3.2 plus terms for those patterns of incomplete data that index the dependent variable.

In Section 4D.4 we simplify results given by the traditional approach in Section 4D.2 for the inverse covariance matrix. Note that it is only by knowing results of Section 4D.3 and by using much algebraic manipulation that we can simplify these equations to those given by the nontraditional approach. The algebraic manipulation is so extensive that numerous human errors occur. Hence, knowing the final result at which to aim is critical. It allows continual checking and correcting of various parts of the equations.

Therefore, the nontraditional approach will be useful in other kinds of problems when the traditional approach gives unwieldy results. Because we are piecing together densities of different dimensions and on different combinations of variables, the notation for the nontraditional method in Section 4D.3 is necessarily complicated. However, for most types of problems, notational difficulties would not exist.

Section 4D.5 concludes the appendix with three examples allowing, with the help of the MACSYMA symbolic computer system, exact solution of the nonlinear system of equations for the asymptotic mean. We also give numerical illustrations for the asymptotic mean, inverse covariance, and covariance matrices. Note that the exact solutions can be used for the Taylor-series approximate posterior mean and the posterior mode, as well as the asymptotic mean, which is the maximum likelihood estimate. Because an exact solution exists only in very special cases and is expensive, however, it cannot generally be used. A general method to solve the nonlinear system of equations is the EM iterative algorithm of Dempster, Laird, and Rubin (1977) discussed in Section 2.3.2.

4D.2 Traditional Approach:

The prior density $g(\underset{\sim}{p})$ in (1.1) is continuous in $\underset{\sim}{p}$ and the likelihood $h(\underset{\sim}{z}|\underset{\sim}{p})$ in (2.8) is regular. Therefore [Cox and Hinkley (1974,p401)], the limiting distribution for the posterior density $f(\underset{\sim}{p}|\underset{\sim}{z})$ is multivariate normal with the vector $\hat{\underset{\sim}{p}}$ for

$$\hat{p}_i = (z_i + \sum_{D \ni i} z_D \hat{p}_i/\hat{p}_D)/n \tag{4D.1}$$

from (2.36) of maximum likelihood estimates for the mean $\underset{\sim}{u}$ and the matrix

$$\{- \partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{p})]/(\partial \underset{\sim}{p} \partial \underset{\sim}{p}')\}_{\underset{\sim}{p}=\hat{\underset{\sim}{p}}} \tag{4D.2}$$

for the inverse covariance matrix.

Recall that the multinomial density is a member of the exponential family, where we define the exponential-family parameters

$$\phi_i = \log(p_i/p_{k+1}). \tag{4D.3}$$

As in Section 2.3.1, let

$$t_i(\underset{\sim}{x}) = z_i + \sum_{D \ni i} z_D^{(i)} \tag{4D.4}$$

for $z_D^{(i)}$ the (unknown) number of the z_D observations that fall in category C_i. Then, as noted in Section 2.3.2, Sundberg (1974) proves that

$$\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial \underset{\sim}{\phi} \partial \underset{\sim}{\phi}') = -\mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{\phi}] + \mathrm{cov}[\underset{\sim}{t}(\underset{\sim}{x})|\underset{\sim}{z},\underset{\sim}{\phi}], \tag{4D.5}$$

where $h(\underset{\sim}{z}|\underset{\sim}{\phi})$ is the likelihood $h(\underset{\sim}{z}|\underset{\sim}{p})$ written in terms of $\underset{\sim}{\phi}$ instead of $\underset{\sim}{p}$.

Since the first partial derivatives are zero at the maximum likelihood estimate, application of the chain rule to the negative of (4D.5) with evaluation at $\underset{\sim}{p}=\hat{\underset{\sim}{p}}\equiv\underset{\sim}{u}$ yields

$$-\left.\frac{\partial^2 \log h}{\partial p_i \partial p_j}\right]_{\underset{\sim}{p}=\underset{\sim}{u}} = \left\{-\sum_{a=1}^{k}\frac{\partial \phi_a}{\partial p_i}\left[\sum_{b=1}^{k}\frac{\partial^2 \log h}{\partial \phi_a \partial \phi_b}\frac{\partial \phi_b}{\partial p_j}\right]\right\}_{\underset{\sim}{p}=\underset{\sim}{u}}$$

$$= \left[\sum_{a=1}^{k}\frac{\partial \phi_a}{\partial p_i}\left\{\sum_{b=1}^{k}[\mathrm{cov}(t_a,t_b|\underset{\sim}{\phi})-\mathrm{cov}(t_a,t_b|\underset{\sim}{z},\underset{\sim}{\phi})]\frac{\partial \phi_b}{\partial p_j}\right\}\right]_{\underset{\sim}{p}=\underset{\sim}{u}} . \tag{4D.6}$$

Now,

$$\left.\mathrm{cov}(t_a,t_b|\underset{\sim}{\phi})\right]_{\underset{\sim}{p}=\underset{\sim}{u}} = \begin{cases} nu_a(1-u_a) & \text{for } b=a \\ -nu_a u_b & \text{for } b\neq a \end{cases} \tag{4D.7}$$

and

$$\left.\mathrm{cov}(t_a,t_b|\underset{\sim}{z},\underset{\sim}{\phi})\right]_{\underset{\sim}{p}=\underset{\sim}{u}} = \begin{cases} \sum_{D\ni a} z_D u_a(u_D-u_a)/u_D^2 & \text{for } b=a \\ -\sum_{D\ni a,b} z_D u_a u_b/u_D^2 & \text{for } b\neq a. \end{cases} \tag{4D.8}$$

From (4D.3)

$$\left.\partial\phi_i/\partial p_j\right]_{\underset{\sim}{p}=\underset{\sim}{u}} = \begin{cases} (u_i+u_{k+1})/(u_i u_{k+1}) & \text{for } j=i \\ 1/u_{k+1} & \text{for } j\neq i. \end{cases} \tag{4D.9}$$

Applying Theorem 8.3.3 of Graybill (1969,p170) to (4D.7) yields that

$$\sigma_{(1)}^{ij} = \begin{cases} (u_i+u_{k+1})/(nu_i u_{k+1}) & \text{for } j=i \\ u_{k+1}/n & \text{for } j\neq i \end{cases} \tag{4D.10}$$

for $\sigma_{(1)}^{ij}$ the i,j^{th} element of $\{[\mathrm{cov}(t|\underset{\sim}{\phi})]_{\underset{\sim}{p}=\underset{\sim}{u}}\}^{-1}$. Hence, we note from

(4D.9) that, for all $1 \le i,j \le k$,

$$\partial\phi_i/\partial p_j \Big]_{\underset{\sim}{p}=\underset{\sim}{u}} = n\sigma_{(1)}^{ij}. \tag{4D.11}$$

Substituting from (4D.7) - (4D.9) and (4D.10) - (4D.11) into (4D.6) and writing $\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial\underset{\sim}{u}\partial\underset{\sim}{u}')$ to mean $\{\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial\underset{\sim}{p}\partial\underset{\sim}{p}')\}_{\underset{\sim}{p}=\underset{\sim}{u}}$, yields that

$$\begin{aligned}
-\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial u_i \partial u_i) &= (u_i+u_{k+1})^2[n(1-u_i) - \sum_{D\ni i}(z_D/u_D^2)(u_D-u_i)]/(u_i u_{k+1}^2) \\
&\quad +2(u_i+u_{k+1})\sum_{a\neq i}^{k}[-nu_a + \sum_{D\ni i,a} z_D u_a/u_D^2]/u_{k+1}^2 + \sum_{a\neq i}^{k} u_a[n(1-u_a) - \sum_{D\ni a}(z_D/u_D^2) \\
&\quad \times(u_D-u_a) + \sum_{b\neq i,a}^{k}(-nu_b + \sum_{D\ni a,b} z_D u_b/u_D^2)]/u_{k+1}^2 \\
&= (u_i+u_{k+1})^2[n(1-u_i) - \sum_{D\ni i} z_D(u_D-u_i)/u_D^2]/(u_i u_{k+1}^2) \qquad (4D.12) \\
&\quad +2(u_i+u_{k+1})[-n(1-u_i-u_{k+1}) + \sum_{a\neq i}^{k}(\sum_{D\ni i,a} u_a z_D/u_D^2)]/u_{k+1}^2 \\
&\quad +\{n(1-u_i-u_{k+1}) - \sum_{a\neq i}^{k} u_a[\sum_{D\ni a} z_D(u_D-u_a)]/u_D^2 \\
&\quad -n(1-u_i-u_{k+1})^2 + \sum_{a\neq i}^{k} u_a[\sum_{b\neq i,a}^{k}(\sum_{D\ni a,b} z_D u_b/u_D^2)]\}/u_{k+1}^2.
\end{aligned}$$

Similarly, for the i,j^{th} element of the asymptotic inverse covariance matrix, we have that

$$
\begin{aligned}
-\partial^2 \log[h(\underset{\sim}{z}|\underset{\sim}{\phi})]/(\partial u_i \partial u_j) = (u_i+u_{k+1})\{[n(1-u_i)-\sum_{D\ni i} z_D(u_D-u_i)/u_D^2] \\
+(u_j+u_{k+1})[-n+\sum_{D\ni i,j} z_D/u_D^2]+\sum_{b\neq i,j}^{k} u_b[-n+\sum_{D\ni i,b} z_D/u_D^2]\}/u_{k+1}^2 \\
+u_j\{u_i[-n+\sum_{D\ni i,j} z_D/u_D^2]+\sum_{b\neq i,j}^{k} u_b[-n+\sum_{D\ni j,b} z_D/u_D^2] \\
+(u_j+u_{k+1})[n(1-u_j)-\sum_{D\ni j} z_D(u_D-u_j)/u_D^2]/u_j\}/u_{k+1}^2 \\
+\sum_{a\neq i,j}^{k} u_a\{u_i[-n+\sum_{D\ni i,a} z_D/u_D^2]+(u_j+u_{k+1})[-n+\sum_{D\ni a,j} z_D/u_D^2] \\
+\sum_{\substack{b\neq i,j \\ b\neq a}} u_b[-n+\sum_{D\ni a,b} z_D/u_D^2]+[n(1-u_a)-\sum_{D\ni a} z_D(u_D-u_a)/z_D^2]\}/u_{k+1}^2 .
\end{aligned}
\tag{4D.13}
$$

4D.3 Nontraditional Approach:

4D.3.1 Theory:

From (2.9) the k-dimensional posterior density of $\underset{\sim}{p}$ given incomplete data $\underset{\sim}{z}$ is

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = g(\underset{\sim}{p})h(\underset{\sim}{z}|\underset{\sim}{p})/\int_{P_k} g(\underset{\sim}{p})h(\underset{\sim}{z}|\underset{\sim}{p})d\underset{\sim}{p} \tag{4D.14}$$

for the Dirichlet prior (1.1)

$$g(\underset{\sim}{p}) = [\Gamma(\sum_{i=1}^{k+1} \nu_i)/\prod_{i=1}^{k+1} \Gamma(\nu_i)] \prod_{i=1}^{k+1} p_i^{\nu_i-1}$$

and the likelihood (2.8)

$$h(\underset{\sim}{z}|\underset{\sim}{p}) = \prod_{P}\{[(\sum_{\mathcal{S}\in P} z_{\mathcal{S},P})!/\prod_{\mathcal{S}\in P} z_{\mathcal{S},P}!] \prod_{\mathcal{S}\in P} p_{\mathcal{S}}^{z_{\mathcal{S},P}}\}.$$

Here $\underset{\sim}{p}$ takes values in the k-dimensional probability simplex $P_k = \{(p_1,\ldots,p_{k+1}): p_i \geq 0, \sum_{i=1}^{k+1} p_i = 1\}$; $\nu_i > 0$; $\mathcal{S}$ is a nonempty subset of $\{1,2,\ldots,k+1\}$; P is a set of mutually exclusive and exhaustive subsets $\mathcal{S}$; $\mathcal{S},P$ is the set element $\mathcal{S}$ in the set P; $\beta_{\mathcal{S},P}$ is the number of elements in $\mathcal{S},P$; $z_{\mathcal{S},P}$ is the number of observations such that each observation falls in one of the $\beta_{\mathcal{S},P}$ categories C_i for $i\in\mathcal{S}$, but is not further classified into a particular one of these $\beta_{\mathcal{S},P}$ categories if $\beta_{\mathcal{S},P}>1$; and $\underset{\sim}{z}$ is the vector of $z_{\mathcal{S}} = \sum_{P\ni\mathcal{S}} z_{\mathcal{S},P}$.

Since $\sum_{P} z_{\mathcal{S},P} = z_{\mathcal{S}}$ and we can cancel from the numerator and denominator of the posterior density (4D.14) any terms that are not a function of $\underset{\sim}{p}$, we can write the posterior density (4D.14) as

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = \prod_{i=1}^{k+1} p_i^{\nu_i-1} \prod_{\mathcal{S}} p_{\mathcal{S}}^{z_{\mathcal{S}}} / \int_{P_k} [\prod_{i=1}^{k+1} p_i^{\nu_i-1} \prod_{\mathcal{S}} p_{\mathcal{S}}^{z_{\mathcal{S}}}] d\underset{\sim}{p}. \tag{4D.15}$$

In (4D.15) the second product is over all distinct $\mathcal{S}$. Thus, the product is over k+1 sets $\mathcal{S}_i=\{i\}$, $1\le i\le k+1$, containing one element and κ sets $\mathcal{S}_{k+1+i}$, $1\le i\le\kappa$, containing more than one element. Each of the κ latter sets correspond to a different pattern of incomplete data. From these $k+1+\kappa$ sets, we will make $\kappa+1$ Dirichlet distributions $D(l)$ for $1\le l\le\kappa+1$.

To do so, reorder the terms in (4D.15) so that the first k+1 multiplicands are those terms $p_i^{z_{\mathcal{S}}+\nu_i-1}$, $1\le i\le k+1$, for which $\mathcal{S}=\{i\}$ contains only one element. Denote the remaining κ sets $\mathcal{S}$, those sets containing more than one element and indexing a unique pattern of incomplete data as $Q(l,1)$ for $2\le l\le\kappa+1$. For $2\le l\le\kappa+1$, multiply $p_{Q(l,1)}^{z_{Q(l,1)}}$ by $p_i^{r_{il}z_{\{i\}}}$ for all $i\notin Q(l,1)$, where the ratio $0<r_{il}<1$, $\sum_{l=2} r_{il}<1$, is to be determined. Define $Q(l,2),\ldots,Q(l,k-q(l)+2)$ as the $k-q(l)+1$ sets indexing the $r_{il}z_{\{i\}}$ where $q(l)$ is the number of categories among which $z_{Q(l,1)}$ is shared. Define $Q(l)$ as the set of these $k+2-q(l)$ mutually exclusive and exhaustive subsets $Q(l,j)$ for $1\le j\le k+2-q(l)$.

For example, for $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12},z_{13})$ we have that $k=2$, $\kappa=2$, $q(2)=q(3)=2$, $Q(2,1)=\{1,2\}$, $Q(2,2)=\{3\}$, $Q(3,1)=\{1,3\}$, $Q(3,2)=\{2\}$, $Q(2)=\{\{1,2\},\{3\}\}$, and $Q(3)=\{\{1,3\},\{2\}\}$.

Now, for $2\le l\le\kappa+1$, we have multiplied $p_{Q(l,1)}^{z_{Q(l,1)}}$ in (4D.15) by $p_i^{r_{il}z_{\{i\}}}$ for $i\notin Q(l,1)$. Accordingly, for each $i\notin Q(l,1)$ multiply $p_i^{z_{\{i\}}+\nu_i-1}$ in the product of the first k+1 terms by $p_i^{-r_{il}z_{\{i\}}}$. This process yields the posterior density (4D.15) as

$$f(\underset{\sim}{p}|\underset{\sim}{z}) = \left[\prod_{i=1}^{k+1} p_i^{(1-\sum_{l=2}^{\kappa+1} r_{il})z_{\{i\}}+\nu_i-1}\right]\prod_{l=2}^{\kappa+1}\left[p_{Q(l,1)}^{z_{Q(l,1)}}\prod_{i\notin Q(l,1)} p_i^{r_{il}z_{\{i\}}}\right] \tag{4D.16}$$

$$/\int_{p_k} (\text{numerator})\, d\underset{\sim}{p}.$$

Multiplying and dividing (4D.16) by

$$\Gamma\{\sum_{i=1}^{k+1}[(1-\sum_{l=2}^{\kappa+1}r_{il})z_{\{i\}}+\nu_i]\}/\{\prod_{i=1}^{k+1}\Gamma[(1-\sum_{l=2}^{\kappa+1}r_{il})z_{\{i\}}+\nu_i]\} \tag{4D.17}$$

for the first set of multiplicands in (4D.16) and, for $2\leq l\leq\kappa+1$, by

$$\Gamma[z_{Q(l,1)}+\sum_{i\notin Q(l,1)}r_{il}z_{\{i\}}+\mathcal{D}(l)+1]/[\Gamma(z_{Q(l,1)}\prod_{i\notin Q(l,1)}\Gamma(r_{il}z_{\{i\}}+1)] \tag{4D.18}$$

for each of the remaining κ sets of multiplicands yields the numerator and integrand of the denominator of (4D.16) as a product of $\kappa+1$ Dirichlet densities, where the lth density has dimension

$$\mathcal{D}(l) = k-q(l)+1 \tag{4D.19}$$

for $q(l)$ again the number of elements in $Q(l,1)$; that is, for num denoting numerator, $j_s\in Q(l,1)$ for $1\leq s\leq k-q(l)$, and, paralleling notation from Wilks (1963,p178), $d(x_1,\ldots,x_k;x_{k+1})$ denoting the k-dimensional density of the Dirichlet distribution $D(x_1,\ldots,x_k;x_{k+1})$,

$$\begin{aligned}\text{num}[f(\underset{\sim}{p}|\underset{\sim}{z})] = d[&(1-\sum_{l=2}^{\kappa+1}r_{11})z_{\{1\}}+\nu_1,\ldots,(1-\sum_{l=2}^{\kappa+1}r_{k1})z_{\{k\}}+\nu_k;(1-\sum_{l=2}^{\kappa+1}r_{k+1,1})z_{\{k+1\}}\\ &+\nu_{k+1}]\times\prod_{l=2}^{\kappa+1}d[z_{Q(l,1)}+1,r_{j_1l}z_{\{j_1\}}+1,\ldots,r_{j_{k-q(l)}l}z_{\{j_{k-q(l)}\}}+1;\\ &r_{k+1,l}z_{\{k+1\}}+1].\end{aligned} \tag{4D.20}$$

In (4D.20) we assume that there is at least one category, say C_{k+1}, on which all data is complete. The completely generalized case of at least some data being incomplete on all k+1 categories is more complicated. Therefore, the general case is deferred until Section 4D.3.3 where we out-

line the basic conversion of that case to the one of this section and give results for the mean and covariance matrices. For shorthand, we refer to the $\kappa+1$ densities on the right-hand side of (4D.20) as simply d(1), d(2), ...,d($\kappa+1$), respectively. Note that r_{j1} in (4D.20) is often zero for most $1 \leq j \leq \mathcal{D}(1)$ and $2 \leq 1 \leq \kappa+1$.

Since the limit of a product is the product of the limits of the multiplicands and (Appendix 4B) the limiting Dirichlet distribution is multivariate normal, the limit of the numerator (4D.20) of the posterior distribution is a product of multivariate normal distributions. That is,

$$\lim_{n\to\infty}[(4D.20)] = \prod_{1=1}^{\kappa+1} N_{\mathcal{D}(1)}(\underset{\sim}{\mu}^{(1)}, \underset{\sim}{\Sigma}^{(1)}), \tag{4D.21}$$

where $\underset{\sim}{\mu}^{(1)}=(\mu_i{}^{(1)})$ and $\underset{\sim}{\Sigma}^{(1)}=(\sigma_{ij}{}^{(1)})$ are the $\mathcal{D}(1)\times 1$-dimensional mean and $\mathcal{D}(1)\times\mathcal{D}(1)$-dimensional covariance matrices of $\underset{\sim}{p}$ given data

$$m^{(1)} = z_{Q(1,1)} + \sum_{i\notin Q(1,1)} [r_{i1} z_{\{i\}} + 1] \tag{4D.22}$$

if $2 \leq 1 \leq \kappa+1$ and

$$m^{(1)} = \sum_{i=1}^{k+1} [(1 - \sum_{1=2}^{\kappa+1} r_{i1}) z_{\{i\}} + \nu_i] \tag{4D.23}$$

if 1=1. Thus Wilks(1963,p179) , for $1 \leq i \leq k+1$,

$$\mu_i{}^{(1)} = [(1 - \sum_{1=2}^{\kappa+1} r_{i1}) z_{\{i\}} + \nu_i]/m^{(1)}; \tag{4D.24}$$

for $2 \leq 1 \leq \kappa+1$,

$$\mu_1{}^{(1)} = (z_{Q(1,1)} + 1)/m^{(1)}; \tag{4D.25}$$

for $2 \leq 1 \leq \kappa+1$, $2 \leq i \leq \mathcal{D}(1)+1$, and $j_{i-1} \notin Q(1,1)$ for $1 \leq j_{i-1} \leq k+1$,

$$\mu_i{}^{(1)} = (r_{j_{i-1}1} z_{\{j_{i-1}\}} + 1)/m^{(1)}; \tag{4D.26}$$

and, for $1 \leq i \leq \mathcal{D}(1)$, $i<j \leq \mathcal{D}(1)$, and $1 \leq 1 \leq \kappa+1$,

$$\sigma_{ii}^{(1)} = \mu_i^{(1)}(1-\mu_i^{(1)})/(m^{(1)}+1) \tag{4D.27}$$

and

$$\sigma_{ij}^{(1)} = -\mu_i^{(1)}\mu_j^{(1)}/(m^{(1)}+1). \tag{4D.28}$$

In most of this appendix we find it more convenient to refer to elements of the 1th mean, covariance, and, particularly, inverse covariance matrices for $2 \leq 1 \leq \kappa+1$ in terms of sets $\mathcal{S}$ and T, for $\mathcal{S}$ and T each one of the $\mathcal{D}(1)+1$ sets $Q(1,1)$, $Q(1,2) \equiv \{j_{i-1}\},..,Q(1,\mathcal{D}(1)+1) \equiv \{j_{\mathcal{D}(1)}\}$; $T \neq \mathcal{S}$. Accordingly, for $2 \leq 1 \leq \kappa+1$, define

$$\mu_{\{i\}}^{(1)} = (r_{i1}z_{\{i\}}+1)/m^{(1)}, \tag{4D.29}$$

$$\mu_{\mathcal{S}}^{(1)} = \begin{cases} \mu_1^{(1)} & \text{for } \mathcal{S}=Q(1,1) \\ \mu_{\{i\}}^{(1)} & \text{for } \mathcal{S}=\{i\}, \end{cases} \tag{4D.30}$$

$$\sigma_{\mathcal{S}\mathcal{S}}^{(1)} = \mu_{\mathcal{S}}^{(1)}(1-\mu_{\mathcal{S}}^{(1)})/(m^{(1)}+1), \tag{4D.31}$$

and, for $T \neq \mathcal{S}$,

$$\sigma_{\mathcal{S}T}^{(1)} = -\mu_{\mathcal{S}}^{(1)}\mu_T^{(1)}/(m^{(1)}+1). \tag{4D.32}$$

Because the multivariate densities in (4D.21) are of differing dimensions and on differing combinations of the same random variable $\underset{\sim}{p}$, we do not immediately have that product (4D.21) is a k-dimensional multivariate normal density on the k components of $\underset{\sim}{p}$. However, we now show that, owing to the special relationship between d(1) and d(1) for 1>1, product (4D.21) is also multivariate normal in the limit.

To begin, sum the $\kappa+1$ exponents in product (4D.21). For (4D.21) to be normal, we must be able to write this sum as the exponent of a normal density; that is, we must be able to write

$$\sum_{l=1}^{\kappa+1}(\underset{\sim}{p}^{(l)}-\underset{\sim}{\mu}^{(l)})\underset{\sim}{\Sigma}_{(l)}^{-1}(\underset{\sim}{p}^{(l)}-\underset{\sim}{\mu}^{(l)})' = (\underset{\sim}{p}-\underset{\sim}{u})\underset{\sim}{S}^{-1}(\underset{\sim}{p}-\underset{\sim}{u})' \tag{4D.33}$$

for $\underset{\sim}{0}\leq\underset{\sim}{u}\leq\underset{\sim}{1}$ and $\underset{\sim}{S}^{-1}$ positive definite and symmetric, where $\underset{\sim}{p}^{(l)}$ is the k-dimensional vector $(p_1,\ldots,p_k)$ if $l=1$ and the $\mathcal{D}(l)$-dimensional vector of $p_{Q(l,1)}, p_{Q(l,2)},\ldots,$ and $p_{Q(l,\mathcal{D}(l))}$ for the lth Dirichlet density in (4D.20) if $l>1$.

Expanding the right-hand side of (4D.33) yields, for S^{ij} the i,jth element of $\underset{\sim}{S}^{-1}$ for $1\leq i,j\leq k$,

$$\begin{aligned}\sum_{i=1}^{k}(p_i-u_i)^2S^{ii}+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}(p_i-u_i)(p_j-u_j)S^{ij} &= \sum_{i=1}^{k}p_i^2S^{ii}+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}p_ip_jS^{ij}\\ -2\Big[\sum_{i=1}^{k}p_iu_iS^{ii}+\sum_{i=1}^{k-1}\sum_{j>i}^{k}(u_ip_j+u_jp_i)S^{ij}\Big]&+\Big[\sum_{i=1}^{k}u_i^2S^{ii}+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}u_iu_jS^{ij}\Big].\end{aligned} \tag{4D.34}$$

Expanding the left-hand side of (4D.33) yields, for $\sigma_{(l)}^{ij}$ the i,jth element of $\underset{\sim}{\Sigma}_{(l)}^{-1}$ for $1\leq i,j\leq k$ and $\sigma_{(l)}^{\mathcal{S}}$ that element of $\underset{\sim}{\Sigma}_{(l)}^{-1}$ referenced by the sets $\mathcal{S}$ and T in $Q(l)$ [recall (4D.30) - (4D.32)],

$$\begin{aligned}\sum_{i=1}^{k}(p_i-\mu_i^{(l)})^2\sigma_{(l)}^{ii}+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}(p_i-\mu_i^{(l)})(p_j-\mu_j^{(l)})\sigma_{(l)}^{ij}&\\ +\sum_{l=2}^{\kappa+1}\Big[\sum_{\substack{\mathcal{S}\in Q(l)\\ \mathcal{S}\neq\{k+1\}}}(p_{\mathcal{S}}-\mu_{\mathcal{S}}^{(l)})^2\sigma_{(l)}^{\mathcal{S}\mathcal{S}}+2\sum_{\mathcal{S}\in Q(l)}\sum_{\substack{T\in Q(l)\\ T(1)>\mathcal{S}(1)\\ T\neq\{k+1\}}}(p_{\mathcal{S}}-\mu_{\mathcal{S}}^{(l)})(p_T-\mu_T^{(l)})\sigma_{(l)}^{\mathcal{S}T}\Big].&\end{aligned} \tag{4D.35}$$

In (4D.35) $\mathcal{S}(1)$ means the first element of the set $\mathcal{S}$. Thus, if $\mathcal{S}=\{4\}$, then

$\mathcal{S}(1)=4$. If $\mathcal{S}=Q(1,1)$, the $\mathcal{S}(1)$ is the first of the $q(1)$ linearly ordered elements of $Q(1,1)$. Hence, if $Q(1,1)=\{2,3,9\}$, then $\mathcal{S}(1)=2$.

Recalling that $p_{\mathcal{S}}=\sum_{i\in\mathcal{S}} p_i$, we expand coefficients of $\sigma_{(1)}^{ij}$ and $\sigma_{(1)}^{\mathcal{S}T}$ for all i, j, $\mathcal{S}$, and T to write (4D.35) as

$$\sum_{i=1}^{k} c_{ii}p_i^2+\sum_{i=1}^{k-1}\sum_{j>i}^{k} c_{ij}p_ip_j+\sum_{i=1}^{k} c_{0i}p_i+c_{00} \tag{4D.36}$$

where, for

$$\delta_{A,B}=\begin{cases}1 & \text{if A and B both true}\\ 0 & \text{otherwise,}\end{cases} \tag{4D.37}$$

the coefficients c_{ij} in (4D.36) are, for $1\le i\le k$,

$$c_{ii}=\sigma_{(1)}^{ii}+\sum_{l=2}^{\kappa+1}\Big[\sum_{\mathcal{S}\in Q(l)}\delta_{\mathcal{S}\ni i,\mathcal{S}\ni i}\sigma_{(l)}^{\mathcal{S}\mathcal{S}}\Big]; \tag{4D.38}$$

for $j>i$,

$$c_{ij}=2\Big[\sigma_{(1)}^{ij}+\sum_{l=2}^{\kappa+1}\sum_{\mathcal{S}\in Q(l)}\Big(\delta_{\mathcal{S}\ni i,\mathcal{S}\ni j}\sigma_{(l)}^{\mathcal{S}\mathcal{S}}+\sum_{\substack{T\in Q(l)\\ T\neq\mathcal{S}}}\delta_{\mathcal{S}\ni i,T\ni j}\sigma_{(l)}^{\mathcal{S}T}\Big)\Big]; \tag{4D.39}$$

$$c_{0i}=-2\Big\{\mu_i^{(1)}\sigma_{(1)}^{ii}+\sum_{\substack{j=1\\ j\neq i}}^{k}\mu_j^{(1)}\sigma_{(1)}^{ij}+\sum_{l=2}^{\kappa+1}\Big[\sum_{\mathcal{S}\in Q(l)}\Big(\delta_{\mathcal{S}\ni i,\mathcal{S}\ni i}\mu_{\mathcal{S}}^{(l)}\sigma_{(l)}^{\mathcal{S}\mathcal{S}} \tag{4D.40}$$

$$+\sum_{\substack{T\in Q(l)\\ T\neq\mathcal{S}}}\delta_{\mathcal{S}\ni i,T\ni i}\mu_T^{(l)}\sigma_{(l)}^{\mathcal{S}T}\Big)\Big]\Big\};$$

and

$$c_{00}=\sum_{i=1}^{k}(\mu_i^{(1)})^2\sigma_{(1)}^{ii}+2\sum_{i=1}^{k-1}\sum_{j>i}^{k}\mu_i^{(1)}\mu_j^{(1)}\sigma_{(1)}^{ij} \tag{4D.41}$$

$$+\sum_{l=2}^{\kappa+1}\Big\{\sum_{\substack{\mathcal{S}\in Q(l)\\ \mathcal{S}\not\ni k+1}}\Big[(\mu_{\mathcal{S}}^{(l)})^2\sigma_{(l)}^{\mathcal{S}\mathcal{S}}+2\sum_{\substack{T\in Q(l)\\ T(1)>\mathcal{S}(1)\\ T\not\ni k+1}}\mu_{\mathcal{S}}^{(l)}\mu_T^{(l)}\sigma_{(l)}^{\mathcal{S}T}\Big]\Big\}$$

where $\sigma^{\mathcal{S}T} = \sigma^{T\mathcal{S}}$.

Then, equating coefficients of p_i in (4D.34) and (4D.36) yields that, from the coefficients of p_i^2 for, $1 \leq i \leq k$:

$$S^{ii} = \sigma_{(1)}^{ii} + \sum_{l=2}^{\kappa+1} \Big[\sum_{\mathcal{S} \in Q(l)} \delta_{\mathcal{S} \supset i, \mathcal{S} \supset i} \sigma_{(l)}^{\mathcal{S}\mathcal{S}} \Big], \tag{4D.42}$$

and, from the coefficients of $p_i p_j$, for $1 \leq i \leq k$, $i < j \leq k$:

$$S^{ij} = \sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \Big[\sum_{\mathcal{S} \in Q(l)} \Big(\delta_{\mathcal{S} \supset i, \mathcal{S} \supset j} \sigma_{(l)}^{\mathcal{S}\mathcal{S}} + \sum_{\substack{T \in Q(l) \\ T \neq \mathcal{S}}} \delta_{\mathcal{S} \supset i, T \supset j} \sigma_{(l)}^{\mathcal{S}T} \Big) \Big]. \tag{4D.43}$$

Hence, elements of the inverse covariance matrix $\underset{\sim}{S}^{-1}$ are finite linear combinations of variances and covariances and thus are of the same order of magnitude as the comprising variances and covariances. Therefore, from (4D.27) - (4D.28) and (4D.31) - (4D.32), elements of $\underset{\sim}{S}$, if $\underset{\sim}{S}$ exists, are $O(n^{-1})$.

Continuing to equate coefficients of p_i, we have that, from the coefficients of $-2p_i$, for $1 \leq i \leq k$,

$$\sum_{j=1}^{k} u_j S^{ij} = \sum_{j=1}^{k} \mu_j^{(1)} \sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \Big[\sum_{\mathcal{S} \in Q(l)} \Big(\delta_{\mathcal{S} \supset i, \mathcal{S} \supset i} \mu_{\mathcal{S}}^{(l)} \sigma_{(l)}^{\mathcal{S}\mathcal{S}} + \sum_{\substack{T \in Q(l) \\ T \neq \mathcal{S}}} \delta_{\mathcal{S} \supset i, T \supset i} \mu_T^{(l)} \sigma_{(l)}^{\mathcal{S}T} \Big) \Big]. \tag{4D.44}$$

Substituting from (4D.42) and (4D.43) into the left-hand side of (4D.44) yields the left-hand side as

$$\sum_{\substack{j=1 \\ j \neq i}}^{k} u_j \Big[\sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \sum_{\mathcal{S} \in Q(l)} \Big(\delta_{\mathcal{S} \supset i, \mathcal{S} \supset j} \sigma_{(l)}^{\mathcal{S}\mathcal{S}} + \sum_{\substack{T \in Q(l) \\ T \neq \mathcal{S}}} \delta_{\mathcal{S} \supset i, T \supset j} \sigma_{(l)}^{\mathcal{S}T} \Big) \Big] \tag{4D.45}$$

$$+ u_i[\sigma_{(1)}^{ii} + \sum_{l=2}^{\kappa+1}\Big(\sum_{\mathcal{S}\in Q(l)} \delta_{\mathcal{S}\ni i,\mathcal{S}\ni i}\sigma_{(l)}^{\mathcal{S}\mathcal{S}}\Big)]$$

$$= \sum_{j=1}^{k} u_j\sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1}\Big\{\sum_{\mathcal{S}\in Q(l)}\Big[\sum_{j=1}^{k} u_j\delta_{\mathcal{S}\ni i,\mathcal{S}\ni j}\sigma_{(l)}^{\mathcal{S}\mathcal{S}}$$

$$+ \sum_{\substack{\mathcal{T}\in Q(l)\\ \mathcal{T}\neq\mathcal{S}}}\Big(\sum_{\substack{j=1\\ j\neq i}}^{k} u_j\delta_{\mathcal{S}\ni i,\mathcal{T}\ni j}\sigma_{(l)}^{\mathcal{S}\mathcal{T}}\Big)\Big]\Big\}.$$

Equating coefficients of $\sigma_{(1)}^{ij}$, $1\le i,j\le k$, $\sigma_{(1)}^{\mathcal{S}\mathcal{S}}$, and $\sigma_{(1)}^{\mathcal{S}\mathcal{T}}$ on the left-hand side (4D.45) of (4D.44) with those on the right-hand side of (4D.44) yields that

from the coefficient of $\sigma_{(1)}^{ij}$

$$u_i = \mu_i^{(1)}; \tag{4D.46}$$

from the coefficient of $\sigma_{(1)}^{\mathcal{S}\mathcal{S}}$

$$\delta_{\mathcal{S}\ni i,\mathcal{S}\ni i}\mu_{\mathcal{S}}^{(1)} = \sum_{j=1}^{k} u_j\delta_{\mathcal{S}\ni i,\mathcal{S}\ni j}; \tag{4D.47}$$

therefore,

$$\mu_{\mathcal{S}}^{(1)} = \sum_{j\in\mathcal{S}} u_j, \tag{4D.48}$$

so that, from (4D.48),

$$\mu_{\mathcal{S}}^{(1)} = \sum_{j\in\mathcal{S}} \mu_j^{(1)}; \tag{4D.49}$$

and

from the coefficient of $\sigma_{(1)}^{\mathcal{S}\mathcal{T}}$

$$\delta_{\mathcal{S}\ni i,\mathcal{T}\not\ni i}\mu_{\mathcal{T}}^{(1)} = \sum_{\substack{j=1\\ j\neq i}}^{k} \delta_{\mathcal{S}\ni i,\mathcal{T}\ni j}u_j = \sum_{j\in\mathcal{T}} u_j,$$

so that, echoing (4D.49),

$$\mu_T^{(1)} = \sum_{j\in T} \mu_j^{(1)}. \tag{4D.50}$$

The last step in equating coefficients of p_i is checking that for values (4D.46) for u_i, (4D.42) for S^{ii}, and (4D.43) for S^{ij}, the constant term $\sum_{i=1}^{k} u_i^2 S^{ii} + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} u_i u_j S^{ij}$ from the right-hand side (4D.34) of desired identity (4D.33) equals the constant term (4D.41) from the left-hand side (4D.36) of (4D.33).

Substituting for u_i, S^{ii}, and S^{ij} in the constant term of (4D.34) yields that

$$\begin{aligned}
\sum_{i=1}^{k} u_i^2 S^{ii} + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} u_i u_j S^{ij} &= \sum_{i=1}^{k} (\mu_i^{(1)})^2 [\sigma_{(1)}^{ii} + \sum_{l=2}^{\kappa+1} (\sum_{\mathcal{S}\in Q(l)} \delta_{\mathcal{S}\ni i,\mathcal{S}\ni i}\sigma_{(1)}^{\mathcal{S}\mathcal{S}})] \\
&\quad + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} \mu_i^{(1)}\mu_j^{(1)} [\sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \sum_{\mathcal{S}\in Q(l)} (\delta_{\mathcal{S}\ni i,\mathcal{S}\ni j}\sigma_{(1)}^{\mathcal{S}\mathcal{S}} \\
&\quad + \sum_{\substack{T\in Q(l)\\ T\neq\mathcal{S}}} \delta_{\mathcal{S}\ni i,T\ni j}\sigma_{(1)}^{\mathcal{S}T})] \qquad (4D.51) \\
&= \sum_{i=1}^{k} (\mu_i^{(1)})^2 \sigma_{(1)}^{ii} + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} \mu_i^{(1)}\mu_j^{(1)}\sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \{ \sum_{S\in Q(l)} \\
&\quad \times [(\sum_{i\in\mathcal{S}} \mu_i^{(1)})^2 \sigma_{(1)}^{\mathcal{S}\mathcal{S}} + 2 \sum_{\substack{T\in Q(l)\\ T(1)>\mathcal{S}(1)}} (\sum_{i\in S} \mu_i^{(1)})(\sum_{j\in T} \mu_j^{(1)})\sigma_{(1)}^{\mathcal{S}T}]\}.
\end{aligned}$$

Thus, since (4D.49) gives that $\sum_{j\in\mathcal{S}} \mu_j^{(1)} = \mu_{\mathcal{S}}^{(1)}$, the constant term (4D.51) from the right-hand side of (4D.33) equals the constant term (4D.41) from the left-hand side of (4D.33).

Remaining in our proof is to show that $\underset{\sim}{0} \leq \underset{\sim}{u} \leq \underset{\sim}{1}$ and that $\underset{\sim}{S}^{-1}$ is positive definite and symmetric. From equation (4D.46) for u_i,

definition (4D.24) of $\mu_i^{(1)}$, and bounds on ratios r_{i1}, $1 \leq i \leq k$, $2 \leq 1 \leq k+1$, we have that $\underset{\sim}{0} \leq \underset{\sim}{u} \leq \underset{\sim}{1}$. Before proceeding to the remaining proof, we note that $\underset{\sim}{S}^{-1}$ being positive definite and symmetric implies that $\underset{\sim}{S}=(\underset{\sim}{S}^{-1})^{-1}$ exists [Graybill (1969,p318)] and is positive definite and symmetric [Anderson (1958,p337)]; hence [Dempster (1969,p41)], $\underset{\sim}{S}$ is a covariance matrix.

The matrix $\underset{\sim}{S}^{-1}$ is symmetric because, from (4D.42) and (4D.43), each element S^{ij} of $\underset{\sim}{S}^{-1}$ is a finite sum of elements from inverse covariance matrices $\underset{\sim}{\Sigma}_{(1)}^{-1}$, each of which is symmetric by definition of covariance. From Dempster (1969,p41), the matrix $\underset{\sim}{S}^{-1}$ is positive definite if and only if $\underset{\sim}{y}\underset{\sim}{S}^{-1}\underset{\sim}{y}'>0$ for all k-dimensional $\underset{\sim}{y} \neq \underset{\sim}{0}$. Thus, let $\underset{\sim}{y}$ be any k-dimensional vector such that $\underset{\sim}{y} \neq \underset{\sim}{0}$. Then,

$$
\begin{aligned}
\underset{\sim}{y}\underset{\sim}{S}^{-1}\underset{\sim}{y}' &= \sum_{i=1}^{k}\sum_{j=1}^{k} y_i y_j S^{ij} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} y_i y_j [\sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \sum_{S \in Q(l)} (\delta_{\$ \ni i, \$ \ni j} \sigma_{(1)}^{\$\$} \\
&\quad + \sum_{\substack{T \in Q(l) \\ T \neq \$}} \delta_{\$ \ni i, T \ni j} \sigma_{(1)}^{\$T})] \qquad (4D.52) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} y_i y_j \sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} [\sum_{i=1}^{k}\sum_{j=1}^{k} y_i y_j \sum_{\$ \in Q(l)} (\delta_{\$ \ni i, \$ \ni j} \sigma_{(1)}^{\$\$} \\
&\quad + \sum_{\substack{T \in Q(l) \\ T \neq \$}} \delta_{\$ \ni i, T \ni j} \sigma_{(1)}^{\$T})].
\end{aligned}
$$

Since $\underset{\sim}{y} \neq \underset{\sim}{0}$ and the inverse covariance matrix $\underset{\sim}{\Sigma}_{(1)}^{-1}$ is positive definite, the first term $\sum_{i=1}^{k}\sum_{j=1}^{k} y_i y_j \sigma_{(1)}^{ij}$ in the last equality in (4D.52) is positive. It therefore remains to show that the remaining

term is nonnegative. We can write this term as

$$\sum_{l=2}^{\kappa+1}\sum_{i=1}^{k}\sum_{j=1}^{k}\Big[\sum_{\mathcal{S}\in Q(l)} y_i y_j \delta_{\mathcal{S}\ni i,\mathcal{S}\ni j}\sigma_{(l)}^{\mathcal{S}\mathcal{S}} + \sum_{\substack{T\in Q(l)\\ T\neq\mathcal{S}}} y_i y_j \delta_{\mathcal{S}\ni i,T\ni j}\sigma_{(l)}^{\mathcal{S}T}\Big]$$

$$= \sum_{l=2}^{\kappa+1}\Big[\sum_{\mathcal{S}\in Q(l)}\Big(\sum_{i\in\mathcal{S}} y_i^2\Big)\sigma_{(l)}^{\mathcal{S}\mathcal{S}} + \sum_{\substack{T\in Q(l)\\ T\neq\mathcal{S}}}\Big(\sum_{i\in\mathcal{S}}\sum_{j\in T} y_i y_j\Big)\sigma_{(l)}^{\mathcal{S}T}\Big] \qquad (4D.53)$$

$$= \sum_{l=2}^{\kappa+1}\Big[\sum_{\mathcal{S}\in Q(l)}\sum_{T\in Q(l)} w_{\mathcal{S}} w_T \sigma^{\mathcal{S}T}\Big] \geq 0,$$

for $w_{\mathcal{S}} = \sum_{i\in\mathcal{S}} y_i$ and $w_T = \sum_{j\in T} y_j$, since every matrix $\underset{\sim}{\Sigma}_{(l)}^{-1}$, $2\leq l\leq\kappa+1$, is positive definite so that the term within brackets is nonnegative [positive unless $w_{\mathcal{S}} = 0$ for all $\mathcal{S}\in Q(l)$] and since the sum of κ nonnegative numbers is again nonnegative. Therefore, the matrix $\underset{\sim}{S}^{-1}$ is positive definite.

Thus, for values (4D.46) for u_i, (4D.42) for S^{ii}, and (4D.43) for S^{ij}, equality (4D.33) holds; that is, we can write the sum of exponents in the limiting numerator (4D.21) of the posterior density as the exponential term of a k-dimensional multivariate normal density.

Now, the limit of the posterior density (4D.16) is the limit (4D.21) of the numerator divided by the limit of the denominator. To calculate the limit of the denominator, first note that a Dirichlet density is continuous in $\underset{\sim}{p}$. Further, from (4B.29), Appendix 4B, a multidimensional Dirichlet density uniformly converges to a multivariate normal density. Therefore, the product (4D.20) of Dirichlet densities, the numerator of the posterior density, is continuous in $\underset{\sim}{p}$ and uniformly converges to a product of multivariate normal densities on the closed and bounded k-dimensional set P_k. Thus [Buck (1965,p186),

Bartle (1966,p67)], we have for the denominator of the posterior density (D.5) that

$$\lim_{n\to\infty} \int_{P_k} \text{num}\,[f(\underset{\sim}{p}|\underset{\sim}{z})]\,d\underset{\sim}{p} = \int_{P_k} \lim_{n\to\infty} \{\text{num}\,[f(\underset{\sim}{p}|\underset{\sim}{z})]\}\,d\underset{\sim}{p}. \qquad (4D.54)$$

Therefore, canceling coefficients $\prod_{l=1}^{\kappa+1} (2\pi)^{-\mathcal{D}(l)/2}[\det(\underset{\sim}{\Sigma}^{(l)})]^{-1/2}$ in the limiting numerator and denominator of the posterior density and multiplying both by $(2\pi)^{-k/2}[\det(S)]^{-1/2}$ yields the limiting denominator as 1 and the limiting numerator, and thus the limiting posterior density, as the density of the k-dimensional $N_k(\underset{\sim}{u},\underset{\sim}{S})$ multivariate normal distribution with elements of the mean and covariance matrices given by (4D.46), (4D.42), and (4D.43), respectively.

Rao [1968,(xv)p104] proves that if the density of a random variable converges to some density, then the distribution of the random variable converges to the distribution for the limiting density. Therefore, we have proved for all cases but that in which all $\kappa+1$ categories have some incomplete data, which case will be considered in 4D.3.3, that the limit of the k-dimensional posterior distribution of $\underset{\sim}{p}$ given incomplete data $\underset{\sim}{z}$ is k-dimensional multivariate normal.

4D.3.2 Special Case:

In Section 4D.3.1, elements (4D.46), (4D.42), and (4D.43) of the asymptotic mean vector $\underset{\sim}{u}$ and inverse covariance matrix $\underset{\sim}{S}^{-1}$, respectively were expressed in terms of unknown ratios r_{il}, $1\leq i\leq k+1$, $2\leq l\leq \kappa+1$, and incomplete data $\underset{\sim}{z}$. In this subsection we eliminate these ratios and derive expressions for elements of the asymptotic mean and covariance

matrices in terms of the asymptotic means and the data $\underset{\sim}{z}$. Again we assume that there exists at least one category, say C_{k+1}, on which all data is complete. The next subsection treats the general case allowing all categories to have some incomplete data.

Recall from Section 4D.3.1 that for κ different patterns of incomplete data, we separated the numerator of the posterior density into $\kappa+1$ Dirichlet densities $d(l)$, $1 \leq l \leq \kappa+1$. For the first density we had complete data on all $k+1$ categories C_i, $1 \leq i \leq k+1$, and for each of the κ remaining Dirichlet densities $d(l)$, $2 \leq k \leq \kappa+1$, we had exactly one of the κ sets of incomplete data. Recalling from (4D.19) that $\mathcal{D}(l)$ is the dimension of the lth Dirichlet density for $2 \leq l \leq \kappa+1$, note in (4D.20) that for each of the last κ Dirichlet densities, there are $\mathcal{D}(l)-1$ unknown ratios $r_{j_i l}$, $1 \leq i \leq \mathcal{D}(l)-1$. Thus, there are a total of $\kappa[\mathcal{D}(l)-1]$ unknown ratios $r_{j_i l}$, $1 \leq i \leq \mathcal{D}(l)-1$, $2 \leq l \leq \kappa+1$.

From (4D.24) - (4D.26), (4D.29), (4D.30), (4D.46), (4D.48), and (4D.49), elements u_i of the asymptotic mean vector u are expressed in terms of these $\kappa[\mathcal{D}(l)-1]$ unknown ratios. Letting l range from 2 to $\kappa+1$, we could derive a system of $\kappa[\mathcal{D}(l)-1]$ nonlinear equations in the $\kappa[\mathcal{D}(l)-1]$ unknown ratios, from which solution the asymptotic means could be evaluated.

However, an easier approach to evaluate these means is to reexpress them in a way that eliminates the ratios altogether and leads to a system of just k nonlinear equations in k unknowns, the unknowns then being the means. In such an approach, we will have evaluated the means in a one-step, rather than two-step, process and the nonlinear system to do so will be much simpler.

Recall that $r_{i1}=0$ for any Dirichlet density $d(1)$ for which category i has incomplete data. From (4D.49), for $2\le 1\le \kappa+1$,

$$\mu_1^{(1)} = \sum_{h\in Q(1,1)} \mu_h^{(1)}. \tag{4D.55}$$

For $m^{(1)}$ and $m^{(1)}$ defined in (4D.22) and (4D.23), respectively, substitution from (4D.24) and (4D.25) into (4D.55) yields that

$$(z_{\mathcal{J}(k+1)}+1)/m^{(1)} = \sum_{h\in Q(1,1)} [(1 - \sum_{g=2}^{\kappa+1} r_{hg})z_{\{h\}}+\nu_h]/m^{(1)}. \tag{4D.56}$$

Similarly, use of (4D.49), (4D.24), (4D.26), and $p_{k+1}=1- \sum_{\mathcal{J}\in Q(1) \atop \mathcal{J}\neq\{k+1\}} p_{\mathcal{J}}$, for all 1, yields that, for $2\le i\le \mathcal{D}(1)+1$,

$$(r_{j_i 1}z_{\{j_i\}}+1)/m^{(1)} = [(1 - \sum_{g=2}^{\kappa+1} r_{j_i g})z_{\{j_i\}}+\nu_{j_i}]/m^{(1)}. \tag{4D.57}$$

Hence, ignoring terms 1 and ν_h that go to zero as the sample size n increases, we have from (4D.56) and (4D.57) that, for $2\le i\le \mathcal{D}(1)+1$, $2\le 1\le \kappa+1$, and $1\le j_i\le k+1$,

$$\begin{aligned} m^{(1)}/m^{(1)} &= z_{\mathcal{J}(k+1)} / \sum_{h\in Q(1,1)} (1 - \sum_{g=2}^{\kappa+1} r_{hg})z_{\{h\}} \\ &= r_{j_i 1}/(1 - \sum_{g=2}^{\kappa+1} r_{j_i g}), \end{aligned} \tag{4D.58}$$

whence, defining $u_{\mathcal{J}} = \sum_{j\in\mathcal{J}} u_j$ for all sets $\mathcal{J}$, we have from (4D.46) that

$$\begin{aligned} m^{(1)} &= z_{Q(1,1)} / \sum_{h\in Q(1,1)} u_h \\ &= z_{Q(1,1)}/u_{Q(1,1)}. \end{aligned} \tag{4D.59}$$

Therefore, from (4D.24), (4D.46), (4D.48), and (4D.58), for $1\le i\le k+1$ and $n= \sum_{i=1}^{k+1} z_{\{i\}} + \sum_{1=2}^{\kappa+1} z_{Q(1,1)}$ the sample size,

$$
\begin{aligned}
u_i &= (1-\sum_{l=2}^{\kappa+1} r_{il})z_{\{i\}}/m^{(1)} \\
&= [z_{\{i\}}/n][n(1-\sum_{l=2}^{\kappa+1} r_{il})]/m^{(1)} \\
&= [z_{\{i\}}/n][(m^{(1)}+\sum_{l=2}^{\kappa+1} m^{(l)})(1-\sum_{l=2}^{\kappa+1} r_{il})/m^{(1)}] \\
&= [z_{\{i\}}/n][(1-\sum_{l=2}^{\kappa+1} r_{il})(1+\sum_{\substack{l=2 \\ Q(l,1)\ni i}}^{\kappa+1} m^{(l)}/m^{(1)}+\sum_{\substack{l=2 \\ Q(l,1)\ni i}}^{\kappa+1} m^{(l)}/m^{(1)})] \\
&= [z_{\{i\}}/n](1-\sum_{l=2}^{\kappa+1} r_{il})\{1 \sum_{\substack{l=2 \\ Q(l,1)\ni i}}^{\kappa+1} z_{Q(l,1)}/\sum_{h\in Q(l,1)}[(1-\sum_{g=2}^{\kappa+1} r_{hg})z_{\{h\}}] \\
&\qquad + \sum_{\substack{l=2 \\ Q(l,1)\not\ni i}}^{\kappa+1} r_{il}/(1-\sum_{l=2}^{\kappa+1} r_{il})\} \qquad (4D.60) \\
&= \{z_{\{i\}}+\sum_{Q(l,1)\ni i}[(1-\sum_{l=2}^{\kappa+1} r_{il})z_{\{i\}}/m^{(1)}] \\
&\qquad /\sum_{h\in Q(l,1)}[(1-\sum_{g=2}^{\kappa+1} r_{hg})z_{\{h\}}/m^{(1)}]z_{Q(l,1)}\}/n \\
&= [z_{\{i\}}+\sum_{D\ni i}(u_i/\sum_{j\in D} u_j)z_D]/n \\
&= [z_{\{i\}}+\sum_{D\ni i} z_D u_i/u_D]/n.
\end{aligned}
$$

Note that (4D.60) is the maximum likelihood estimate (4D.1).

Successively setting $i=1,\ldots,k$ in equation (4D.60) yields a system of k nonlinear equations to solve for the k unknowns u_i, $1\le i\le k$, where we also have the constraints $0\le u_i\le 1$ and $\sum_{i=1}^{k+1} u_i=1$. Some of the numerous approaches for finding a numerical solution are outlined in Scheid (1968,chpt.25). As discussed in Section 2.3.2, Dempster, Laird, and

Rubin describe an algorithm for an iterative solution. Examples allowing exact solution are given in Section 4D.5.

Noting from Graybill (1969,p170) the form of elements of the inverse covariance matrix $\underset{\sim}{S}^{-1}$ and recalling (4D.27), (4D.28), (4D.31), (4D.32), and (4D.42), we have for large n that

$$\begin{aligned} S^{ii} &= \sigma_{(1)}^{ii} + \sum_{Q(1,1)\ni i} \sigma_{(1)}^{Q(1,1),Q(1,1)} + \sum_{Q(1,1)\not\ni i} \sigma_{(1)}^{\{i\},\{i\}} \\ &= (m^{(1)} + \sum_{Q(1,1)\not\ni i} m^{(1)})(u_i+u_{k+1})/(u_i u_{k+1}) + \sum_{Q(1,1)\ni i} m^{(1)}(u_{Q(1,1)}+u_{k+1}) \\ &\qquad /(u_{Q(1,1)}u_{k+1}) \qquad (4D.61) \\ &= n(u_i+u_{k+1})/(u_i u_{k+1}) - \sum_{D\ni i}(z_D/u_D)(u_D-u_i)/(u_i u_D) \end{aligned}$$

since

$$m^{(1)} = n - \sum_{l=2}^{K+1} m^{(l)} = n - \Big(\sum_{D\ni i} z_D/u_D + \sum_{D\not\ni i} z_D/u_D\Big). \qquad (4D.62)$$

Similarly, from (4D.43) and for "$Q(l,1)\ni i,j$" beneath a summation sign meaning $Q(l,1)$ containing both i and j,

$$\begin{aligned} S^{ij} &= m^{(1)}/u_{k+1} + \sum_{Q(1,1)\ni i,j} m^{(1)}(u_{Q(1,1)}+u_{k+1})/(u_{Q(1,1)}u_{k+1}) + \sum_{Q(1,1)\not\ni i,j} m^{(1)}/u_{k+1} \\ &= (n - \sum_{D\ni i,j} z_D/u_D)/u_{k+1} + \sum_{D\ni i,j}(z_D/u_D)(u_D+u_{k+1})/(u_D u_{k+1}) \qquad (4D.63) \\ &= n/u_{k+1} + \sum_{D\ni i,j}(z_D/u_D)/u_D. \end{aligned}$$

Note how simple final results in (4D.62) and (4D.63) are, especially compared with corresponding equations (4D.12) and (4D.13) from the traditional approach in 4D.2. Furthermore, final results in (4D.62) and (4D.63) parallel results (given by their first term) for complete data [See Graybill (1969,p171).]

4D.3.3 General Case:

As long as there is at least one category having no incompletely specified data, we can apply the methods of the preceding sections. That is, if category C_{k+1} has some incomplete data, we can change the dependent variable from p_{k+1} to any variable p_i for which category C_i has only complete data. However, there are cases in which no category has only complete data; i.e., all k+1 categories have some incomplete data, so that such a variable does not exist. In this section, we extend theory from the preceding sections to this remaining case.

The only time there are problems using the theory of the preceding subsections is when Q(1,1) contains that element, say k+1, that indexes the dependent variable for d(1). To handle these instances, we have two approaches. In the first approach, we write $p_{Q(1,1)}$ as $1- \sum_{j \in Q(1,1)} p_j$ and then proceed with the methods of 4D.3.1 of equating coefficients of powers of p_i on the left- and right-hand sides of (4D.33). A simpler approach is making $p_{Q(1,1)}$ the dependent variable and then proceeding as in 4D.3.1 and 4D.3.2.

The first approach requires more types of cases than the second approach and, unlike the second approach, requires transformation of formula for the inverse covariance matrix before allowing proof that this matrix is positive definite. Hence, we pursue the second approach. Therefore, if Q(1,1) contains k+1, then we make $p_{Q(1,1)}$, instead of $p_{Q(1,k+q+1)}$, the dependent variable.

Following this approach and the procedures of 4D.3.1 and 4D.3.2 yields for elements u_i, S^{ii}, and S^{ij}, respectively, of the mean and covariance matrices

$$u_i = [z_{\{i\}} + \sum_{D \ni i} z_D (u_i/u_D)]/n, \tag{4D.64}$$

$$\begin{aligned} S^{ii} &= \sigma_{(1)}^{ii} + \sum_{l=2}^{\kappa+1} \{ \sum_{Q(l,1) \not\ni i} \sigma_{(l)}^{\{i\},\{i\}} + \sum_{\substack{Q(l,1) \ni i \\ Q(l,1) \not\ni k+1}} \sigma_{(l)}^{Q(l,1),Q(l,1)} \\ &= m^{(1)}(u_i+u_{k+1})/(u_i u_{k+1}) + \sum_{l=2}^{\kappa+1} \{ \sum_{\substack{Q(l,1) \not\ni i \\ Q(l,1) \not\ni k+1}} m^{(l)}(u_i+u_{k+1})/(u_i u_{k+1}) \\ &\quad + \sum_{\substack{Q(l,1) \ni i \\ Q(l,1) \not\ni k+1}} m^{(l)}(u_{Q(l,1)}+u_{k+1})/(u_{Q(l,1)} u_{k+1}) + \sum_{\substack{Q(l,1) \ni k+1 \\ Q(l,1) \not\ni i}} m^{(l)} \\ &\quad \times (u_{Q(l,1)}+u_i)/(u_i u_{Q(l,1)}) \} \\ &= n(u_i+u_{k+1})/(u_i u_{k+1}) - \sum_{\substack{D \not\ni i \\ D \ni k+1}} (z_D/u_D)(u_D-u_{k+1})/(u_D u_{k+1}) - \sum_{\substack{D \ni i \\ D \not\ni k+1}} (z_D/u_D) \\ &\quad (u_D-u_i)/(u_D u_i) - \sum_{\substack{D \ni i \\ D \ni k+1}} (z_D/u_D)(u_i+u_{k+1})/(u_i u_{k+1}), \end{aligned} \tag{4D.65}$$

and

$$\begin{aligned} S^{ij} &= \sigma_{(1)}^{ij} + \sum_{l=2}^{\kappa+1} \{ \sum_{\substack{Q(l,1) \ni i \\ Q(l,1) \not\ni j}} \sigma_{(l)}^{Q(l,1),\{j\}} + \sum_{\substack{Q(l,1) \not\ni i \\ Q(l,1) \not\ni j}} \sigma_{(l)}^{\{i\},\{j\}} \\ &\quad + \sum_{\substack{Q(l,1) \not\ni i \\ Q(l,1) \ni j}} \sigma_{(l)}^{Q(l,1),\{i\}} + \sum_{Q(l,1) \ni i,j} \sigma_{(l)}^{Q(l,1),Q(l,1)} \} + \sum_{\substack{l=2 \\ Q(l,1) \ni k+1 \\ Q(l,1) \not\ni i,j}}^{\kappa+1} \sigma_{(l)}^{\{i\},\{j\}} \\ &= m^{(1)}/u_{k+1} + \sum_{l=2}^{\kappa+1} [\sum_{Q(l,1) \not\ni i,j} m^{(l)}/u_{k+1} + \sum_{Q(l,1) \ni i,j} m^{(l)}(u_{Q(l,1)}+u_{k+1}) \end{aligned}$$

$$/(u_{Q(1,1)}u_{k+1})]+\sum_{\substack{Q(1,1)\ni k+1\\ Q(1,1)\not\ni i,/j}} m^{(1)}/u_{Q(1,1)} \tag{4D.66}$$

$$= n/u_{k+1}+\sum_{\substack{Q(1,1)\not\ni k+1\\ Q(1,1)\ni i,j}} m^{(1)}/u_{Q(1,1)}+\sum_{Q(1,1)\ni k+1}\Big[\sum_{Q(1,1)\not\ni i,\not\ni j} m^{(1)}(1/u_{Q(1,1)}$$

$$-1/u_{k+1})-m^{(1)}\Big(\sum_{Q(1,1)\ni i,j} 1+\sum_{\substack{Q(1,1)\ni i\\ Q(1,1)\not\ni j}} 1+\sum_{\substack{Q(1,1)\not\ni i\\ Q(1,1)\ni j}} 1\Big)/u_{k+1}\Big]$$

$$= n/u_{k+1}+\sum_{\substack{D\not\ni k+1\\ D\ni i,j}} (z_D/u_D)/u_D-\sum_{D\ni k+1} (z_D/u_D)/u_{k+1}+\sum_{\substack{D\ni k+1\\ D\not\ni i,\not\ni j}} (z_D/u_D)/u_D,$$

where "$D\ni i,j$" means D containing <u>both</u> i and j, "$D\not\ni i,j$" means D not containing i and j together (ie, D can contain one or neither of i and j but not both), and all conditions under a summation sign are to be met simultaneously, since, as in Section 4D.3.2, the procedure yields that, for $2\le l\le K+1$,

$$m^{(l)} = z_{Q(l,1)}/u_{Q(l,1)} \tag{4D.67}$$

and

$$m^{(1)} = n-\sum_{l=2}^{K+1} m^{(l)} \tag{4D.68}$$

for

$$u_{Q(l,1)} = \sum_{j\in Q(l,1)} u_j. \tag{4D.69}$$

Proof of positive definiteness of $\underset{\sim}{S}^{-1}$ will parallel that given in (4D.52) and (4D.53) of the last section with the following modification. Note from the first equality of (4D.65) and (4D.66) for S^{ii} and S^{ij},

respectively, that no direct contribution is made to S^{ii} and S^{ij} from those sets Q(1,1) simultaneously containing both i and k+1. Thus, we must modify (4D.53) by adding $\ell \neq k+1$ under both $\ell \in Q(1)$ and $T \in Q(1)$ everywhere in (4D.53). Therefore, the sums within brackets in (4D.53) for these particular sets Q(1) will involve only that submatrix of $\underset{\sim}{\Sigma}_{(1)}^{-1}$ referring to those variables not indexed in $Q(1,1) \ni i,k+1$. But since this submatrix is also a covariance matrix, it is positive definite; thus, the remaining proof will follow like that of (4D.53).

Remaining proofs for the limiting posterior distribution are identical to those of Sections 4D.3.1 and 4D.3.2. Therefore, for all cases the limiting posterior distribution of $\underset{\sim}{p}$ given incomplete multinomial data $\underset{\sim}{z}$ is multivariate normal with expressions for elements of the mean and inverse covariance matrix given by (4D.64) - (4D.66). Note that, as in Section 4D.3.2, expressions (4D.65) and (4D.66) for elements of the inverse covariance matrix are simple and parallel those for complete data.

4D.4 Equivalence of Results:

In this section, we show how results (4D.12) and (4D.13) for the asymptotic inverse covariance matrix given in 4D.2 by the traditional approach can be simplified to those, (4D.65) and (4D.66), respectively, given by the nontraditional approach in Section 4D.3. Because of the large amount of algebraic manipulation (and, thus, possible error) involved, knowing results (4D.65) and (4D.66) to work toward is very important.

To show that (4D.12) equals (4D.65) for S^{ii}, divide (4D.12) into the four groupings - complete-data term, sum of all terms for which $D \not\ni i, D \ni k+1$, sum of all terms for which $D \ni i, D \not\ni k+1$, and sum of all terms for which $D \ni i, D \ni k+1$ - given in (4D.65). Note in making this division that there are no terms for which $D \not\ni i, D \not\ni k+1$; the one combination for which there is no contribution in (4D.65).

In (4D.12) we can rewrite the complete-data term as

$$[n(u_i+u_{k+1})/(u_i u_{k+1})]\{[(u_i+u_{k+1})(1-u_i)-2u_i(1-u_i-u_{k+1})+u_i(1-u_i-u_{k+1})]/u_{k+1}\} \tag{4D.70}$$
$$= n(u_i+u_{k+1})/(u_i u_{k+1})$$

since the term inside braces is one.

For the sum of those terms in (4D.12) over those sets $D \not\ni i$, we have

$$\sum_{a\neq i}^{k} u_a\Big[-\sum_{\substack{D\ni a,\not\ni i\\ D\not\ni k+1}} z_D(u_D-u_a)/u_D^2+\sum_{b\neq i,a}^{k}\sum_{\substack{D\ni a,b\\ D\not\ni i,\not\ni k+1}} z_D u_b/u_D^2-\sum_{\substack{D\ni a,\not\ni i\\ D\ni k+1}} z_D(u_D-u_a)u_D^2$$
$$+\sum_{b\neq i,a}^{k}\Big(\sum_{\substack{D\ni a,b\\ D\not\ni i,\ni k+1}} z_D u_b/u_D^2\Big)\Big]/u_{k+1}^2 \tag{4D.71}$$
$$= -\sum_{a\neq i}^{k} u_a\Big[\sum_{\substack{D\ni a,\not\ni i\\ D\ni k+1}}(z_D/u_D^2)\Big]/u_{k+1}$$

$$= - \sum_{\substack{D \not\ni i \\ D \ni k+1}} [(z_D/u_D^2)(\sum_{\substack{a \in D \\ a \neq k+1}} u_a)]/u_{k+1}$$

$$= - \sum_{\substack{D \not\ni i \\ D \ni k+1}} (z_D/u_D)(u_D-u_{k+1})/(u_D u_{k+1})$$

since, inside the brackets in the first line in (4D.71), the first term is the negative of the second because, for the restrictions $D \not\ni i,k+1$ on D for these two terms,

$$u_D - u_a = \sum_{\substack{b \in D \\ b \neq i,a}} u_b. \tag{4D.72}$$

From the last two terms inside these brackets, we pick up

$$- \sum_{\substack{D \ni a,k+1 \\ D \not\ni i}} z_D u_{k+1}/u_D^2 \tag{4D.73}$$

since, for the restrictions $D \not\ni i, D \ni k+1$ on D for these two terms,

$$u_D - u_a = \sum_{\substack{b \in D \\ b \neq i,a}}^{k} u_b + u_{k+1}. \tag{4D.74}$$

For the sum of those terms in (4D.12) over those sets $D \ni i$, we have

$$\begin{aligned}
&\{-(u_i+u_{k+1})^2/u_i \sum_{D \ni i} z_D(u_D-u_i)/u_D^2 + 2(u_i+u_{k+1}) \sum_{a \neq i}^{k} \sum_{D \ni i,a} u_a z_D/u_D^2 \\
&+ \sum_{a \neq i}^{k} u_a [- \sum_{\substack{D \ni a \\ D \ni i}} z_D(u_D-u_a)/u_D^2 + \sum_{b \neq i,a}^{k} (\sum_{\substack{D \ni a,b \\ D \ni i}} z_D u_b/u_D^2)]\}/u_{k+1}^2 \\
&= \{ \sum_{\substack{D \ni i \\ D \ni k+1}} z_D/u_D^2 [-(u_D-u_i)(u_i+u_{k+1})^2/u_i + 2(u_i+u_{k+1})(u_D-u_i-u_{k+1}) \\
&-(u_i+u_{k+1})(u_D-u_i-u_{k+1})] + \sum_{\substack{D \ni i \\ D \not\ni k+1}} z_D/u_D^2 [-(u_D-u_i)(u_i+u_{k+1})^2/u_i
\end{aligned} \tag{4D.75}$$

$$+2(u_D-u_i)(u_i+u_{k+1})-u_i(u_D-u_i)]\}/u_{k+1}^2$$

$$=-\sum_{\substack{D\ni i\\ D\ni k+1}}(z_D/u_D)(u_i+u_{k+1})/(u_iu_{k+1})-\sum_{\substack{D\ni i\\ D\not\ni k+1}}(z_D/u_D)(u_D-u_i)/(u_iu_D)$$

since, inside the braces in the first line of (4D.75) we can divide the first term into a sum over sets $D\ni k+1$ and a sum over sets $D\not\ni k+1$, we can write the second term as

$$2(u_i+u_{k+1})\Big[\sum_{\substack{D\ni i\\ D\ni k+1}} z_D/u_D^2\Big(\sum_{\substack{a\in D\\ a\neq i,k+1}} u_a\Big)+\sum_{\substack{D\ni i\\ D\not\ni k+1}} z_D/u_D^2\Big(\sum_{\substack{a\in D\\ a\neq i}} u_a\Big)\Big] \tag{4D.76}$$

$$=2(u_i+u_{k+1})\Big[\sum_{\substack{D\ni i\\ D\ni k+1}} z_D(u_D-u_i-u_{k+1})/u_D^2+\sum_{\substack{D\ni i\\ D\not\ni k+1}} z_D(u_D-u_i)/u_D^2\Big],$$

and we can write the last two terms as

$$\sum_{a\neq i} u_a\Big[-\sum_{\substack{D\ni i,a\\ D\not\ni k+1}} z_D(u_D-u_a)/u_D^2+\sum_{b\neq i,a}^{k}\sum_{\substack{D\ni a,b,i\\ D\not\ni k+1}} z_Du_b/u_D^2$$

$$-\sum_{\substack{D\ni i,a\\ D\ni k+1}} z_D(u_D-u_a)/u_D^2+\sum_{b\neq i,a}^{k}\sum_{\substack{D\ni a,b,i\\ D\ni k+1}} z_Du_b/u_D^2\Big]/u_{k+1}^2 \tag{4D.77}$$

$$=-\sum_{a\neq i}^{k} u_a\Big[\sum_{\substack{D\ni a,i\\ D\not\ni k+1}} z_Du_i/u_D^2+\sum_{\substack{D\ni a,i\\ D\ni k+1}} z_D(u_i+u_{k+1})/u_D^2\Big]/u_{k+1}^2$$

$$=-\Big[u_i\sum_{\substack{D\ni i\\ D\not\ni k+1}} z_D(u_D-u_i)/u_D^2+(u_i+u_{k+1})\sum_{\substack{D\ni i\\ D\ni k+1}} z_D(u_D-u_i-u_{k+1})/u_D^2\Big]/u_{k+1}^2$$

since the first two terms inside brackets in the first line of (4D.77) combine through

$$u_D - u_a = \sum_{\substack{b\in D \\ b\neq a}}^{k} u_b \tag{4D.78}$$

and the last two terms combine through

$$u_D - u_a = \sum_{\substack{b\in D \\ b\neq a}}^{k} u_b + u_{k+1}. \tag{4D.79}$$

Therefore, from (4D.70), (4D.71), and (4D.75), we have for S^{ii} that (4D.12) from the traditional approach does simplify to (4D.65) from the nontraditional approach.

Similarly breaking up terms given in (4D.13) for S^{ij} from the traditional approach, we have that, since

$$\begin{aligned}\sum_{a\neq i,j}^{k} u_a\Big(-n \sum_{\substack{b\neq i,j \\ b\neq a}}^{k} u_b\Big) &= -n \sum_{a\neq i,j}^{k} u_a(1-u_a-u_i-u_j-u_{k+1}) \\ &= n\Big[\sum_{a\neq i,j}^{k} u_a^2-(1-u_i-u_j-u_{k+1})^2\Big],\end{aligned} \tag{4D.80}$$

the complete-data term in (4D.13) is

$$\begin{aligned}&n\{(u_i+u_{k+1})[(1-u_i)-(u_j+u_{k+1})-(1-u_i-u_j-u_{k+1}]+u_j[-u_i-(1-u_i-u_j-u_{k+1}) \\ &\quad +(u_j+u_{k+1})(1-u_j)/u_j]+(1-u_i-u_j-u_{k+1})[-u_i-(u_j+u_{k+1})-(1-u_i-u_j-u_{k+1})+1]\}/u_{k+1}^2 \\ &= n/u_{k+1}.\end{aligned} \tag{4D.81}$$

Noting that we will find no contribution for the case $D\not\ni k+1,i,j$, we divide each of the ten sums over sets D in (4D.13) into the five cases: $D\not\ni k+1$, $D\ni i,j$; $D\ni k+1$, $D\ni i,j$; $D\ni k+1,i$, $D\not\ni j$; $D\ni k+1,j$, $D\not\ni i$; $D\ni k+1$, $D\not\ni i,j$.

Doing so and then combining results for the ninth and tenth sums, we have for the five cases that

$$\sum_{\substack{D\not\ni k+1\\ D\ni i,j}} (z_D/u_D)(u_{k+1}{}^2u_D)^{-1}[-(u_i+u_j)(u_D-u_i-u_j)+(u_j+u_{k+1})(u_D-u_i-u_j)$$

$$+u_i(u_D-u_i-u_j)-(u_j+u_{k+1})(u_D-u_j)+u_j(u_D-u_i-u_j)+u_iu_j$$

$$+(u_i+u_{k+1})(u_D-u_i-u_j)+(u_i+u_{k+1})(u_j+u_{k+1})-(u_i+u_{k+1})(u_D-u_i)]$$

$$= \sum_{\substack{D\not\ni k+1\\ D\ni i,j}} (z_D/u_D)/u_D, \tag{4D.82}$$

$$\sum_{\substack{D\ni k+1\\ D\ni i,j}} z_D/(u_Du_{k+1})^2[-(u_i+u_j+u_{k+1})(u_D-u_i-u_j-u_{k+1})+(u_j+u_{k+1})(u_D-u_i-u_j-u_{k+1})$$

$$+u_i(u_D-u_i-u_j-u_{k+1})-(u_j+u_{k+1})(u_D-u_j)+u_j(u_D-u_i-u_j-u_{k+1})+u_iu_j$$

$$+(u_i+u_{k+1})(u_D-u_i-u_j-u_{k+1})+(u_i+u_{k+1})(u_j+u_{k+1})-(u_i+u_{k+1})(u_D-u_i)]$$

$$= -\sum_{\substack{D\ni k+1\\ D\ni i,j}} (z_D/u_D)/u_{k+1}, \tag{4D.83}$$

$$\sum_{\substack{D\ni k+1,i\\ D\not\ni j}} z_D/(u_Du_{k+1})^2[-(u_i+u_{k+1})(u_D-u_i-u_{k+1})+u_i(u_D-u_i-u_{k+1})$$

$$+(u_i+u_{k+1})(u_D-u_i-u_{k+1})-(u_D-u_i)(u_i+u_{k+1})]$$

$$= -\sum_{\substack{D\ni k+1,i\\ D\not\ni j}} (z_D/u_D)/u_{k+1}, \tag{4D.84}$$

$$\sum_{\substack{D\ni k+1,j\\ D\not\ni i}} z_D/(u_D u_{k+1})^2[-(u_D-u_j-u_{k+1})(u_j+u_{k+1})+(u_j+u_{k+1})(u_D-u_j-u_{k+1})$$

$$-(u_j+u_{k+1})(u_D-u_j)+u_j(u_D-u_j-u_{k+1})] \qquad (4D.85)$$

$$= -\sum_{\substack{D\not\ni i,\ j\\ D\ni k+1}} (z_D/u_D)/u_{k+1},$$

and, for the last case,

$$\sum_{a\neq i,j}^{k} u_a\Big[\sum_{\substack{b\neq i,j\\ b\neq a}}^{k} u_b\Big(\sum_{\substack{D\ni a,b\\ D\not\ni i,\not\ni j\\ D\ni k+1}} z_D/u_D^2\Big)-\sum_{\substack{D\ni a,k+1\\ D\not\ni i,\not\ni j}} z_D(u_D-u_a)/u_D^2\Big]/u_{k+1}^2 \qquad (4D.86)$$

$$= -\sum_{\substack{D\not\ni i,\not\ni j\\ D\ni k+1}} (z_D/u_D)(u_D-u_{k+1})/(u_D u_{k+1}).$$

Therefore, from (4D.80) - (4D.86), we have for s^{ij} that (4D.13) from the traditional approach simplifies to (4D.66) from the nontraditional approach.

4D.5 Examples Allowing Exact Solution:

When the nonlinear system (4D.64) of equations for the asymptotic mean involves a polynomial in the mean components of degree less than 5, then an exact algebraic solution exists for the asymptotic mean and, hence, for the asymptotic posterior covariance matrix. In this section, we give three examples. For the first two examples, we give exact algebraic solutions as well as numerical evaluation for a data set. The second example requires use of the MACSYMA symbolic computer system. We conclude the section with a numerical example for the most general case for the trinomial distribution. This general case requires solution of a 5-degree polynomial. For one data set, we use MACSYMA to evaluate the five roots. The usual probability constraints $0 \leq p_i \leq 1$ and $\sum_{i=1}^{3} p_i = 1$, along with the nature of the data, preclude all solutions but one.

Note that the analysis in this section holds for the posterior mode and the Taylor-series approximate posterior mean as well as for the asymptotic posterior mean, which is the maximum likelihood estimate. In general, we do not use the exact solutions because they are too expensive and, as just discussed, hold only for special cases. Instead, we use the EM iterative algorithm of Dempster, Laird, and Rubin (1977) discussed in Section 2.3.2 to evaluate elements of the maximum likelihood estimate (hence, the asymptotic posterior mean), posterior mode, and Taylor-series approximate posterior mean.

For this section we drop the braces in the set notations $\{i\}$. Hence, we write z_i rather than $z_{\{i\}}$.

For the first example, we calculate the asymptotic mean and covariance matrix of $\underset{\sim}{p}$ given incomplete trinomial data $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12})$. Expression

(4D.64) gives two equations

$$u_1 = (z_1+z_{12}u_1/u_{12})/n$$
$$u_2 = (z_2+z_{12}u_2/u_{12})/n \tag{4D.87}$$

to solve for the two unknowns u_1 and u_2. Note that $u_1+u_2=1-z_3/n$ and that $u_3=1-u_1-u_2=z_3/n$. Solving (4D.87) for u_1 and u_2 yields that

$$u_1 = z_1[1+z_{12}/(z_1+z_2)]/n$$
and
$$u_2 = z_2[1+z_{12}/(z_1+z_2)]/n. \tag{4D.88}$$

From (4D.65) and (4D.66), elements of the asymptotic inverse posterior covariance matrix are

$$S^{11} = n(u_1+u_3)/(u_1u_3)-z_{12}u_2/[u_1(u_1+u_2)^2],$$
$$S^{12} = n/u_3+z_{12}/(u_1+u_2)^2, \tag{4D.89}$$
and
$$S^{22} = n(u_2+u_3)/(u_2u_3)-z_{12}u_1/[u_2(u_1+u_2)^2].$$

For data having values $z_1=105$, $z_2=98$, $z_3=200$, and $z_{12}=200$, evaluation of (4D.88) and (4D.89) yields

$$u_1 = .35, \quad u_2 = .32, \quad u_3 = .33,$$
and
$$S^{11} = 3,142.9, \quad S^{12} = 2,272.8, \quad \text{and} \quad S^{22} = 3,224.3. \tag{4D.90}$$

Hence, elements S_{ij} of the asymptotic posterior covariance matrix $\underset{\sim}{S}$ have values

$$S_{11} = 6.4900^{-4}, \quad S_{12} = 4.5748^{-4}, \quad \text{and} \quad S_{22} = 6.3261^{-4}. \qquad (4D.91)$$

From (4D.91) the standard deviations $\sqrt{S_{11}}=.025$ and $\sqrt{S_{22}}=.025$ are 7.1% and 7.6% of u_1 and u_2, respectively.

For the next example, consider incomplete trinomial data $\underset{\sim}{z}=(z_1,z_2, z_3,z_{12},z_{13})$. From (4D.64) the asymptotic mean is the solution of the following nonlinear system of equations

$$u_1 = (z_1+z_{12}u_1/u_{12}+z_{13}u_1/u_{13})/n$$

and (4D.92)

$$u_2 = (z_2+z_{12}u_2/u_{12})/n$$

where $u_3=1-u_1-u_2$ $[=(z_3+z_{13}u_3/u_{13})/n]$. Substituting for u_3 in (4D.92) and solving with MACSYMA yields in the following Table 4D.1 the three algebraic solutions for u_1 and u_2.

Substitution of data $z_1=100$, $z_2=200$, $z_3=200$, $z_{12}=200$, $z_{13}=200$, and $n=900$ into the three solution sets yields the three solutions $u_1=u_2=0$; $u_1=u_2=1/3$; and $u_1=-1/3$, $u_2=2/3$. Consideration of the constraint $u_i\geq 0$ eliminates the third solution. Consideration of the data eliminates the first solution. Therefore, there is only one satisfactory solution; $u_1=u_2=u_3=1/3$.

Note that results given in Table 4D.1 were expensive to obtain and utilized the maximum amount of computer memory available. Yet, these results were for only two patterns of incomplete data. Further, each of these patterns ({1,2} and {1,3}) involved only two categories (C_1,C_2 and C_1,C_3, respectively). A total of only three variables (p_1, p_2, and p_3) was involved. Hence, an algebraic solution can be obtained only in very special cases.

TABLE 4D.1

ALGEBRAIC SOLUTION FOR POSTERIOR MEAN

SOLUTION 1: $u_1=0, \quad u_2=0.$

SOLUTION 2:

$$u_1 = -((nz_2-z_1z_{12}+nz_1)\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2n)z_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2+2nz_1z_{12}z_{13}+n^2z_1^2)+nz_2^3+(nz_{13}+(z_1+n)z_{12}+2nz_1-n^2)z_2^2+(((z_1+n)z_{12}+nz_1)z_{13}+z_1z_{12}^2+(z_1^2+2nz_1)z_{12}+nz_1^2-2n^2z_1)z_2+(z_1z_{12}^2-nz_1z_{12})z_{13}+nz_1^2z_{12}-n^2z_1^2)/((nz_2+nz_{13}+nz_1)\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2n)z_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2+2nz_1z_{12}z_{13}+n^2z_1^2)+nz_2^3+(2nz_{13}-nz_{12}+2nz_1-n^2)z_2^2+(nz_{13}^2+(-2nz_{12}+2nz_1-n^2)z_{13}-nz_1z_{12}+nz_1^2-2n^2z_1)z_2-nz_{12}z_{13}^2+(-nz_1z_{12}-n^2z_1)z_{13}-n^2z_1^2)$$

$$u_2 = (-\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2nz_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2+2nz_1z_{12}z_{13}+n^2z_1^2)+z_2^2+z_{12}(z_2+z_{13})+n(z_2+z_1)+z_{13}z_2+z_1z_2)/(n(2z_2+2z_{13}+2z_1))$$

TABLE 4D.1 (continued)

SOLUTION 3:

$$u_1 = -((nz_2-z_1z_{12}+nz_1)\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2n)z_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2+2nz_1z_{12}z_{13}+n^2z_1^2)-nz_2^3+(-nz_{13}+(-z_1-n)z_{12}-2nz_1+n^2)z_2^2+(((-z_1-n)z_{12}-nz_1)z_{13}-z_1z_{12}^2+(-z_1^2-2nz_1)z_{12}-nz_1^2+2n^2z_1)z_2+(nz_1z_{12}-z_1z_{12}^2)z_{13}-nz_1^2z_{12}-n^2z_1^2)/((nz_2+nz_{13}+nz_1)\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2n)z_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2+2nz_1z_{12}z_{13}+n^2z_1^2)-nz_2^3+(-2nz_{13}+nz_{12}-2nz_1+n^2)z_2^2+(-nz_{13}^2+(2nz_{12}-2nz_1+n^2)z_{13}+nz_1z_{12}-nz_1^2+2n^2z_1)z_2-nz_{12}z_{13}^2+(nz_1z_{12}+n^2z_1)z_{13}+n^2z_1^2)$$

$$u_2 = (\mathrm{SQRT}(z_2^4+(2z_{13}+2z_{12}+2z_1-2n)z_2^3+(z_{13}^2+(4z_{12}+2z_1-2n)z_{13}+z_{12}^2+(2z_1-2n)z_{12}+z_1^2-4nz_1+n^2)z_2^2+(2z_{12}z_{13}^2+(2z_{12}^2+(2z_1-2n)z_{12}-2nz_1)z_{13}-2nz_1z_{12}-2nz_1^2+2n^2z_1)z_2+z_{12}^2z_{13}^2-2nz_1z_{12}z_{13}+n^2z_1^2)+z_2^2+z_{12}(z_2+z_{13})+n(z_2+z_1)+z_{13}z_2+z_1z_2)/(n(2z_2+2z_{13}+2z_1))$$

For the last example, consider incomplete trinomial data $\underset{\sim}{z}=(z_1,z_2,z_3, z_{12},z_{13},z_{23})$. For elements of the asymptotic mean and inverse covariance matrices, equations (4D.64) - (4D.66) yield

$$\begin{aligned} u_1 &= (z_1+z_{12}u_1/u_{12}+z_{13}u_1/u_{13})/n, \\ u_2 &= (z_2+z_{12}u_2/u_{12}+z_{23}u_2/u_{23})/n, \quad \text{and} \end{aligned} \tag{4D.93}$$

$$\begin{aligned} S^{11} &= nu_{13}/(u_1u_3)-z_{23}u_2/(u_3u_{23}^2)-z_{12}u_2/(u_1u_{12}^2)-z_{13}/(u_1u_3), \\ S^{12} &= n/u_3+z_{12}/u_{12}^2-z_{13}/(u_3u_{13})-z_{23}/(u_3u_{23}), \\ S^{22} &= nu_{23}/(u_2u_3)-z_{13}u_1/(u_3u_{13}^2)-z_{12}u_1/(u_2u_{12}^2)-z_{23}/(u_2u_3), \end{aligned} \tag{4D.94}$$

respectively. Note that (4D.93) is a nonlinear system of equations involving fifth powers of the means. Therefore, we do not obtain the exact algebraic solution. However, suppose that $z_1=3{,}000$, $z_2=4{,}400$, $z_3=10{,}000$, $z_{12}=5{,}000$, $z_{13}=3{,}400$, and $z_{23}=4{,}000$. Then, substituting these values into (4D.93) and setting $u_3=1-u_1-u_2$ yields, with the aid of MACSYMA, the five sets of solutions:

$u_1=u_2=0$; $u_1=0.8151925$, $u_2=0.852957431$; $u_1=-0.52547874$, $u_2=0.75930824$; $u_1=0.20089479$, $u_2=0.29789739$; and $u_1=0.73063732$, $u_2=-0.51744858$. Consideration of the constraints $u_i\geq 0$ and $\sum_{i=1}^{3} u_i=1$ eliminates all solutions except the first and fourth. Consideration of the data eliminates the first solution. Therefore, the only satisfactory solution to (4D.93) is

$$u_1=.2008948, \quad u_2=.2978974, \quad u_3=.5012078. \tag{4D.95}$$

Substituting solution (4D.95) into (4D.94) yields for the asymptotic inverse covariance matrix $\underset{\sim}{S}^{-1}$ the elements

$$S^{11} = 1.4050^{5}, \quad S^{12} = 5.9904^{4}, \quad \text{and} \quad S^{22} = 1.1638^{5}, \tag{4D.96}$$

whence elements of the asymptotic covariance matrix $\underset{\sim}{S}$ are

$$S_{11} = 9.1184^{-6}, \quad S_{12} = -4.6934^{-6}, \quad \text{and} \quad S_{22} = 1.1008^{-5}. \tag{4D.97}$$

APPENDIX 4E

ERROR PROPAGATION

In this appendix, we study the error incurred when the iterative solution to an approximation

$$\dot{\underset{\sim}{p}} = G(\dot{\underset{\sim}{p}}) \tag{4E.1}$$

is considered as a solution to the function

$$\dot{\underset{\sim}{p}} = g(\dot{\underset{\sim}{p}}) \tag{4E.2}$$

being approximated. In particular, we consider the approximation (4E.1) where

$$G_i(\dot{\underset{\sim}{p}}) = (z_i+\nu_i+\sum_{D\ni i} z_D\dot{p}_i/\dot{p}_D)/m, \qquad 1\leq i\leq k, \tag{4E.3}$$

for the function (4E.2) where

$$\begin{aligned} g_i(\dot{\underset{\sim}{p}}) &= (z_i+\nu_i)/m+\sum_{D\ni i} z_D/m\{\dot{p}_i/\dot{p}_D+\underset{\sim}{\Sigma}[\partial^2(r_{1Q})/(\partial\underset{\sim}{p}\partial\underset{\sim}{p}')]_{\dot{\underset{\sim}{p}}}+\text{h.o.t.}\} \\ &= G_i(\dot{\underset{\sim}{p}}) + \varepsilon_i. \end{aligned} \tag{4E.4}$$

In (4E.4), "h.o.t." denotes higher order terms in the Taylor-series expansion of $\underset{\sim}{p}$ about the exact posterior mean $\tilde{\underset{\sim}{p}}$ [see Appendix 3B], where, however, evaluation of the partial derivatives is now made at $\dot{\underset{\sim}{p}}$, not $\tilde{\underset{\sim}{p}}$. The term (r_{1Q}) denotes the matrix of ratios $r_{1Q}=p_1/p_Q$.

Note that no element of the matrices of partial derivatives is a function of the sample size n. For example, elements of $[\partial^2(r_{1Q})/(\partial\underset{\sim}{p}\partial\underset{\sim}{p}')]_{\dot{\underset{\sim}{p}}}$ are given by, where 1 and j are elements of the set Q,

$$\partial^2 r_{1Q}/\partial p_1^{\,2} = -2(p_Q-p_1)/p_Q^{\,3},$$

$$\partial^2 r_{lQ}/(\partial p_l \partial p_j) = (2p_l - p_Q)/p_Q^3,$$

and

$$\partial^2 r_{lQ}/\partial p_j^2 = -2p_l/p_Q^3.$$

For $j \notin Q$, $\partial^2 r_{lQ}/(\partial p_q \partial p_j) = 0$ for any q.

From (4.13), elements of the posterior covariance matrix $\underset{\sim}{\Sigma}$ are of magnitude $O(n^{-1})$ and from Lemma 3B.2, elements of the higher order terms are of successively decreasing order of magnitude. Therefore, the error ε_i in (4E.4) is of order $O(n^{-1})$ for all i; i.e.,

$$\varepsilon_i = O(n^{-1}), \qquad 1 \leq i \leq k. \tag{4E.5}$$

We use the following lemma and proof derived from Theorem 3, page 92, and Theorem 2, page 111, of Isaacson and Keller (1966):

Lemma 4E.1: Suppose, for $1 \leq i \leq k$, that we have approximated $\dot{\tilde{p}}_i = g_i(\dot{\underset{\sim}{p}})$ by a function $G_i(\dot{\underset{\sim}{p}})$ in such a way that the error $\varepsilon_i(\dot{\underset{\sim}{p}})$ in $G_i(\dot{\underset{\sim}{p}})$ is bounded by some value $\delta > 0$. Suppose, further, that we use the iterative scheme given by

$$\dot{\underset{\sim}{p}}^{(s+1)} = G(\dot{\underset{\sim}{p}}^{(s)}) \tag{4E.6}$$

to calculate a root of $G(\dot{\underset{\sim}{p}})$. Note that (4E.6) can also be written as

$$\dot{\underset{\sim}{p}}^{(s+1)} = g(\dot{\underset{\sim}{p}}^{(s)}) + \underset{\sim}{\varepsilon}^{(s)} \tag{4E.7}$$

where $|\varepsilon_i^{(s)}| \leq \delta$ for all i.

From Appendix 3B, one root of $g_i(\dot{\underset{\sim}{p}})$ in (4E.4) is the exact posterior mean $\tilde{\underset{\sim}{p}}$, which we now study. Suppose that in all intervals $\|\dot{\underset{\sim}{p}} - \tilde{\underset{\sim}{p}}\|_\infty < \rho$, where $\|\dot{\underset{\sim}{p}} - \tilde{\underset{\sim}{p}}\|_\infty \equiv \max_{1 \leq i \leq k} |\dot{\tilde{p}}_i - \tilde{\tilde{p}}_i|$ and $\rho > 0$, $g(\dot{\underset{\sim}{p}})$ satisfies

$$\max_i \sum_{j=1}^{k} |\partial g_i(\dot{\underset{\sim}{p}})/\partial \dot{\tilde{p}}_j| \leq \lambda < 1. \tag{4E.8}$$

Let the initial iterative estimate $\overset{*}{\underset{\sim}{p}}^{(0)}$ be any point in the ρ_0 sphere $\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty \leq \rho_0$ for $0<\rho_0 \leq \rho-\delta/(1-\lambda)$. Then the iterates $\overset{*}{\underset{\sim}{p}}^{(s)}$ of (4E.7) lie in the interval $\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty \leq \rho$ and

$$\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty \leq \delta/(1-\lambda) + \lambda^s[\rho_0-\delta/(1-\lambda)] \tag{4E.9}$$

where $\lambda^s \to 0$ as $s\to\infty$.

proof (by induction):

By assumption, $\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}^{(0)}\|_\infty \leq \rho_0$. Therefore, $\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}^{(0)}\|_\infty \leq \rho_0+\delta/(1-\lambda) \leq \rho$. Assume that $\overset{*}{\underset{\sim}{p}}^{(l)}$ for $1\leq l\leq s-1$ are in $\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}\|_\infty \leq \rho$. Then,

$$\begin{aligned}\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}^{(s)}\|_\infty &\leq \|g(\tilde{\underset{\sim}{p}})-[g(\overset{*}{\underset{\sim}{p}}^{(s-1)})+\underset{\sim}{\varepsilon}^{(s-1)}]\|_\infty \\ &\leq \|[g(\tilde{\underset{\sim}{p}})-g(\overset{*}{\underset{\sim}{p}}^{(s-1)})]\|_\infty + \delta.\end{aligned} \tag{4E.10}$$

Now, for any two points $\tilde{\underset{\sim}{p}}$ and $\overset{*}{\underset{\sim}{p}}^{(s-1)}$ in $\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\| \leq \rho$, Taylor's theorem yields that

$$g_i(\overset{*}{\underset{\sim}{p}})-g_i(\tilde{\underset{\sim}{p}}) = \sum_{j=1}^{k} \partial g_i(\xi^{(i)})/\partial \overset{*}{p}_j \,(\overset{*}{p}_j-\tilde{p}_j), \quad \text{for } 1\leq i\leq k, \tag{4E.11}$$

where $\xi^{(i)}$ is a point on the open line segment joining $\overset{*}{\underset{\sim}{p}}$ and $\tilde{\underset{\sim}{p}}$. Thus, $\xi^{(i)}$ is in $\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty$ and

$$\begin{aligned}|g_i(\overset{*}{\underset{\sim}{p}})-g_i(\tilde{\underset{\sim}{p}})| &\leq \sum_{j=1}^{k} |\partial g_i(\xi^{(i)})/\partial \overset{*}{p}_j| \times |\overset{*}{p}_j-\tilde{p}_j| \\ &\leq \|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty \sum_{j=1}^{k} |\partial g_i(\xi^{(i)})/\partial \overset{*}{p}_j| \\ &\leq \lambda \,\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty.\end{aligned} \tag{4E.12}$$

Since the inequality holds for each i,

$$\|g(\overset{*}{\underset{\sim}{p}})-g(\tilde{\underset{\sim}{p}})\|_\infty \leq \lambda \ \|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty. \tag{4E.13}$$

Therefore, from (4E.10) and (4E.13),

$$\begin{aligned}
\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}^{(s)}\|_\infty &\leq \lambda \|\overset{*}{\underset{\sim}{p}}^{(s-1)}-\tilde{\underset{\sim}{p}}\|_\infty + \delta \\
&\leq \lambda^2 \|\overset{*}{\underset{\sim}{p}}^{(s-2)}-\tilde{\underset{\sim}{p}}\|_\infty + \lambda\delta + \delta \\
&\leq \lambda^3 \|\overset{*}{\underset{\sim}{p}}^{(s-3)}-\tilde{\underset{\sim}{p}}\|_\infty + \lambda^2\delta + \lambda\delta + \delta \\
&\ \ \vdots \\
&\leq \lambda^s \|\overset{*}{\underset{\sim}{p}}^{(0)}-\tilde{\underset{\sim}{p}}\|_\infty + \lambda^{s-1}\delta + \cdots + \lambda\delta + \delta \qquad (4E.14) \\
&\leq \lambda^s\rho_0 + \delta[(1-\lambda^s)/(1-\lambda)] \\
&\leq \lambda^s\rho_0 + \delta/(1-\lambda) - \lambda^s\delta/(1-\lambda) \\
&\leq \rho_0 + \delta/(1-\lambda) \\
&\leq \rho.
\end{aligned}$$

Therefore, all the iterates $\overset{*}{\underset{\sim}{p}}^{(1)}$ lie in $\|\overset{*}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty \leq \rho$ and the iteration process is defined. Finally, from the last inequality involving s,

$$\|\tilde{\underset{\sim}{p}}-\overset{*}{\underset{\sim}{p}}^{(s)}\|_\infty \leq \delta/(1-\lambda) + \lambda^s[\rho_0-\delta/(1-\lambda)]; \tag{4E.15}$$

i.e.,

$$|\tilde{p}_i-\overset{*}{p}_i^{(s)}| \leq \delta/(1-\lambda) + \lambda^s[\rho_0-\delta/(1-\lambda)] \tag{4E.16}$$

for all $1 \leq i \leq k$.

This lemma shows that the exact posterior mean $\tilde{\underset{\sim}{p}}$ satisfying (4E.4) can be approximated by the Taylor-series approximate posterior mean $\overset{*}{\underset{\sim}{p}}$ from (4E.3) to an accuracy determined essentially by the accuracy of the

errors $\delta > \varepsilon_i(\dot{\underset{\sim}{p}}) = g_i(\dot{\underset{\sim}{p}}) - G_i(\dot{\underset{\sim}{p}})$ in (4E.4). Thus, $\varepsilon_i(\dot{\underset{\sim}{p}})$ small for all i implies that $\dot{\tilde{p}}_i - \tilde{p}_i$ is small.

From (4E.5), $\varepsilon_i(\dot{\underset{\sim}{p}}) = O(n^{-1})$ for all i. Thus, $\delta = O(n^{-1})$. From (4E.4),

$$\begin{aligned}
\max_i \sum_{j=1}^{k} |\partial g_i(\dot{\underset{\sim}{p}})/\partial \dot{p}_j| &= \max_i \{ \sum_{D \ni i} z_D/m [\sum_{j=1}^{k} |\partial(\dot{p}_i/\dot{p}_D)/\partial \dot{p}_j| + O(n^{-1})]\} \\
&= \max_i \sum_{D \ni i} z_D/m [|(\dot{p}_D - \dot{p}_i)/\dot{p}_D^2| + \sum_{\substack{j \in D \\ j \neq i}} |-\dot{p}_i/\dot{p}_D^2|] + O(n^{-1}) \\
&= \max_i \sum_{D \ni i} z_D/m [(\dot{p}_D - \dot{p}_i)/\dot{p}_D^2 + \sum_{\substack{j \in D \\ j \neq i}} \dot{p}_i/\dot{p}_D^2] + O(n^{-1}) \\
&= \max_i \sum_{D \ni i} z_D/m \{ 1/\dot{p}_D + [\beta(D) - 2]\dot{p}_i/\dot{p}_D^2 \} + O(n^{-1})
\end{aligned} \tag{4E.17}$$

for $\beta(D)$ the number of elements in D.

In general, there is no guarantee that there exists a neighborhood of $\tilde{\underset{\sim}{p}}$ in which (4E.8) is satisfied everywhere within the neighborhood. If there is, we call the largest such neighborhood the epm (exact-posterior-mean) convergence region. [See the following Figure 4E.1 for an illustration of an epm convergence region.]

Note, however, that for the trinomial distribution $\beta(D) = 2$. Hence, the second term $[\beta(D) - 2]\dot{p}_i/\dot{p}_D^2$ in (4E.17) is zero and

$$\max_{1 \leq i \leq k} \sum_{j=1}^{k} |\partial g_i(\dot{\underset{\sim}{p}})/\partial \dot{p}_j| = \max_{1 \leq i \leq k} \sum_{D \ni i} (z_D/m)/\dot{p}_D + O(n^{-1}).$$

Further, recall from Sections 1.2, 2.2.3, and 4D.3 that $\underset{\sim}{z}$ can be considered as coming from related multinomial populations. For example, $\underset{\sim}{z} = (z_1, z_2, z_3, z_{12})$ can be considered as coming from a trinomial distribution with $v_1 = z_1$, $v_2 = z_2$, v_3 and a binomial distribution with $y_1 = z_{12}$ and y_3,

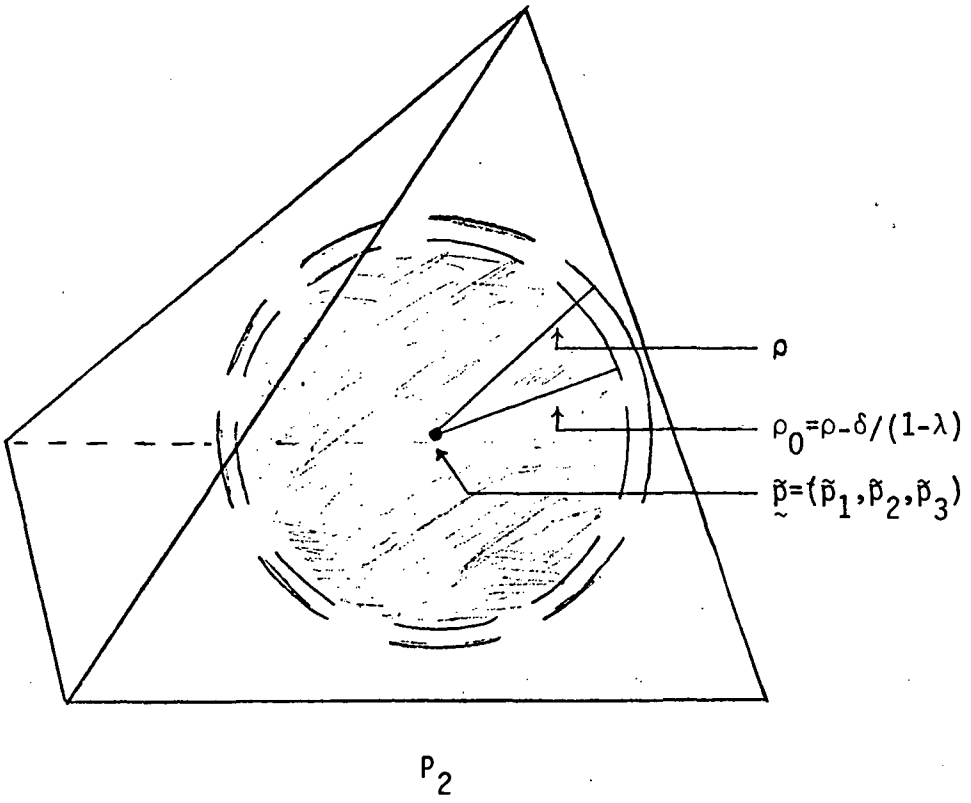


Figure 4E.1 GUARANTEED-CONVERGENCE REGION FOR THE EXACT POSTERIOR MEAN

where $v_3+y_3=z_3$. For $n_{12}=z_{12}+y_3$, then, $z_{12}/n = (z_{12}/n_{12})(n_{12}/n) = (n_{12}/n) \times\hat{p}_{12} = (n_{12}/n)\tilde{p}_{12}+O_p(n^{-1/2})$, where $\hat{p}_{12}$ is the maximum likelihood estimate. Therefore, for any incompletely specified data z_D,

$$z_D/m = (n_D/n)\tilde{p}_D + O_p(n^{-1/2}), \tag{4E.18}$$

so that, for the trinomial distribution, we can write (4E.17) as

$$\max_{1\leq i\leq k} \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial\dot{\tilde{p}}_j| = \max_i \sum_{D\ni i} (n_D/n)\tilde{p}_D/\dot{\tilde{p}}_D + O_p(n^{-1/2}). \tag{4E.19}$$

Because $\sum_{D\ni i} n_D<n$, $\sum_{D\ni i} n_D/n<1$. Therefore, for large enough sample size n, the bound $\lambda \doteq \max_i \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial\dot{\tilde{p}}_j|$ is less than 1 if $\dot{\tilde{p}}_D$ is close enough to $\tilde{p}_D$. Since for all values of $\tilde{p}$ (which never has zero components because the prior parameter ν never has zero components) there does exist a neighborhood such that $\tilde{p}_D/\dot{\tilde{p}}_D$ will be close to 1 for all values of $\dot{\tilde{p}}_D$ in this neighborhood, for the trinomial distribution there exists an epm convergence region. For higher dimensions, however, there need not exist an epm convergence region and we give an example of such a case in the main text, Section 4.3.2.

Observe that, anytime (4E.19) is satisfied, the term $\delta/(1-\lambda)$ in (4E.15) is $O(n^{-1})$. Since ρ_0 is a constant, $\delta/(1-\lambda)=O(n^{-1})$, $\lambda<1$, and, in particular, s can be assumed as large as desired, the term $\lambda^s[\rho_0-\delta/(1-\lambda)]$ in (4E.15) can be assumed to be zero. In particular, s can be assumed large enough that λ^s is small enough that this term is of magnitude no greater than $O(n^{-1})$.

Therefore, if there exists a neighborhood $\|\dot{\tilde{p}}-\tilde{p}\|_\infty<\rho$ around $\tilde{p}$ such that λ in (4E.8) is satisfied and, further, the initial iterative esti-

mate $\overset{\bullet}{\underset{\sim}{p}}^{(0)}$ is chosen in the neighborhood $\|\overset{\bullet}{\underset{\sim}{p}}-\tilde{\underset{\sim}{p}}\|_\infty < \rho_0 \leq \rho-\delta/(1-\lambda) \doteq \rho$, then the error in the Taylor-series approximate posterior mean $\overset{\bullet}{\underset{\sim}{p}}$ is $O(n^{-1})$, i.e.,

$$\overset{\bullet}{p}_i = \tilde{p}_i + O(n^{-1}). \qquad (4E.20)$$

Comments: Since $\delta=O(n^{-1})$, for large enough sample sizes, the ρ_0 neighborhood can be closely approximated by the ρ neighborhood. In turn, we can determine whether the iterates can be expected to be within the epm convergence region bounded by ρ, where condition (4E.8) must hold, by checking, first, whether the following inequality

$$\max_i \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial \dot{p}_j| \doteq \max_i \sum_{D \ni i} z_D/m\{1/\dot{p}_D^{(s)} + [\beta(D)-2]\dot{p}_i^{(s)}/[\dot{p}_D^{(s)}]^2\} < 1 \quad (4E.21)$$

holds for every iterate $\dot{p}_i^{(s)}$, $q \leq s \leq t$, for $t+1$ the number of iterations required for the convergence condition to be met and q the number of the first iteration that begins an unbroken succession of iterations satisfying (4E.21). If (4E.21) does not successively hold after some number of iterations, then different initial estimates can be tried and inequality (4E.21) reevaluated.

Second, if (4E.21) holds for sets of iterates converging to different values [i.e., to different roots of (4E.4)], more than one of which is in P_k, we must determine which root, if any, is in the epm convergence region. [See Section 4D.5 for two examples of multiple roots, one having three roots and the other having five roots, for the asymptotic posterior mean for incomplete trinomial data.] As discussed in the main text, Section 4.3.2, the global maximum within P_k is conjectured to be the root that is in the epm convergence region or at least closest to $\tilde{p}$. Hence, of those iteration sequences satisfying (4E.21) and converging to different roots in P_k, we choose that one for which the likelihood function

$$\dot{p}_1^{z_1+\nu_1-1} \dot{p}_2^{z_2+\nu_2-1} \cdots \dot{p}_{k+1}^{z_{k+1}+\nu_{k+1}-1} \prod_D \dot{p}_D^{z_D}$$

is a maximum.

Note that the conditions on the partial derivatives and initial iterative estimate are sufficient but not necessary. Finally, we give three examples that show that Lemma 4E.1 gives very conservative bounds on the error $\|\dot{\underset{\sim}{p}}^{(t)}-\underset{\sim}{\tilde{p}}\|_\infty$ and on the guaranteed-convergence neighborhood of $\underset{\sim}{\tilde{p}}$. For these examples we use the data $\underset{\sim}{z}=(2,5,6,4,2,0)$ given in Section 2.2.3 where we calculated the exact posterior mean as $\underset{\sim}{\tilde{p}}=(.2412,.3849,.3739)$.

For the first example, consider the neighborhood $\|\dot{\underset{\sim}{p}}-\underset{\sim}{\tilde{p}}\|_\infty<\rho=.11$ of $\underset{\sim}{\tilde{p}}$. For all probabilities $\dot{\underset{\sim}{p}}$ in this ρ neighborhood, $\max_i \sum_{j\ni i} |\partial g_i(\dot{\underset{\sim}{p}})/\partial\dot{p}_j|=(4/22)/\dot{p}_{12}+(2/22)/\dot{p}_{13} < .56 < 1$ and a bound on the error made by approximating the exact posterior mean by a Taylor-series expansion is $\delta=0.035$. Thus, $\delta/(1-\lambda)=0.080$. Suppose that we choose an initial iterative estimate $\dot{\underset{\sim}{p}}^{(0)}$ in the region bounded by $\rho_0\leq\rho-\delta/(1-\lambda)=.11-.08=.05$. Then the iteration process is guaranteed to converge to within $\delta/(1-\lambda)=.08$ of the exact posterior mean. However, for any initial iterative probability (including that one whose three components each differ from the three corresponding components of $\underset{\sim}{\tilde{p}}$ by .11) chosen within this ρ_0 neighborhood, the maximum difference between the converged iterative estimate and the exact posterior mean was 0.003, more than 25 times smaller than the $\delta/(1-\lambda)=.080$ error bound given by Theorem 4E.1.

Now consider as an initial iterative estimate for $\underset{\sim}{\tilde{p}}=(.2412,.3849,.3739)$ the value $\dot{\underset{\sim}{p}}^{(0)}=(.05,.10,.85)$. For this value,

$$\sum_{j=1}^{2} |\partial g_i(\dot{\underset{\sim}{p}})/\partial\dot{p}_j| = (4/22)/.15 + (2/22)/.90 = 1.21 + .10 + 1.31 > 1.$$

Hence, conditions of Lemma 4E.1 are not satisfied. However, use of this

initial iterative estimate gives successive iterates, as shown in the following Table 4E.1, that do converge to within a small error of $\tilde{p}$.

The initial iterative estimate $\dot{\tilde{p}}^{(0)}$ failed condition (4E.8) because $\dot{p}_{12}$=.15 was smaller than $z_{12}/m \doteq (n_{12}/m)(\tilde{p}_{12}/\dot{\tilde{p}}_{12})$. Note from this example that small values of $\dot{\tilde{p}}_D$ will be particularly troublesome in keeping the term $(z_D/m)/\dot{\tilde{p}}_D \doteq (n_D/m)\tilde{p}_D/\dot{\tilde{p}}_D$ less than 1.

In this example $\|\dot{\tilde{p}}^{(0)}-\tilde{p}\|_\infty$=max(.19,.28,.48)=.48. Thus, the largest value of ρ for a guaranteed-convergence neighborhood of $\tilde{p}$ must be smaller than .48. In the next example we choose as an initial iterative estimate $\dot{\tilde{p}}^{(0)}$ a probability $\dot{\tilde{p}}^{(0)}$=(.90,.07,.03) that is even further away from $\tilde{p}$. For this estimate, $\|\dot{\tilde{p}}^{(0)}-\tilde{p}\|_\infty$=max(.66,.31,.34)=.66. [See also Figure 4E.2.] Since .66>.48 of the last example, this initial iterative estimate cannot be in a guaranteed-convergence neighborhood of $\tilde{p}$. Yet, for this estimate,

$$\max_i \sum_{j=1}^{2} |\partial g_i(\dot{\tilde{p}})/\partial \dot{p}_j| = (4/22)/.97+(2/22).93 = .29 < 1.$$

Futher, as shown in Table 4E.1, the sequence of iterates arising from this initial iterative estimate also converges to within a small error of the exact posterior mean.

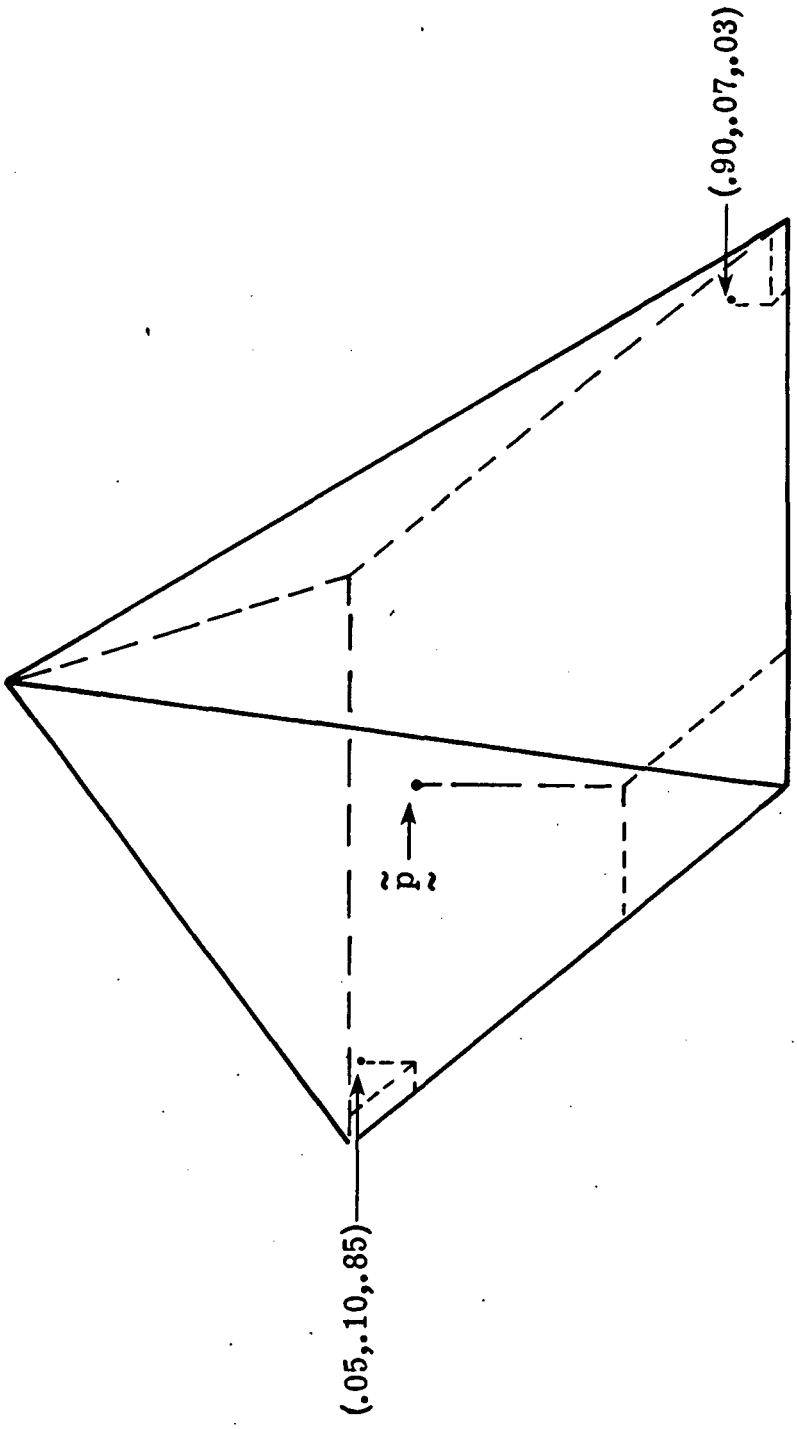


FIGURE 4E.2

GOOD INITIAL ESTIMATES OUTSIDE GUARANTEED-CONVERGENCE SPHERE

TABLE 4E.1

CONVERGENCE EXAMPLES FOR OUTSIDE[1] INITIAL ESTIMATES

	SECOND EXAMPLE		THIRD EXAMPLE	
β^2	$(\dot{\tilde{p}}_1)_\beta$	$(\dot{\tilde{p}}_2)_\beta$	$(\dot{\tilde{p}}_1)_\beta$	$(\dot{\tilde{p}}_2)_\beta$
0	.0500	.1000	.9000	.0700
1	.2020	.3939	.3930	.2858
2	.2283	.3929	.2917	.3493
3	.2374	.3877	.2598	.3718
4	.2391	.3870	.2488	.3798
5	.2413	.3851	.2448	.3826
6	.2421	.3845	.2433	.3836
7	.2420	.3846	.2428	.3839

[1] Initial iterative estimates chosen outside the guaranteed-convergence sphere of Lemma 4E.1 for the exact posterior mean $\tilde{p}$=(.2412,.3849,.3739).

[2] Iteration number

CHAPTER 5

SMALL-SAMPLE STUDIES OF APPROXIMATIONS FOR POSTERIOR MOMENTS AND OF ESTIMATORS FOR MINIMIZING QUADRATIC LOSS

5.1 Introduction:

In the last chapter, we showed that for large sample sizes the Taylor-series approximations should be very close to corresponding exact posterior moments. We now consider how well these asymptotic properties hold in small- and medium-size samples. We also compare the Taylor-series approximations with the posterior mode and maximum-likelihood estimate to determine which best approximates the exact posterior mean for these smaller sample sizes. Although all three approximations will be very close for very large sample sizes, we expect differences in the most commonly encountered sample sizes.

We then turn to our main interest and report which of these three estimators best minimizes expected quadratic loss (risk) for specified values of the Dirichlet probabilities. Note that we do not include the exact posterior mean in the risk study. Results from the approximation part of this small-sample study showed that there was no difference between the Taylor-series approximation and the exact posterior mean that would alter conclusions from using the Taylor-series approximation for the exact posterior mean. Since the exact posterior mean becomes increasingly expensive as the sample size and/or percentage of incomplete data increases, we used the Taylor-series approximation for the exact posterior mean. Therefore, in our mnemonics, we refer to the Taylor-series approximation as APM (approximate posterior mean).

In Figure 5.0 we give expressions for the three approximations for the exact posterior mean and estimators for minimizing quadratic loss. These equations were presented in Chapter 1 or derived in Chapter 3. The mnemonics APM, PMD (posterior mode), and MLE (maximum likelihood estimate) in parenthesis are used throughout these next three chapters. They are especially useful in presenting results in the next two chapters. For the risk study, we attach suffixes R0, R1, and R2 to these mnemonics to denote the three robustness studies for use of the original, uniform, and perturbed priors, respectively, in the Bayesian estimators.

In summary, we are interested in four main questions: (1) how well the Taylor-series expansions approximate the exact posterior mean and covariance matrices; (2) which of three estimators (Taylor-series, posterior mode, and maximum likelihood estimate) best approximates the exact posterior mean; (3) which of these three estimators best minimizes risk; and (4) how robust results from (3) are to use of the wrong prior in the Bayesian estimators. Because we were unable to solve these problems theoretically, we used Monte-Carlo simulation studies. Hence, results will be only indicative, not conclusive.

In this chapter, we discuss designs and computational procedures for two Monte-Carlo studies. In the next two chapters, we discuss results from these studies.

FIGURE 5.0

APPROXIMATIONS FOR EXACT POSTERIOR MEAN

AND ESTIMATORS FOR QUADRATIC LOSS

Taylor-Series (APM)

$$\mathring{\tilde{p}}_i = [z_{\{i\}}+\nu_i+\sum_{D\ni i}(\mathring{\tilde{p}}_i/\mathring{\tilde{p}}_D)\, z_D]/[n+\sum_{j=1}^{k+1}\nu_j]$$

Posterior Mode (PMD)

$$\hat{\tilde{p}}_i = [z_{\{i\}}+\nu_i-1+\sum_{D\ni i}(\hat{\tilde{p}}_i/\hat{\tilde{p}}_D)\, z_D]/[n+\sum_{j=1}^{k+1}\nu_j-(k+1)]$$

Maximum Likelihood (MLE)

$$\hat{p}_i = [z_{\{i\}}+\sum_{D\ni i}(\hat{p}_i/\hat{p}_D)\, z_D]/n$$

Note that $k=2$ for trinomial simulation study. Also note that braces in $z_{\{i\}}$ are henceforth dropped.

5.2 Special Notation and Mnemonics:

Notation:

p_i	i^{th} element of Dirichlet generator probability vector
$\tilde{p}_i$	i^{th} element of exact posterior mean
$\dot{p}_i$	i^{th} element of Taylor-series (T.S.) approximate posterior mean (APM)
$\hat{p}_i$	i^{th} element of maximum likelihood estimate (MLE)
$\hat{\tilde{p}}_i$	i^{th} element of posterior mode (PMD)
$\ddot{p}_i$	i^{th} element of complete-data maximum likelihood estimate (used mainly for variance reduction in estimating mean squared error)
$\dot{p}_i$	i^{th} element of dummy estimator $\dot{p}$, which is used when describing properties or formula that pertain to more than one of the above estimators
$\dot{e}_i$	$\dot{p}_i - p_i$ for $\dot{p}_i$ any of above estimators $\dot{\tilde{p}}_i$, $\hat{p}_i$, $\hat{\tilde{p}}_i$, and $\ddot{p}_i$
$\dot{\varepsilon}_i$	$\dot{p}_i - \tilde{p}_i$ for $\dot{p}_i$ any of above estimators $\dot{\tilde{p}}_i$, $\hat{p}_i$, and $\hat{\tilde{p}}_i$

Note that we are using $\underset{\sim}{p}$ to denote both any value of the simplex $P_2 = \{(p_1,p_2,p_3): 0 \leq p_1,p_2,p_3 \leq 1;\ p_1+p_2+p_3=1\}$ and a particular value of P_2. The context in which $\underset{\sim}{p}$ is used should make clear the particular meaning. Further, note that both $\underset{\sim}{p}$ and $\tilde{\underset{\sim}{p}}$ are Dirichlet probabilities. The $\underset{\sim}{p}$ either is set to the expected value of the Dirichlet distribution of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}$ (note Design 1 in following Section 5.4) or is generated from this distribution (Design 2). In either case, we shall refer to $\underset{\sim}{p}$ as the generator. The $\tilde{\underset{\sim}{p}}$ refers to the posterior mean of the Dirichlet

distribution of $\underset{\sim}{p}$ given the incomplete data $\underset{\sim}{z}$. Thus, for Design 1, $\underset{\sim}{p}$ is the prior Dirichlet mean and $\tilde{\underset{\sim}{p}}$, the posterior Dirichlet mean.

Mnemonics: Note that the following mnemonics might appear in lower-case, as well as capital, letters:

APC	Taylor-series approximate posterior covariance
APM	Taylor-series approximate posterior mean
EPC	exact posterior covariance
EPM	exact posterior mean
MLE	maximum likelihood estimate
MSE	mean squared error
PID	percentage of incomplete data
PMD	posterior mode
SS	sample size

5.3 Criteria of Goodness:

To determine how good an estimator was either for estimating an exact posterior moment or for minimizing quadratic loss, we used several criteria. The main criterion for judging the accuracy of approximations for the exact posterior moments was percent relative difference. To judge among the estimators for estimating the exact posterior mean, we also used mean squared error $E[(\tilde{\underset{\sim}{p}}-\dot{\underset{\sim}{p}})'(\tilde{\underset{\sim}{p}}-\dot{\underset{\sim}{p}})]$. Of course, for judging which estimator best minimized quadratic loss, the criterion was the mean squared error $E[(\underset{\sim}{p}-\dot{\underset{\sim}{p}})'(\underset{\sim}{p}-\dot{\underset{\sim}{p}})]$. Estimates of mean squared error (mse) are discussed in Section 5.9.

Additional measures of goodness were also considered in Chapter 6 where we studied the estimators in detail. For example, among additional calculations were the frequency distributions of the number of iterations, deviations, and percentage relative difference. Criterion of goodness are included in the listing of tables in Chapters 6 and 7.

5.4 Computer:

Computers used for the simulation were a CDC (Control Data Corporation) 6600 and Cyber 175 with 60-bit words. Single-precision calculations were accurate to about 14.5 significant figures; double-precision calculations, to 29. The programing language was Fortran Extended, Version 4.6. To minimize execution cost, recommendations from the NASA, Langley Research Center "Computer Programing Manual", (1975,v1,sect.8) were incorporated.

Main incorporations were the passing of parameters among programs through COMMON rather than calling sequences and a reduction in a number of otherwise large DO-LOOP indices. Owing to the latter, program statements and number of variables increased. Number of dimensions on a variable decreased. Among other inclusions were use of "IF (A-B) 10,20,20" instead of "IF (A.GE.B) 20,10", collapsed dimensioning for array initializations, and special procedures for arithmetic operations.

Unless otherwise noted, all programs were written by the author. A listing of most of these programs is given in Credeur (1978). An index precedes the listing.

5.5 Factors in Experiments:

In investigating the four issues outlined in the Introduction, we were interested in the effects of variation in prior parameter $\underset{\sim}{\nu}$, the Dirichlet probabilities arising from the distribution of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}$, sample size (SS), and percentage of incomplete data (PID).

Owing to cost constraints, we limited the number of these variations. For percentage of incomplete data (PID) we chose 15 and 40. We already knew from Chapter 3 that for 0% incomplete data, the Taylor-series approximations (APM and APC) exactly equaled the posterior mean and posterior covariance, respectively, whereas the posterior-mode (PMD) and maximum-likelihood-estimate (MLE) approximations did not. Thus, for investigating the first two introductory questions concerning estimators for the exact posterior mean and covariance (EPM and EPC) matrices, we essentially had PID for values 0, 15, and 40.

For sample size (SS) we chose 25 and 50. For these values and ranges of PID we were able to calculate the exact posterior mean and covariance matrices. As noted in Chapter 2, for sample sizes much larger than 50, calculations for the exact values would be expensive, especially for those cases in which PID=40.

To set values of the prior parameter $\underset{\sim}{\nu}$, we first considered values we wanted for the Dirichlet probabilities arising from the distribution of $\underset{\sim}{p}$ given the prior. We wanted roughly to cover the range of probabilities from (0,0,1) to (1/3,1/3,1/3). We picked four values (.01,.01,.98), (.10,.10,.80), (.20,.30,.50), and (1/3,1/3,1/3) as focal points to be

investigated. Now, usually one has a prior because one has a prior sample. If the size of the prior sample is small relative to the size of the current sample, then the prior has little effect on the estimators. If the prior-sample size is relatively large, the current data has little effect. Therefore, because we chose current sample sizes of 25 and 50, we set the size of the prior data at 10, two-fifths and one-fifth the current information, respectively. Thus, values of the prior parameter $\underset{\sim}{\nu}$ were chosen as 10 times the prior mean we wanted.

That is, since $E(p_i|\underset{\sim}{\nu})=\nu_i/\sum_{j=1}^{3}\nu_j$ and $\sum_{j=1}^{3}\nu_j=10$, then

$$\nu_i = 10 \times E(p_i|\underset{\sim}{\nu}). \qquad (5.1)$$

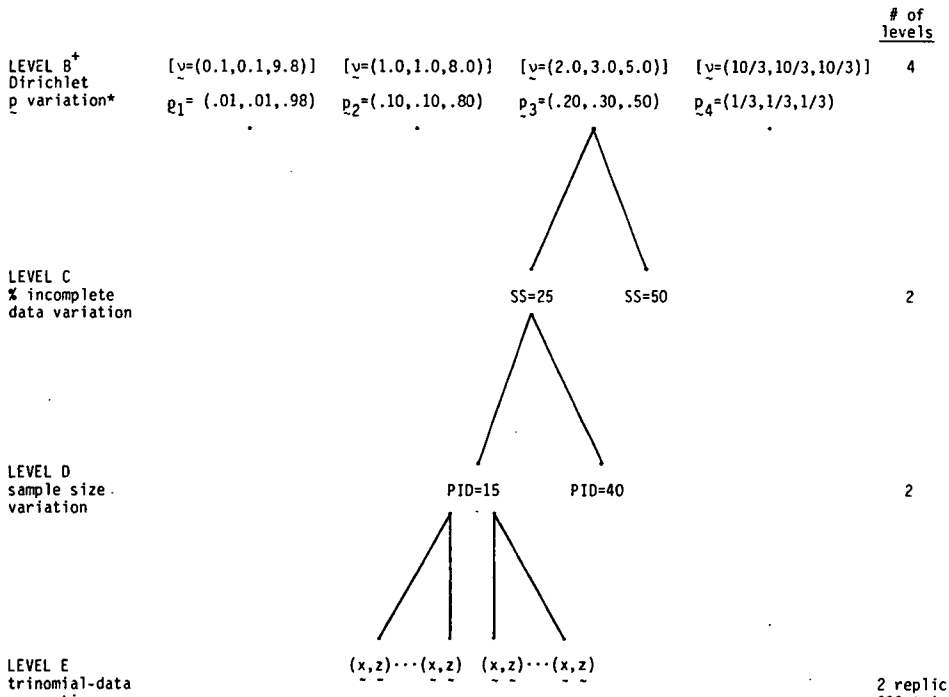
Setting $E(\underset{\sim}{p}|\underset{\sim}{\nu})$ to the four focal points gave values of $\underset{\sim}{\nu}$ as (.1,.1,9.8), (1,1,8), (2,3,5), and (10/3,10/3,10/3).

The simulation study was done in two stages, as follows. In the first stage, which we called Design 1, we fixed the value of the Dirichlet probability at the expected value of the distribution of $\underset{\sim}{p}$ given each one of the four prior parameters $\underset{\sim}{\nu}$. In the second, Design 2, we generated 10 values of the Dirichlet probability from each of the fixed values of $\underset{\sim}{\nu}$. Designs 1 and 2 are illustrated in Figures 5.1 and 5.2, respectively. A summary design is given as Figure 5.3.

Results from Design 1 allowed at least some of the four Introductory questions, especially those concerning the exact-posterior-moments comparisons to be satisfactorily answered. Because cost was less, more details were studied. The second design, Design 2, allowed us to determine how Design 1 results were affected by our choosing a special probability, the expected value of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}$. As we moved away from the

FIGURE 5.1

DESIGN 1



+level A is not present in this design (see Design 2 and Summary Design)

*p is expected value of Dirichlet probability distribution given ν

This design yields 6400 [=4×2×2×200×2repl] data sets and requires generation of 240,000 [=4×2×(25+50)×200×2repl] uniform random numbers.

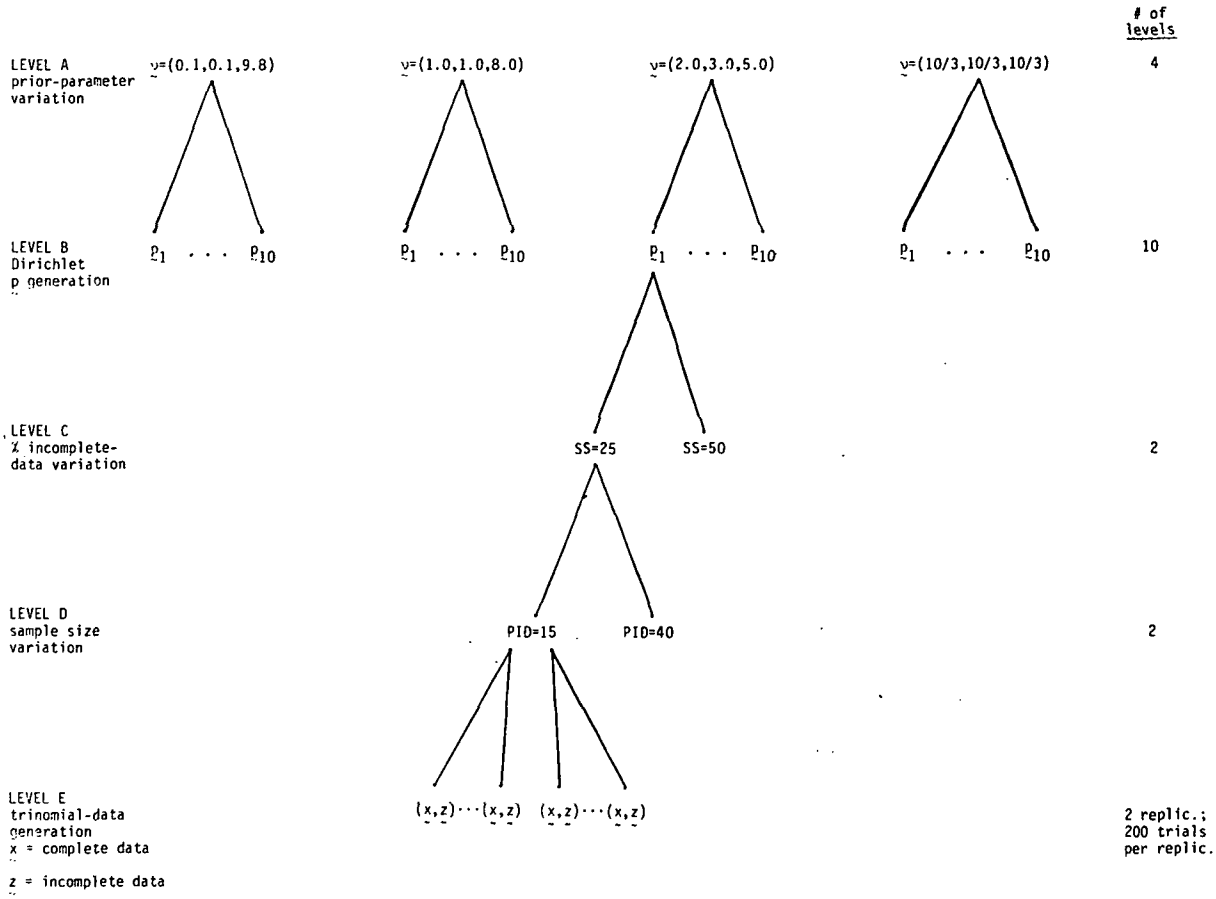
This design constitutes sets of full factorials:

a. for epm comparisons: $4\times2^2\times3$ with 2 replications per cell (last factor level 3 refers to estimators apm, pmd, and mle)

b. for quadratic-loss comparisons: $4\times2^2\times3$ with two replications per cell (last factor level 3 refers to estimators apm, pmd, and mle)

FIGURE 5.2

DESIGN 2



This design yields 64,000 [=4×10×2×2×200×2repl] data sets and requires generation of 2,400,120 random numbers [ie.; 120=4×3*×10 gamma random variables for 40 3-dimensional Dirichlet random variables + 2,400,000 = 4×10×2×(25+50)×200×2repl. uniform random numbers].

*each Dirichlet $\underset{\sim}{p}$ requires generation of 3 gamma random variables

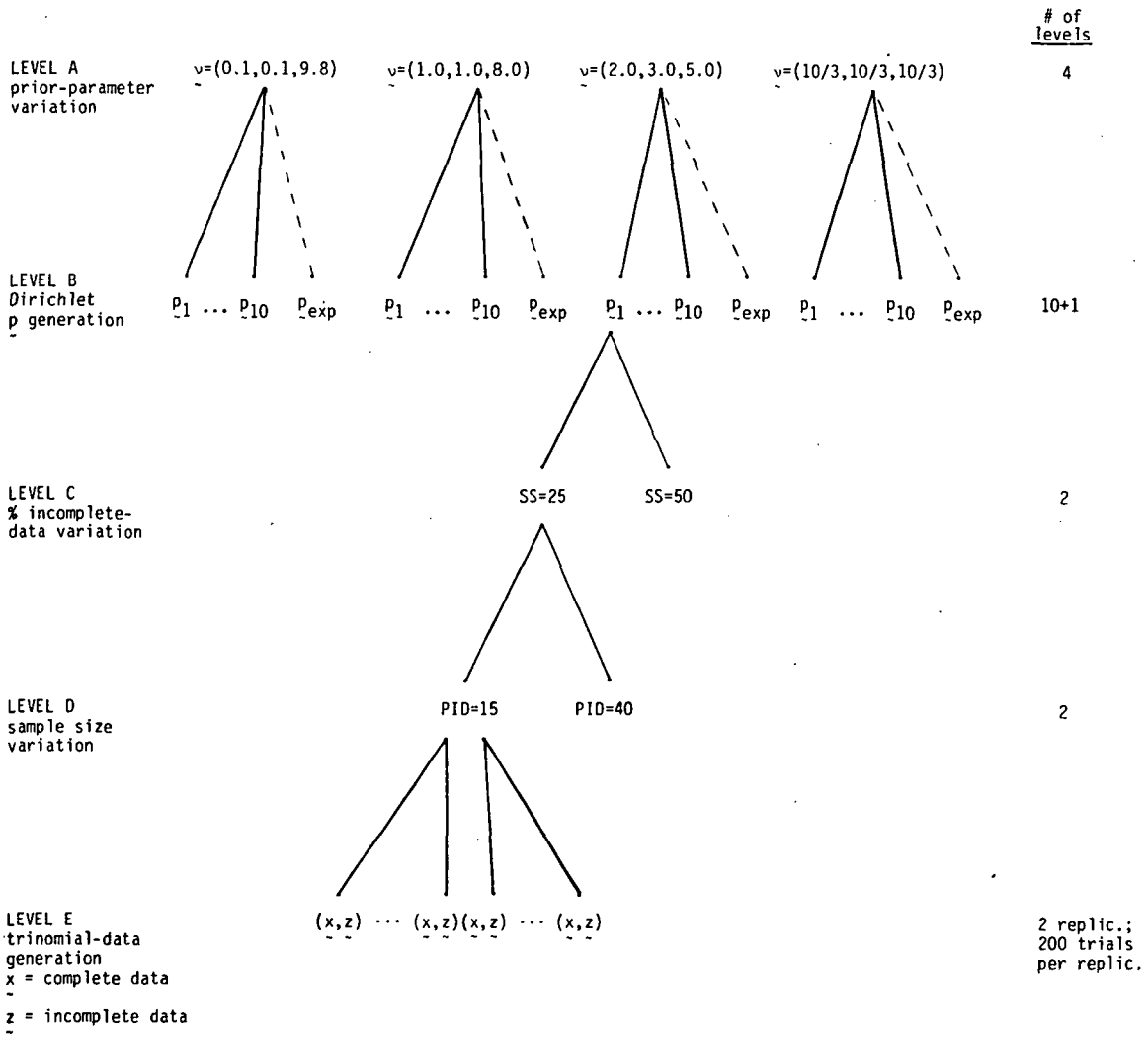
This design constitutes sets of nested factorials:

a. for epm comparisons: $4\text{-}10\times2^2\times3$ with two replications per cell

b. for quadratic loss: $4\text{-}10\times2^2\times3$ with two replications per cell

FIGURE 5.3

SUMMARY DESIGN



Total designs yield 70,400 [=4×11×2×2×200×2replic] data sets and require generation of 2,640,120 random numbers [ie.; 120=4×3*×10 gamma random variables for 40 3-dimensional Dirichlets +2,640,000 = 4×11×2×(25+50)×200×2replic uniform random numbers].

*each Dirichlet p requires generation of 3 gamma random variables

expected value in Design 2, using probabilities randomly generated from fixed $\underset{\sim}{\nu}$, how did Design 1 results change?

To measure the variation in the probabilities $\underset{\sim}{p}$ associated with a prior $\underset{\sim}{\nu}$, define a centrality norm

$$C(\underset{\sim}{p}) = \sum_{i=1}^{2} \sum_{j>i}^{3} (p_i - p_j)^2. \qquad (5.2)$$

Values of $C(\underset{\sim}{p})$ for Design 1 are given in the following Table 5.1. Note from Table 5.1 that as $\underset{\sim}{p}$ moves from a corner of P_2 toward its center, $C(\underset{\sim}{p})$ decreases from 2.00 to 0. Centrality measures for generated Dirichlet probabilities in Design 2 are given in Table 7.1 in Chapter 7.

TABLE 5.1

CENTRALITY MEASURES FOR DESIGN 1

$\underset{\sim}{\nu}$	$E(\underset{\sim}{p}\|\underset{\sim}{\nu})$	$C(\underset{\sim}{p})$
(0.1,0.1,9.8)	(.01,.01,.98)	1.88
(1.0,1.0,8.0)	(.10,.10,.80)	.98
(2.0,3.0,5.0)	(.20,.30,.50)	.14
(10/3,10/3,10/3)	(1/3,1/3,1/3)	.00

Factors SS and PID were quantitative; we considered $\underset{\sim}{\nu}$ and $\underset{\sim}{p}$ to be qualitative. In Design 1 all factors were fixed. In Design 2, $\underset{\sim}{p}$ was random and remaining factors were fixed.

Once we fixed the factor levels, we generated the trinomial data. In the next section, we discuss how we chose the number of trinomial simulations and, in Section 5.7, how we generated the data. To allow a

control variate and thus a better mean-squared-error (mse) estimate for the risk study, we generated complete as well as incomplete data.

For the exploratory robustness study, the two priors used besides the original $\underset{\sim}{\nu}$ prior were the uniform prior and a perturbed prior. Values of both are given in the following Table 5.2. The uniform prior is frequently used when one is uncertain of previous information. It gives equal weight to all three trinomial categories. The perturbed prior not only differs in magnitude from the correct prior $\underset{\sim}{\nu}$ but does so in a skewed manner. The change to the first component is

TABLE 5.2

PRIORS FOR ROBUSTNESS STUDY

Robustness Set	Type	Value
R0	original	$\underset{\sim}{\nu}$
R1	uniform	(1,1,1)
R2	perturbed	$10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$

approximately twice that to the second component and two-thirds that to the third component. The first two components increase; the third decreases.

5.6 Determination of Number of Simulation Trials and Replications:

Because mean squared error was the major overall "goodness" measure, the main criterion for choosing the number of trinomial simulation trials was that the standard errors of the average estimated mean squared errors be small relative to the difference between the mean squared errors. For this purpose, 200 trials was enough; for just comparisons among approximations for the exact posterior mean, fewer trials would have been needed.

As noted in Section 5.3, we were also interested in the deviations of the estimators from the exact posterior mean (the "EPM deviations"). One deviation, or error, measure was the average. However, the number of simulation trials needed to make the standard errors of the average deviations small relative to the difference between the average deviations was prohibitively expensive. Results of Design 1 gave that the average APM deviation was a couple orders of magnitude smaller than the average PMD and MLE deviations. Hence, the difference between it and either of the average PMD or MLE errors approximately equaled the PMD or MLE deviation, respectively. However, even for a number of trials as large as 200, the standard error of the average deviation roughly equaled the respective average deviation. (All EPM deviations averaged zero to varying number of decimal places.) Therefore, the APM-PMD≐PMD and APM-MLE≐MLE differences were not always larger than the standard errors of the PMD and MLE average deviations. (They were, however, much larger than the standard errors for the average APM deviation.)

To estimate the experimental error in estimating averages, including the mean squared errors, we repeated each set of 200 trinomial simulation trials once. [Recall Figures 5.1 - 5.3.] Cost considerations precluded more than two replications. Although each of the 200 trinomial simulations can be called a replication, for differentiation, we reserve this term for these two repetitions. The two replications also provided another check that 200 trinomial simulations were enough. There was little difference between results for each of the two replications.

To determine the number of simulation trials to use in generating Dirichlet probabilities in Design 2, we were guided mainly by cost constraints. We took only 10 trials. Results of Chapter 7 show how surprisingly good 10 trials were in terms of theoretical expectations.

5.7 Data Generation:

For Design 1 we must generate complete and incomplete trinomial data. For Design 2 we must also generate Dirichlet probabilities.

5.7.1 Uniform Random-Number Generator:

As do most other generator algorithms, algorithms to generate trinomial and Dirichlet data depend on a uniform random-number generator. For this generator, we used the multiplicative congruential generator

$$x_i = 43490275647445\ x_{i-1} \bmod(2^{48}) \qquad (5.3)$$

from Ahrens and Dieter (1974,p223). Uniformly distributed variables u_i were then calculated by

$$u_i = x_i/2^{48}. \qquad (5.4)$$

The multiplier 43490275647445 is congruent 5 mod(8); therefore, from Knuth (1969,p18,93), the generator (5.3) has maximum period of 2^{46} and we can apply the Spectral test of Coveyou and Macpherson (1967). The Spectral test is currently the most powerful test of the randomness of a random-number generator. By using a computer program written by Golder (1976,p173) with corrections by Hoaglin and King (1978), we calculated the Spectral Numbers c_i, $2 \leq i \leq 5$, as

$$c_2=2.839, \quad c_3=2.095, \quad c_4=1.819, \quad c_5=0.987. \qquad (5.5)$$

Since c_2, c_3, and c_4 all exceed 1 and c_5 is almost 1, the generator is very good in terms of the Spectral test, a theoretical test. Therefore, it is most likely good in terms of any empirical tests.

As an empirical check on the generator, however, we ran a number of 95% Confidence-Interval tests on the sample means and standard deviations, chi-square tests, serial-correlation tests, Kolmogorov-Smirnov tests on the cumulative frequency, and did plots on the density and the cumulative distribution. The generator did well on all.

5.7.2 Dirichlet Random-Number Generator:

To generate a Dirichlet random vector, we used the following theorem from Wilks (1963,p179):

"If $x_1,\ldots,x_{k+1}$ are independent random variables having gamma distributions $G(\nu_1),\ldots,G(\nu_{k+1})$, then for

$$y_i = x_i/(x_1+\ldots+x_{k+1}), \quad 1 \leq i \leq k, \tag{5.6}$$

$(y_1,\ldots,y_k)$ has the k-variate Dirichlet distribution $D(\nu_1,\ldots,\nu_k;\nu_{k+1})$."

Therefore, to obtain one random vector $\underset{\sim}{p}$, from a Dirichlet distribution with k=2, we must generate three independent gamma random variables. To do so, we used algorithm GT from Ahrens and Dieter (1974,p229).

We checked the Ahrens-Dieter GT algorithm by doing 95% confidence limits on the sample means and standard deviations, plots on the density and cumulative frequency, and Kolmogorov-Smirnov tests on the cumulative distribution. We then performed these same tests on the Dirichlet probabilities $\underset{\sim}{p}$ calculated from these gammas.

Other than the standard deviations, the generators performed well. As shown in the following Table 5.3, for the gamma random variables, the

TABLE 5.3

PERCENT REJECTIONS IN 95% NORMAL[1] CONFIDENCE INTERVALS FOR GAMMA AND MARGINAL DIRICHLET RANDOM VARIABLES

(300 TRIALS, 200 OBSERVATIONS PER TRIAL, EXPECTED PERCENTAGE IS 5)

PRIOR PARAMETER	SEED	GENERATOR	M1	M2	M3	SD1	SD2	SD3
(0.1,3.5,6.4)	21153	Gamma	5%[2]	3%	4%	77%	11%[4]	9%
		Dirichlet	5[3]	4	4	69	3	3
(0.1,0.1,9.8)	21197	Gamma	5	4	4	71	70	9
		Dirichlet	5	5	5	68	69	45
(0.1,0.1,9.8)	21153	Gamma	4	6	5	73	72	9
		Dirichlet	3	5	5	68	68	56
(0.5,0.5,9.0)	21143	Gamma	4	4	4	44	44	10
		Dirichlet	5	5	5	34	31	17
(1.0,1.0,8.0)	31153	Gamma	2	5	4	26	35	13
		Dirichlet	3	4	4	19	20	9
(10/3,10/3,10/3)	21153	Gamma	5	5	4	14	16	17
		Dirichlet	2	5	3	3	4	5
(2.5,3.0,4.5)	22213	Gamma	3	4	3	20	17	15
		Dirichlet	4	7	6	3	3	3
	21113	Gamma	6	6	2	17	19	13
		Dirichlet	5	4	3	6	6	5
	21111	Gamma	5	6	4	18	21	10
		Dirichlet	5	5	5	6	5	2
	21313	Gamma	3	6	3	14	16	13
		Dirichlet	3	2	2	3	4	1
	21153	Gamma	2	4	4	15	13	11
		Dirichlet	5	3	5	4	2	3

[1] Normal approximation is used for the confidence intervals.

[2] In 300 trials (sets of generations), 200 observations per trial, from gamma(0.1), the sample mean M1 (calculated over 200 observations) fell outside the 95% normal confidence interval 5% of the time (approximately 15 of the 300 trials).

[3] In 300 trials, 200 observations per trial, from beta(0.1,3.5+6.4)=beta(0.1,9.9), the sample mean M1 fell outside the 95% normal confidence intervals 5% of the time.

[4] In 300 trials, 200 observations per trial, from gamma(3.5), the sample standard deviation (calculated over 200 observations) fell outside the 95% normal confidence intervals 11% of the time.

standard deviations routinely exceeded the 95% confidence limits more than 5% of the time and became increasingly worse as we moved from 10/3 in the center of the 2-dimensional simplex P_2 to 0.1 or 9.8 at a corner. The same trend was observed for standard deviations of the marginal Dirichlets except that the percentage of rejections was much smaller and the standard deviations were good for points away from the boundary. This behavior may be due either to (1) a poor fit of the generated data to the theoretical curve or (2) to the normal approximation, which we used, for the confidence intervals for the standard deviations being poor for the sample sizes we used.

For three reasons, we accepted the latter explanation. The first reason is that, as noted, for probabilities away from the boundary, marginals from those Dirichlet random vectors generated from these gammas did have standard deviations falling in the 95% confidence intervals all but 5% of the time. [See results in Table 5.3 for non-boundary probabilities corresponding to prior parameters 3.5, 6.4, 10/3, 2.5, 3.0, and 4.5.] The second reason is that the gamma and the Dirichlet marginals performed well on the other tests (and the gamma generator had been studied by Ahrens and Dieter). The third reason is that the sample kurtosis for those random variables near a boundary was very high. Therefore, from Snedecor and Cochran (1968,p.89), the variance of the sample variance was much larger than it would have been had the population been normal. One calculation gave that the variance of the sample variance over 300 trials, 200 observations per trial, for a gamma generation of 0.1, was about 16.5 times as large as it would be in a normal population. Hence, we could not expect the normal approximation

for the 95% confidence limits for the standard deviations to be good in these cases.

Therefore, we accept reason (2) [that the normal approximation for the 95% confidence intervals for the standard deviations was poor] for so many standard deviations falling outside the 95% confidence intervals, especially for boundary probabilities corresponding to prior parameters 0.1 and 9.8. We did not calculate the exact standard deviations (and then test them). However, since the remaining tests (and Ahrens and Dieter's work for the gamma) showed that the gamma and, more important, the resulting Dirichlet variables were well generated, we considered the Dirichlet random-number generator to be good.

5.7.3 Trinomial Random-Number Generator:

Given some value of $\underset{\sim}{p}$ and some percentage PID of incomplete data, we next generated the trinomial complete data $\underset{\sim}{x}=(x_1,x_2,x_3)$ and incomplete data $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12},z_{13},z_{23})$.

We first recalled that PID/100 is simply the probability that an observation was incompletely classified. Second, given that an observation was incomplete, the probability that it was unclassified between C_1 and C_2 (i.e., the observation fell in C_{12}) was $(p_1+p_2)/[(p_1+p_2)+(p_1+p_3)+(p_2+p_3)]=(p_1+p_2)/2$. Therefore, the probability that an observation was incomplete and simultaneously fell in C_{12} was $(PID/100)(p_1+p_2)/2$. Similarly, probabilities for C_{13} and C_{23} were $(PID/100)(p_1+p_2)/2$ and $(PID/100)(p_2+p_3)/2$, respectively, and the probability that an observation was completely specified and fell in C_1, C_2, or C_3 was $(1-PID/100)p_1$,

$(1\text{-PID}/100)p_2$, or $(1\text{-PID}/100)p_3$, respectively.

Therefore, to generate incomplete data $\underset{\sim}{z}$ according to the likelihood equation (2.8) for n observations, we could draw n uniform random numbers u_i, $1 \leq i \leq n$. We would then use these six probabilities to establish intervals that determined where an observation fell. For example, if $0 \leq u_i < (1\text{-PID}/100)p_1$, we would increment z_1 by one and, if $(1\text{-PID}/100) \leq u_i < (1\text{-PID}/100)+(p_1+p_2)/2\times\text{PID}/100$, we would increment z12 by one.

However, we also wanted to generate complete trinomial data $\underset{\sim}{x}$ for use in Section 5.9. Therefore, we had to divide z12, z13, and z23 into proportions that fell into completely specified categories C_1, C_2, and C_3. To do so, we noted that if an observation fell in C_{12}, then with probability $p_1/(p_1+p_2)$ it belonged in C_1; similarly, for C_{13} and C_{23}. Therefore, we divided the z12, z13, and z23 intervals, exampled in the last paragraph, into two by the ratios $p_1/(p_1+p_2)$, $p_1/(p_1+p_3)$, and $p_2/(p_2+p_3)$, respectively.

Finally, we set to 0 each element of the complete data $\underset{\sim}{x}$ and the incomplete data $\underset{\sim}{z}$. We then created dummy variables y_1, y_2, w_1, w_3, v_2, and v_3 and initialized them also to 0. From the uniform-random-number generator described at the beginning of this section, we drew n uniform random numbers u_i, $1 \leq i \leq n$. Then, letting h=PID/100 and $p_{ij}=p_i+p_j$,

if	add 1 to
$0 \leq u_i < p_1(1-h)$	z_1
$p_1(1-h) \leq u_i < p_{12}(1-h)$	z_2
$p_{12}(1-h) \leq u_i < 1-h$	z_3
$1-h \leq u_i < 1-h(1-p_1/2)$	y_1
$1-h(1-p_1/2) \leq u_i < 1-h(1-p_{12}/2)$	y_2
$1-h(1-p_{12}/2) \leq u_i < 1-h(1-p_1-p_2/2)$	w_1
$1-h(1-p_1-p_2/2) \leq u_i < 1-h/2(1-p_1)$	w_3
$1-h/2(1-p_1) \leq u_i < 1-h/2(1-p_{12})$	v_2
$1-h/2(1-p_{12}) \leq u_i < 1$	v_3.

At the end of this process we calculated the complete data $\underset{\sim}{x}$ as

$$\begin{aligned} x_1 &= z_1+y_1+w_1 \\ x_2 &= z_2+y_2+v_2 \\ x_3 &= z_3+w_3+v_3, \end{aligned} \tag{5.7}$$

and the incomplete data $\underset{\sim}{z}$ as

$$\begin{aligned} z_1 &= z_1 & z_{12} &= y_1+y_2 \\ z_2 &= z_2 & z_{13} &= w_1+w_3 \\ z_3 &= z_3 & z_{23} &= v_2+v_3 . \end{aligned} \tag{5.8}$$

This trinomial random-number generator performed well on the same kind of empirical tests used for the uniform, gamma, and Dirichlet random-number generators. Note that to perform empirical tests on these four generators, we used routines from the NASA, Langley Research Center, and the IMSL (International Mathematical and Statistical Libraries, Inc.) computer-program libraries.

5.8 Iteration Considerations:

In this section we discuss the following considerations concerning the iterative algorithms: initial estimate, convergence criterion, problems, and conditions for convergence.

Note here that we used the method that is noniterative in $\dot{\tilde{\sigma}}$ for approximating elements of the exact posterior covariance matrix.

5.8.1 Initial Estimate:

To use iterative algorithms for the maximum likelihood estimate, posterior mode, and Taylor-series approximated posterior mean, we needed initial estimates. Because a major concern of this work was approximating the exact posterior mean, we used the exact posterior mean for the initial estimate. Thus, the number of iterations for convergence was another measure of which estimator best approximated the exact posterior mean.

5.8.2 Convergence Criterion:

In general, the convergence criterion was

$$\text{abs}(\dot{p}_i^{(\ell+1)} - \dot{p}_i^{(\ell)})/\dot{p}_i^{(\ell)} \leq 0.0001 \quad \text{for } i=1,2 \qquad (5.9)$$

for $\dot{p}_i$ denoting one of APM, PMD, and MLE and ℓ denoting the number of iterations.

This criterion gave stability in the $\dot{p}_i$ estimate to at least three significant figures for all cases and to at least four significant figures for nearly all cases. The expected $\underset{\sim}{p}$ for Design 1 were ordered

so that the first two of the three components were less than 0.50. Hence, for most cases the absolute difference between successive iterations for the first two components of an estimate was less than $0.0001 \times 0.50 = 0.00005$ and thus the estimate was stable to the fourth significant figure. The exceptions, which were accurate to the third significant figure, involved those relatively rare cases resulting from generated trinomial data yielding estimators having one of their first two components greater than 0.50.

An artificial example of these exceptions would be trinomial data generated from $\underset{\sim}{p}_2=(.20,.30,.50)$ that yielded an estimator $\dot{\underset{\sim}{p}}^{(\ell)}=(.10,.60,.30)$. The largest absolute difference (acceptable for convergence) between $\dot{p}_2^{(\ell)}$ and $\dot{p}_2^{(\ell+1)}$, the second component of $\dot{\underset{\sim}{p}}^{(\ell+1)}$, would be $.00006 \doteq .0001$; ie, the fourth significant figure would be off by at most 1.

To avoid division by 0 (infinite result) and other small numbers (possibly long iterations), whenever $\dot{p}_i^{(\ell)}$ was less than or equal to 0.10, we used the convergence criterion

$$\text{abs}(\dot{p}_i^{(\ell+1)} - \dot{p}_i^{(\ell)}) \leq 0.00001 \quad \text{for } i=1,2. \tag{5.10}$$

This criterion was equivalent to the first one (5.9) for $\dot{p}_i^{(\ell)} = 0.10$.

5.8.3 Conditions for Convergence:

Recall from Sections 2.3 and 4.3.2 that the EM algorithm converges in P_2 to a solution of the likelihood equation if the eigenvalues of the covariance matrix of the complete-data sufficient statistics are bounded

above zero. Hence, under these conditions the posterior mode and maximum likelihood estimate converge to at least a local maximum. Since the Taylor-series approximate posterior mean can be written as a posterior mode (for the prior $\underset{\sim}{\beta}=\underset{\sim}{\nu}+1$), it also converges to at least a local maximum. The question, however, for the Taylor-series approximate posterior mean is whether it converges to the exact posterior mean. This question also applies to the maximum likelihood estimate and to the posterior mode when they are used as approximations of the exact posterior mean.

In Appendix 4E we addressed this question and determined conditions under which an iterative solution to the Taylor-series approximate posterior mean $\dot{\underset{\sim}{\tilde{p}}}$ agrees with the exact posterior mean $\underset{\sim}{\tilde{p}}$ within a small bounded error. We proved that if there exists a neighborhood $\|\dot{\underset{\sim}{\tilde{p}}}-\underset{\sim}{\tilde{p}}\|_\infty \equiv \max_{1\leq i\leq k} |\dot{\tilde{p}}_i-\tilde{p}_i|<\rho$, for $\rho>0$, of the exact posterior mean such that for all values $\dot{\underset{\sim}{\tilde{p}}}$ in this neighborhood

$$\max_i \sum_{j=1}^{k} |\partial g_i(\dot{\underset{\sim}{\tilde{p}}})/\partial \dot{\tilde{p}}_j| \leq \lambda < 1,$$

where

$$g_i(\dot{\underset{\sim}{\tilde{p}}}) = (z_i+\nu_i+\sum_{D\ni i} z_D\dot{\tilde{p}}_i/\dot{\tilde{p}}_D)/(n+\sum_{j=1}^{k+1}\nu_j),$$

and an initial iterative estimate $\dot{\tilde{p}}_i^{(0)}$ is chosen within the inner sphere $\|\dot{\underset{\sim}{\tilde{p}}}-\underset{\sim}{\tilde{p}}\|_\infty \leq \rho_0$, for $0<\rho_0\leq\rho-\delta/(1-\lambda)$, of this neighborhood, then the Taylor-series approximation will converge to within $\delta/(1-\lambda)$ of the exact posterior mean, where δ is a bound on the error in approximating the exact posterior mean by a first-order Taylor-series expansion. We also

showed how to determine, in practice, whether these conditions can be expected to hold. Note that the same conditions apply to the posterior mode and maximum likelihood estimate except that $g_i(\dot{\underset{\sim}{p}})$ is replaced by the appropriate function.

5.8.4 Problems:

Number of iterations - For some cases having components near zero, convergence took a large number of iterations for the maximum likelihood estimate and the posterior mode. A few cases took over 200 iterations. As noted in Section 6.3, the largest number of iterations was 293 for the maximum likelihood estimate.

Multiple solutions - As discussed in Chapters 3 and 4, equations for the maximum likelihood estimate, posterior mode, and approximate posterior mean are generally expected to have multiple roots. However, as noted in Section 5.8.3, whenever the eigenvalues of the covariance matrix of the complete-data sufficient statistics are bounded above zero, an iterative solution for any of these three estimates converges to a local maximum. Therefore, to insure that the local maximum is a global maximum, we should choose that root that maximizes the likelihood. For the approximate posterior mean, we should choose that root that maximizes the posterior density given the prior $\underset{\sim}{\beta}=\underset{\sim}{\nu}+1$; i.e., that root $\dot{\underset{\sim}{p}}$ for which the likelihood function

$$\dot{p}_1^{z_1+\nu_1-1}\dot{p}_2^{z_2+\nu_2-1}\cdots\dot{p}_{k+1}^{z_{k+1}+\nu_{k+1}-1}\prod_D \dot{p}_D^{z_D}$$

is a maximum. Although it has not been proved, from the complete-data

relationship between the posterior mode and posterior mean, we intuitively expect the global maximum to be in the convergence region of the exact posterior mean $\tilde{\underset{\sim}{p}}$, or at least be the closest root to $\tilde{\underset{\sim}{p}}$.

As illustrated by examples in Section 4D.5 and discussed in Section 4.3.2, however, for trinomial data we usually expect only one root to satisfy the constraints $0 \leq \dot{p}_i \leq 1$, for all $1 \leq i \leq 3$, and $\sum_{i=1}^{3} \dot{p}_i = 1$ for $\dot{p}_i$ any one of the three estimators. Further, exploratory calculations showed that the iterative algorithm for the approximate posterior mean converged to the same solution for a wide range of initial estimates. Finally, all three iterative estimates were close enough to the exact posterior mean and the generator Dirichlet probability that we did not expect a different root as the global maximum. Thus, we did not seek more than one solution.

5.9 Estimates of Mean Squared Error:

Recall that we defined the error $\dot{e}_i=\dot{p}_i-p_i$ for $1\leq i\leq 3$, $\dot{p}_i$ referring to one of estimators APM, PMD, and MLE and p_i referring to the generator Dirichlet probability vector. We want to estimate the mean squared error

$$\text{mse}(\cdot) = E[\dot{e}_1^2+\dot{e}_2^2+\dot{e}_3^2] \tag{5.11}$$

of estimator $\dot{\underset{\sim}{p}}$.

For N denoting the number of simulation trials, the most common estimate of the mean squared error (5.11) is

$$\overline{\text{mse}(\cdot)} = \sum_{j=1}^{N}\sum_{i=1}^{3} \dot{e}_{ij}^2/N, \tag{5.12}$$

where $\dot{e}_{ij}$ is $\dot{e}_i$ on the jth simulation trial. We called (5.12) the "regular" or "usual" mean-squared-error estimate.

For estimating mean squared errors of estimators for minimizing expected quadratic loss, we used two Monte-Carlo techniques to reduce the estimate's variance. In both, we took advantage of any covariance of the quadratic-loss estimators APM, PMD, and MLE with the complete-data maximum-likelihood estimate $\ddot{p}_i=x_i/n$, for x_i denoting the number of the n (25 or 50) observations falling in category i. We called the two resulting estimates the control-variate mean-squared-error estimate and the regression mean-squared-error estimate. Both are discussed by Kleijnen (1975,Part I,Chpt.III).

Let $\ddot{e}_{ij}$ denote $\ddot{e}_i=\ddot{p}_i-p_i$ on the jth simulation trial and, paralleling (5.12), define $\overline{\text{mse}(\cdot\cdot)}=\sum_{j=1}^{N}\sum_{i=1}^{3} \ddot{e}_{ij}$. Then all three mean-

squared-error estimates can be represented in the form:

$$\text{mse\_est}(\cdot) = \overline{\text{mse}(\cdot)} + b\{E[\overline{\text{mse}(\cdot\cdot)}] - \overline{\text{mse}(\cdot\cdot)}\} \tag{5.13}$$

where

$$E_{\underset{\sim}{x}}[\overline{\text{mse}(\cdot\cdot)}] \equiv E_{\underset{\sim}{x}}\left(\sum_{i=1}^{3} \ddot{e}_i^2\right) = \sum_{i=1}^{3} [p_i(1-p_i)]/n = [1 - \sum_{i=1}^{3} p_i^2]/n. \tag{5.14}$$

For the regular mean-squared-error estimate, b=0. For the control-variate mean-squared-error estimate, b=1. For the regression mean-squared-error estimate, b is the regression coefficient b_{re} in the linear regression of $\sum_{i=1}^{3} \dot{e}_i^2$ on $\sum_{i=1}^{3} \ddot{e}_i^2$. Kleijnen (1975) discusses the general case for a constant b not necessarily equal to 1.

Note that, in terminology of Kleijnen (1975), the regression mean-squared-error estimate is also a control-variate estimate. However, the latter term is often used to denote our b=1 case and, to differentiate between the b=1 and $b=b_{re}$ case, we follow this practice.

If the regular estimate of the mean squared error and the regular estimate of the complete-data maximum-likelihood-estimate mean squared error are positively correlated such that

$$\text{var}[\overline{\text{mse}(\cdot\cdot)}] < 2\ \text{cov}[\overline{\text{mse}(\cdot\cdot)}, \overline{\text{mse}(\cdot)}] < \text{var}[\overline{\text{mse}(\cdot\cdot)}] + \text{var}[\overline{\text{mse}(\cdot)}], \tag{5.15}$$

then the control-variate estimate $\overline{\overline{\text{mse}(\cdot)}}$ of the mean squared error will have smaller variance than the regular estimate because

$$\text{var}[\overline{\overline{\text{mse}(\cdot)}}] = \text{var}[\overline{\text{mse}(\cdot)}] + \text{var}[\overline{\text{mse}(\cdot\cdot)}] - 2\ \text{cov}[\overline{\text{mse}(\cdot)}, \overline{\text{mse}(\cdot\cdot)}]. \tag{5.16}$$

Note that both the regular and the control-variate mean-squared-error estimates are unbiased.

The value of b that minimizes the variance of (5.13) is the regression coefficient

$$b_{re} = \text{cov}[\overline{\text{mse}(\cdot)}, \overline{\text{mse}(\cdot\cdot)}]/\text{var}[\overline{\text{mse}(\cdot\cdot)}] \qquad (5.17)$$

used in the regression estimate $\bar{\bar{\text{mse}}}$. For ρ the correlation coefficient

$$\text{var}[\bar{\bar{\text{mse}}}(\cdot)] = \text{var}[\overline{\text{mse}(\cdot)}]\ \{1-\rho^2[\overline{\text{mse}(\cdot)}, \overline{\text{mse}(\cdot\cdot)}]\}. \qquad (5.18)$$

Hence, the variance of the regression estimate is less than the variance of the usual estimate (5.12) by a factor depending on the correlation between $\sum_{i=1}^{3} \dot{e}_i^2$ and $\sum_{i=1}^{3} \ddot{e}_i^2$. We estimate b_{re} by the least squares estimate

$$\hat{b}_{re} = \sum_{j=1}^{N} \{[\sum_{i=1}^{3} \dot{e}_{ij}^2 - \overline{\text{mse}(\cdot)}] \times [\sum_{i=1}^{3} \ddot{e}_{ij}^2 - \overline{\text{mse}(\cdot\cdot)}]\} / \sum_{j=1}^{N} [\sum_{i=1}^{3} \ddot{e}_{ij}^2 - \overline{\text{mse}(\cdot\cdot)}]^2. \qquad (5.19)$$

Although the regression estimate of the mean squared error has minimum variance, it is biased, because

$$E[\bar{\bar{\text{mse}}}(\cdot)] = E[\overline{\text{mse}(\cdot)}] + E(\hat{b}_{re})\ E(\sum_{i=1}^{3} \ddot{e}_i^2) - E[\hat{b}_{re}\ \overline{\text{mse}(\cdot\cdot)}], \qquad (5.20)$$

and, since $\hat{b}_{re}$ is a function of $\overline{\text{mse}(\cdot\cdot)}$, the last term in (5.20) does not equal the second term.

As Cochran (1967) notes, the amount of bias in the regression estimate is difficult to determine. Kleijnen (1975) reviews ways to decrease or remove the bias. However, implementation of these methods can be expensive. More important, 200 simulation trials was enough to remove most of the bias. Results showed that in most cases the regression estimate of the mean squared error lay between the unbiased control-variate mse estimate and the unbiased regular mse estimate. Hence, in

all situations but one, we used the regression-estimate mse since it had the smallest variance.

The one situation in which we did not use the regression estimate was in Design 2 for cases in which the denominator in (5.19) was zero. This sometimes happened when two components of the generated Dirichlet probability were zero to at least three decimal places. In these cases the complete-data maximum likelihood estimate was the same for all 200 trinomial simulations.

5.10 Evaluation of Exact Posterior Mean and Covariance Matrices:

Recall from Section 2.2.4, the dimension, range, precision, and cost problems that generally make numerical evaluation of the exact posterior moments unfeasible.

In our simulation work, however, we

(1) had the smallest-dimension case, the trinomial,

(2) designed the simulation study to have sample sizes small enough for the percentage of incomplete data, and

(3) were able to use a computer with good enough range and significant-figure accuracy

to allow numerical evaluation of these exact moments.

For the trinomial case, the number of terms in each numerator and denominator of the exact posterior moments is

$$\text{number of terms} = (z_{12}+1)\times(z_{13}+1)\times(z_{23}+1). \qquad (5.21)$$

For sample sizes of 25 and 50, percentages of incomplete data of 15 and 40, and probabilities roughly ranging from (0,0,1) to (1/3,1/3,1/3), the number of terms (5.21) ranged from a low near 1 to a high of approximately 512.

For the CDC 6600 and Cyber 175 computers described in Section 5.4, the magnitude range is 10^{-294} to 10^{322}. This range is unusually large for a computer, many of which have ranges more like 10^{-76} to 10^{76}. Therefore, with these special-purpose CDC scientific and engineering computers, we could directly evaluate exact solutions for SS/PID combinations as large as 402/50 or 335/60. The maximum SS/PID

combinations that most other computers can handle is considerably smaller. By "directly" we mean without much extra programing, execution and storage cost, and additional rounding error for scaling down the magnitude of the terms. As noted in Section 5.4, these CDC computers have single-precision accuracy of about 14.5 significant figures. For this machine accuracy, use of 11 significant figures for the gamma $\Gamma(\ )$ functions, and the SS and PID used in this study, our evaluations of the exact posterior moments were accurate to at least 6 significant figures.

Because they could be evaluated directly, equations for the exact posterior moments were programed in a straightforward manner. We used

$$\sum_{a=0}^{z_{12}} \binom{z_{12}}{a} \left\{ \sum_{b=0}^{z_{13}} \binom{z_{13}}{b} \Gamma(z_1+\nu_1+a+b) \left[\sum_{c=0}^{z_{23}} \binom{z_{23}}{c} \Gamma(z_2+\nu_2+z_{12}-a+c) \times \Gamma(z_3+\nu_3+z_{13}-b+z_{23}-c) \right] \right\} \tag{5.22}$$

as a base for all moment calculations, increasing various of the inner and outer sums to obtain numerators for the different desired moments.

For each set of data we called a function GAM once to evaluate $\Gamma(z_1+\nu_1)$, $\Gamma(z_2+\nu_2+z_{12})$, and $\Gamma(z_3+\nu_3+z_{13}+z_{23})$. GAM returned the gamma value from (1) exact values, (2) Abramowitz and Stegun (1970) tables (accurate to 11 significant figures), or (3) from Stirling's Formula for those cases in which the formula gave an approximation accurate to 11 significant figures. [Since Stirling's Formula is an asymptotic formula, there exists some number of terms beyond which the accuracy decreases. For example, $\Gamma(3)$ can not be accurately approximated by Stirling's formula to more than six significant figures; the accuracy decreases beginning with the seventh term.]

From then on, gamma terms in formula (5.22) and its variations were evaluated by the relationship

$$\Gamma(y+1) = y\,\Gamma(y)$$

for both integer and non-integer values of y. Note that for approximately half the cases, the argument to the gamma function was non-integer.

The coefficient in (5.22) was calculated as

$$\binom{z_{ij}}{a} = \frac{z_{ij} - (a-1)}{a} \binom{z_{ij}}{a-1}$$

where $\binom{z_{ij}}{0}$ was set to 1.

CHAPTER 6

RESULTS OF DESIGN 1

6.1 Introduction:

In this chapter, we present results from Design 1. In the following second section, we list special mnemonics common to these next two chapters. In the third section, we discuss characteristics of the estimators arising from the trinomial simulations. In the fourth section, we review results from approximations for elements of the posterior mean and covariance matrices. As part of this review, we discuss which of the Taylor-series approximation, posterior mode, and maximum likelihood estimate best approximates the posterior mean. Finally, we investigate results from which estimator best minimizes quadratic loss. A summary section concludes the chapter.

6.2 Special Mnemonics:

In addition to the mnemonics defined in Section 5.2, we will also use the following in these next two chapters:

APM	APMRO (used in discussions concerning approximations for EPM, for which there was no robustness study)
APMRO	approximate posterior mean APM for robustness set 0 (original prior used in Bayesian estimators)
APMR1	approximate posterior mean APM for robustness set 1 (uniform prior used in Bayesian estimators)
APMR2	approximate posterior mean APM for robustness set 2 (perturbed prior used in Bayesian estimators)
EST	estimator
MLECD	maximum likelihood estimate for complete data (used as control variate in risk study)
NU	prior parameter $\underset{\sim}{\nu}$
OPID	observed percentage of incomplete data
P	$\underset{\sim}{p}$
PMD	PMDRO (used in discussions concerning approximations for EPM, for which there was no robustness study)
PMDRO	posterior mode PMD for robustness set 0 (original prior used in Bayesian estimators)
PMDR2	posterior mode PMD for robustness set 2 (perturbed prior used in Bayesian estimators)

6.3 Estimators:

In this section, we discuss a few properties of estimators from the simulated trinomial data. Recall that for each combination of $\underset{\sim}{p}$, SS, and PID we simulated 200 sets of complete and incomplete trinomial data. From each set of incomplete data, we calculated the estimators EPM, APMRO, PMDRO, MLE, APMR1 [recall that PMDR1≡MLE], APMR2, and PMDR2. The RO, R1, and R2 suffixes refer to robustness sets RO, R1, and R2, respectively. From each set of complete trinomial data, we calculated the complete-data maximum likelihood estimate MLECD.

To examine the sampling distribution of the estimators, we calculated data summaries (extremes, hinges, and median), central values (mean, median, and trimean), and spreads (midspread and range) over the 200 trinomial simulations. Prominent features were that the exact posterior mean and Taylor-series approximate posterior mean had almost identical distributions. So also did the complete-data and incomplete-data maximum likelihood estimates. Since the priors were nonzero, EPM and APM always had nonzero values. However, PMDRO, MLE, and MLECD had a large number of zero values when $\underset{\sim}{p}$=(.01,.01,.98).

The number of iterations for convergence is given in Table 6.1. As expected, the number of iterations increased as the percentage of incomplete data PID increased. The largest change was for $\underset{\sim}{p}_2$; for the original prior, the number of iterations approximately doubled. Direction of sample-size effect was consistent only for APMR1. For this estimator, the average number of iterations decreased from 2% to 15% as SS increased.

One factor affecting the average number of iterations for estimators at $\underset{\sim}{p}_1$ was that 169 of the 9,600 (48×200) sets of six iterative estimators for $\underset{\sim}{p}_1$ required more than 15 iterations. The maximum likelihood estimate constituted most of this 2%. The largest number of iterations was 293 for the maximum likelihood estimate. The large number of iterations occurred when one or more components of the simulated incomplete data $\underset{\sim}{z}$ was zero.

6.4 Approximating Posterior Moments:

6.4.1 Posterior Mean:

Our most important measure of the goodness of an approximation was the percentage absolute realtive difference. In Table 6.2 we give the proportion of 200 trinomial simulations for which the percent absolute relative difference for each of the three components of an approximation was less than specified amounts.

With a few exceptions at $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$, for all cases the percentage absolute relative difference between the Taylor-series approximate posterior mean (APM) and the exact posterior mean (EPM) was less than 1%. That is,

$$|\dot{\tilde{p}}_i - \tilde{p}_i| / \tilde{p}_i \times 100 < 1 \qquad \text{for } 1 \leq i \leq 3,$$

so that

$$|\dot{\tilde{p}}_i - \tilde{p}_i| < 0.01 \times \tilde{p}_i$$

for all three components $\dot{\tilde{p}}_i$, $1 \leq i \leq 3$. Hence, the approximation was accurate to at least two significant figures. The few exceptions are studied later in this section.

Moreover, when PID=15, the APM approximation was accurate to at least three significant figures for nearly all cases and to at least four significant figures for the majority of cases. When PID=40, the approximation was accurate to at least three significant figures for most cases.

As sample size increased from 25 to 50, the APM approximation generally improved for $\underset{\sim}{p}_2$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$. For $\underset{\sim}{p}_1$ it slightly worsened.

The reason is that there were a number of cases for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$ where APM was identical to EPM for SS=25. As the amount of sample data increased, the possibility of a perfect approximation lessened. As already indicated, as PID increased from 15 to 40, the APM approximation worsened, least for $\underset{\sim}{p}_1$ and most for $\underset{\sim}{p}_4$ in terms of three- and four-significant figure accuracy.

In general, the posterior-mode (PMD) and maximum-likelihood-estimate (MLE) approximations were not accurate to even two significant figures. The main exception was at $\underset{\sim}{p}_4$ when SS was 50. There the posterior mode agreed to two significant figures for approximately one-third of the 200 trinomial simulations.

Analyses later in this section showed that even in the few problem cases for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$, APM was a much better EPM approximation than either PMD or MLE. Also, analyses found no bias, mean-squared-error, iteration, or other problems favoring PMD or MLE over APM. Finally, Table 6.2 showed that, except possibly for the APM problem cases, APM was far superior to PMD and MLE in approximating the exact posterior mean in terms of percentage relative difference. Therefore, following a few comments in the next paragraph, we henceforth concentrate only on APM as an approximation for EPM.

Because the exact posterior mean (EPM) was never zero and PMD and MLE were, PMD and MLE were poorest approximations for $\underset{\sim}{p}_1=(.01,.01,.98)$. The better of PMD and MLE was MLE for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$ and PMD for $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$. However, note that even for $\underset{\sim}{p}_4$, when PMD improves in its approximation, it is far inferior to APM. Plots given later in this section illustrate these comparisons.

In Table 6.3 we present the bias for the first component of the three approximations to the exact posterior mean. For PID=15 and PID=40, the bias was estimated over the 200 trinomial simulations by $\sum_{j=1}^{200} (\dot{p}_{1j}-\tilde{p}_{1j})$ $/200 \doteq E(\dot{p}_1)-E(\tilde{p}_1)$ for $\dot{p}_1$ the first component of one of the three approximations APM, PMD, and MLE. The complete-data (PID=0) bias is given for all estimators except for the posterior mode at $\underset{\sim}{p}_1 \equiv (.01,.01,.98)$. Recall that the prior used in the Bayesian estimators for $\underset{\sim}{p}_1$ in Design 1 was $\underset{\sim}{\nu}_1 \equiv (0.1,0.1,9.8)$. For the pair p_i,ν_i having such small values for i=1,2, a solution to the likelihood equations usually does not exist in P_2. [Note that $E[(x_i+\nu_i-1)/(n+\Sigma\nu_j-3)] < 0$ for $\nu_i=0.1$, $p_i=.01$, and n=25 or 50, since $E(x_i)=np_i$.] In this case, the posterior mode occurs on a boundary. Hence, $\tilde{p}_i=0$ for i=1,2. Thus, the likelihood equations are not used to define the posterior mode; therefore, the bias can not be analytically calculated from the i^{th} solution (2.43) to the likelihood equations.

Although the bias was small for all approximations, it was one to three orders of magnitude smaller for APM. For APM, the bias was smallest in absolute value for $\underset{\sim}{p}_4$. For PID=15, it was largest in absolute value for $\underset{\sim}{p}_1$ or $\underset{\sim}{p}_2$; for PID=40, it was largest in absolute value for $\underset{\sim}{p}_1$ or $\underset{\sim}{p}_3$. As sample size increased, the bias generally decreased. The bias was positive for $\underset{\sim}{p}_2$ and $\underset{\sim}{p}_3$ and of both signs for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_4$. Note that, for $\underset{\sim}{p}_1$, $\underset{\sim}{p}_2$, and $\underset{\sim}{p}_4$, results for the second component of the bias were the same as those for the first component of the bias. Results for the second component, .30, of $\underset{\sim}{p}_3$ were similar to those for the first and second components, both .33, of $\underset{\sim}{p}_4$.

Although the estimated biases were small, the individual errors constituting the estimated biases could be large. To investigate this possibility, we calculated data summaries, central values, and spreads over 200 trinomial simulations for the errors of APM, PMD, and MLE in approximating EPM. In general, as sample size increased, error decreased; as PID increased error increased; and as $\underset{\sim}{p}$ moved from the corner $\underset{\sim}{p}_1 \equiv$ (.01,.01,.98) to the center $\underset{\sim}{p}_4 \equiv (1/3,1/3,1/3)$ of the P_2 simplex, the error decreased.

The central values, especially the mean, often differed because the distribution of the errors was not symmetric. To examine this asymmetry, we studied the proportion of the 200 simulations in which the first component of the error was of a given sign. Results showed, for SS=25, that for $\underset{\sim}{p}_2$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$ approximately one half of the APM errors were negative. The remaining half were zero or positive. For $\underset{\sim}{p}_1$, however, almost three fourths of the errors were negative. As sample size increased to 50, the errors remained roughly split as half negative and half positive for $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$. For $\underset{\sim}{p}_2$, however, the error was approximately two-thirds negative and one-third positive. For $\underset{\sim}{p}_1$, it was close to 92% negative, 4% positive, and 4% zero. As expected, the distribution of the APM error was much tighter than those for the PMD and MLE errors.

Finally, the smallness of the midspread relative to the range for all but the {SS=50,PID=40} case (as well as values of the hinges relative to those of the extremes), indicated that most of the APM errors clustered close to zero and that the extreme values were few and unusual.

We next studied these extreme values. In particular, we investigated those cases in Table 6.2 that showed a percentage absolute relative

difference greater than 15. All these cases occurred at $\underset{\sim}{p}_1 \equiv (.01,.01,.98)$.

First, all these cases had empty (0) cells for z_1 and z_2. Second, all these cases occurred for PID=40. However, the observed PID is not necessarily 40. We called this observed percentage of incomplete data OPID. Those cases having high percentage relative difference usually had very high OPID (often in the 50%). Finally, for these cases, the incomplete data was "inconsistent" with the completely specified data and, perhaps less important, with the sampling model. That is, under the sampling model with $z_1=z_2=0$ and z_3 large, we would expect z_{12} small and $z_{13} \doteq z_{23}$. Examples are shown in Figure 6.1, where the estimators are given in successive order as the exact posterior mean, Taylor-series approximate posterior mean, maximum likelihood estimate, and posterior mode and where, again, $\underset{\sim}{z}=(z_1,z_2,z_3,z_{12},z_{13},z_{23})$.

In all three examples, the percentage of incomplete data is very high, 60%, 56%, and 50%, respectively. Further, the data are inconsistent. To see the inconsistency, compare the generated data $\underset{\sim}{z}$ with the expected value of the data given the sampling model. Recall, from Chapter 5, especially Section 5.7.3, that the sampling model is a function of $\underset{\sim}{p}$, PID, and SS. Expected values of $\underset{\sim}{z}$ are given in each example. The most noticeable discrepancy between the expected and generated data is in the relationship between z_{13} and z_{23}. The expected values are identical. The observed values, however, differ greatly. In example 1, z_{13} is approximately one-half z_{23}; in example 2, z_{13} is more than three times z_{23}; and in example 3, z_{13} is almost twice z_{23}. Thus, the probability of observing any data set in these examples, given the sampling model, is small.

FIGURE 6.1

WORST APM APPROXIMATIONS FOR EXACT POSTERIOR MEAN

1. OPID=60*, SS=25, $\underset{\sim}{z}$=(0,0,10,1,5,9), E($\underset{\sim}{z}$)=(.15,.15,14.7,.1,4.95,4.95);

Estimators	Error	% abs rel diff
$\underset{\sim}{\tilde{p}}$=(.0188,.0246,.9566)		
$\underset{\sim}{\dot{\tilde{p}}}$=(.0138,.0304,.9558)	(-.0050, .0058,-.0008)	(26.6, 23.6,0.1)
$\underset{\sim}{\hat{p}}$=(.0001,.0625,.9374)	(-.0187, .0379,-.0192)	(99.5,154.1,2.0)
$\underset{\sim}{\hat{\tilde{p}}}$=(0, 0, 1)	(-.0188,-.0246, .0434)	(100.0,100.0,4.5)

2. OPID=56*, SS=25, $\underset{\sim}{z}$=(0,0,11,1,10,3), E($\underset{\sim}{z}$)=(.15,.15,14.7,.1,4.95,4.95);

Estimators	Error	% abs rel diff
$\underset{\sim}{\tilde{p}}$=(.0266,.0168,.9566)		
$\underset{\sim}{\dot{\tilde{p}}}$=(.0351,.0102,.9547)	(.0085,-.0066, .0019)	(32.0, 39.3,0.2)
$\underset{\sim}{\hat{p}}$=(.0666,0, .9334)	(.0400,-.0168,-.0232)	(150.4,100.0,2.4)
$\underset{\sim}{\hat{\tilde{p}}}$=(0, 0, 1)	(-.0266,-.0168, .0434)	(100.0,100.0,4.5)

3. OPID=50*, SS=50, $\underset{\sim}{z}$=(0,0,25,2,15,8), E($\underset{\sim}{z}$)=(.3,.3,29.4,.2,9.9,9.9);

Estimators	Error	% abs rel diff
$\underset{\sim}{\tilde{p}}$=(.0275,.0186,.9539)		
$\underset{\sim}{\dot{\tilde{p}}}$=(.0369,.0105,.9526)	(.0094,-.0081,-.0013)	(34.2, 43.6,0.1)
$\underset{\sim}{\hat{p}}$=(.0571,.0001,.9428)	(.0296,-.0185,-.0111)	(107.6, 99.5,1.2)
$\underset{\sim}{\hat{\hat{p}}}$=(.0262,0, .9738)	(-.0013,-.0186, .0199)	(4.7,100.0,2.1)

*PID=40 for all three examples

In practice, one does not know the population model and thus can not check for consistency in the same manner. However, if one calculates the expected value of the data using the estimator $\dot{\underset{\sim}{p}}$ and OPID (the observed percentage of incomplete data), rather than $\underset{\sim}{p}$ and PID, one finds the same discrepancy between z_{13} and z_{23} (even though $\dot{\underset{\sim}{p}}$ is a function of $\underset{\sim}{z}$). These expected values for the three examples are:

1. $E(\underset{\sim}{z}|\tilde{\underset{\sim}{p}},OPID)=(.28,.37,14.35,.22,4.88,4.91)$
   $E(\underset{\sim}{z}|\dot{\underset{\sim}{p}},OPID)=(.21,.46,14.34,.22,4.85,4.93)$
   $E(\underset{\sim}{z}|\hat{\underset{\sim}{p}},OPID)=(1.00,0.,14.00,.33,5.00,4.67)$
   $E(\underset{\sim}{z}|\hat{\tilde{\underset{\sim}{p}}},OPID)=(0.00,0.,15.00,0.0,5.00,5.00)$

2. $E(\underset{\sim}{z}|\tilde{\underset{\sim}{p}},OPID)=(.29,.18,10.52,.30,6.88,6.81)$
   $E(\underset{\sim}{z}|\dot{\underset{\sim}{p}},OPID)=(.17,.17,10.78,.31,6.93,6.75)$
   $E(\underset{\sim}{z}|\hat{\underset{\sim}{p}},OPID)=(.73,0.0,10.27,.47,7.00,6.53)$
   $E(\underset{\sim}{z}|\hat{\underset{\sim}{p}},OPID)=(0.0,0.0,11.00,0.0,7.00,7.00)$

3. $E(\underset{\sim}{z}|\tilde{\underset{\sim}{p}},OPID)=(0.69,.47,23.85,.58,12.27,12.16)$
   $E(\underset{\sim}{z}|\dot{\underset{\sim}{p}},OPID)=(0.92,.26,23.82,.59,12.37,12.04)$
   $E(\underset{\sim}{z}|\hat{\underset{\sim}{p}},OPID)=(1.43,.00,23.57,.72,12.50,11.79)$
   $E(\underset{\sim}{z}|\hat{\tilde{\underset{\sim}{p}}},OPID)=(0.66,0.0,24.35,.33,12.50,12.17),$

respectively. Therefore, to indicate whether data are inconsistent, an approach that can be used in practice is to compare the data with the expected value of the data given OPID and any of these four estimators.

For the Taylor-series approximate posterior mean (APM), the second and third examples had the highest percentage absolute relative difference of all cases. The second example is the one case keeping the proportion

from being 1.00 in column 7 of Table 6.2 for "% abs rel diff < 25". Note that, as also found in the remaining cases, the posterior mode and maximum likelihood estimate were even worse approximations than was the Taylor-series approximate posterior mean.

As an extra check that the Taylor-series approximate posterior mean was the best approximation for the exact posterior mean, even in the rare cases just illustrated when the percentage relative difference was high, we calculated the proportion of 200 trinomial simulations when an estimator was best. Because it is possible, especially with three estimator components $\dot{p}_j$, $1 \leq j \leq 3$, for an approximation to be minimum with respect to one criteria but not with respect to another, we used two different criteria to determine when an estimator was best. For a squared-error criterion, for each of the $1 \leq i \leq 200$ trinomial simulations, we chose the approximation that had the smallest squared error, $\sum_{j=1}^{3} (\dot{p}_{ij} - \tilde{p}_{ij})^2$. For a relative-difference criterion, for each of these 200 simulations, we chose the approximation having the smallest absolute relative difference $\sum_{j=1}^{3} |\dot{p}_{ij} - \tilde{p}_{ij}| / \tilde{p}_{ij}$. Note that the divisor in the latter criterion was never zero. By both criteria, for all sets of $\underset{\sim}{p}$, PID, and SS variations, and for both replications, APM was always a better approximation for EPM than were PMD and MLE.

Relating to the squared-error criterion, we next investigated in Table 6.4 the mean squared errors of the approximations. Since APM always had the smallest squared error for each of the 200 trinomial simulations, it also had to have the smallest mean squared error [often called the average mean squared error]. However, we were also interested in order-of-magnitude comparisons among estimators and how

mean squared error varied with $\underset{\sim}{p}$, SS, and PID. Mean squared error (mse) is often used for comparison among estimators because it measures estimator variance as well as bias.

As for the bias, we also calculated the mean squared errors for the complete-data (PID=0) estimators. Note that, as discussed for Table 6.3, we could not analytically calculate the mean squared error for the posterior mode at $\underset{\sim}{p}_1$. For PID=15 and PID=40, mean squared error was estimated by the "usual" estimate $\sum_{j=1}^{200} (\dot{p}_i - \tilde{p}_i)^2/200$. We did not use any variance-reduction techniques, such as discussed in Section 5.9, in estimating these mean squared errors because the control variate $\ddot{\underset{\sim}{p}}$ for the risk study was not expected to be helpful for the exact-posterior-mean study. Further, the mean squared error was not as important in the exact-posterior-mean study as it was in the risk study. Hence, the greater care in its estimation was not necessary. Finally, the difference between the regular APM mean-squared-error estimate and either of the regular PMD or MLE mean-squared-error estimates was so large that use of a variate-reduction technique was not expected to alter results concerning differences.

Results of Table 6.4 show that the APM mean-squared-error estimate was 1½ to 6 orders of magnitude smaller than those for PMD and MLE. Mean squared error increased 1 to 2 orders of magnitude as PID increased from 15 to 40. It usually decreased as SS doubled. For easier comparison of APM with PMD and MLE, average bias and mean-squared-error ratios are given in Table 6.5. Note from Table 6.2 that the bias ratios are only for the first component of an estimator.

Finally, recall Table 6.1 showing the number of iterations required for convergence of the iterative approximations. Since the initial iterative estimate was the exact posterior mean, the number of iterations was some measure of which approximation was best. By this measure also, the Taylor-series approximate posterior mean was the superior approximation.

We next performed an analysis of variance (ANOVA) on the bias and on the mean-squared-error data. From theory in Steel and Torrie (1960, p157) and Snedecor and Cochran (1968,p324-5,329) and from examples of Dempster, Schatzoff, and Wermuth (1977,p77) and Gunst and Mason (1977, p616), we expected errors from an ANOVA on the original mean-squared-error data to exhibit enough nonnormality and inequality of variances to yield too many false significant F tests. Therefore, for protection against this occurrence, along with improved additivity of the model, we transformed the mean squared errors to natural logarithms. Doing so, however, meant that all mean-squared-error results are interpreted in terms of the log(mse) rather than more naturally in terms of the original data. However, for the risk study we do give an approximate translation of results from logarithms back to the original data.

Note that, although an ANOVA is concerned with all factors affecting bias and log(mse), we are interested only in those significant effects involving the estimators. Note also that usually one studies residuals from the ANOVA model to detect failure to meet assumptions and to learn whether any transformation might correct the failure. However, Scheffé (1967,p363) generally recommends against transforming data to reduce nonnormality in analyzing means. He does so because

interpretation of results concerning transformed data is often difficult. We have already transformed to logs. Thus, further transformation, even if warranted, would lose more in ease of interpretation than would be gained in improving assumptions, especially since the F-test is already fairly robust against assumptions. Therefore, we do not analyze the residuals.

Results from the bias ANOVA, along with significant values, are given in Table 6.6A. The presence of high-order significant interactions affects conclusions about lower-order interactions and the main effects. For the EPM bias ANOVA in Table 6.6A, the main effects for P ($\underset{\sim}{p}$) and estimator EST and the two-factor interaction P×EST are so highly significant relative to the remaining effects that, together with previous bias results, we expect the remaining significant two-factor and three-factor interactions to mean only that effects of EST, P, and P×EST vary with SS and PID.

Plots in 6.6B confirm this hypothesis. As sample size SS increases or PID decreases, the average bias (summed over those factors not appearing in the plot) slightly decreases. Approximation APM has zero average bias. So also, approximately, does MLE. The most striking effect of these two plots is the poorness of the posterior mode as an approximation for the exact posterior mean for all but $\underset{\sim}{p}_4$, and especially for $\underset{\sim}{p}_2$, in terms of average bias.

In Table 6.7A we present F values in the ANOVA for natural logarithms of the estimated mean squared errors given in Table 6.4. Since estimator EST has such huge significance relative to other factors, it will be at least partly responsible for the significant

higher-order interactions. The significant estimator two-factor interactions are plotted in Table 6.7B. The larger the negative value of $\log_e(\text{mse})$, the better the approximation. Thus, plots in 6.7B show that mean squared error decreases slightly as SS increases or PID decreases and that APM is the superior approximation. Approximation APM is poorer at $\underset{\sim}{p}_1$ than at the remaining values of $\underset{\sim}{p}$.

The three-factor interaction P×PID×EST was significant at the 10% level. The effect of PID on the P×EST plot given in 6.7B was that, as PID increased from 15 to 40, differences between APM and either of MLE and PMD decreased and all approximations slightly worsened.

6.4.2 Posterior Covariance Matrix:

In this subsection, we discuss results from Design 1 concerning how well the truncated Taylor-series expansion approximated elements of the posterior covariance matrix.

Note first that elements of the Taylor-series approximate posterior covariance matrix were calculated by the method that is noniterative in elements of the posterior covariance matrix. This method was described in Section 3.2.8. After convergence of components of the approximate posterior mean vector, we solved a linear system of equations for the approximate variances and covariances. These approximations are functions of the approximate posterior means. Thus, the accuracy of the posterior variance and covariance approximations is a function of the accuracy of the posterior mean approximations.

Data summaries, central values, and spreads over 200 trinomial simulations were calculated for the covariance approximations for the

first replication. In general, results indicated very good agreement between sampling distributions of the Taylor-series approximations and the exact posterior covariances. Agreements improved as $\underset{\sim}{p}$ moved from the corner $\underset{\sim}{p}_1$ to the center $\underset{\sim}{p}_4$ of the P_2 simplex. Central values agreed well for all values of $\underset{\sim}{p}$ except $\underset{\sim}{p}_1$, where, as noted in the last section, the distribution of values was heavily skewed because we were at a lower bound for the first two components.

As for the posterior mean, the most important measure of the accuracy of an approximation for an element of the posterior covariance matrix was the percentage of absolute relative difference. In Table 6.8, we give the proportion of 200 trinomial simulations in which the percentage absolute relative difference of the Taylor-series approximation is less than specified amounts. The column headings C11, C12, and C22 denote $\text{var}(p_1|\underset{\sim}{z})$, $\text{cov}(p_1,p_2|\underset{\sim}{z})$, and $\text{var}(p_2|\underset{\sim}{z})$, respectively.

Results show that the variance approximation was correct to at least two significant figures for nearly all 200 trinomial simulations when PID=15. When PID=40, the proportion of 200 variance approximations accurate to at least two significant figures ranged from .83 to 1.00. Further, for the majority of cases, the variance approximation was accurate to at least three significant figures.

Excluding $\underset{\sim}{p}_1$, we find that the approximation remained excellent or improved as $\underset{\sim}{p}$ moved toward the center of the P_2 simplex. Except for $\underset{\sim}{p}_4$, the approximation worsened as PID increased. Sample size SS had little effect when PID=15 because the approximation was already excellent when SS=25. When PID=40, the approximation remained excellent for $\underset{\sim}{p}_4$, slightly improved for $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_2$, and slightly worsened for $\underset{\sim}{p}_1$ as SS

increased.

We next investigated the accuracy of the Taylor-series approximation for the posterior covariance. Results in columns headed by "C12" show that it was not as good an approximation as that for the variance. Even so, for nearly all trinomial simulations, the covariance approximation was correct to at least two significant figures. As for the variance approximation, the covariance approximation remained excellent or improved for $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$ and became poorer for $\underset{\sim}{p}_1$ as the sample size increased. As the percentage of incomplete data increased, the approximation worsened.

To examine relatively poorer results for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$, we investigated averages (over 200 trinomial simulations), percentage average relative difference, average percentage relative difference, and ratio of square root of the estimated mean squared error to the average exact value.

For $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$, the covariance averages were approximately an order of magnitude smaller than the variance averages. In particular, for $\underset{\sim}{p}_1$, the exact posterior covariances ranged in value from -0.4×10^{-5} to $-.2\times10^{-4}$. It could be that values so close to zero were more difficult to approximate. To support this hypothesis, we noticed that when covariances roughly equaled variances, then the average percentage relative differences were also roughly equal. For example, average percentage relative differences for the approximate posterior covariance of p_1 and p_2 given $\underset{\sim}{z}$ at $\underset{\sim}{p}_3$ and the approximate posterior variance of p_1 given $\underset{\sim}{z}$ at $\underset{\sim}{p}_1$, both at PID=15, were of the same magnitude and their average percentage relative differences were also of the same magnitude.

For all but one case, the square-root ratio was less than 1. Finally, the standard errors of the average variance and covariance

approximations were large relative to the averages. Therefore, statistically, we could not differentiate between the approximations and the corresponding exact values.

We also investigated biases of the Taylor-series approximations by examining data summaries, central values, and spreads over 200 trinomial simulations. These results showed that the sampling distributions were tight. Not only were the means, median, trimeans, and midspreads zero, but also the ranges were zero to at least three, and usually four, decimal places.

In general, as sample size increased, the bias decreased. As percentage of incomplete data increased, bias increased. As $\underset{\sim}{p}$ moved toward the center of the P_2 simplex, bias decreased. Exceptions again occurred at $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$ because of the larger number of perfect approximations at those values of $\underset{\sim}{p}$.

To determine whether the Taylor-series approximations generally over approximated, we next investigated the proportion of biases having a positive sign. A positive bias is preferable for a variance approximation because we have defined bias as "approximation - exact". Thus, if most of the biases are positive, then the approximation generally provides an upper bound on the exact posterior variance.

Results showed that the proportion of positive biases was, as has been for other measures, a function of the position of $\underset{\sim}{p}$ in the P_2 simplex. When $\underset{\sim}{p}$ was near the center of the simplex, most of the biases were positive. As $\underset{\sim}{p}$ moved toward a corner of the simplex, the proportion of negative biases increased. At a corner, negative biases dominated.

Although the large proportion of negative variance biases at $\underset{\sim}{p}_1$ is not preferred, it is not of concern as long as the percentage relative difference of the approximation is small. At the end of this section, we investigate cases where the percentage relative difference is greater than 15.

For the covariance approximation, a negative bias was preferred. Since the covariances were negative, a negative bias meant that the approximate covariance was larger in absolute value than the exact covariance. The controlling factor for the proportion of negative biases roughly correlated with the sum of the two covariance elements. When the sum of the two generator $\underset{\sim}{p}$ components was less than 0.75, the proportion of negative biases was larger than that of positive biases. When the sum was higher than .75, between .75 and 1.00, the opposite occurred. For example, the proportion of negative biases for $\text{cov}(p_1,p_2|\underset{\sim}{z})$ for $\underset{\sim}{p}_1 \equiv$ (.01,.01,.98) was near 1; that for $\text{cov}(p_1,p_3|\underset{\sim}{z})$ was near 0. As SS or PID increased, the proportion of negative biases generally increased.

We now investigate those variance and covariance approximations differing in percentage absolute relative value from the exact values by more than 15%. In approximately one-third of these cases, the Taylor-series approximation and the exact posterior mean also differed in percentage absolute relative value by more than 15%. We expected this correlation since elements of the posterior covariance matrix were functions of the approximate posterior means. In these situations, approximations for the posterior variances were usually equal to or 1 - 10% points better than the posterior mean approximation; approximations

for the posterior covariance, usually equal to or 1 - 10% points worse. Part of the reason the covariance approximation was worse than the variance approximation seems, again, to be that the closer the exact value was to zero, the harder it was to estimate.

Recall from Section 6.4.1 that, in these cases of poorer approximations for the posterior mean as well as for the posterior covariance matrix, the incomplete data $\underset{\sim}{z}$ had zero observations for z_1 and z_2, the true percentage of incomplete data TPID was usually very high, and the incompletely specified observations z_{12}, z_{13}, and z_{23} were inconsistent with the completely specified observations z_1, z_2, and z_3 and with the sampling model.

Of the remaining two-thirds cases, three-fifths also had zero observations for z_1 and z_2 and had inconsistent data. Most also had high percentage of incomplete data. Of the last two-fifths of the cases, all but two had percentage absolute relative difference less than 24. These percentages were

(32,20,15),20,19,21,21,22,17,23,19,(41,24,18),15,17,19,16,23,22,16.

Two values of 15 are present because they were greater than 15.000. Numbers in parenthese apply to the same set of data. All percentages, except the 20,15 and 24,18 in parenthesis, are for $\text{cov}(p_1,p_2|\underset{\sim}{z})$, the covariance of two very small values, each varying around 0.01. The 20,15 and 24,18 were values for $\text{var}(p_2|\underset{\sim}{z}),\text{cov}(p_2,p_3|\underset{\sim}{z})$. Nearly all of these cases occurred for data sets having one of z_1 and z_2 equal to 0 and the remaining value equal to 1.

The values 32 and 41 enclosed in parenthesis were of concern. The data for these values was $\underset{\sim}{z}=(1,0,12,1,3,8)$ and $\underset{\sim}{z}=(1,0,26,1,6,16)$, respectively. Observed percentages of incomplete data (OPID) were high, 48% and 46%, respectively. Further, the data was inconsistent for these two cases for a sampling model yielding $E(\underset{\sim}{z})\doteq(0,0,15,0,5,5)$ and $E(\underset{\sim}{z})\doteq(0,0,30,0,10,10)$, respectively. As for the three "problem" examples given in the last section, under this sampling model, with $E(z_1)=E(z_2)=0$ and $E(z_3)$ large, we would expect $z_{13}\doteq z_{23}$. Yet, both cases had z_{23} approximately three times as large as z_{13}.

In essence, when the posterior means of p_1 and p_2, respectively, were very small, we expected the posterior covariances of p_1 and p_2 to be very small. Trying to approximate very small covariances, or covariances of very small values, was relatively difficult, especially when at least one of the two corresponding completely specified observations z_1 and z_2 was zero.

6.4.3 Conclusions:

The Taylor-series approximation for the exact posterior mean was excellent. In most cases it was accurate to at least three significant figures; in many cases, to at least four. In the few exceptions, where the percentage absolute relative difference ranged between 15% and 40%, the data had zero values for two of the three completely specified cells, the percentage of incomplete data was usually very high (40% - 60%), and the incompletely specified data was inconsistent with the completely specified data and with the sampling model. Even in these cases, however, the Taylor-series approximate posterior mean was a better

approximation than the posterior mode or maximum likelihood estimate. The posterior mode and maximum likelihood estimates were nearly always very poor approximations for the exact posterior mean.

The posterior variance and covariance Taylor-series approximations were functions of the Taylor-series approximate posterior mean. Therefore, they were not quite as excellent as approximations in terms of percentage relative difference; the error of the Taylor-series approximate posterior mean was built into their errors. Nonetheless, they were very good. In nearly all cases, they were accurate to at least two significant figures; in most cases, to at least three. As for the posterior mean, exceptions occurred for inconsistent incomplete data having zero values for any two of the three completely specified cells, especially when the percentage of incomplete data was high. Exceptions also occurred for the posterior covariance approximation of two components both having values near zero when the incomplete data had zero observations for either one of the corresponding completely specified cells.

In general, the Taylor-series approximate posterior variance was a slightly better approximation than the Taylor-series approximate posterior covariance, which was usually of values closer to zero. Results indicated that the closer a value was to absolute zero, the harder it was to approximate.

As expected, all approximations generally improved as sample size increased or percentage of incomplete data decreased. An exception were values near a boundary of the P_2 simplex, where, for a sample size of 25,

a number of approximations were perfect. As the sample size increased, the possibility of a perfect fit lessened.

As $\underset{\sim}{p}$ moved from a corner toward the center of the P_2 simplex, approximations generally improved in terms of all the measures that were considered, except for those cases near the P_2 boundaries already having a perfect or near-perfect fit.

6.5 Minimizing Risk for Quadratic Loss:

6.5.1 Introduction:

In this section, we report results from determining which of three estimators best minimized risk, expected quadratic loss, for specified values of $\underset{\sim}{p}$. Two of the estimators were the maximum likelihood estimate MLE and the posterior mode PMD. The remaining estimator was the Taylor-series approximate posterior mean APM. Except at the end of this introductory section, we do not report results from using the exact posterior mean EPM because these results were the same as those from using the approximate posterior mean. We report APM results instead of EPM results because we expect the Taylor-series approximation to be more often used in practice.

As discussed in the introductory chapter, Chapter 1, we were particularly interested in whether the maximum likelihood estimate was best for probabilities at the boundaries of the P_2 simplex; the posterior mean, otherwise. Therefore, the generators were chosen to represent one extreme probability $\underset{\sim}{p}_1 \equiv (.01,.01,.98)$, a probability near a corner of the simplex, and one probability $\underset{\sim}{p}_4 \equiv (1/3,1/3,1/3)$ at the center. The remaining two probabilities $\underset{\sim}{p}_2 \equiv (.10,.10,.80)$ and $\underset{\sim}{p}_3 \equiv (.20,.30,.50)$ lay between the boundary and the center. Hence, if the maximum likelihood estimate <u>is</u> best for $\underset{\sim}{p}_1$ and the posterior mean, for $\underset{\sim}{p}_4$, we will be particularly interested in whether $\underset{\sim}{p}_2$ or $\underset{\sim}{p}_3$ or some probability between them is a crossover point for which estimator best minimizes risk.

As discussed in Chapter 1, we compare the three estimators by using two wrong priors, as well as the correct, original, prior in their

calculations. Note that the maximum likelihood estimate, not being a Bayesian estimate, was the same for all three studies. We labeled these three studies as RO (robustness study 0), R1 (robustness study 1), and R2 (robustness study 2).

For the first wrong prior, in robustness study R1, we chose the uniform prior (1,1,1) because of its common use when one is uncertain of prior knowledge. The uniform prior gives equal weight to all components of $\underset{\sim}{p}$. For this prior, the posterior mode equals the maximum likelihood estimate. For the second wrong prior, in robustness study R2, we chose $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$, where $\underset{\sim}{\nu}$ is the original prior. This prior perturbs the three components of $\underset{\sim}{p}$ by .09, .05, and -.14, respectively. Hence, we called it the perturbed prior. Values of the original-prior mean $\underset{\sim}{p}$ versus the wrong-prior means are given in Figure 6.2.

FIGURE 6.2

PRIOR MEANS FOR THREE ROBUSTNESS STUDIES

R1 prior mean for uniform prior	RO prior mean for original prior	R2 prior mean for perturbed prior
(1/3,1/3,1/3)	(.01,.01,.98)	(.100,.060,.840)
(1/3,1/3,1/3)	(.10,.10,.80)	(.190,.150,.660)
(1/3,1/3,1/3)	(.20,.30,.50)	(.290,.350,.360)
(1/3,1/3,1/3)	(1/3,1/3,1/3)	(.423,.383,.193)

The situation of having previous data but data that yields the wrong prior is more realistically addressed by the perturbed prior in the R2 study. In this study, we picked a wrong prior that was extreme relative

to the correct prior. For example, if the original prior mean in Figure 6.2 is $\underset{\sim}{p}_1$=(.01,.01,.98), then prior data giving a mean of (.10,.06,.84) is unlikely [but not impossible].

Results of robustness study R0 (original prior used in Bayesian estimators) are given in the next section, Section 6.5.2. Those for robustness study R1 (uniform prior used in Bayesian estimators) are given in Section 6.5.3. Results for robustness study R2 (perturbed prior used in Bayesian estimators) are given in Section 6.5.4. Section 6.5.5 summarizes these results for minimizing risk for quadratic loss.

Before leaving this section, we briefly discuss the mean-squared-error (mse) estimates. Risk for quadratic loss is also called mean squared error. As described in Section 5.9, we had three estimates of mean squared error. These were the regular, control-variate, and regression estimates. We found in Section 5.9 that the regression mse estimate had the smallest variance. Nonetheless, for 200 trinomial simulations, the regression-estimate sample variance usually did not differ greatly from that of the regular or control-variate estimates. The differences were nearly always within one order of magnitude. The main exception was a two orders-of-magnitude difference between the control-variate and regression estimates for PMDRO at $\underset{\sim}{p}_1$ when SS=25.

Recall that the regression estimate is biased; the other two are not. However, in almost all cases the biased regression estimate lay between the unbiased regular and control-variate estimates. In the few exceptions, it was close to one of the two unbiased estimates. Hence, its bias was negligible. Therefore, since the regression estimate had

the smallest variance, we used it as the estimate of mean squared error, the estimator's risk.

6.5.2 Original Prior in Bayesian Estimators:

In this subsection, we discuss results from robustness study RO where we used the original, correct, prior $\underset{\sim}{\nu}$ in the Bayesian estimators.

In Table 6.9 we give values for the regression-estimate mean squared error (risk) over 200 trinomial simulations for both replications. For all PID and SS variations, the posterior mean has the smallest mean squared error for $\underset{\sim}{p}_2$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$; the posterior mode, for $\underset{\sim}{p}_1$. Therefore, results indicate that when the correct prior is used in the Bayesian estimators, the posterior mean is the best estimator for all probabilities except those on a boundary of the P_2 simplex. For these boundary probabilities, the posterior mode is the best estimator, although the difference between the posterior mode and the posterior mean decreases as sample size increases. The maximum likelihood estimate is always the worst estimator.

To determine significant effects in Table 6.9, we next present results from analysis of variance on the natural logarithms of these mean squared errors. Table 6.10A shows a huge F value (22,461) for the main effect of $\underset{\sim}{p}$, very large F values (2172 and 1654, respectively) for main effects of sample size and estimator, and a high F value (92, 6df) for the P×EST interaction. Hence, as Snedecor and Cochran (1968,p344) and Steel and Torrie (1960,p207) imply, the significant three-factor interaction P×SS×EST might mean only that there is a minor change in P×EST as

SS varies. Similarly, the large F value for EST relative to that for the two-factor interaction PID×EST might mean only that there is a minor change in EST as PID varies.

Plots of PID×EST and P×EST×SS in Part B of Table 6.10 generally show this premise to be true. Values for the plots were calculated by summing over nonpresent factors (including replication) in Table 6.9 after natural logarithms had been taken. The PID×EST plot shows that, summed over all factors but PID, APM is the best estimator and MLE, the worst. As PID increases, all three estimators become worse. The P×EST×SS plots show that the posterior mean is best for $\underset{\sim}{p}_2$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$ when SS=25. The posterior mode is best for $\underset{\sim}{p}_1$. However, it does not differ greatly from the posterior mean. When sample size increases to 50, the posterior mode and posterior mean become approximately equal at $\underset{\sim}{p}_1$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$. The maximum likelihood estimate is everywhere the worst estimator.

To determine how much risk in Table 6.10B is reduced by using the best estimator, we made a rough translation from $\log_e$(mse) back to mse in the following way. Let v1 and v2 denote the risk of an estimator for replications 1 and 2 (r1 and r2), respectively. Then,

$$\log_e(v1) + \log_e(v2) = \log_e(v1 \times v2).$$

Let w1 and w2 denote the corresponding risk of a second estimator. Then the difference between the summed natural logarithms in the plots of these two estimators is

$$\log_e(v1 \times v2) - \log_e(w1 \times w2) = \log_e[(v1 \times v2)/(w1 \times w2)]$$

$$= \log_e[(v1/w1)(v2/w2)]$$

and

$$\exp[\log_e(v1\times v2)-\log_e(w1\times w2)] = (v1/w1)(v2/w2).$$

Table 6.9 shows that risk differs little between replications r1 and r2; i.e., $v1 \doteq v2$ and $w1 \doteq w2$. Therefore, we can approximate the ratio of the risk of one estimator to that of another estimator by the square root of the last equation; i.e., by

$$\surd\{\exp[\log_e(v1\times v2)-\log_e(w1\times w2)]\}.$$

Again note that $\log_e(v1\times v2)$ is the value plotted in Table 6.10B for an estimator.

Using this basis, then, to roughly translate results from $\log_e(\text{risk})$ to risk, we find that plots in Table 6.10B show that use of the correct estimator reduced risk by about one-fourth (almost one-half at $\underset{\sim}{p}_2$) over use of the next best estimator and by slightly more than one-half over use of the worst estimator when the sample size (SS) was 25. Corresponding reduction in risk when the sample size was 50 was 10% - 15% (25% - 32% at $\underset{\sim}{p}_2$) and 35% - 40%, respectively.

To study further the mean squared errors at $\underset{\sim}{p}_1$, we broke the mean squared error into its 200 components corresponding to the individual trinomial simulations. We then calculated which estimator had the smallest squared error for each of these simulations. From results of the last plot, we would expect the proportion at $\underset{\sim}{p}_1$ to be highest for the posterior mode. However, the proportion of simulations in which the posterior mean had the smallest squared error was two to four times higher than that for the posterior mode! This discrepancy indicates that when the posterior mode <u>is</u> best, it is best by a much greater

amount than when the posterior mean is best.

Finally, we investigated bias for the first two components for those estimators having approximately equal risk, where bias was estimated by $\sum_{j=1}^{200} (\dot{p}_{ij}-p_i)/200$ for i=1,2. Results showed that, except for the trimean for $\underset{\sim}{p}_3$, all central values (mean, median, and trimean) for the errors $\dot{p}_{ij}-p_i$ were smallest for the posterior mean, even at $\underset{\sim}{p}_1$. For each of the first two components, all estimators at $\underset{\sim}{p}_4$ had approximately one-half negative and one-half positive errors. Except at $\underset{\sim}{p}_4$, proportions of negative errors were noticeably higher for the posterior mode than for the posterior mean or maximum likelihood estimate. Proportions for the latter two were always close and were often identical. As $\underset{\sim}{p}$ moved from the center toward the corner of the P_2 simplex or as PID increased, the proportion of negative errors for each component usually increased. As sample size increased, proportions moved toward a 50/50 ratio.

6.5.3 Uniform Prior in Bayesian Estimators:

In this subsection, we discuss results from robustness study R1 where we used the uniform prior (1,1,1), instead of the correct prior $\underset{\sim}{\nu}$, in the Bayesian estimators. For this uniform prior, the posterior mode equals the maximum likelihood estimate. Hence, we have only two estimators for this robustness study.

In Table 6.9 we give values for the regression-estimate mean squared error (risk) over 200 trinomial simulations for both replications. For $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$ for both levels of sample size and both levels of percentage of incomplete data, the posterior mean has the smaller mean

squared error, although differences tend to be small. For $\underset{\sim}{p}_1$, the posterior mode (maximum likelihood estimate) has the smaller mean squared error. For $\underset{\sim}{p}_2$, the posterior mode (mle) has the smaller mean squared error for PID=15 and the posterior mean, for PID=40. However, the difference at $\underset{\sim}{p}_2$ between the two estimators is very small.

Therefore, results indicate that when a uniform prior is used in the Bayesian estimators, the posterior mode (mle) is the better estimator for probabilities at or near a boundary of the P_2 simplex. The posterior mean is the better estimator for all other probabilities.

To determine significant effects in Table 6.9, we next present in Table 6.11 results from an analysis of variance on the natural logarithms of these mean squared errors. As for the original prior, F values for P×EST and SS were so large relative to those for P×SS×EST that we expected the significance for the latter to reflect mainly a variation in P×EST for the two levels of sample size. The plot in Part B of Table 6.11 shows this to be true. Estimators have larger negative $\log_e$(mse) at SS=50, but curves at the two sample sizes are similar. The plot also shows that the difference between the two estimators at $\underset{\sim}{p}_1$ is large relative to the difference at the other three values of $\underset{\sim}{p}$. Finally, as expected, differences between estimators decrease as sample size increases.

Using the rough translation given in Section 6.5.3 for $\log_e$(mse), we find that plots in Table 6.11 show that the largest reduction in risk occurred at the corner probability $\underset{\sim}{p}_1$. At $\underset{\sim}{p}_1$ the risk of the posterior mean was almost six times larger than that of the posterior mode (mle) when the sample size was 25; almost four times larger, when the sample

size was 50. At $\underset{\sim}{p}_2$, however, the risks of the estimators were almost equal. For $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$, the risk of the posterior mean was about 25% smaller than that of the posterior mode (mle) when the sample size was 25 and 15% smaller, when the sample size was 50.

As noted in Section 6.4.2, it is possible for an estimator to have smaller mse but not have smaller squared error for most of the 200 trinomial simulations. Hence, we next studied several estimator characteristics for each of the 200 trials. Since $\underset{\sim}{p}_2$ seemed to be a crossover probability for which estimator was better, results for $\underset{\sim}{p}_2$ were of special interest. They showed that each estimator was better approximately 50% of the time in terms of squared error. However, in terms of percentage relative difference, the posterior mean was the better estimator for two-thirds of the trials.

An investigation of the estimated bias found that central values for the individual errors were smaller for the posterior mode (mle) than for the posterior mean at $\underset{\sim}{p}_2$. Further, in all cases, the posterior mode was slightly closer to a 50/50 ratio of positive errors to negative errors than was the posterior mean. The posterior mean had a higher proportion of positive errors.

6.5.4 Perturbed Prior in Bayesian Estimators:

In this subsection, we discuss results from robustness study R2 where we used the perturbed prior $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$, instead of the correct prior $\underset{\sim}{\nu}$, in the Bayesian estimators.

Table 6.9 gives values for the regression-estimate mean squared error (risk) over 200 trinomial simulations for both replications.

Results are similar to those for the uniform prior. The posterior mean is best for $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$. The posterior mode is best for $\underset{\sim}{p}_1$ and usually best for $\underset{\sim}{p}_2$. The maximum likelihood estimate is the worst estimator at $\underset{\sim}{p}_2$, $\underset{\sim}{p}_3$, and $\underset{\sim}{p}_4$; the posterior mean, at $\underset{\sim}{p}_1$.

Therefore, results indicate that when a very wrong prior is used, the posterior mode is the best estimator for probabilities at or near a boundary. However, the posterior mean will still be the best estimator for probabilities away from the boundary.

To determine significant effects among variables in Table 6.9, we next performed an analysis of variance on the natural logarithms of the mean squared errors. Significant F values are given in Table 6.12. Plots of the significant PID×EST and P×EST×SS interactions are given in Part B of Table 6.12. The PID×EST plot shows that, when summed over $\underset{\sim}{p}$, SS, and replication, the posterior mode PMD is the best estimator, followed by APM and MLE. [However, when this analysis was done on the original mean squared errors rather than on $\log_e(\text{mse})$, the posterior mean, not the posterior mode, was best.] As expected, all estimators worsen as the percentage of incomplete data increases. However, the difference between estimators is almost constant as PID changes.

The plot of P×EST×SS shows that, when summed over PID and replication, the posterior mode is best for $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$, the posterior mean is best for $\underset{\sim}{p}_4$, and the posterior mode and posterior mean are equally best for $\underset{\sim}{p}_3$. Except at $\underset{\sim}{p}_1$, the maximum likelihood estimate is the worst estimator. Estimators improve and differences between estimators decrease as sample size increases.

By using the $\log_e$(mse) translation given in Section 6.5.3, we find that plots in Table 6.12 show that, as for the uniform prior, the largest reduction in risk occurred at the corner probability $\underset{\sim}{p}_1$. At $\underset{\sim}{p}_1$ the risk of the posterior mean was over four times larger than that of the posterior mode when the sample size was 25 and almost three times larger, when the sample size was 50. The risk of the maximum likelihood estimate was twice larger than that of the posterior mode when the sample size was 25 and about 64% larger, when the sample size was 50. At $\underset{\sim}{p}_2$, the risk of the posterior mode and posterior mean were almost equal. The risk of the maximum likelihood estimate was close to one-half that of the posterior mode when the sample size was 25 and about 72% that of the posterior mode when the sample size was 50. At $\underset{\sim}{p}_3$, the risk of the posterior mean was only slightly smaller than that of the posterior mode but was about one-half that of the maximum likelihood estimate when the sample size was 25 and about 70% that of the maximum likelihood estimate when the sample size was 50. At $\underset{\sim}{p}_4$, the risk was reduced about 20% by using the posterior mean instead of the posterior mode when the sample size was 25; about 12%, when the sample size was 50. The relationship between the risk for the posterior mean and maximum likelihood estimate was the same as it was for $\underset{\sim}{p}_3$.

As for the original and uniform priors, we next examined several additional properties of the estimators. The most important result was that, when MLE or PMD had smallest risk, it was generally because, when it had smallest squared error for one of the 200 trinomial simulations, the difference between it and APM's squared error was much larger than the difference when APM was best. This larger difference usually owed to APM, having nonzero prior, never being zero.

6.5.5 Conclusions:

We now conclude results from the three studies for minimizing risk, and we recommend an operating rule. In this section, we are interested in choosing which of the three estimators is best for minimizing risk (expected quadratic loss). As anticipated from the introductory discussion in Section 6.5.1, this minimizing estimator was a function of the probability (or probability mean in the Bayesian framework) that was being estimated.

Summary results from these three studies are given in Tables 6.13 and 6.14. In Table 6.13 we give the ratio of the estimated mean squared errors for the posterior mean to those of the posterior mode and to those of the maximum likelihood estimate. A ratio of less than 1 means that the posterior mean is best. In Table 6.14, we condense results from Table 6.13 and give the estimator having the smallest mean squared error (risk).

If we use the correct prior, results indicate that the posterior mean is best for all values of $\underset{\sim}{p}$ except those very near a boundary of the P_2 simplex. Even very near a boundary, results for the posterior mean differed little from those for the best estimator, the posterior mode, especially for a sample size of 50. [See Plot 6.10B and Table 6.13.] When the sample size was 25, risk was usually reduced by one-fourth if the best estimator was used instead of the next best estimator and by one-half if the best estimator was used instead of the worst estimator, the maximum likelihood estimate. These reductions decreased to about 12% and 38%, respectively, when the sample size doubled.

If we do not have, or want to use, past knowledge for estimating a prior and instead use a uniform prior, in which case the posterior mode equals the maximum likelihood estimate, then results indicate that the posterior mode (mle) is best for points very near a boundary and, for PID=15, those near a boundary. The posterior mean is best everywhere else. The crossover point is approximately $\underset{\sim}{p}_2$, where estimated mean squared errors for the posterior mean and posterior mode are almost equal. [See Plot 6.11B and summary tables, Tables 6.13 and 6.14.] In this robustness study, the largest reduction in risk occurred at the corner probability $\underset{\sim}{p}_1$ where risk was reduced by five-sixths if the posterior mode (mle) was used instead of the posterior mean when the sample size was 25. When the sample size was 50, the reduction was three-fourths. For $\underset{\sim}{p}_3$ and $\underset{\sim}{p}_4$, risk was reduced about one-fourth by using the posterior mean instead of the posterior mode (mle) when the sample size was 25; by one-seventh, when the sample size was 50.

For an estimate of the prior that is very poor, conclusions are similar to those for the uniform prior. The posterior mode is best at or near a boundary; the posterior mean, elsewhere. The main difference is that the crossover point is a little closer toward the center of P_2. [See Plot 6.12B and summary tables, Tables 6.13 and 6.14. In particular, observe how similar curves in Plot 6.12B are to those in Plot 6.10B.] In this robustness study also, the largest reduction in risk occurred at the corner probability $\underset{\sim}{p}_1$. At $\underset{\sim}{p}_1$, risk was reduced by three-fourths when the posterior mode was used instead of the posterior mean when the sample size was 25. When the sample size was 50, the reduction was

two-thirds. At the center $\underset{\sim}{p}_4$ of P_2, risk was reduced by about 20% when the posterior mean was used instead of the posterior mode when the sample size was 25; about 12% when the sample size was 50. Otherwise, at $\underset{\sim}{p}_2$ and $\underset{\sim}{p}_3$, the risk of the posterior mean and posterior mode differed little. Use of the best estimator instead of the maximum likelihood estimate usually reduced risk by one-half when the sample size was 25 and one-third when the sample size was 50.

Recall the centrality measure $C(\underset{\sim}{p})$ that we defined in equation (5.2). This norm is a measure of the distance a probability is from the center of the P_2 simplex. For the four values, $\underset{\sim}{p}_1 \equiv (.01,.01,.98)$, $\underset{\sim}{p}_2 \equiv (.10,.10,.80)$, $\underset{\sim}{p}_3 \equiv (.20,.30,.50)$, and $\underset{\sim}{p}_4 \equiv (1/3,1/3,1/3)$ of $\underset{\sim}{p}$ in the simulation study, centrality measures were 1.88, .98, .14, and 0, respectively. In general, probabilities nearest a boundary have a centrality measure larger than 1.

When we used the uniform prior or the badly estimated prior in the robustness studies, the crossover point for which estimator was best lay between $\underset{\sim}{p}_2$ and $\underset{\sim}{p}_3$. Between $\underset{\sim}{p}_2$ and the crossover point, however, there was little difference between results for the posterior mode, the best estimator, and those for the posterior mean. Further, for priors that are not as badly estimated as were those in the second robustness study, we expect the crossover point to be closer to $\underset{\sim}{p}_2$ or, based on Plot 6.10B for the correct prior, possibly between $\underset{\sim}{p}_1$ and $\underset{\sim}{p}_2$.

Since $\underset{\sim}{p}_2$ has a centrality measure of .98, we recommend, as an operating rule, use of the posterior mean if the centrality measure of $\underset{\sim}{p}$ is less than 1 and the posterior mode, otherwise. This operating rule is a function of $\underset{\sim}{p}$ and in practice, of course, we do not know $\underset{\sim}{p}$. Hence, we can

not calculate the exact centrality measure. However, for any estimate $\hat{\nu}$ of the prior, we can approximate the centrality measure by $C(\underset{\sim}{p}) \doteq C(\hat{\nu}/\sum_{j=1}^{k+1} \hat{\nu}_j)$. In those cases having no estimate of the prior, we could use a uniform prior and, thus, approximate $C(\underset{\sim}{p})$ by 0.

Note that the maximum likelihood estimate was everywhere the worst estimator when the correct prior was used in the Bayesian estimates. Even when a very poor estimate of the correct prior was used [robustness study R2], the maximum likelihood estimate was the worst estimator everywhere except very near a boundary where it was second best.

As sample size increased, the difference between the estimators decreased. A sample size of 50 was large enough for some of the estimators in some cases to be approximately equal. As the percentage of incomplete data increased, all estimators worsened. However, the difference between estimators did not significantly change.

6.6 Summary:

In this chapter we gave results of Design 1 in the simulation study. In the first half we discussed Taylor-series approximations for elements of the posterior mean and covariance matrices. These approximations were needed for the second half of the study. In the second half, we reported which of the posterior mean, posterior mode, and maximum likelihood estimate best minimized risk for quadratic loss at specified values of the population probability (or probability mean in the Bayesian framework). Conclusions and recommendations were given at the end of each of these discussions.

Briefly, the Taylor-series approximations were excellent except for some of those cases simultaneously having inconsistent data, zero observations for two of the three completely specified cells, and high percentage (40% - 60%) of incomplete data. Even in these rare cases, the approximations are probably satisfactory considering the inherent uncertainty associated with estimating nonzero probabilities from zero data. In nearly all cases, the approximations were accurate to at least two significant figures. The approximation for elements of the posterior mean vector was even better. In most cases, it was accurate to at least four significant figures.

The risk study indicated that the posterior mean is the best estimator for all values of the probability $\underset{\sim}{p}$ except those very near a boundary of the P_2 simplex if we use the correct prior in the Bayesian estimates. The posterior mode is best at a boundary. However, it does not

differ much from the posterior mean. If, however, we use a uniform prior or a bad estimate of the correct prior, then the posterior mode is the best estimator for values at or near a boundary of the P_2 simplex; the posterior mean, elsewhere. By using the best estimator, risk was usually reduced by one-fourth over that of the next best estimator and by one-half over that of the worst estimator (nearly always the maximum likelihood estimate) when the sample size was 25. Corresponding reductions when the sample size was 50 were one-eighth and three-eighths, respectively. At a corner $\underset{\sim}{p}_1 \equiv (.01,.01,.98)$, however, the reduction was much larger when an incorrect prior was used in the Bayesian estimators; the risk was reduced by as much as five-sixths when the posterior mode was used instead of the posterior mean.

In the last section we gave the following operating rule for determining which estimator to use in practice: use the posterior mean if the centrality measure calculated from your estimate of the prior is less than 1; otherwise, use the posterior mode.

TABLE 6.1

NUMBER OF ITERATIONS FOR CONVERGENCE AVERAGED OVER 200 TRINOMICAL SIMULATIONS. DESIGN 1. CONVERGENCE CRITERION ABSOLUTE DIFFERENCE IN SUCCESSIVE ITERATIONS RELATIVE TO LAST ITERATION IS LESS THAN 0.0001.[1]

Sample Size		SS = 25												SS = 50											
Prior[2]		original						uniform[3]		perturbed				original						uniform[3]		perturbed			
Estimator		apm		pmd		mle		apm		apm		pmd		apm		pmd		mle		apm		apm		pmd	
Replic. No.		r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2
PID	PROBAB.																								
15	p_1	1.4	1.3	2.9	2.6	5.9	6.1	5.0	5.1	4.5	4.7	3.5	3.5	1.3	1.7	3.4	3.5	5.2	7.5	4.2	4.8	4.4	4.5	3.6	3.7
	p_2	1.3	1.5	4.9	5.3	5.0	5.4	4.8	5.1	4.5	4.8	4.2	4.6	1.5	1.4	4.9	4.9	4.7	4.7	4.7	4.7	4.6	4.6	4.4	4.3
	p_3	1.4	1.5	4.4	4.5	5.1	5.2	4.7	4.8	4.5	4.7	4.4	4.6	1.3	1.3	4.3	4.2	4.6	4.5	4.3	4.3	4.4	4.4	4.3	4.2
	p_4	1.3	1.3	3.9	3.7	5.0	4.8	4.5	4.4	4.4	4.4	4.6	4.6	1.2	1.1	3.5	3.5	4.3	4.3	3.9	4.0	4.3	4.3	4.4	4.4
40	p_1	1.8	2.9	3.1	3.1	7.0	11.5	7.4	7.5	6.3	6.4	4.9	5.0	3.0	3.7	3.7	3.9	13.2	20.0	6.9	7.0	6.5	6.6	5.2	5.3
	p_2	3.1	3.1	10.8	10.3	12.6	12.0	7.7	7.7	6.8	6.8	7.1	7.3	2.9	3.0	8.3	8.6	8.5	8.9	7.1	7.2	6.9	6.9	7.0	7.1
	p_3	2.8	2.8	6.7	6.6	10.3	8.8	7.3	7.4	6.5	6.6	6.5	6.6	2.7	2.7	6.6	6.5	7.7	7.6	6.9	6.9	6.7	6.6	6.4	6.5
	p_4	2.5	2.4	5.5	5.4	8.3	8.0	6.9	6.8	6.5	6.4	6.9	6.8	2.3	2.2	5.2	5.0	7.0	6.8	6.3	6.1	6.5	6.4	6.8	6.6

[1]For any component of the last iteration that is less than or equal to 0.10, use 'absolute difference is less than 0.00001'.

[2]Prior that is used in estimators. Prior assumed for data is 10 p.

[3]For uniform prior, pmd = mle from original prior (columns 5 and 6).

TABLE 6.2

PROPORTION OF 200 TRINOMIAL SIMULATIONS FOR WHICH PERCENT ABSOLUTE RELATIVE DIFFERENCE[1] FOR EACH COMPONENT IS LESS THAN SPECIFIED AMOUNTS.
EPM DIFFERENCE. DESIGN 1.

Sample Size		SS=25												SS=50									
% Inc. Data		PID=15						PID=40						PID=15						PID=40			
Estimator		apm		pmd		mle		apm		pmd		mle		apm		pmd		mle		apm		pmd		mle	
Replic. No.		r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2
Prob	% Rel. Diff.																								
p_1	.01	.98	.97	.00	.00	.00	.00	.95	.89	.00	.00	.00	.00	.95	.92	.00	.00	.00	.00	.82	.77	.00	.00	.00	.00
	.10	.98	.97	.00	.00	.00	.00	.95	.89	.00	.00	.00	.00	.95	.93	.00	.00	.00	.00	.82	.77	.00	.00	.00	.00
	1.00	.98	.98	.00	.00	.00	.00	.97	.90	.00	.00	.00	.00	.97	.95	.00	.00	.00	.00	.87	.81	.00	.00	.00	.00
	5.00	.99	.99	.00	.00	.00	.00	.97	.92	.00	.00	.00	.00	.99	.96	.00	.00	.00	.00	.93	.89	.00	.00	.00	.00
	10.00	.99	1.00	.00	.00	.00	.00	.97	.94	.00	.00	.00	.00	.99	.98	.00	.00	.01	.00	.94	.93	.00	.00	.01	.00
	15.00	1.00	1.00	.00	.00	.00	.00	.98	.97	.00	.00	.00	.00	1.00	1.00	.00	.00	.09	.11	.95	.96	.00	.00	.03	.00
	20.00	1.00	1.00	.00	.00	.00	.00	.98	.98	.00	.00	.00	.00	1.00	1.00	.00	.00	.12	.17	.97	.99	.00	.00	.06	.06
	25.00	1.00	1.00	.00	.00	.00	.01	.99	.98	.00	.00	.01	.00	1.00	1.00	.00	.00	.12	.17	.98	1.00	.00	.00	.07	.06
p_2	.01	.79	.71	.00	.00	.00	.00	.24	.23	.00	.00	.00	.00	.59	.63	.00	.00	.00	.00	.18	.11	.00	.00	.00	.00
	.10	.92	.90	.00	.00	.00	.00	.49	.43	.00	.00	.00	.00	.96	.96	.00	.00	.00	.00	.53	.54	.00	.00	.00	.00
	1.00	1.00	1.00	.00	.00	.00	.00	.93	.94	.00	.00	.01	.01	1.00	1.00	.00	.00	.00	.01	.98	.97	.00	.00	.00	.01
	5.00	1.00	1.00	.00	.00	.04	.04	1.00	1.00	.00	.00	.05	.03	1.00	1.00	.00	.00	.20	.21	1.00	1.00	.00	.00	.11	.10
	10.00	1.00	1.00	.01	.00	.19	.20	1.00	1.00	.00	.01	.12	.09	1.00	1.00	.05	.07	.62	.61	1.00	1.00	.02	.02	.34	.29
	15.00	1.00	1.00	.03	.03	.27	.28	1.00	1.00	.01	.03	.23	.17	1.00	1.00	.28	.25	.69	.70	1.00	1.00	.13	.12	.58	.53
	20.00	1.00	1.00	.15	.12	.39	.45	1.00	1.00	.04	.07	.32	.32	1.00	1.00	.55	.50	.89	.88	1.00	1.00	.33	.28	.71	.62
	25.00	1.00	1.00	.22	.20	.48	.54	1.00	1.00	.12	.15	.45	.47	1.00	1.00	.68	.67	.89	.88	1.00	1.00	.50	.45	.73	.64
p_3	.01	.60	.54	.00	.00	.00	.00	.04	.08	.00	.00	.00	.00	.67	.74	.00	.00	.00	.00	.09	.05	.00	.00	.00	.00
	.10	1.00	.96	.00	.00	.00	.00	.65	.62	.00	.00	.00	.00	1.00	1.00	.00	.00	.00	.00	.74	.80	.00	.00	.00	.00
	1.00	1.00	1.00	.01	.00	.01	.01	1.00	1.00	.00	.00	.00	.00	1.00	1.00	.01	.00	.03	.04	1.00	1.00	.01	.01	.02	.01
	5.00	1.00	1.00	.27	.23	.12	.13	1.00	1.00	.15	.18	.11	.08	1.00	1.00	.61	.66	.46	.47	1.00	1.00	.49	.55	.32	.29
	10.00	1.00	1.00	.66	.66	.29	.35	1.00	1.00	.52	.59	.29	.26	1.00	1.00	.92	.96	.81	.84	1.00	1.00	.86	.86	.67	.65
	15.00	1.00	1.00	.83	.82	.55	.60	1.00	1.00	.74	.81	.42	.44	1.00	1.00	1.00	.99	.93	.96	1.00	1.00	.97	.95	.82	.84
	20.00	1.00	1.00	.95	.91	.73	.73	1.00	1.00	.85	.89	.56	.63	1.00	1.00	1.00	1.00	.98	.98	1.00	1.00	.99	1.00	.92	.90
	25.00	1.00	1.00	.96	.94	.79	.81	1.00	1.00	.96	.96	.69	.73	1.00	1.00	1.00	1.00	.99	.99	1.00	1.00	.99	1.00	.94	.93
p_4	.01	.63	.65	.00	.00	.00	.00	.04	.08	.00	.00	.00	.00	.82	.84	.00	.00	.00	.00	.10	.09	.00	.00	.00	.00
	.10	1.00	1.00	.00	.00	.00	.00	.75	.81	.00	.00	.00	.00	1.00	1.00	.01	.00	.00	.00	.85	.93	.00	.00	.00	.00
	1.00	1.00	1.00	.10	.10	.00	.00	1.00	1.00	.06	.09	.00	.00	1.00	1.00	.40	.36	.06	.05	1.00	1.00	.30	.31	.03	.02
	5.00	1.00	1.00	.82	.88	.12	.13	1.00	1.00	.75	.80	.07	.10	1.00	1.00	1.00	.99	.58	.59	1.00	1.00	.95	.96	.38	.40
	10.00	1.00	1.00	.98	.98	.35	.50	1.00	1.00	.96	.98	.29	.31	1.00	1.00	1.00	1.00	.90	.92	1.00	1.00	1.00	1.00	.72	.77
	15.00	1.00	1.00	1.00	1.00	.64	.72	1.00	1.00	1.00	1.00	.51	.52	1.00	1.00	1.00	1.00	.99	.97	1.00	1.00	1.00	1.00	.91	.93
	20.00	1.00	1.00	1.00	1.00	.78	.84	1.00	1.00	1.00	1.00	.68	.70	1.00	1.00	1.00	1.00	1.00	.99	1.00	1.00	1.00	1.00	.95	.96
	25.00	1.00	1.00	1.00	1.00	.88	.91	1.00	1.00	1.00	1.00	.77	.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.98	.98

[1] $|\hat{p}_j - p_j| / p_j \times 100$ for $1 \le j \le 3$ for $\hat{p}_j$ denoting the j^{th} component of one of the estimators apm, pmd, and mle and for p_j denoting the j^{th} component of the exact posterior mean; note that p_j was never zero in Design 1.

TABLE 6.3

BIAS OF APPROXIMATIONS FOR EXACT POSTERIOR MEAN. FIRST COMPONENT. DESIGN 1.

Dirichlet Prob.	Approximation	SS=25 PID=0	SS=25 PID=15 r1	SS=25 PID=15 r2	SS=25 PID=40 r1	SS=25 PID=40 r2	SS=50 PID=0	SS=50 PID=15 r1	SS=50 PID=15 r2	SS=50 PID=40 r1	SS=50 PID=40 r2
p_1	APM	0	.10-4(.24-4)	.18-4(.98-5)	.59-4(.67-4)	-.35-4(.80-4)	0	.19-4(.12-4)	-.38-5(.16-4)	.90-4(.69-4)	.34-4(.41-4)
	PMD	*	-.77-2(.72-3)	-.85-2(.74-3)	-.68-2(.63-3)	-.80-2(.70-3)	*	-.79-2(.54-3)	-.90-2(.54-3)	-.78-2(.55-3)	-.80-2(.54-3)
	MLE	0	-.15-3(.52-3)	.39-3(.51-3)	-.11-2(.52-3)	-.39-3(.76-3)	0	.23-3(.21-3)	.37-3(.25-3)	.65-3(.40-3)	.62-3(.43-3)
p_2	APM	0	.78-5(.27-5)	.59-5(.46-5)	.70-4(.27-4)	.46-4(.23-4)	0	.35-5(.19-5)	.24-5(.16-5)	.58-4(.12-4)	.46-4(.16-4)
	PMD	.22-1	-.24-1(.41-3)	-.23-1(.42-3)	-.28-1(.91-3)	-.26-1(.94-3)	.12-1	-.13-1(.18-3)	-.13-1(.20-3)	-.16-1(.41-3)	-.16-1(.42-3)
	MLE	0	-.19-2(.15-2)	.31-3(.15-2)	-.48-2(.22-2)	.34-3(.21-2)	0	.29-4(.55-3)	-.40-3(.60-3)	-.50-3(.93-3)	-.12-2(.90-3)
p_3	APM	0	.73-5(.31-5)	.14-4(.45-5)	.59-4(.22-4)	.79-4(.19-4)	0	.46-5(.17-5)	.41-5(.14-5)	.79-4(.13-4)	.78-4(.12-4)
	PMD	.13-1	-.13-1(.47-3)	-.14-1(.47-3)	-.17-1(.60-3)	-.16-1(.57-3)	.70-2	-.81-2(.22-3)	-.77-2(.21-3)	-.99-2(.29-3)	-.94-2(.31-3)
	MLE	0	.21-2(.20-2)	-.65-3(.20-2)	-.25-2(.28-2)	.24-2(.26-2)	0	-.80-3(.83-3)	.45-3(.78-3)	.49-4(.11-2)	.17-2(.12-2)
p_4	APM	0	-.33-5(.29-5)	.94-6(.28-5)	.27-5(.17-4)	.32-4(.18-4)	0	.90-6).15-5)	-.61-6(.13-5)	-.73-5(.11-4)	-.45-5(.84-5)
	PMD	0	.70-3(.56-3)	.14-3(.50-3)	-.24-3(.60-3)	.36-3(.61-3)	0	.53-3(.23-3)	-.15-3(.24-3)	-.16-3(.32-3)	.25-3(.33.3)
	MLE	0	.30-2(.25-2)	.55-3(.22-2)	-.93-3(.29-2)	.17-2(.29-2)	0	.20-2(.90-3)	-.56-3(.92.3)	-.58-3(.13-2)	.10-2(.13.2)

Note that values are given in scientific notation; e.g., .000010 is written as .10-4. Values in parenthesis are standard errors.
For PID=15 and PID=40, bias is estimated by averaging deviations over 200 trinomial simulations [see Section 6.4.]

*Values were not calculated (see main text)

TABLE 6.4

MEAN SQUARED ERROR OF APPROXIMATIONS FOR EXACT POSTERIOR MEAN. DESIGN 1.

Dirichlet Prob.	Approximation	SS=25 PID=0	SS=25 PID=15 r1	SS=25 PID=15 r2	SS=25 PID=40 r1	SS=25 PID=40 r2	SS=50 PID=0	SS=50 PID=15 r1	SS=50 PID=15 r2	SS=50 PID=40 r1	SS=50 PID=40 r2
p_1	APM	0	.22-6(.16-6)	.36-7(.21-7)	.17-5(.79-6)	.27-5(.97-6)	0	.52-7(.36-7)	.98-7(.39-7)	.18-5(.87-6)	.68-6(.19-6)
	PMD	*	.79-3(.77-4)	.82-3(.84-4)	.82-3(.88-4)	.86-3(.90-4)	*	.56-3(.40-4)	.64-3(.42-4)	.53-3(.42-4)	.59-3(.40-4)
	MLE	.13-3	.17-3(.31-4)	.16-3(.19-4)	.25-3(.32-4)	.38-3(.55-4)	.22-2	.31-4(.34-5)	.42-4(.45-5)	.79-4(.11-4)	.90-4(.95-5)
p_2	APM	0	.27-8(.71-9)	.83-8(.29-8)	.29-6(.62-7)	.27-6(.70-7)	0	.16-8(.43-9)	.13-8(.31-9)	.74-7(.11-7)	.13-6(.36-7)
	PMD	.29-2	.34-2(.63-4)	.34-2(.59-4)	.46-2(.98-4)	.47-2(.10-3)	.92-3	.11-2(.14-4)	.11-2(.17-4)	.15-2(.28-4)	.15-2(.28-4)
	MLE	.11-2	.15-2(.12-3)	.14-2(.96-4)	.29-2(.22-3)	.30-2(.22-3)	.19-3	.23-3(.19-4)	.27-3(.20-4)	.55-3(.42-4)	.54-3(.39-4)
p_3	APM	0	.41-8(.70-9)	.77-8(.20-8)	.20-6(.69-7)	.18-6(.23-7)	0	.16-8(.43-9)	.10-8(.12-9)	.81-7(.10-7)	.78-7(.12-7)
	PMD	.29-2	.64-3(.30-4)	.64-3(.29-4)	.94-3(.41-4)	.89-3(.41.4)	.15-3	.19-3(.74-5)	.20-3(.80-5)	.31-3(.14.4)	.29-3(.12-4)
	MLE	.20-2	.28-2(.19-3)	.29-2(.21-3)	.51-2(.40-3)	.49-2(.38-3)	.34-3	.51-3(.35-4)	.48-3(.38-4)	.97-3(.76-4)	.89-3(.59-4)
p_4	APM	0	.62-8(.27-8)	.46-8(.12-8)	.15-6(.23-7)	.14-6(.29-7)	0	.11-8(.17-9)	.11-8(.20-9)	.71-7(.91-8)	.56-7(.73-8)
	PMD	.12-3	.17-3(.11-4)	.15-3(.11-4)	.23-3(.15-4)	.21-3(.14-4)	.26-4	.31-4(.19-5)	.34-4(.22-5)	.62-4(.46-5)	.61-4(.50-5)
	MLE	.22-2	.34-2(.24-3)	.28-2).22-3)	.53-2(.36-3)	.46-2(.33-3)	.37-3	.47-3(.29-4)	.50-3(.33-4)	.10-2(.89-4)	.10-2(.82-4)

Note that values are given in scientific notation; e.g., .00000022 is written as .22-6. Values in parenthesis are standard errors.
For PID=15 and PID=40, mean squared error is estimated by the "usual" estimate [see Section 5.9]

*Values were not calculated (see main text)

TABLE 6.5

BIAS AND MSE RATIOS FOR EPM COMPARISONS. DESIGN 1.

Sample Size		SS=25				SS=50			
% Inc. Data		PID=15		PID=40		PID=15		PID=40	
Replic. No.		r1	r2	r1	r2	r1	r2	r1	r2
$\underset{\sim}{p}^*$	Approx.								
		A. RATIO OF BIAS(APM) TO BIAS(PMD) AND BIAS(MLE) FOR EPM COMPARISONS							
$\underset{\sim}{p}_1$	PMD	-.13 -2	-.21 -2	-.87 -2	.44 -2	-.24 -2	.43 -3	-.11 -1	-.42 -2
	MLE	-.69 -1	.45 -1	-.52 -1	.89 -1	.81 -1	-.10 -1	.14. 0	.55 -1
$\underset{\sim}{p}_2$	PMD	-.33 -3	-.25 -3	-.25 -2	-.18 -2	-.26 -3	-.18 -3	-.37 -2	-.30 -2
	MLE	-.41 -2	.19 -1	.14 -1	-.14 0	.12 0	-.60 -2	-.12 0	-.38 -1
$\underset{\sim}{p}_3$	PMD	-.54 -3	-.10 -2	-.35 -2	-.48 -2	-.57 -3	-.53 -3	-.81 -2	-.83 -2
	MLE	.35 -2	-.22 -1	-.24 -1	.32 -1	-.58 -2	.92 -2	.16 1	.45 -1
$\underset{\sim}{p}_4$	PMD	-.48 -2	.70 -2	-.12 -1	.89 -1	.17 -2	.41 -2	.47 -1	-.18 -1
	MLE	-.11 -2	.17 -2	-.30 -2	.19 -1	.44 -3	.11 -2	.13 -1	-.45 -2
		B. RATIO OF MSE(APM) TO MSE(PMD) AND MSE(MLE) FOR EPM COMPARISONS							
$\underset{\sim}{p}_1$	PMD	.28 -3	.45 -4	.20 -2	.31 -2	.93 -4	.15 -3	.34 -2	.12 -2
	MLE	.13 -2	.23 -3	.68 -2	.71 -2	.17 -2	.24 -2	.23 -1	.76 -2
$\underset{\sim}{p}_2$	PMD	.78 -6	.24 -5	.64 -4	.58 -4	.15 -5	.12 -5	.50 -4	.84 -4
	MLE	.18 -5	.59 -5	.10 -3	.92 -4	.71 -5	.48 -5	.13 -3	.24 -3
$\underset{\sim}{p}_3$	PMD	.64 -5	.12 -4	.22 -3	.20 -3	.87 -5	.53 -5	.27 -3	.27 -3
	MLE	.15 -5	.27 -5	.40 -4	.37 -4	.32 -5	.22 -5	.84 -4	.87 -4
$\underset{\sim}{p}_4$	PMD	.36 -4	.31 -4	.67 -3	.69 -3	.37 -4	.32 -4	.11 -2	.92 -3
	MLE	.18 -5	.16 -5	.29 -4	.31 -4	.25 -5	.21 -5	.68 -4	.57 -4

*Dirichlet probability (expected value of the Dirichlet distribution of $\underset{\sim}{p}$ given prior parameters $\underset{\sim}{\nu}_1$, $\underset{\sim}{\nu}_2$, $\underset{\sim}{\nu}_3$, and $\underset{\sim}{\nu}_4$, respectively)

TABLE 6.6A
ANALYSIS OF VARIANCE FOR ESTIMATED EPM BIAS

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
P	3	.658393 -3	.219464 -3	201.924 ***
SS	1	.539909 -4	.539909 -4	49.676 ***
PID	1	.794404 -5	.794404 -5	7.309 ***
EST	2	.211893 -2	.105946 -2	974.787 ***
P×SS	3	.686198 -4	.228733 -4	21.045 ***
P×PID	3	.634564 -5	.211521 -5	1.946
P×EST	6	.104191 -2	.173652 -3	159.773 ***
SS×PID	1	.219305 -5	.219305 -5	2.018
SS×EST	2	.879171 -4	.439586 -4	40.445 ***
PID×EST	2	.642627 -5	.321314 -5	2.956 *
P×SS×PID	3	.452496 -6	.150832 -6	.139
P×SS×EST	6	.114745 -3	.191242 -4	17.596 ***
P×PID×EST	6	.125370 -4	.208950 -5	1.922 *
SS×PID×EST	2	.135775 -5	.678876 -6	.625
P×SS×PID×EST	6	.621923 -6	.103654 -6	.095
ERROR	48	.521696 -4	.108687 -5	
TOTAL	95	.423455 -2		

* Significant at 10% level.
*** Significant at 1% level.

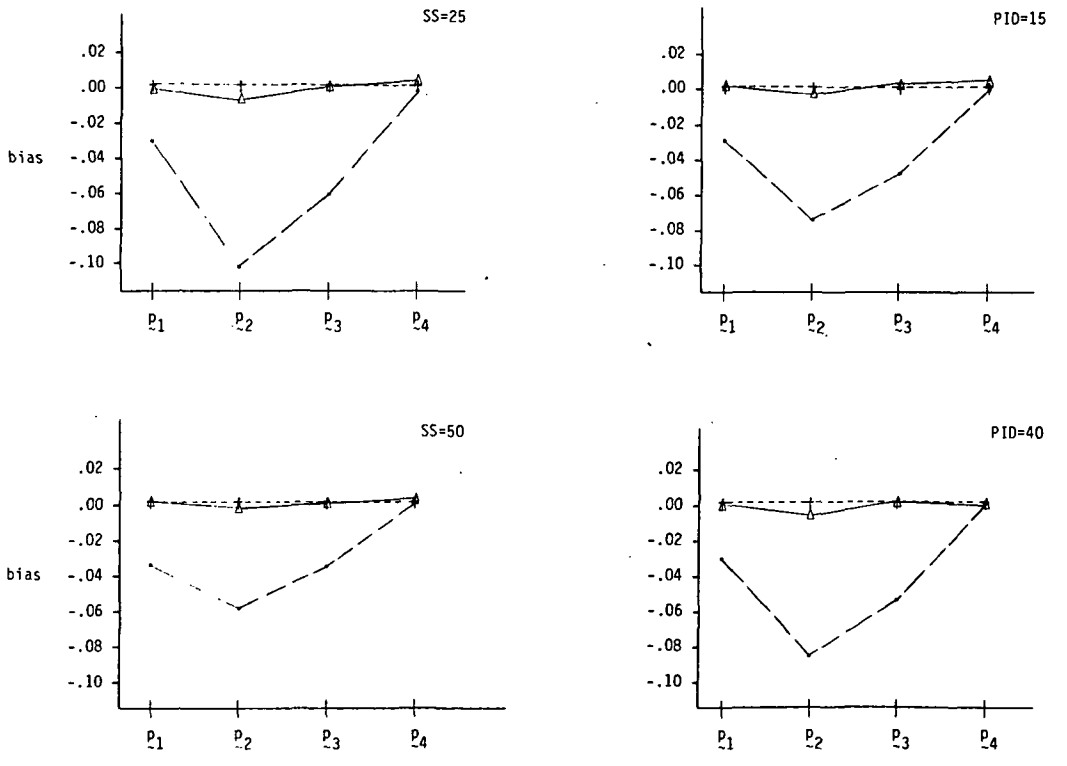
Note that exponential notation is used for the third and fourth columns; for example, .00423455 is written as .423455 -2.

6.6B PLOTS OF P×EST×SS AND P×EST×PID INTERACTIONS*.

P×EST×SS

P×EST×PID

† APM
• PMD
Δ MLE



*Values are sums over nonpresent factors, including replication.

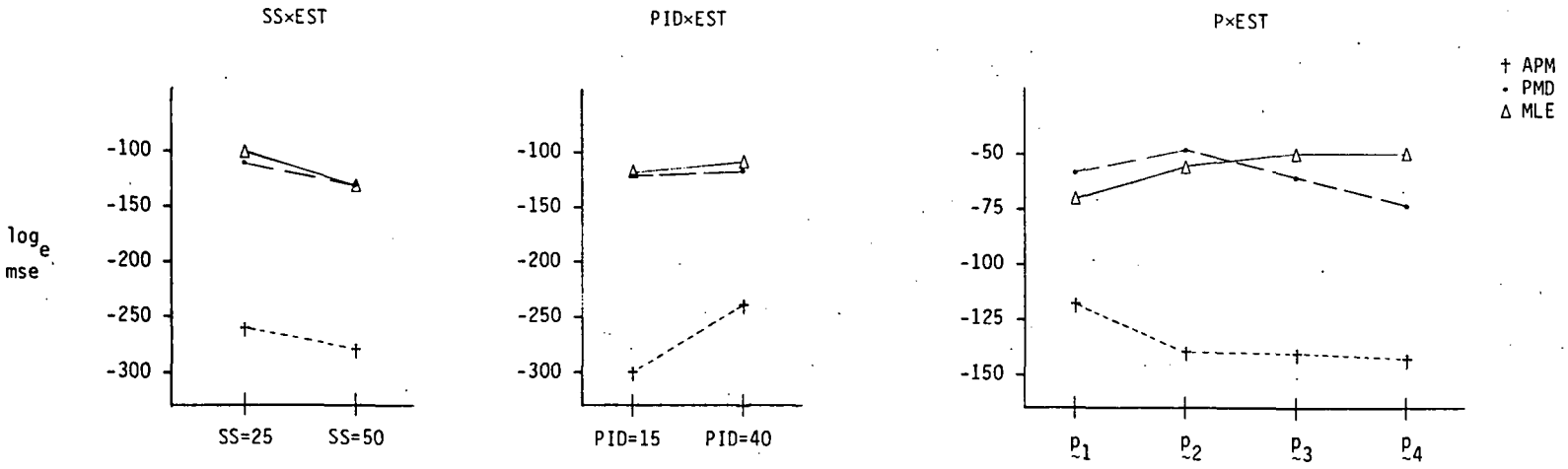
TABLE 6.7A

ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED EPM MEAN SQUARED ERROR

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
P	3	.127603 2	.425343 1	52.797 ***
SS	1	.354935 2	.354935 2	440.575 ***
PID	1	.569041 2	.569041 2	706.342 ***
EST	2	.199362 4	.996812 3	12,373.269 ***
P×SS	3	.191251 1	.637503 0	7.913 ***
P×PID	3	.886197 0	.259399 0	3.667 **
P×EST	6	.121670 3	.202784 2	251.713 ***
SS×PID	1	.162900 0	.162900 0	2.022
SS×EST	2	.219450 1	.109725 1	13.620 ***
PID×EST	2	.540537 2	.270269 2	335.480 ***
P×SS×PID	3	.242087 0	.806956 -1	1.002
P×SS×EST	6	.317632 0	.529387 -1	.657
P×PID×EST	6	.984097 0	.164016 0	2.036 *
SS×PID×EST	2	.287500 -1	.143750 -1	.178
P×SS×PID×EST	6	.228965 0	.381608 -1	.474
ERROR	48	.386696 1	.805617 -1	
TOTAL	95	.228533 4		

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

6.7B PLOTS OF ESTIMATOR TWO-WAY INTERACTIONS*



*Values are sums over nonpresent factors, including replication.

TABLE 6.S

PROPORTION OF 200 TRINOMIAL SIMULATIONS IN WHICH PERCENT ABSOLUTE RELATIVE DIFFERENCE[1] BETWEEN TAYLOR-SERIES APPROXIMATION (T.S. APC) AND EXACT (EPC) POSTERIOR COVARIANCES IS LESS THAN SPECIFIED AMOUNTS. DESIGN 1.

Sample Size		SS=25												SS=50									
% Inc. Data		PID=15						PID=40						PID=15						PID=40			
Covariance		C11		C12		C22		C11		C12		C22		C11		C12		C22		C11		C12		C22	
Replic. No.		r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2	r1	r2
Prob	% Rel. Diff.																								
p_1	.01	.98	.97	.63	.61	.98	.98	.94	.88	.00	.02	.95	.87	.95	.92	.61	.60	.95	.92	.82	.77	.00	.00	.82	.77
	.10	.98	.97	.98	.97	.98	.98	.95	.89	.45	.36	.95	.90	.95	.92	.95	.92	.95	.92	.83	.78	.32	.33	.82	.77
	1.00	.99	.97	.98	.97	.98	.99	.96	.91	.95	.89	.96	.91	.98	.94	.96	.92	.97	.94	.88	.83	.82	.77	.86	.83
	5.00	.99	.99	.98	.98	.99	1.00	.96	.92	.97	.90	.97	.93	.98	.97	.97	.94	.99	.96	.90	.89	.86	.78	.90	.86
	10.00	1.00[2]	1.00	1.00[2]	.99	1.00	1.00	.97	.95	.98	.92	.98	.94	.99	.99	.99	.99	.99	.99	.93	.93	.89	.81	.95	.90
	15.00	1.00[2]	1.00	1.00[2]	1.00	1.00	1.00	.98	.98	1.00[2]	.95	.98	.96	1.00	1.00	.99	1.00	1.00	1.00	.96	.96	.92	.84	.97	.96
p_2	.01	.76	.73	.56	.51	.77	.70	.22	.23	.00	.03	.22	.23	.49	.57	.34	.37	.46	.59	.13	.10	.00	.00	.13	.13
	.10	.89	.84	.78	.72	.94	.83	.47	.40	.24	.19	.43	.52	.87	.89	.48	.56	.87	.86	.34	.28	.10	.07	.31	.32
	1.00	1.00	1.00[2]	.97	.93	1.00	1.00[2]	.85	.84	.49	.48	.89	.85	1.00	1.00[2]	.95	.95	1.00	1.00	.87	.88	.41	.36	.88	.89
	5.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	1.00	.93	.92	1.00[2]	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00[2]	1.00	.95	.95	1.00	1.00[2]
	10.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00[2]	1.00	1.00
	15.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
p_3	.01	.51	.50	.41	.43	.69	.71	.07	.08	.04	.05	.15	.19	.45	.53	.30	.34	.72	.70	.04	.06	.01	.01	.11	.13
	.10	.94	.90	.87	.85	.99	1.00[2]	.53	.46	.27	.29	.75	.75	.98	.97	.89	.93	1.00	1.00	.46	.45	.18	.21	.86	.80
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.98	.98	.96	.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00[2]	1.00[2]	1.00[2]	1.00	1.00
	5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	15.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
p_4	.01	.74	.75	.52	.60	.72	.74	.21	.21	.11	.11	.23	.22	.82	.79	.71	.66	.82	.80	.19	.25	.12	.11	.23	.30
	.10	.99	1.00[2]	.99	1.00[2]	1.00	1.00	.91	.89	.71	.73	.88	.91	1.00	1.00	1.00	1.00	1.00	1.00	.93	.96	.80	.88	.91	.96
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	10.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	15.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[1] $|APCII-EPCII|/EPCII \times 100$ for II = 11, 12, and 22; APC denoting approximated posterior covariance; and EPC denoting exact posterior covariance; note that EPCII was never zero in Design 1. CIJ is the I,Jth element of the posterior covariance matrix.

[2] .995 rounded

TABLE 6.9

MEAN SQUARED ERRORS FOR RISK STUDY. DESIGN 1.

Dirichlet Prob.	Estimator	SS=25 PID=0	SS=25 PID=15 r1	SS=25 PID=15 r2	SS=25 PID=40 r1	SS=25 PID=40 r2	SS=50 PID=0	SS=50 PID=15 r1	SS=50 PID=15 r2	SS=50 PID=40 r1	SS=50 PID=40 r2
p_1	APMR0	.80-3	.81-3(.42-4)	.82-3(.31-4)	.82-3(.47-4)	.80-3(.75-4)	.55-3	.54-3(.23-4)	.59-3(.28-4)	.64-3(.37-4)	.63-3(.49-4)
	PMDR0	*	.60-3(.17-4)	.57-3(.88-5)	.58-3(.12-4)	.62-3(.28-4)	*	.52-3(.19-4)	.50-3(.18-4)	.58-3(.20-4)	.56-3(.26-4)
	MLE	.16-2	.17-2(.88-4)	.17-2(.66-4)	.19-2(.12-3)	.19-2(.17-3)	.79-3	.81-3(.34-4)	.88-3(.43-4)	.11-2(.63-4)	.10-2(.81-4)
	APMR1	.85-2	.95-2(.30-3)	.96-2(.28-3)	.12-1(.39-3)	.12-1(.51-3)	.27-2	.30-2(.13-3)	.33-2(.14-3)	.38-2(.19-3)	.40-2(.20-3)
	APMR2	.33-2	.35-2(.14-3)	.36-2(.13-3)	.41-2(.16-3)	.40-2(.22-3)	.14-2	.15-2(.76-4)	.17-2(.82-4)	.18-2(.11-3)	.19-2(.12-3)
	PMDR2	*	.86-3(.47-4)	.87-3(.31-4)	.81-3(.39-4)	.92-3(.67-4)	. *	.53-3(.27-4)	.55-3(.29-4)	.70-3(.35-4)	.67-3(.54-4)
p_2	APMR0	.69-2	.71-2(.22-3)	.67-2(.21-3)	.80-2(.40-3)	.70-2(.37-3)	.47-2	.51-2(.15-3)	.50-2(.15-3)	.67-2(.33-3)	.59-2(.27-3)
	PMDR0	.11-1	.12-1(.63-3)	.11-1(.67-3)	.15-1(.83-3)	.15-1(.76-3)	.61-2	.65-2(.30-3)	.68-2(.32-3)	.93-2(.49-3)	.91-2(.47-3)
	MLE	.14-1	.15-1(.49-3)	.14-1(.44-3)	.20-1(.11-2)	.18-1(.94-3)	.68-2	.76-2(.22-3)	.75-2(.22-3)	.11-1(.54-3)	.99-2(.45-3)
	APMR1	.15-1	.15-1(.98-3)	.16-1(.86-3)	.20-1(.15-2)	.17-1(.14-2)	.71-2	.82-2(.35-3)	.76-2(.40-3)	.11-1(.69-3)	.91-2(.56-3)
	APMR2	.94-2	.95-2(.61-3)	.98-2(.53-3)	.11-1(.80-3)	.99-2(.78-3)	.56-2	.62-2(.27-3)	.58-2(.31-3)	.80-2(.50-3)	.65-2(.40-3)
	PMDR2	.84-2	.87-2(.30-3)	.83-2(.54-3)	.10-1(.26-3)	.91-2(.48-3)	.53-2	.57-2(.16-3)	.56-2(.17-3)	.77-2(.37-3)	.68-2(.31-3)
p_3	APMR0	.13-1	.14-1(.46-3)	.13-1(.47-3)	.14-1(.73-3)	.14-1(.82-3)	.86-2	.96-2(.33-3)	.94-2(.37-3)	.11-1(.58-3)	.10-1(.42-3)
	PMDR0	.16-1	.18-1(.64-3)	.17-1(.63-3)	.19-1(.90-3)	.18-1(.11-2)	.97-2	.11-1(.36-3)	.11-1(.43-3)	.13-1(.71-3)	.12-1(.50-3)
	MLE	.25-1	.29-1(.10-2)	.28-1(.11-2)	.37-1(.19-2)	.36-1(.20-2)	.12-1	.14-1(.50-3)	.14-1(.57-3)	.19-1(.99-3)	.17-1(.71-3)
	APMR1	.20-1	.23-1(.85-3)	.23-1(.92-3)	.27-1(.15-2)	.27-1(.16-2)	.11-1	.13-1(.50-3)	.12-1(.54-3)	.16-1(.85-3)	.14-1(.64-3)
	APMR2	.15-1	.17-1(.79-3)	.17-1(.84-3)	.18-1(.12-2)	.19-1(.12-2)	.95-2	.11-1(.52-3)	.10-1(.52-3)	.12-1(.71-3)	.11-1(.59-3)
	PMDR2	.16-1	.18-1(.75-3)	.18-1(.81-3)	.20-1(.12-2)	.20-1(.13-2)	.99-2	.12-1(.48-3)	.11-1(.50-3)	.13-1(.73-3)	.12-1(.57-3)
p_4	APMR0	.14-1	.15-1(.48-3)	.14-1(.44-3)	.15-1(.73-3)	.15-1(.68-3)	.93-2	.92-2(.30-3)	.10-1(.31-3)	.12-1(.55-3)	.11-1(.64-3)
	PMDR0	.16-1	.18-1(.58-3)	.17-1(.54-3)	.19-1(.93-3)	.18-1(.86-3)	.10-1	.10-1(.33-3)	.11-1(.35-3)	.14-1(.64-3)	.13-1(.74-3)
	MLE	.27-1	.32-1(.11-2)	.29-1(.10-2)	.38-1(.19-2)	.36-1(.17-2)	.13-1	.14-1(.45-3)	.15-1(.47-3)	.21-1(.95-3)	.19-1(.11-2)
	APMR1	.21-1	.24-1(.80-3)	.23-1(.75-3)	.28-1(.14-2)	.26-1(.12-2)	.12-1	.12-1(.39-3)	.13-1(.41-2)	.17-1(.79-3)	.16-1(.91-3)
	APMR2	.16-1	.18-1(.90-3)	.17-1(.76-3)	.20-1(.11-2)	.19-1(.11-2)	.10-1	.11-1(.41-3)	.11-1(.43-3)	.14-1(.68-3)	.13-1(.82-3)
	PMDR2	.19-1	.22-1(.11-2)	.21-1(.93-3)	.25-1(.14-2)	.24-1(.13-2)	.11-1	.12-1(.46-3)	.12-1(.48-3)	.16-1(.80-3)	.15-1(.94-3)

Note that values are given in scientific notation; e.g., .00080 is written as .80-3. Values in parenthesis are standard errors.

For PID=15 and 40, mean squared error is estimated by the regression estimate [see Section 5.9]

*Values were not calculated (see main text)

TABLE 6.10A

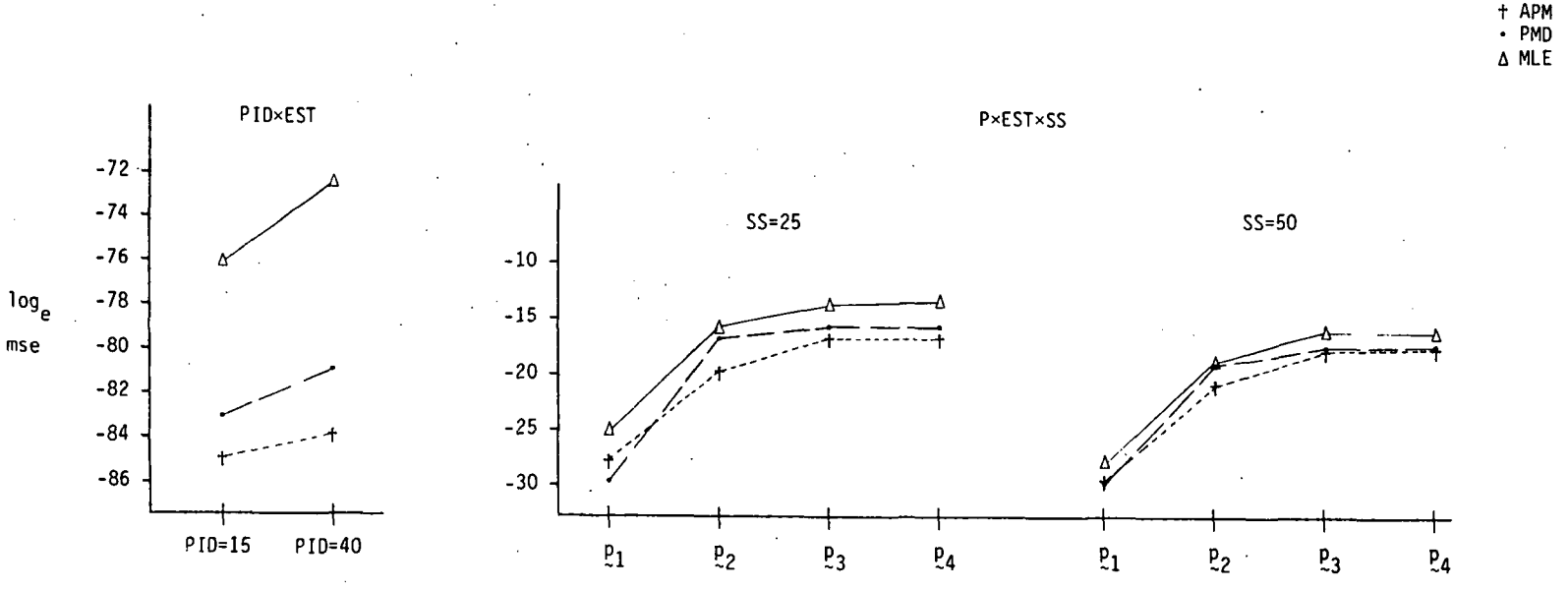
ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED RISK

FOR ROBUSTNESS SET 0

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
P	3	.147185 3	.490617 2	22,460.893 ***
SS	1	.474330 1	.474330 1	2,171.526 ***
PID	1	.645378 0	.645678 0	295.460 ***
EST	2	.722603 1	.361302 1	1,654.071 ***
P×SS	3	.697938 -1	.232646 -1	10.651 ***
P×PID	3	.717094 -1	.239031 -1	10.943 ***
P×EST	6	.120826 1	.201377 0	92.192 ***
SS×PID	1	.535130 -1	.535130 -1	24.499 ***
SS×EST	2	.628491 0	.314245 0	143.864 ***
PID×EST	2	.768419 -1	.384210 -1	17.589 ***
P×SS×PID	3	.883219 -2	.294406 -2	1.348
P×SS×EST	6	.155608 0	.259346 -1	11.873 ***
P×PID×EST	6	.824202 -2	.137367 -2	.629
SS×PID×EST	2	.433219 -2	.216610 -2	.992
P×SS×PID×EST	6	.127261 -2	.212102 -3	.097
ERROR	48	.104847 0	.218432 -2	
TOTAL	95	.162192 3		

*** Significant at 1% level.

6.10B PLOTS OF PID×EST AND P×EST×SS INTERACTIONS*



*Values are sums over nonpresent factors, including replication.

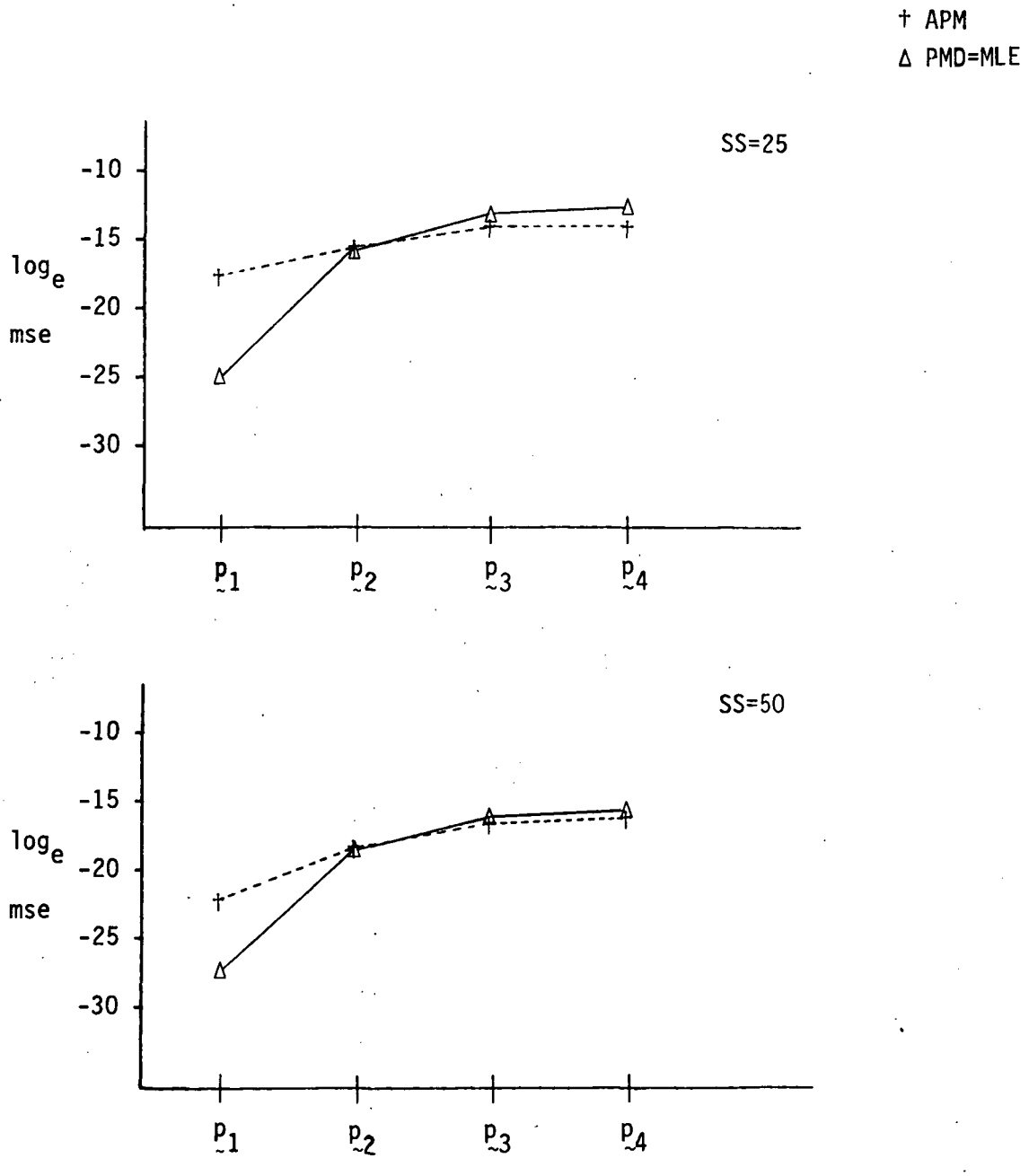
TABLE 6.11A

ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED RISK FOR ROBUSTNESS SET 1

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
P	3	.427955 2	.142652 2	4967.065 ***
SS	1	.779571 1	.779571 1	2714.429 ***
PID	1	.766894 0	.766894 0	267.029 ***
EST	1	.129177 1	.129177 1	449.789 ***
P×SS	3	.196884 0	.656280 -1	22.851 ***
P×PID	3	.115092 -1	.383639 -2	1.336
P×EST	3	.866720 1	.288907 1	1005.959 ***
SS×PID	1	.107157 -1	.107157 -1	3.731 *
SS×EST	1	.122873 -1	.122873 -1	4.278 **
PID×EST	1	.570937 -2	.570937 -2	1.988
P×SS×PID	3	.802564 -2	.267521 -2	.931
P×SS×EST	3	.244437 0	.814791 -1	28.371 ***
P×PID×EST	3	.131641 -1	.438803 -2	1.528
SS×PID×EST	1	.326823 -3	.326823 -3	.114
P×SS×PID×EST	3	.523357 -2	.174452 -2	.607
ERROR	32	.919024 -1	.287195 -2	
TOTAL	63	.619173 2		

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

6.11B PLOT OF P×EST×SS INTERACTION*



*Values are sums over PID and replication

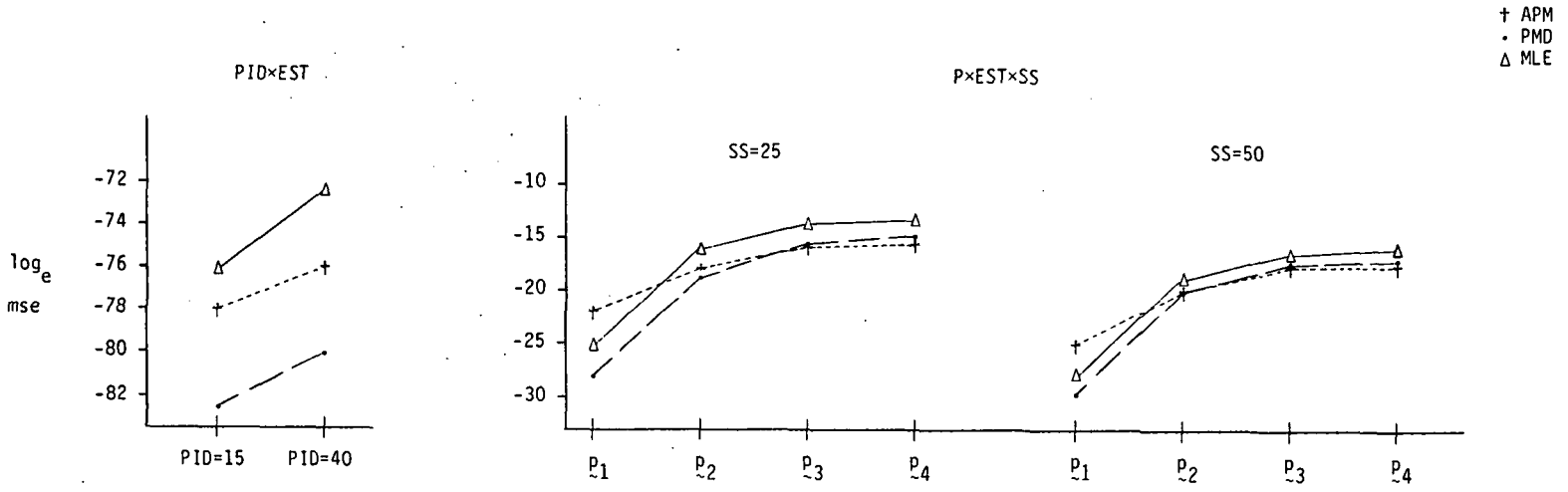
TABLE 6.12A

ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED RISK FOR ROBUSTNESS SET 2

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
P	3	.106983 3	.356610 2	13981.332 ***
SS	1	.697137 1	.697137 1	2733.214 ***
PID	1	.755904 0	.755904 0	296.361 ***
EST	2	.320158 1	.160079 1	627.610 ***
P×SS	3	.577776 -1	.192592 -1	7.551 ***
P×PID	3	.266412 -1	.888039 -2	3.482 **
P×EST	6	.613926 1	.102321 1	401.162 ***
SS×PID	1	.422372 -1	.422372 -1	16.560 ***
SS×EST	2	.249801 0	.124901 0	48.969 ***
PID×EST	2	.449282 -1	.224641 -1	8.807 ***
P×SS×PID	3	.118694 -1	.395645 -2	1.551
P×SS×EST	6	.198784 0	.331307 -1	12.989 ***
P×PID×EST	6	.121991 -1	.203318 -2	.797
SS×PID×EST	2	.569427 -2	.284714 -2	1.116
P×SS×PID×EST	6	.943908 -2	.157318 -2	.617
ERROR	48	.122429 0	.255061 -2	
TOTAL	95	.124833 3		

** Significant at 5% level
*** Significant at 1% level

6.12B PLOTS OF PID×EST AND P×EST×SS INTERACTIONS*



*Values are sums over nonpresent factors, including replication.

TABLE 6.13

RATIO OF MSE(APM) TO MSE(PMD) AND MSE(MLE) FOR QUADRATIC-LOSS COMPARISONS.

Sample Size		SS=25				SS=50			
% Inc. Data		PID=15		PID=40		PID=15		PID=40	
Replic. No.		r1	r2	r1	r2	r1	r2	r1	r2
$\underset{\sim}{P}$†	Estimator								
		A. ROBUSTNESS SET 0 (ORIGINAL PRIOR USED IN BAYESIAN ESTIMATORS)							
$\underset{\sim}{P}_1$	PMD	.14 1	.14 1	.14 1	.13 1	.10 1	.12 1	.11 1	.11 1
	MLE	.47 0	.48 0	.43 0	.41 0	.67 0	.66 0	.60 0	.60 0
$\underset{\sim}{P}_2$	PMD	.58 0	.60 0	.52 0	.48 0	.78 0	.73 0	.73 0	.65 0
	MLE	.48 0	.48 0	.39 0	.39 0	.67 0	.67 0	.61 0	.60 0
$\underset{\sim}{P}_3$	PMD	.80 0	.80 0	.75 0	.77 0	.89 0	.87 0	.84 0	.85 0
	MLE	.47 0	.47 0	.39 0	.40 0	.66 0	.66 0	.60 0	.60 0
$\underset{\sim}{P}_4$	PMD	.82 0	.82 0	.79 0	.79 0	.89 0	.89 0	.87 0	.87 0
	MLE	.47 0	.47 0	.40 0	.40 0	.67 0	.67 0	.59 0	.60 0
		B. ROBUSTNESS SET 1 (UNIFORM PRIOR USED IN BAYESIAN ESTIMATORS)							
$\underset{\sim}{P}_1$	PMD*	.55 1	.56 1	.64 1	.63 1	.37 1	.38 1	.36 1	.38 1
$\underset{\sim}{P}_2$	PMD	.10 1	.11 1	.98 0	.95 0	.11 1	.10 1	.10 1	.92 0
$\underset{\sim}{P}_3$	PMD	.79 0	.80 0	.73 0	.76 0	.90 0	.88 0	.85 0	.86 0
$\underset{\sim}{P}_4$	PMD	.77 0	.77 0	.72 0	.72 0	.88 0	.88 0	.84 0	.85 0
		C. ROBUSTNESS SET 2 (PERTURBED PRIOR USED IN BAYESIAN ESTIMATORS)							
$\underset{\sim}{P}_1$	PMD	.41 1	.41 1	.50 1	.44 1	.28 1	.31 1	.25 1	.28 1
	MLE	.20 1	.21 1	.21 1	.21 1	.18 1	.19 1	.17 1	.18 1
$\underset{\sim}{P}_2$	PMD	.11 1	.12 1	.11 1	.11 1	.11 1	.10 1	.10 1	.96 1
	MLE	.64 0	.69 0	.55 0	.55 0	.82 0	.78 0	.72 0	.66 0
$\underset{\sim}{P}_3$	PMD	.92 0	.92 0	.90 0	.92 0	.95 0	.94 0	.92 0	.94 0
	MLE	.56 0	.59 0	.49 0	.52 0	.76 0	.71 0	.65 0	.68 0
$\underset{\sim}{P}_4$	PMD	.82 0	.82 0	.79 0	.79 0	.89 0	.89 0	.87 0	.87 0
	MLE	.58 0	.58 0	.52 0	.53 0	.76 0	.72 0	.68 0	.70 0

*For uniform prior, PMD=MLE

†Dirichlet probability

TABLE 6.14

ESTIMATOR HAVING SMALLEST AVERAGE ESTIMATED MEAN SQUARED ERROR FOR QUADRATIC-LOSS COMPARISON.

Sample Size	SS=25				SS=50			
% Inc. Data	PID=15		PID=40		PID=15		PID=40	
Replic. No.	r1	r2	r1	r2	r1	r2	r1	r2
Dir. Prob.								
	A. ROBUSTNESS SET 0 (ORIGINAL PRIOR IN ESTIMATORS)							
$\underset{\sim}{p}_1$	pmd	pmd	pmd	pmd	pmd	pmd	pmd	pmd
$\underset{\sim}{p}_2$	apm	apm	apm	apm	apm	apm	apm	apm
$\underset{\sim}{p}_3$	apm	apm	apm	apm	apm	apm	apm	apm
$\underset{\sim}{p}_4$	apm	apm	apm	apm	apm	apm	apm	apm
	B. ROBUSTNESS SET 1 (UNIFORM PRIOR IN ESTIMATORS)							
$\underset{\sim}{p}_1$	pmd[1]	pmd	pmd	pmd	pmd	pmd	pmd	pmd
$\underset{\sim}{p}_2$[2]	pmd	pmd	apm	apm	pmd	pmd	apm	apm
$\underset{\sim}{p}_3$	apm	apm	apm	apm	apm	apm	apm	apm
$\underset{\sim}{p}_4$	apm	apm	apm	apm	apm	apm	apm	apm
	C. ROBUSTNESS SET 2 (PERTURBED PRIOR IN ESTIMATORS)							
$\underset{\sim}{p}_1$	pmd	pmd	pmd	pmd	pmd	pmd	pmd	pmd
$\underset{\sim}{p}_2$[2]	pmd	pmd	pmd	pmd	pmd	pmd	pmd	apm
$\underset{\sim}{p}_3$	apm	apm	apm	apm	apm	apm	apm	apm
$\underset{\sim}{p}_4$	apm	apm	apm	apm	apm	apm	apm	apm

[1] pmd = mle for uniform prior

[2] pmd and apm are nearly equal for all conditions for $\underset{\sim}{p}_2$ for Robustness sets 1 and 2. Recall Table 6.13.

CHAPTER 7

RESULTS OF DESIGN 2

7.1 Introduction:

In this chapter, we report results from Design 2. We want to know whether risk results from Design 1 depend on the very special choice used there for the trinomial generator probabilities. Recall that each of the four trinomial generators was the mean of a prior Dirichlet distribution. This chapter reports what happens when, instead, we choose probabilities randomly generated from the Dirichlet distribution for these trinomial generators. [See Figures 5.1 - 5.3 for a comparison of Design 2 with Design 1.]

Note that, except for a brief discussion in the next section, we do not report work on the Taylor-series approximations. Results of Design 1 show that risk conclusions depend on the value of the generator $\underset{\sim}{p}$. However, the accuracy of the Taylor-series approximations, although depending slightly on the value of the generator $\underset{\sim}{p}$, was good for all values of $\underset{\sim}{p}$. Rare exceptions occurred at some of those boundary values that gave empty cells for the completely specified data when the percentage of incomplete data was high. Although some of the calculations discussed in Section 6.4 for Design 1 were repeated for Design 2, results were identical to those already reported.

Other than the generator probabilities $\underset{\sim}{p}$, factors in Design 2 were the same as those in Design 1. There were four values of the prior parameter $\underset{\sim}{\nu}$: $\underset{\sim}{\nu}_1 \equiv (.1,.1,9.8)$, $\underset{\sim}{\nu}_2 \equiv (1,1,8)$, $\underset{\sim}{\nu}_3 \equiv (2,3,5)$, and $\underset{\sim}{\nu}_4 \equiv (10/3,10/3,10/3)$. Sample sizes were SS=25 and SS=50. The percentage of

incomplete data PID varied around PID=15 and PID=40. The three estimators were the posterior mean (approximated by the Taylor-series expansion), the posterior mode, and the maximum likelihood estimate.

As in Design 1, we had three robustness studies, one each for use in the Bayesian estimates of the original prior $\underset{\sim}{\nu}$, the uniform prior (1,1,1), and the perturbed prior $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$. Note, again, that the maximum likelihood estimate was the same for all three studies.

Recall from Section 5.6 that cost constraints limited to 10 the number of Dirichlet generations of $\underset{\sim}{p}$ given each of the four values of $\underset{\sim}{\nu}$. Values of these probabilities, generated by the procedures described in Section 5.7.2, are given in Table 7.1. As expected, the generated values varied around the means (.01,.01,.98), (.10,.10,.80), (.20,.30,.50), and (.33,.33,.33) of the prior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}_1$, $\underset{\sim}{\nu}_2$, $\underset{\sim}{\nu}_3$, and $\underset{\sim}{\nu}_4$, respectively. Table 7.1 also gives the centrality measure $C(\underset{\sim}{p})$ for each generated value of $\underset{\sim}{p}$. In Design 1, this centrality measure became the basis for deciding which estimator to use for minimizing risk. Recall from Table 5.1 that centrality measures for the prior means of the distribution of $\underset{\sim}{p}$ given the four values of $\underset{\sim}{\nu}$ are 1.88, .98, .14, and 0, respectively. Note, then, in Table 7.1 that centrality measures for $\underset{\sim}{\nu}_1$ ranged from 1.39 to 2.00 (the highest possible value). Those for $\underset{\sim}{\nu}_2$ ranged from .06 to 1.94; for $\underset{\sim}{\nu}_3$, from .05 to 1.05; and, for $\underset{\sim}{\nu}_4$, from 0 to .38. Centrality measures for the prior mean of the distribution of $\underset{\sim}{p}$ given the four values of the perturbed prior $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$ are 1.16, .48, .01, and .09. [Recall Figure 6.2 for perturbed-prior means.]

Results from the three robustness studies are reported in the next section and concluded in the last section. Appendix 7A gives the complete-data (PID=0) risks and estimated risks (with associated standard errors) for PID=15 and PID=40 for the three estimators, three robustness studies, four values of $\underset{\sim}{\nu}$, and ten Dirichlet probabilities $\underset{\sim}{p}$. Tukey data summaries, central values, and spreads were calculated for the risk estimates over the ten Dirichlet generations. The averages, with standard errors, are given in Table 7A.7. We note here that the posterior mean had the smallest average, even at $\underset{\sim}{\nu}_1$, when the original prior was used in the Bayesian estimators. This result is important because it means that the sampling distribution, even though based on only ten probabilities, agreed with the theoretical distribution at least in terms of which estimator minimized average risk. [Recall Section 1.2.] Appendix 7A also gives the analyses-of-variance results for the three robustness studies and plots of two of their interactions.

In the remainder of this section, we briefly discuss computational aspects peculiar to Design 2. Since we were investigating which estimator best minimized risk for quadratic loss, the criterion for choosing among estimators was the estimated mean squared error (risk). After we discussed estimated mean squared error in Design 1, we studied the estimators in detail, especially at those values of $\underset{\sim}{p}$ for which two or more estimators had approximately equal risk. In Design 2, we studied only estimated mean squared error and results from the analysis of variance on its natural logarithms.

In Design 1, we used the regression estimate for the mean squared error. Where we could, we also used the regression estimate in Design 2.

However, we could not calculate the regression estimate for some cases when the prior was $\underset{\sim}{\nu}_1 \equiv (.1,.1,9.8)$. In these cases, the complete-data maximum likelihood estimate was the same for all 200 trinomial simulations. Hence, its sample variance was zero. Therefore, the denominator for the regression mean-squared-error estimate in (5.19) was undefined.

The problem cases were those in which $C(\underset{\sim}{p})$ was 1.9993, 1.9999, and 2.0000. These three cases were those in which the generated Dirichlet probabilities were approximately (0,0,1). The probabilities were (.00000 004,.00011,.99989), (.00000 9,.00000 00000 0003,.99999 1), and (.00000 3,.00000 00000 001,.99999 7), respectively. Note that there were no problems in calculating regression estimates of the mean squared error for the generated Dirichlet probability $\underset{\sim}{p}=(.01,0,.99)$, for which $C(\underset{\sim}{p})=1.9182$. Further, for the case in which $C(\underset{\sim}{p})=1.9993$, the regression estimate was undefined for only half the cases. Hence, it was only when the population probability was almost identically (0,0,1) that the regression estimate did not exist.

In these cases of undefined regression mse estimate, we used the control-variate estimate. However, the control-variate estimate was negative several times for the posterior mode when the generated Dirichlet probability was approximately (0,0,1). Although this happened only in cases where the regression estimate was defined, it happened for a Dirichlet probability which had an undefined regression estimate for most of the SS, PID, and replication variations. Hence, the control-variate estimate was used for most variations and, for consistency, would have been a better choice than the regression estimate for the

remaining cases. [See the fourth generation for PMDRO in Table 7A.3 for inconsistent mean squared errors resulting from use of the two different estimates.] The control-variate estimate was negative because the posterior mode had a regular mse estimate a couple orders of magnitude smaller than either the true or regular-estimate complete-data mean squared error and the regular estimate was larger than the true value. [See equation (5.13).]

The inconsistent mse estimates for PMDRO at $\underset{\sim}{p}=(0,0,1)$ affected results in the ANOVA. Variations among PID and SS levels were as large as 100. This large variation gave rise to unnaturally large effects of SS and, particularly, PID relative to those for estimator. Further, the ANOVA model had an additional factor $\underset{\sim}{\nu}$, ten levels (instead of one) for $\underset{\sim}{p}$ within $\underset{\sim}{\nu}$, and $\underset{\sim}{p}$ as a random factor instead of a fixed factor. Therefore, the ANOVA model was more complicated than that for Design 1. Hence, its results were more subject to error.

Therefore, as a precaution against reaching wrong conclusions, we studied certain interactions, especially the PwNU interaction (PwNU)×SS ×PID×EST, independent of their significant effects in an ANOVA. An additional reason for investigating this particular interaction was that we wanted to insure that any lack of significant effect for PID was accurate. Even more important, we wanted to know how any lack of significant effect related to absence of any change in $\log_e$(mse) for the two levels of PID. That is, PID could show no significant effect in the ANOVA model strictly because the other factors had huge effects relative to PID. In this case, there could still be a large change in $\log_e$(mse) for the two levels of PID.

Recall that in the last chapter we found that the Taylor-series approximation APM for the posterior mean was usually accurate to at least four significant figures. In these cases, the APM mean-squared-error estimate was a good approximation for the EPM mean-squared-error estimate. It was usually accurate to at least three significant figures. The rare cases in which the Taylor-series approximation was not as good, however, were cause of concern for how well the APM mse estimate approximated the EPM mse estimate. These cases occurred several times in Design 2 at $\underset{\sim}{\nu}_1$ when the generated Dirichlet probability was approximately (0,0,1). However, even though a few of the 200 trinomial simulations yielded poor approximations for the posterior mean, the APM mse estimate was an unusually good approximation for the EPM mse estimate. It was nearly always accurate to at least five significant figures. The reason is that in those cases (the majority of the 200 trinomial simulations) in which the approximation was not poor, the approximated posterior mean agreed extremely well with the exact posterior mean. Therefore, the APM mse estimate, an average over the 200 trinomial simulations, was a very good approximation.

7.2 Results:

In this section we briefly give results from Design 2. We begin by giving, in Table 7.2, the risks averaged over the ten Dirichlet generations of $\underset{\sim}{p}$. For estimated risks, those for PID=15 and PID=40, we also averaged over the two replications. Note that averaged risks for PID=0 are not given for the posterior mode (PMD) at $\underset{\sim}{\nu}_1$. As in Design 1, we could not analytically calculate the complete-data risk for values of $\underset{\sim}{p}$ containing one or more very small components when $\underset{\sim}{\nu}=(0.1,0.1,9.8)$. In these cases, a solution to the likelihood equations may not exist in P_2. [Note that $(x_i+\nu_i-1)/(n+\Sigma\nu_j-3)$ is negative when $x_i=0$ if $\nu_i=0.1$.] If not, the posterior mode occurs on the boundary. Hence, $\hat{\tilde{p}}_i$ may equal 0 or 1 but the i^{th} solution (2.43) to the likelihood equation can not be used to calculate the risk.

We are interested in how much risk increases as the data becomes incomplete. Table 7.2 shows that in 34 out of 44 cases, the averaged risk increased between 5% and 12% as the percentage of incomplete data (PID) increased from 0 to 15. The highest increase, 20%, was at a sample size of 50 for $\underset{\sim}{\nu}_3$ for the posterior mean (APM) when the perturbed prior was used in the Bayesian estimators. As the percentage of incomplete data increased from 0 to 40, the averaged risk increased between 17% and 50%. Individual values showed greater variation than the averages given in Table 7.2. Occasionally, the complete-data risk was even greater than the risk when approximately 15% of the data was incomplete. In these cases, however, the complete-data exact value was nearly always within a standard error of the PID=15 estimated value. These cases probably occurred

when the observed percentage of incomplete data was on the low side of 15%. [Recall that, for a sample size of 25, when PID=15 the observed percentage of incomplete data could be 0%, 4%, 8%, 16%, 20%, or 24%,...] Finally, note that, as sample size decreased, the averaged risk decreased by roughly one-half.

To compare the difference between estimators, we divided the averaged risk for the posterior mean by that for the posterior mode and that for the maximum likelihood estimate. Results given in Table 7.3 show that, with small exception, the averaged risk was smallest for the posterior mean for all variations in prior parameter, percentage of incomplete data, and sample sizes when the correct prior was used in the Bayesian estimators. The exception is that the posterior mode had, to two significant figures, the same risk for $\underset{\sim}{\nu}_3$ and almost equal risk for $\underset{\sim}{\nu}_4$. [Recall Table 7.2.] As $\underset{\sim}{p}$ moved from the center of the P_2 simplex toward a corner (from $\underset{\sim}{\nu}_4$ to $\underset{\sim}{\nu}_1$), the advantage in using the posterior mean over the posterior mode increased. The advantage in using the posterior mean over the maximum likelihood estimate was greatest at the center or a corner of P_2. At $\underset{\sim}{\nu}_1$, the risk of the posterior mean was almost one half that for the posterior mode or maximum likelihood estimate for a sample size of 25. For other values of the prior parameter $\underset{\sim}{\nu}$, percentage of incomplete data PID, and sample size SS, the averaged risk of the posterior mean lay between 70% and 100% of that for the posterior mode and maximum likelihood estimate.

When a uniform prior was used in the Bayesian estimators, the posterior mode equaled the maximum likelihood estimate. For this case, results of Table 7.3 show that, in terms of averaged risk, the posterior

mean was the best estimator at $\underset{\sim}{\nu}_3$ and $\underset{\sim}{\nu}_4$, in the middle and near the center of P_2. The maximum likelihood estimate (≡posterior mode) was the best estimator near a boundary of P_2; i.e., at $\underset{\sim}{\nu}_1$ and $\underset{\sim}{\nu}_2$. The maximum difference near the center of P_2 was only 70% (relative to the smallest value). At a boundary, however, the averaged risk was between two and three times smaller for the maximum likelihood estimate (≡posterior mode) than for the posterior mean.

When the perturbed prior $10\times[\underset{\sim}{\nu}/10+(.09,.05,-.14)]$ was used in the Bayesian estimators, the posterior mode had the smallest averaged risk, except at the center of the P_2 simplex, where the posterior mean was slightly better. The largest difference between estimators was at $\underset{\sim}{\nu}_1$ where the risk of the posterior mean was 40% larger than that for the posterior mode.

Note that for all three priors (correct, uniform, and perturbed), there was very little difference between estimators as the percentage of incomplete data changed. As sample size increased, the ratios moved toward 1; i.e., the difference between estimators decreased.

As discussed in Chapter 1, however, we were most interested in difference in risk as a function of the individual values of $\underset{\sim}{p}$. To investigate this relationship, we first performed an analysis of variance on the natural logarithms of the estimated mean squared errors (risks). The F values from these analyses are given in Tables 7.8A, 7.9A, and 7.10A for use of the correct, uniform, and perturbed prior (robustness study R0, R1, and R2), respectively, in the Bayesian estimators. By far the most significant effect in all three ANOVAs was that of $\underset{\sim}{p}$ within $\underset{\sim}{\nu}$

(Pw.NU), usually followed by (Pw.NU)×EST. Further, in all three robustness studies, (Pw.NU)×EST×SS was significant at the 1% level and there was a three-way (×PID) or four-way (×PID×SS) significant interaction of (Pw.NU)×EST with PID in each analysis.

Following the ANOVA tables in Appendix 7A are plots of significant or otherwise important (recall Introduction) interactions. These plots indicated that there was little change in the difference between estimators as the percentage of incomplete data (PID) increased from 15 to 40. Further, although the difference between estimators decreased as sample size increased, the shape of the estimator curves for the two sample sizes was nearly the same. Therefore, we summarize results from these analyses by giving in Tables 7.4, 7.5, and 7.6 plots of the (Pw.NU)×SS ×PID×EST interactions for only SS=25 and PID=15. Note that the horizontal axis is the centrality measure of the generated $\underset{\sim}{p}$. The vertical axis is $\log_e(\text{risk})$ [$\equiv\log_e(\text{mse})$]. Recall from Chapter 6 that, because we used two replications, the square root of the exponential of a difference between logarithms approximately equals the ratio of the risk of the two estimators. Thus, any difference of 6 between two estimators in the $\log_e$ scale in Plots 7.4, 7.5, and 7.6 means that one of the two estimators had a risk about twenty times larger than that of the other estimator.

There are three important factors to consider in these three plots: the distribution from which the generated $\underset{\sim}{p}$ comes, the value of the generated $\underset{\sim}{p}$, and the value of the prior parameters used in the Bayesian estimators. In all three plots, the distribution from which $\underset{\sim}{p}$ comes is the Dirichlet distribution given the prior $\underset{\sim}{\nu}$. The centrality measure of the mean of this distribution is marked on the three plots by the arrow

(↑) for the four values of $\underset{\sim}{\nu}$. We call this prior mean the *ν-prior mean* or the correct-prior mean. We call the mean of the prior distribution given the prior parameters used in the Bayesian estimators the *estimator-prior mean*. The centrality measure of the estimator-prior mean is marked on the plots by an "x" when, in Plots 7.5 and 7.6, it differs from the $\underset{\sim}{\nu}$-prior mean. [Recall Figure 6.2 for estimator-prior means.]

Notice that the closer the $\underset{\sim}{\nu}$-prior mean is to a corner or to the center of the P_2 simplex, the tighter the distribution of the generated values of $\underset{\sim}{p}$. Away from these points, the distribution is fairly wide; for example, the distribution of $\underset{\sim}{p}$ given $\underset{\sim 2}{\nu}$ covers almost the entire $C(\underset{\sim}{p})$ axis.

Denote the estimator-prior mean by $\underset{\sim}{\bar{p}}$. Plots 7.4 - 7.6 show that, except for $\underset{\sim 1}{\nu}$, there is a neighborhood of $C(\underset{\sim}{\bar{p}})$ in which the posterior mean is the best estimator for minimizing risk, often followed by an outer one-sided neighborhood toward 2.00 in which the posterior mode is best. Finally, in the tails of the distribution of $\underset{\sim}{p}$ given the prior parameters used in the Bayesian estimators, the maximum likelihood estimate is best.

Thus, the posterior mean was the best estimator most of the time. In these cases, the posterior mode was usually next best. Other than cross-over probabilities, the smallest difference between the posterior mode and mean was at the center of the P_2 simplex. There, the risk of the posterior mean was reduced only 14% to 23% from that of the posterior mode, whereas it was reduced 22% to 42% from that of the maximum likelihood estimate.

Except near the tails of the estimator-prior distribution or near the $\underset{\sim}{p}=(0,0,1)$ corner of P_2, the difference in $\log_e(\text{mse})$ for the estimators was usually between -1.4 and 0.8; difference in risk ranged from 0 to a 50% decrease. Use of the correct estimator most often reduced the risk by about one-third. At the tails, the maximum difference between $\log_e(\text{mse})$ for the three estimators ranged from 0.8 (1/3 increase in risk) in Table 7.4 at $C(\underset{\sim}{p})=0.06$ for $\underset{\sim}{\nu}_2$ to 1.9 (risk almost tripled) at $C(\underset{\sim}{p})=1.05$ for $\underset{\sim}{\nu}_3$ in Table 7.6 to 5.6 (risk increased more than 16 times) at $C(\underset{\sim}{p})=1.94$ for $\underset{\sim}{\nu}_2$ in Table 7.6. However, the largest difference between estimators occurred for $\underset{\sim}{\nu}_1$ at the corner $\underset{\sim}{p}=(0,0,1)$ where $C(\underset{\sim}{p})=2.00$. At this probability $\underset{\sim}{p}$, values of $\log_e(\text{mse})$ for the maximum likelihood estimate and posterior mode were equal. The large difference in $\log_e(\text{mse})$ between this value and that for the posterior mean was 10.6, 20.8, and 18.9 for use in the Bayesian estimators of the correct, uniform, and perturbed prior, respectively. These differences correspond to an increase in risk of 200 times, 33,000 times, and 13,000 times the risk for the maximum likelihood estimate or posterior mode. Note, however, that this enormous difference occurred only exactly at the (0,0,1) corner. For example, the probability $\underset{\sim}{p}=(.4^{-7},.1^{-3},.99989)$ also had, rounded off, $C(\underset{\sim}{p})=2.00$ but the multiplicative increase in risk in using the posterior mean instead of the posterior mode was by a factor of 77.5, 992, and 854, respectively, for the three robustness studies. Thus, the increase was huge but not of the order found when the first two components had more zeros. As $\underset{\sim}{p}$ moved further from the (0,0,1) corner, the difference in risks continued to drop sharply.

In Figure 7.1 we give the ranges for each value of the estimator-prior in which the posterior mean, posterior mode, and maximum likelihood estimate was best. Note that the limits sometimes differ slightly from the plots. In these cases, the difference between the limit and the correct value was small. We used the wrong value to give limits in .05 increments and to give agreement between slightly different values for the limits for those estimator-prior means having centrality measures of 0 and .01. [Recall that the estimator-prior mean for all four plots in Table 7.5 (use of the uniform prior) has centrality measure $C(\bar{\underset{\sim}{p}})$ of 0 as well as one estimator-prior mean in Table 7.4.] Note in Figure 7.1 that, for the uniform prior, the region in which the posterior mean is best is $0 \leq C(\underset{\sim}{p}) < .70$. Also note that the posterior-mode range for $C(\bar{\underset{\sim}{p}})$ =0.09 was unusually short; that for the maximum likelihood estimate began sooner than results from neighboring values of $C(\bar{\underset{\sim}{p}})$ would indicate.

Results from Design 2 indicate that if one is even reasonably confident in the prior, then the best estimator to use is the posterior mean unless the prior mean is at the corner of the P_2 simplex, in which case the posterior mode is better. Hence, we recommend, for an initial try, use of the posterior mean if $C(\bar{\underset{\sim}{p}}) \leq 1.5$; the posterior mode, otherwise.

In practice, one can replace $\underset{\sim}{p}$ in Figure 7.1 by the estimator $\dot{\underset{\sim}{p}}$ and interpolate in the intervals in Figure 7.1 to refine the estimation process. That is, if one uses the prior $\underset{\sim}{\beta}$ with prior mean $\bar{p}_i = \beta_i / \Sigma \beta_j$ in an estimator $\dot{\underset{\sim}{p}}$, then one can compare $C(\dot{\underset{\sim}{p}})$ with the regions given for $C(\bar{\underset{\sim}{p}})$ to determine if the best estimator was used. If not, then $\dot{\underset{\sim}{p}}$ can

FIGURE 7.1

INTERPOLATION TABLE

if $C(\tilde{p})$ was	and $C(p)$ was	then best estimator was
0[a] and .01[b]	$0 \leq C(p) < .20$	posterior mean
	$.20 \leq C(p) < .70$	posterior mode[c]
	otherwise	maximum likelihood estimate
.09[b]	$0 \leq C(p) < .20$	posterior mean
	$.20 \leq C(p) < .35$	posterior mode[d]
	otherwise	maximum likelihood estimate
.14[a]	$0 \leq C(p) < .45$	posterior mean
	$.45 \leq C(p) < 1.05$	posterior mode
	otherwise	maximum likelihood estimate
.48[b]	$0 \leq C(p) < .85$	posterior mean
	$.85 \leq C(p) < 1.60$	posterior mode
	otherwise	maximum likelihood estimate
.98[a]	$0 \leq C(p) < .10$	maximum likelihood estimate
	$.10 \leq C(p) < 1.45$	posterior mean
	$1.45 \leq C(p) < 1.94$	posterior mode
	$[[1.94 \leq C(p) \leq 2.00$	posterior mode]][e]
1.16[b]	$[[.0 \leq C(p) < .30$	maximum likelihood estimate]][e]
	$[[.30 \leq C(p) < 1.25$	posterior mean]][e]
	$1.25 \leq C(p) < 1.55$	posterior mean
	$1.55 \leq C(p) \leq 2.00$	posterior mode
1.88[a]	$[[0 \leq C(p) < .90$	maximum likelihood estimate]][e]
	$[[.90 \leq C(p) < 1.25$	posterior mean]][e]
	$1.25 \leq C(p) < 1.55$	posterior mean
	$1.55 \leq C(p) \leq 2.00$	posterior mode

[a] See plots in Table 7.4

[b] See plots in Table 7.6

[c] When uniform prior was used in Bayesian estimators, best estimator was the posterior mean instead of the posterior mode; see Table 7.5

[d] Risk of posterior mode differs little from that of posterior mean or max. likelihood est.

[e] extrapolated

be discarded and the recommended estimator used. For example, if one has a prior $\underset{\sim}{\beta}=(12,1,2)$, then the estimator-prior mean $\bar{\underset{\sim}{p}}$ is (.80,.07,.13) which has centrality measure $C(\bar{\underset{\sim}{p}})=(.80-.07)^2+(.80-.13)^2+(.07-.13)^2=.99$. Results of Figure 7.1 indicate that if use of the posterior mean gives an estimator $\dot{\underset{\sim}{p}}$ with $C(\dot{\underset{\sim}{p}})$ between .10 and 1.45, then the posterior mean is the best estimator to use. If, however, $C(\dot{\underset{\sim}{p}})$ is greater than 1.45, then we should discard the posterior mean and use the posterior mode. Similarly, if $C(\dot{\underset{\sim}{p}})$ is less than .10, we should replace the posterior mean by the maximum likelihood estimate.

Note that results of Designs 1 and 2 indicate that the maximum likelihood estimate, posterior mode, and posterior mean will usually be close enough that their centrality measures will differ little. That is, $C(\dot{\underset{\sim}{p}})$ should not differ greatly for the three estimators. Finally, we emphasize that the regions in Figure 7.1 are not exact. Further, replacing $\underset{\sim}{p}$ by the estimator $\dot{\underset{\sim}{p}}$ in Figure 7.1 makes the regions even less exact. Hence, regions in Figure 7.1 should be considered only as rough guidelines. Even so, their use can still be expected to reduce risk by 1/4 to 1/2 in most cases and by substantially more in many cases.

7.3 Conclusions:

Based on results from Design 2 summarized in the last section, we revised the operating guideline from Design 1 as shown in the following Figure 7.2:

FIGURE 7.2

OPERATING GUIDELINES

Given data $\underset{\sim}{z}$ and prior parameter $\underset{\sim}{\nu}$

Calculate prior mean $\underset{\sim}{\bar{p}}$ with component $\bar{p}_i = \nu_i / \Sigma \nu_j$

Calculate $C(\underset{\sim}{\bar{p}}) = \sum_{i=1}^{k} \sum_{j>i}^{k+1} (\bar{p}_i - \bar{p}_j)^2$

if	for estimator $\underset{\sim}{\dot{p}}$ to minimize risk, use
$0 \leq C(\underset{\sim}{\bar{p}}) < 1.50$	posterior mean (Taylor-series approx.)
$1.50 \leq C(\underset{\sim}{\bar{p}}) \leq 2.00$	posterior mode

calculate $C(\underset{\sim}{\dot{p}})$

compare $C(\underset{\sim}{\dot{p}})$ with $C(\underset{\sim}{p})$ intervals in Figure 7.1 for prior $\underset{\sim}{\bar{p}}$

if $C(\underset{\sim}{\dot{p}})$ is not in recommended interval, recalculate estimator as recommended in Figure 7.1

The gain in using the estimator recommended by these procedures is usually a 1/4 to 1/2 reduction in risk. In many cases, however, the reduction can be very large. The largest reduction in risk in this study occurred when $\underset{\sim}{p}=(0,0,1)$. For this corner probability, the risk of the posterior mean was as much as 33,000 times larger than the risk for the posterior mode or maximum likelihood estimate.

TABLE 7.1

THE GENERATED DIRICHLET PROBABILITIES

A. VALUES AND CENTRALITY MEASURES

Gen. No.	$\underset{\sim}{\nu}_1 \equiv (0.1,0.1,9.8)$		$\underset{\sim}{\nu}_2 \equiv (1.0,1.0,8.0)$		$\underset{\sim}{\nu}_3 \equiv (2.0,3.0,5.0)$		$\underset{\sim}{\nu}_4 \equiv (3.3,3.3,3.3)$	
	$\underset{\sim}{p}$	$C(\underset{\sim}{p})$	$\underset{\sim}{p}$	$C(\underset{\sim}{p})$	$\underset{\sim}{p}$	$C(\underset{\sim}{p})$	$\underset{\sim}{p}$	$C(\underset{\sim}{p})^*$
1	(.00,.15,.85)	1.25	(.01,.01,.99)	1.94	(.04,.61,.35)	.48	(.53,.08,.39)	.31
2	(.02,.00,.98)	1.89	(.08,.06,.85)	1.22	(.26,.24,.50)	.13	(.32,.40,.28)	.02
3	(.00,.00,1.0)	2.00	(.01,.12,.87)	1.33	(.14,.13,.73)	.72	(.33,.29,.39)	.01
4	(.00,.00,1.0)	2.00	(.18,.00,.82)	1.10	(.15,.37,.48)	.17	(.38,.27,.35)	.02
5	(.07,.04,.89)	1.39	(.27,.19,.54)	.20	(.04,.40,.57)	.44	(.11,.44,.45)	.22
6	(.04,.00,..96)	1.75	(.16,.01,.84)	1.17	(.23,.42,.35)	.05	(.45,.28,.28)	.06
7	(.05,.03,.91)	1.51	(.01,.08,.91)	1.51	(.04,.15,.81)	1.05	(.32,.32,.36)	.00
8	(.00,.07,.93)	1.61	(.08,.01,.91)	1.53	(.17,.49,.34)	.15	(.10,.31,.60)	.38
9	(.01,.00,.99)	1.92	(.31,.24,.45)	.06	(.26,.29,.45)	.06	(.21,.24,.54)	.20
10	(.00,.00,1.0)	2.00	(.17,.15,.67)	.52	(.08,.61,.31)	.44	(.28,.47,.26)	.08

$$^*C(\underset{\sim}{p}) = \sum_{i=1}^{2} \sum_{j>i}^{3} (p_i - p_j)^2$$

Note that the three $C(\underset{\sim}{p})$ values of 2.00 correspond to 1.9999 for $\underset{\sim}{p}=(.9^{-5},.3^{-13},1.)$, 1.9993 for $\underset{\sim}{p}=(.4^{-7},.1^{-3},1.)$, and 2.0000 for $\underset{\sim}{p}=(.3^{-5},.1^{-17},1.)$, respectively.

TABLE 7.2

AVERAGED[1] RISKS

Sample Size		SS=25						SS=50					
Est. Prior[2]			original		uniform[3]	perturbed			original		uniform[3]	perturbed	
Estimator		MLE	APM	PMD	APM	APM	PMD	MLE	APM	PMD	APM	APM	PMD
NU[4]	PID												
ν_1	0	.36 -2	.23 -2	*	.95 -2	.38 -2	*	.18 -2	.14 -2	*	.34 -2	.19 -2	*
	15[5]	.39 -2	.25 -2	.41 -2	.11 -1	.42 -2	.30 -2	.20 -2	.15 -2	.23 -2	.39 -2	.21 -2	.17 -2
	40[5]	.47 -2	.27 -2	.45 -2	.14 -1	.49 -2	.34 -2	.24 -2	.17 -2	.26 -2	.50 -2	.25 -2	.20 -2
ν_2	0	.13 -1	.10 -1	.13 -1	.14 -1	.12 -1	.99 -2	.63 -2	.57 -2	.66 -2	.67 -2	.63 -2	.55 -2
	15[5]	.14 -1	.11 -1	.15 -1	.16 -1	.13 -1	.11 -1	.70 -2	.60 -2	.71 -2	.74 -2	.68 -2	.59 -2
	40[5]	.17 -1	.13 -1	.18 -1	.19 -1	.15 -1	.12 -1	.87 -2	.74 -2	.90 -2	.92 -2	.83 -2	.73 -2
ν_3	0	.22 -1	.17 -1	.17 -1	.19 -1	.19 -1	.18 -1	.11 -1	.95 -2	.95 -2	.10 -1	.10 -1	.97 -2
	15[5]	.24 -1	.18 -1	.18 -1	.21 -1	.21 -1	.19 -1	.12 -1	.10 -1	.10 -1	.11 -1	.12 -1	.11 -1
	40[5]	.31 -1	.21 -1	.21 -1	.25 -1	.25 -1	.23 -1	.15 -1	.12 -1	.12 -1	.14 -1	.14 -1	.13 -1
ν_4	0	.25 -1	.16 -1	.17 -1	.20 -1	.21 -1	.22 -1	.12 -1	.99 -2	.10 -1	.11 -1	.11 -1	.12 -1
	15[5]	.28 -1	.18 -1	.19 -1	.23 -1	.23 -1	.25 -1	.14 -1	.11 -1	.11 -1	.12 -1	.12 -1	.13 -1
	40[5]	.37 -1	.21 -1	.22 -1	.27 -1	.28 -1	.31 -1	.18 -1	.13 -1	.13 -1	.15 -1	.16 -1	.16 -1

[1]Averaged over ten Dirichlet generations of $\underset{\sim}{p}$ [see Figure 5.2 and Tables 7A.1 - 7A.6]

[2]Prior used in Bayesian estimators APM (Taylor-series approximate posterior mean) and PMD (posterior mode)

[3]For uniform prior, PMD≡MLE

[4]NU is prior parameter [see Level A in Figure 5.2]

[5]Values are the average over two replications in addition to an average over ten Dirichlet generations. Risk is estimated by the regression or control-variate estimate (see Tables 7A.1 - 7A.6 and Sections 5.9 and 7.1).

*Values were not calculated (see main text)

TABLE 7.3

RATIO OF AVERAGED[1] RISK(APM) TO AVERAGED[1] RISK(PMD) AND AVERAGED[1] RISK(MLE)

Sample Size		SS=25					SS=50				
Est. Prior[2]		original		uniform	perturbed		original		uniform	perturbed	
Estimator		MLE	PMD	MLE[3]	MLE	PMD	MLE	PMD	MLE[3]	MLE	PMD
NU	PID										
1	0	.6	-	2.6	1.1	-	.8	-	1.9	1.1	-
	15	.6	.6	2.8	1.1	1.4	.8	.7	2.0	1.1	1.2
	40	.6	.6	3.0	1.0	1.4	.7	.7	2.1	1.0	1.3
2	0	.8	.8	1.1	.9	1.2	.9	.9	1.1	1.0	1.1
	15	.8	.7	1.1	.9	1.2	.9	.9	1.1	1.0	1.2
	40	.8	.7	1.1	.9	1.3	.9	.8	1.1	1.0	1.1
3	0	.8	1.0	.9	.9	1.1	.9	1.0	.9	.9	1.0
	15	.8	1.0	.9	.9	1.1	.8	1.0	.9	1.0	1.1
	40	.7	1.0	.8	.8	1.1	.8	1.0	.9	.9	1.1
4	0	.6	.9	.8	.8	1.0	.8	1.0	.9	.9	.9
	15	.6	1.0	.8	.8	.9	.8	1.0	.9	.9	.9
	40	.6	1.0	.7	.8	.9	.7	1.0	.8	.9	1.0

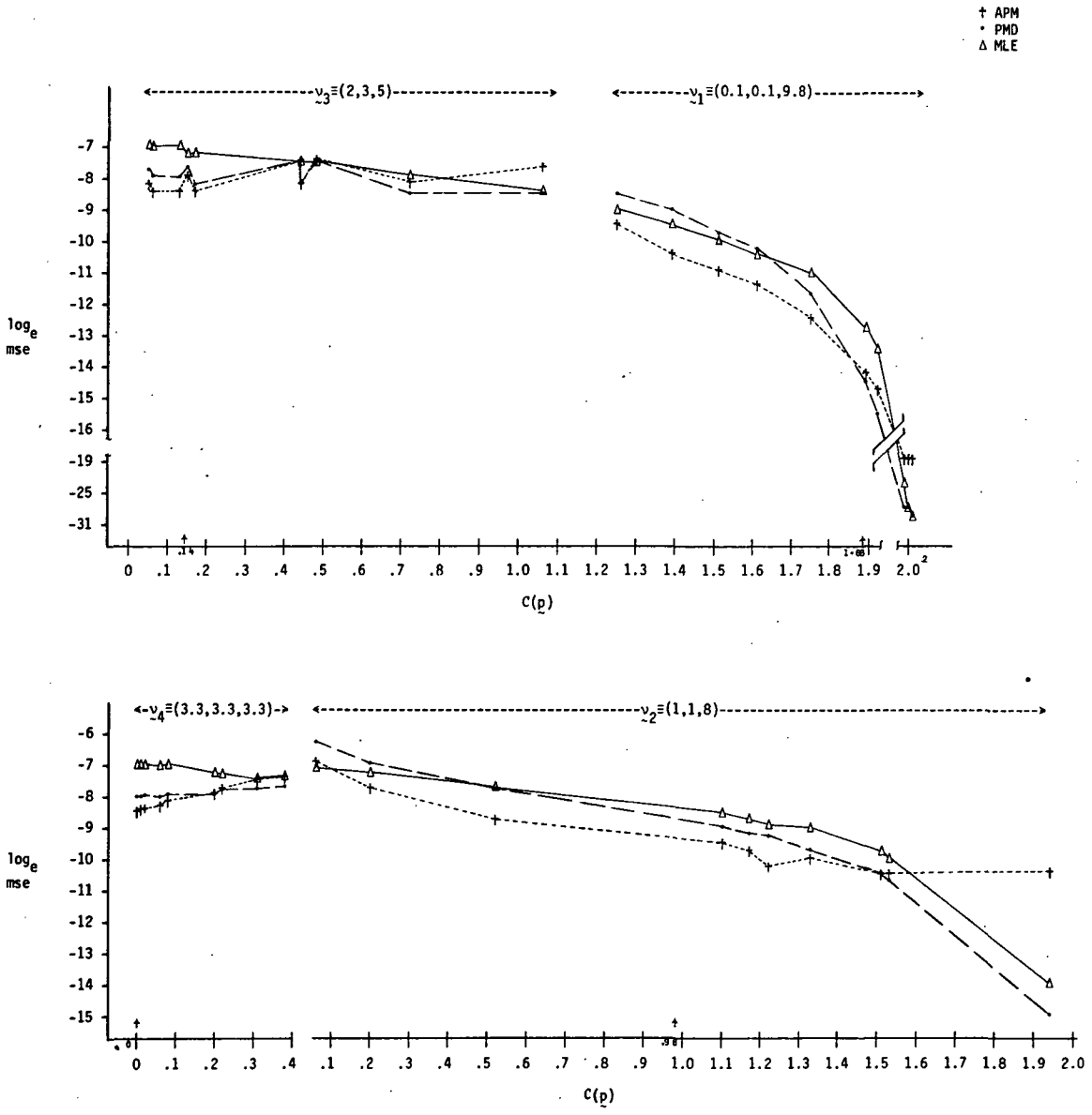
[1]Ratios are of averaged risks given in Table 7.2

[2]Prior used in Bayesian estimators APM (Taylor-series approximate posterior mean) and PMD (posterior mode)

[3]For uniform prior, PMD≡MLE

TABLE 7.4

PLOT OF (Pw.NU)×EST×SS×PID INTERACTION[1]. SS=25, PID=15. ORIGINAL PRIOR USED IN BAYESIAN ESTIMATORS.

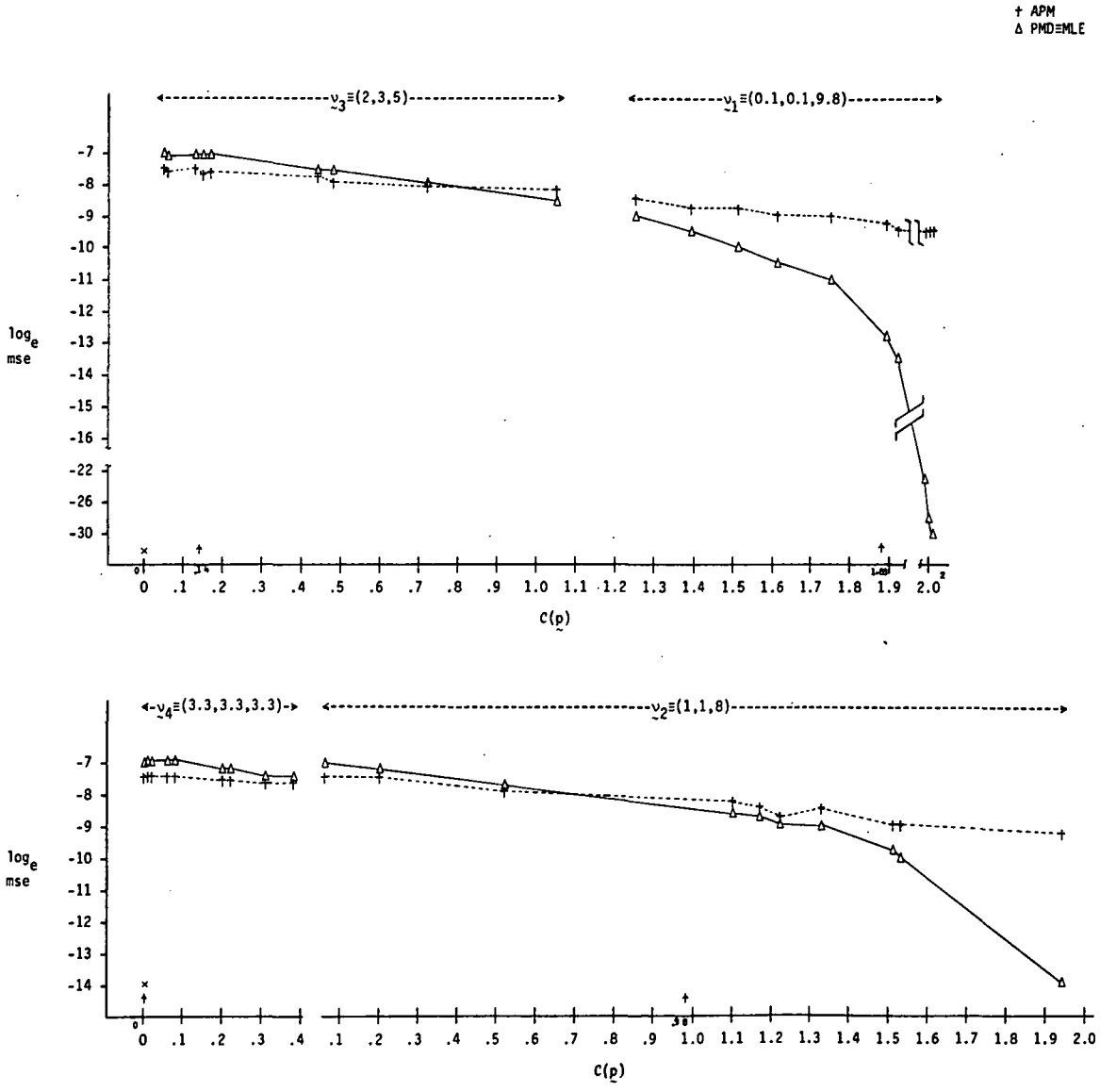


[1] Values are sums over replication. Arrow (†) denotes centrality measure of expected value of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}$ [See Table 5.1].

[2] Horizontal axis for the three sets of values plotted at 2.0 for $\underset{\sim}{\nu}_1$ is rescaled to have values 1.9993, 1.9999, 2.0000. Note that vertical axis is also rescaled.

TABLE 7.5

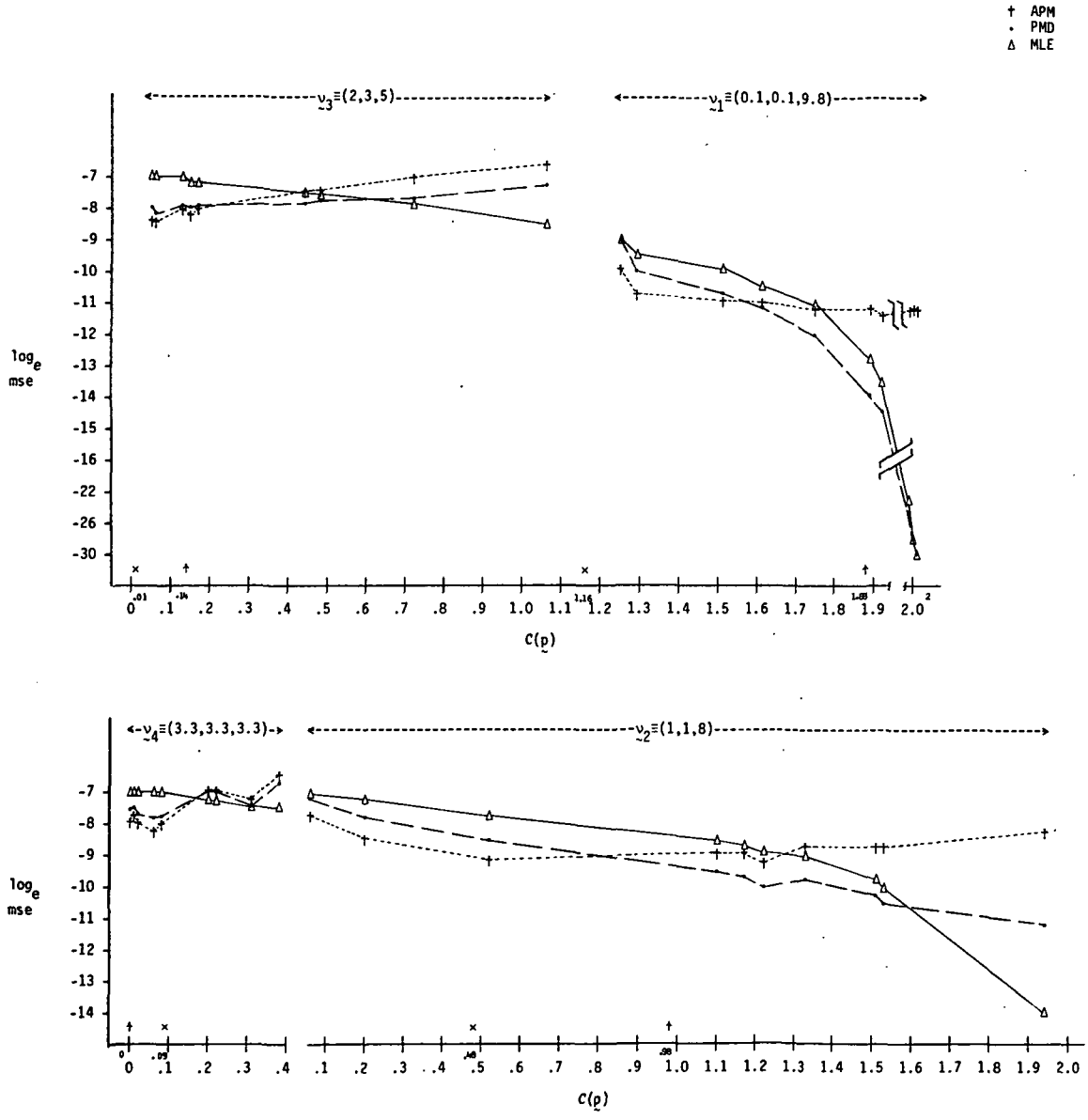
PLOT OF (Pw.NU)×EST×SS×PID INTERACTION[1]. SS=25, PID=15. UNIFORM PRIOR USED IN BAYESIAN ESTIMATORS.



[1]Values are sums over replication. Arrow (†) denotes the centrality measure of the expected value of p given ν (see Table 5.1]. Cross (x) denotes the centrality measure of the expected value of p given the uniform prior (1,1,1) [see Table 5.2 and Figure 6.2].
[2]Horizontal axis for the three sets of values plotted at 2.0 for ν_1 is rescaled to have values 1.9993, 1.9999, 2.0000. Note that vertical axis is also rescaled.

TABLE 7.6

PLOT OF (Pw.NU)×EST×SS×PID INTERACTION[1]. SS=25, PID=15. PERTURBED PRIOR USED IN BAYESIAN ESTIMATORS.



[1] Values are sums over replication. Arrow (†) denotes the centrality measure of the expected value of p given ν [see Table 5.1]; (x) denotes the centrality measure of the expected value of p given the perturbed prior 10×[ν/10+(.09,.05,-.14)] [see Table 5.2, Fig. 6.2].

[2] Horizontal axis for the three sets of values plotted at 2.0 for ν_1 is rescaled to have values 1.9993, 1.9999, 2.0000. Note that vertical axis is also rescaled.

APPENDIX 7A

DATA FOR DESIGN 2

TABLE 7A.1

RISK FOR MLE, THE MAXIMUM LIKELIHOOD ESTIMATE

NU	Dir. Gen.	SS=25 PID=0	SS=25 PID=15 r1	SS=25 PID=15 r2	SS=25 PID=40 r1	SS=25 PID=40 r2	SS=50 PID=0	SS=50 PID=15 r1	SS=50 PID=15 r2	SS=50 PID=40 r1	SS=50 PID=40 r2
ν_1	1	.99-2	.11-1(.41-3)	.11-1(.42-3)	.12-1(.70-3)	.12-1(.81-3)	.50-2	.54-2(.21-3)	.52-2(.20-3)	.60-2(.30-3)	.59-2(.41-3)
	2	.15-2	.16-2(.69-4)	.16-2(.51-4)	.17-2(.13-3)	.19-2(.12-3)	.76-3	.83-3(.33-4)	.78-3(.32-4)	.99-3(.57-4)	.10-2(.52-4)
	3	.75-6	.75-6(.8-12)	.75-6(.9-12)	.74-6(.3-11)	.74-6(.2-11)	.37-6	.37-6(.8-12)	.37-6(.8-12)	.37-6(.2-11)	.37-6(.2-11)
	4	.88-5	.88-5(.9-11)	.88-5(.1-10)	.88-5(.4-10)	.88-5(.4-10)	.44-5	.44-5(.7-11)	.48-5(.9-11)	.34-5(.44-5)	.70-5(.23-6)
	5	.82-2	.88-2(.26-3)	.91-2(.31-3)	.12-1(.68-3)	.11-1(.49-3)	.41-2	.46-2(.15-3)	.47-2(.12-3)	.59-2(.27-3)	.59-2(.33-3)
	6	.33-2	.39-2(.14-3)	.37-2(.24-3)	.43-2(.27-3)	.43-2(.34-3)	.17-2	.18-2(.60-4)	.18-2(.66-4)	.20-2(.12-3)	.22-2(.14-3)
	7	.65-2	.72-2(.20-3)	.73-2(.30-3)	.93-2(.59-3)	.93-2(.49-3)	.32-2	.35-2(.14-3)	.36-2(.13-3)	.48-2(.28-3)	.49-2(.29-3)
	8	.53-2	.56-2(.24-3)	.56-2(.24-3)	.69-2(.46-3)	.64-2(.46-3)	.26-2	.28-2(.12-3)	.28-2(.14-3)	.33-2(.20-3)	.34-2(.24-3)
	9	.11-2	.12-2(.57-4)	.12-2(.30-4)	.16-2(.83-4)	.18-2(.87-4)	.55-3	.63-3(.24-4)	.61-3(.19-4)	.67-3(.48-4)	.75-3(.52-4)
	10	.27-6	.27-6(.2-12)	.27-6(.2-12)	.27-6(.7-12)	.27-6(.7-12)	.14-6	.14-6(.2-12)	.14-6(.2-12)	.14-6(.4-12)	.14-6(.5-12)
ν_2	1	.85-3	.98-3(.64-4)	.93-3(.27-4)	.11-2(.98-4)	.11-2(.11-3)	.42-3	.47-3(.13-4)	.48-3(.82-5)	.54-3(.43-4)	.58-3(.21-4)
	2	.10-1	.11-1(.27-3)	.13-1(.49-3)	.15-1(.68-3)	.14-1(.66-3)	.52-2	.58-2(.19-3)	.60-2(.20-3)	.72-2(.32-3)	.86-2(.40-3)
	3	.90-2	.11-1(.40-3)	.10-1(.40-3)	.11-1(.68-3)	.12-1(.64-3)	.45-2	.47-2(.14-3)	.46-2(.15-3)	.52-2(.28-3)	.61-2(.33-3)
	4	.12-1	.13-1(.49-3)	.14-1(.48-3)	.16-1(.11-2)	.15-1(.95-3)	.60-2	.63-2(.22-3)	.68-2(.25-3)	.68-2(.46-3)	.71-2(.41-3)
	5	.24-1	.28-1(.90-3)	.27-1(.11-2)	.36-1(.18-2)	.36-1(.19-2)	.12-1	.13-1(.47-3)	.13-1(.44-3)	.17-1(.87-3)	.18-1(.99-3)
	6	.11-1	.13-1(.41-3)	.12-1(.51-3)	.14-1(.78-3)	.15-1(.95-3)	.55-2	.60-2(.19-3)	.62-2(.19-3)	.64-2(.40-3)	.74-2(.51-3)
	7	.65-2	.73-2(.26-3)	.76-2(.32-3)	.87-2(.49-3)	.83-2(.43-3)	.33-2	.34-2(.11-3)	.36-2(.13-3)	.43-2(.22-3)	.48-2(.29-3)
	8	.63-2	.69-2(.31-3)	.69-2(.29-3)	.81-2(.51-3)	.85-2(.67-3)	.32-2	.32-2(.90-4)	.33-2(.95-4)	.39-2(.18-3)	.36-2(.21-3)
	9	.26-1	.29-1(.97-3)	.29-1(.98-3)	.38-1(.18-2)	.36-1(.19-2)	.13-1	.15-1(.44-3)	.14-1(.41-3)	.18-1(.92-3)	.19-1(.10-2)
	10	.20-1	.21-1(.64-3)	.22-1(.84-3)	.27-1(.14-2)	.29-1(.16-2)	.99-2	.11-1(.36-3)	.11-1(.35-3)	.16-1(.93-3)	.14-1(.66-3)
ν_3	1	.20-1	.23-1(.89-3)	.22-1(.83-3)	.25-1(.13-2)	.28-1(.16-2)	.10-1	.11-1(.47-3)	.11-1(.44-3)	.15-1(.87-3)	.11-1(.50-3)
	2	.25-1	.29-1(.99-3)	.28-1(.83-3)	.37-1(.19-2)	.37-1(.20-2)	.12-1	.14-1(.40-3)	.15-1(.53-3)	.17-1(.81-3)	.18-1(.92-3)
	3	.17-1	.19-1(.69-3)	.20-1(.65-3)	.25-1(.14-2)	.25-1(.13-2)	.86-2	.10-1(.37-3)	.10-1(.33-3)	.12-1(.60-3)	.12-1(.59-3)
	4	.24-1	.28-1(.98-3)	.28-1(.82-3)	.36-1(.18-2)	.33-1(.16-2)	.12-1	.14-1(.42-3)	.14-1(.48-3)	.17-1(.84-3)	.18-1(.86-3)
	5	.21-1	.23-1(.85-3)	.23-1(.88-3)	.26-1(.16-2)	.28-1(.18-2)	.10-1	.11-1(.40-3)	.11-1(.38-3)	.13-1(.80-3)	.14-1(.76-3)
	6	.26-1	.30-1(.95-3)	.30-1(.10-2)	.37-1(.20-2)	.38-1(.22-2)	.13-1	.14-1(.50-3)	.14-1(.48-3)	.18-1(.82-3)	.19-1(.92-3)
	7	.13-1	.14-1(.52-3)	.14-1(.45-3)	.19-1(.11-2)	.17-1(.10-2)	.63-2	.69-2(.22-3)	.71-2(.22-3)	.95-2(.45-3)	.89-2(.50-3)
	8	.25-1	.28-1(.83-3)	.27-1(.93-3)	.34-1(.17-2)	.34-1(.16-2)	.12-1	.14-1(.40-3)	.14-1(.50-3)	.18-1(.96-3)	.17-1(.85-3)
	9	.26-1	.28-1(.97-3)	.29-1(.97-3)	.37-1(.21-2)	.37-1(.19-2)	.13-1	.14-1(.39-3)	.14-1(.46-3)	.18-1(.88-3)	.19-1(.10-2)
	10	.21-1	.23-1(.70-3)	.23-1(.84-3)	.28-1(.13-2)	.29-1(.16-2)	.10-1	.12-1(.38-3)	.12-1(.38-3)	.13-1(.63-3)	.14-1(.70-3)
ν_4	1	.22-1	.26-1(.11-2)	.23-1(.75-3)	.29-1(.14-2)	.32-1(.15-2)	.11-1	.12-1(.43-3)	.13-1(.42-3)	.16-1(.79-3)	.15-1(.89-3)
	2	.26-1	.31-1(.92-3)	.29-1(.11-2)	.44-1(.23-2)	.38-1(.22-2)	.13-1	.15-1(.43-3)	.15-1(.45-3)	.19-1(.80-3)	.21-1(.11-2)
	3	.26-1	.31-1(.94-3)	.30-1(.11-2)	.37-1(.20-2)	.41-1(.23-2)	.13-1	.15-1(.46-3)	.15-1(.46-3)	.19-1(.82-3)	.18-1(.89-3)
	4	.26-1	.31-1(.12-2)	.29-1(.11-2)	.38-1(.20-2)	.41-1(.20-2)	.13-1	.16-1(.46-3)	.15-1(.56-3)	.19-1(.92-3)	.20-1(.90-3)
	5	.24-1	.27-1(.10-2)	.26-1(.85-3)	.35-1(.19-2)	.37-1(.20-2)	.12-1	.13-1(.43-3)	.14-1(.45-3)	.17-1(.84-3)	.16-1(.86-3)
	6	.26-1	.30-1(.10-2)	.29-1(.90-3)	.39-1(.21-2)	.39-1(.19-2)	.13-1	.14-1(.44-3)	.14-1(.44-3)	.17-1(.95-3)	.19-1(.95-3)
	7	.27-1	.31-1(.96-3)	.29-1(.97-3)	.37-1(.20-2)	.35-1(.18-2)	.13-1	.15-1(.47-3)	.14-1(.48-3)	.18-1(.82-3)	.18-1(.92-3)
	8	.22-1	.25-1(.80-3)	.25-1(.71-3)	.34-1(.17-2)	.30-1(.19-2)	.11-1	.12-1(.42-3)	.13-1(.50-3)	.15-1(.65-3)	.15-1(.74-3)
	9	.24-1	.27-1(.10-2)	.26-1(.91-3)	.37-1(.18-2)	.33-1(.18-2)	.12-1	.13-1(.43-3)	.13-1(.39-3)	.17-1(.80-3)	.18-1(.84-3)
	10	.26-1	.30-1(.11-2)	.30-1(.10-2)	.38-1(.20-2)	.40-1(.20-2)	.13-1	.15-1(.51-3)	.14-1(.58-3)	.20-1(.10-2)	.19-1(.10-2)

Note that values are given in scientific notation; e.g., .0099 is written as .99-2. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate if the regression coefficient b exists (i.e.; b≠number/zero). Otherwise, risk is estimated by the control-variate mse estimate. [See Sections 5.9 and 7.1.] Undefined regression estimate occurs in some cases when generated Dirichlet probability p approximately equals (0,0,1).

TABLE 7A.2

RISK FOR APMRO, THE TAYLOR-SERIES APPROXIMATE POSTERIOR MEAN FOR ROBUSTNESS STUDY 0 (CORRECT PRIOR USED IN BAYESIAN ESTIMATORS)

NU	Dir. Gen.	SS=25 PID=0	SS=25 PID=15 r1	SS=25 PID=15 r2	SS=25 PID=40 r1	SS=25 PID=40 r2	SS=50 PID=0	SS=50 PID=15 r1	SS=50 PID=15 r2	SS=50 PID=40 r1	SS=50 PID=40 r2
ν_1	1	.79-2	.85-2(.57-3)	.79-2(.55-3)	.90-2(.65-3)	.93-2(.66-3)	.44-2	.42-2(.28-3)	.46-2(.31-3)	.55-2(.32-3)	.49-2(.38-3)
	2	.79-3	.81-3(.31-4)	.79-3(.23-4)	.75-3(.49-4)	.83-3(.50-4)	.53-3	.56-3(.20-4)	.53-3(.21-4)	.62-3(.33-4)	.64-3(.30-4)
	3	.49-4	.56-4(.21-6)	.56-4(.22-6)	.68-4(.42-6)	.68-4(.37-6)	.17-4	.19-4(.62-7)	.19-4(.67-7)	.24-4(.12-6)	.24-4(.12-6)
	4	.53-4	.62-4(.20-6)	.59-4(.20-6)	.74-4(.41-6)	.74-4(.41-6)	.20-4	.22-4(.56-7)	.22-4(.60-7)	.25-4(.28-5)	.29-4(.17-6)
	5	.52-2	.58-2(.25-3)	.54-2(.25-3)	.65-2(.31-3)	.58-2(.27-3)	.32-2	.36-2(.15-3)	.35-2(.14-3)	.40-2(.18-3)	.40-2(.21-3)
	6	.18-2	.19-2(.76-4)	.21-2(.11-3)	.21-2(.11-3)	.20-2(.13-3)	.12-2	.13-2(.44-4)	.13-2(.51-4)	.14-2(.72-4)	.15-2(.78-4)
	7	.39-2	.42-2(.16-3)	.40-2(.19-3)	.44-2(.23-3)	.43-2(.20-3)	.24-2	.23-2(.11-3)	.26-2(.11-3)	.34-2(.18-3)	.32-2(.17-3)
	8	.32-2	.34-2(.19-3)	.33-2(.19-3)	.40-2(.23-3)	.36-2(.22-3)	.20-2	.22-2(.10-3)	.23-2(.12-3)	.24-2(.16-3)	.25-2(.16-3)
	9	.57-3	.59-3(.27-4)	.60-3(.11-4)	.69-3(.33-4)	.75-3(.26-4)	.38-3	.42-3(.16-4)	.42-3(.13-4)	.41-3(.26-4)	.47-3(.32-4)
	10	.49-4	.55-4(.20-6)	.55-4(.23-6)	.68-4(.41-6)	.68-4(.43-6)	.17-4	.19-4(.63-7)	.19-4(.61-7)	.24-4(.11-6)	.24-4(.12-6)
ν_2	1	.48-2	.54-2(.11-3)	.54-2(.96-4)	.65-2(.19-3)	.64-2(.15-3)	.18-2	.22-2(.59-4)	.21-2(.54-4)	.24-2(.88-4)	.28-2(.81-4)
	2	.57-2	.52-2(.20-3)	.62-2(.32-3)	.61-2(.36-3)	.63-2(.36-3)	.37-2	.39-2(.16-3)	.43-2(.17-3)	.44-2(.23-3)	.54-2(.30-3)
	3	.58-2	.65-2(.23-3)	.63-2(.24-3)	.68-2(.35-3)	.69-2(.30-3)	.35-2	.36-2(.11-3)	.36-2(.13-3)	.41-2(.19-3)	.47-2(.23-3)
	4	.74-2	.79-2(.27-3)	.83-2(.29-3)	.90-2(.49-3)	.94-2(.45-3)	.46-2	.49-2(.17-3)	.52-2(.18-3)	.54-2(.30-3)	.58-2(.27-3)
	5	.21-1	.22-1(.12-2)	.21-1(.13-2)	.27-1(.15-2)	.27-1(.16-2)	.11-1	.12-1(.76-3)	.11-1(.67-3)	.16-1(.93-3)	.17-1(.11-2)
	6	.67-2	.78-2(.22-3)	.71-2(.25-3)	.80-2(.34-3)	.89-2(.41-3)	.42-2	.45-2(.13-3)	.47-2(.14-3)	.50-2(.27-3)	.55-2(.32-3)
	7	.51-2	.53-2(.28-3)	.55-2(.27-3)	.60-2(.36-3)	.60-2(.37-3)	.29-2	.29-2(.13-3)	.30-2(.16-3)	.36-2(.20-3)	.40-2(.25-3)
	8	.50-2	.54-2(.31-3)	.52-2(.28-3)	.57-2(.37-3)	.59-2(.46-3)	.28-2	.28-2(.13-3)	.27-2(.14-3)	.33-2(.19-3)	.30-2(.18-3)
	9	.29-1	.33-1(.19-2)	.32-1(.16-2)	.38-1(.23-2)	.38-1(.20-2)	.14-1	.16-1(.90-3)	.15-1(.84-3)	.19-1(.12-2)	.19-1(.11-2)
	10	.12-1	.12-1(.54-3)	.12-1(.63-3)	.13-1(.70-3)	.14-1(.76-3)	.75-2	.78-2(.34-3)	.81-2(.36-3)	.10-1(.56-3)	.94-2(.47-3)
ν_3	1	.22-1	.26-1(.15-2)	.23-1(.14-2)	.31-1(.16-2)	.29-1(.18-2)	.11-1	.11-1(.87-3)	.14-1(.76-3)	.17-1(.10-2)	.14-1(.80-3)
	2	.13-1	.14-1(.53-3)	.14-1(.51-3)	.16-1(.82-3)	.16-1(.82-3)	.89-2	.96-2(.33-3)	.10-1(.37-3)	.11-1(.50-3)	.11-1(.56-3)
	3	.16-1	.15-1(.99-3)	.18-1(.94-3)	.21-1(.14-2)	.20-1(.13-2)	.83-2	.93-2(.57-3)	.10-1(.56-3)	.12-1(.73-3)	.94-2(.68-3)
	4	.13-1	.14-1(.54-3)	.14-1(.45-3)	.15-1(.83-3)	.14-1(.69-3)	.87-2	.99-2(.32-3)	.92-2(.35-3)	.11-1(.54-3)	.11-1(.55-3)
	5	.14-1	.16-1(.53-3)	.15-1(.51-3)	.17-1(.78-3)	.18-1(.84-3)	.84-2	.89-2(.31-3)	.88-2(.32-3)	.11-1(.54-3)	.11-1(.54-3)
	6	.16-1	.18-1(.82-3)	.16-1(.90-3)	.19-1(.11-2)	.20-1(.14-2)	.10-1	.11-1(.55-3)	.11-1(.50-3)	.13-1(.68-3)	.13-1(.68-3)
	7	.18-1	.20-1(.12-2)	.23-1(.12-2)	.26-1(.15-2)	.25-1(.15-2)	.85-2	.95-2(.58-3)	.97-2(.57-3)	.12-1(.78-3)	.12-1(.82-3)
	8	.18-1	.17-1(.98-3)	.20-1(.10-2)	.19-1(.13-2)	.23-1(.14-2)	.10-1	.10-1(.54-3)	.12-1(.60-3)	.14-1(.81-3)	.13-1(.79-3)
	9	.14-1	.14-1(.54-3)	.14-1(.48-3)	.15-1(.83-3)	.16-1(.70-3)	.92-2	.93-2(.29-3)	.97-2(.33-3)	.11-1(.53-3)	.12-1(.61-3)
	10	.23-1	.24-1(.15-2)	.25-1(.16-2)	.32-1(.20-2)	.30-1(.18-2)	.11-1	.13-1(.77-3)	.12-1(.82-3)	.14-1(.86-3)	.15-1(.91-3)
ν_4	1	.20-1	.23-1(.97-3)	.22-1(.87-3)	.27-1(.12-2)	.26-1(.11-2)	.11-1	.11-1(.50-3)	.13-1(.50-3)	.15-1(.79-3)	.15-1(.77-3)
	2	.14-1	.16-1(.53-3)	.14-1(.64-3)	.19-1(.99-3)	.16-1(.80-3)	.94-2	.10-1(.35-3)	.10-1(.34-3)	.12-1(.55-3)	.13-1(.69-3)
	3	.14-1	.15-1(.48-3)	.14-1(.58-3)	.15-1(.75-3)	.17-1(.85-3)	.93-2	.10-1(.32-3)	.98-2(.33-3)	.11-1(.49-3)	.11-1(.54-3)
	4	.14-1	.15-1(.57-3)	.15-1(.59-3)	.16-1(.87-3)	.17-1(.86-3)	.93-2	.11-1(.34-3)	.10-1(.38-3)	.12-1(.58-3)	.12-1(.52-3)
	5	.18-1	.20-1(.81-3)	.20-1(.87-3)	.24-1(.13-2)	.26-1(.12-2)	.10-1	.11-1(.43-3)	.12-1(.48-3)	.13-1(.71-3)	.13-1(.71-3)
	6	.15-1	.16-1(.61-3)	.15-1(.64-3)	.18-1(.10-2)	.19-1(.97-3)	.95-2	.98-2(.37-3)	.95-2(.38-3)	.11-1(.62-3)	.12-1(.72-3)
	7	.14-1	.15-1(.48-3)	.14-1(.47-3)	.15-1(.80-3)	.14-1(.69-3)	.93-2	.97-2(.32-3)	.95-2(.33-3)	.11-1(.50-3)	.11-1(.56-3)
	8	.21-1	.24-1(.11-2)	.24-1(.12-2)	.32-1(.17-2)	.29-1(.18-2)	.11-1	.12-1(.63-3)	.12-1(.73-3)	.17-1(.88-3)	.16-1(.91-3)
	9	.18-1	.19-1(.12-2)	.19-1(.11-2)	.25-1(.14-2)	.22-1(.14-2)	.10-1	.11-1(.59-3)	.11-1(.49-3)	.13-1(.77-3)	.14-1(.84-3)
	10	.15-1	.18-1(.89-3)	.16-1(.77-3)	.19-1(.11-2)	.18-1(.10-2)	.96-2	.11-1(.48-3)	.11-1(.57-3)	.14-1(.76-3)	.13-1(.72-3)

Note that values are given in scientific notation; e.g., .0079 is written as .79-2. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate [see Section 5.9]

TABLE 7A.3

RISK FOR PMDRO, THE POSTERIOR MODE FOR ROBUSTINESS STUDY RO (CORRECT PRIOR USED IN BAYESIAN ESTIMATORS)

		SS=25					SS=50				
NU	Gen.	PID=0	PID=15		PID=40		PID=0	PID=15		PID=40	
			r1	r2	r1	r2		r1	r2	r1	r2
ν_1	1	.18-1	.14-1(.88-3)	.13-1(.88-3)	.15-1(.10-2)	.16-1(.10-2)	.76-2	.58-2(.43-3)	.64-2(.47-3)	.79-2(.48-3)	.71-2(.55-3)
	2	.64-2	.67-3(.75-5)	.70-3(.17-4)	.70-3(.81-5)	.75-3(.24-4)	.23-2	.59-3(.21-4)	.57-3(.19-4)	.68-3(.34-4)	.61-3(.20-4)
	3	.47-2	.75-6(0)	.75-6(0)	.75-6(0)	.75-6(0)	.15-2	.37-6(0)	.37-6(0)	.37-6(0)	.37-6(0)
	4	.48-2	.88-5(0)	.74-7(0)	.88-5(0)	.88-5(0)	.15-2	.44-5(0)	.56-7(0)	.24-7(0)	.70-7(.14-8)
	5	.15-1	.11-1(.42-3)	.10-1(.42-3)	.12-1(.45-3)	.11-1(.44-3)	.62-2	.65-2(.31-3)	.61-2(.32-3)	.67-2(.31-3)	.67-2(.33-3)
	6	.85-2	.28-2(.14-3)	.29-2(.12-3)	.29-2(.81-4)	.30-2(.13-3)	.33-2	.20-2(.10-3)	.19-2(.10-3)	.20-2(.11-3)	.23-2(.11-3)
	7	.12-1	.74-2(.27-3)	.73-2(.27-3)	.75-2(.26-3)	.72-2(.26-3)	.52-2	.41-2(.22-3)	.46-2(.22-3)	.55-2(.26-3)	.50-2(.24-3)
	8	.11-1	.59-2(.29-3)	.57-2(.30-3)	.64-2(.29-3)	.65-2(.29-3)	.45-2	.35-2(.20-3)	.38-2(.23-3)	.37-2(.25-3)	.39-2(.24-3)
	9	.59-2	.41-3(.21-4)	.41-3(.15-4)	.43-3(.19-4)	.41-3(.97-5)	.21-2	.36-3(.23-4)	.37-3(.14-4)	.36-3(.13-4)	.38-3(.16-4)
	10	.47-2	.27-6(0)	.27-6(0)	.27-6(0)	.27-6(0)	.15-2	.14-6(0)	.14-6(0)	.14-6(0)	.14-6(0)
ν_2	1	.52-3	.58-3(.35-4)	.56-3(.13-4)	.59-3(.46-4)	.61-3(.58-4)	.33-3	.35-3(.90-5)	.37-3(.56-5)	.40-3(.30-4)	.41-3(.12-4)
	2	.79-2	.91-2(.43-3)	.97-2(.39-3)	.11-1(.51-3)	.99-2(.50-3)	.45-2	.50-2(.22-3)	.48-2(.21-3)	.63-2(.31-3)	.66-2(.32-3)
	3	.70-2	.77-2(.48-3)	.78-2(.45-3)	.80-2(.57-3)	.88-2(.60-3)	.39-2	.43-2(.21-3)	.39-2(.20-3)	.47-2(.29-3)	.50-2(.31-3)
	4	.10-1	.11-1(.76-3)	.11-1(.80-3)	.12-1(.10-2)	.13-1(.97-3)	.56-2	.58-2(.35-3)	.63-2(.36-3)	.64-2(.48-3)	.69-2(.43-3)
	5	.30-1	.32-1(.18-2)	.32-1(.19-2)	.43-1(.23-2)	.42-1(.23-2)	.14-1	.16-1(.10-2)	.14-1(.89-3)	.21-1(.12-2)	.23-1(.14-2)
	6	.92-2	.11-1(.67-3)	.97-2(.65-3)	.11-1(.80-3)	.11-1(.78-3)	.50-2	.59-2(.30-3)	.55-2(.29-3)	.60-2(.45-3)	.56-2(.42-3)
	7	.47-2	.54-2(.27-3)	.54-2(.28-3)	.61-2(.33-3)	.58-2(.33-3)	.27-2	.30-2(.13-3)	.31-2(.14-3)	.35-2(.21-3)	.39-2(.23-3)
	8	.45-2	.48-2(.25-3)	.49-2(.25-3)	.54-2(.33-3)	.58-2(.36-3)	.26-2	.28-2(.11-3)	.29-2(.11-3)	.33-2(.16-3)	.32-2(.20-3)
	9	.38-1	.44-1(.25-2)	.43-1(.21-2)	.54-1(.31-2)	.54-1(.28-2)	.17-1	.19-1(.11-2)	.18-1(.10-2)	.23-1(.14-2)	.23-1(.13-2)
	10	.20-1	.21-1(.11-2)	.20-1(.13-2)	.25-1(.14-2)	.25-1(.15-2)	.10-1	.10-1(.58-3)	.11-1(.63-3)	.14-1(.86-3)	.13-1(.76-3)
ν_3	1	.20-1	.24-1(.16-2)	.20-1(.13-2)	.27-1(.17-2)	.26-1(.19-2)	.10-1	.11-1(.81-3)	.13-1(.72-3)	.16-1(.97-3)	.12-1(.76-3)
	2	.16-1	.18-1(.70-3)	.17-1(.63-3)	.21-1(.11-2)	.20-1(.11-2)	.99-2	.11-1(.38-3)	.11-1(.39-3)	.13-1(.62-3)	.13-1(.65-3)
	3	.13-1	.13-1(.71-3)	.14-1(.65-3)	.16-1(.12-2)	.16-1(.11-2)	.73-2	.82-2(.41-3)	.87-2(.40-3)	.10-1(.59-3)	.85-2(.53-3)
	4	.16-1	.17-1(.66-3)	.17-1(.53-3)	.20-1(.11-2)	.17-1(.87-3)	.96-2	.11-1(.38-3)	.10-1(.40-3)	.12-1(.62-3)	.13-1(.65-3)
	5	.14-1	.15-1(.64-3)	.15-1(.60-3)	.16-1(.85-3)	.17-1(.10-2)	.84-2	.87-2(.32-3)	.87-2(.33-3)	.10-1(.61-3)	.10-1(.62-3)
	6	.19-1	.21-1(.98-3)	.20-1(.11-2)	.24-1(.14-2)	.26-1(.17-2)	.11-1	.13-1(.61-3)	.12-1(.55-3)	.15-1(.78-3)	.15-1(.77-3)
	7	.12-1	.13-1(.84-3)	.14-1(.83-3)	.16-1(.12-2)	.15-1(.12-2)	.62-2	.67-2(.39-3)	.70-2(.39-3)	.87-2(.57-3)	.83-2(.62-3)
	8	.20-1	.20-1(.10-2)	.22-1(.11-2)	.22-1(.15-2)	.26-1(.15-2)	.11-1	.11-1(.54-3)	.12-1(.61-3)	.15-1(.84-3)	.14-1(.82-3)
	9	.17-1	.19-1(.78-3)	.18-1(.66-3)	.20-1(.12-2)	.22-1(.10-2)	.10-1	.11-1(.37-3)	.11-1(.40-3)	.13-1(.64-3)	.14-1(.75-3)
	10	.21-1	.23-1(.15-2)	.23-1(.16-2)	.29-1(.20-2)	.27-1(.19-2)	.11-1	.12-1(.69-3)	.11-1(.74-3)	.13-1(.80-3)	.14-1(.85-3)
ν_4	1	.19-1	.21-1(.94-3)	.21-1(.81-3)	.25-1(.12-2)	.24-1(.11-2)	.10-1	.11-1(.47-3)	.12-1(.46-3)	.14-1(.77-3)	.14-1(.76-3)
	2	.16-1	.18-1(.59-3)	.17-1(.71-3)	.23-1(.12-2)	.19-1(.99-3)	.10-1	.11-1(.37-3)	.11-1(.35-3)	.13-1(.60-3)	.14-1(.77-3)
	3	.16-1	.18-1(.55-3)	.17-1(.66-3)	.18-1(.93-3)	.21-1(.10-2)	.10-1	.11-1(.35-3)	.11-1(.35-3)	.13-1(.56-3)	.13-1(.62-3)
	4	.16-1	.18-1(.66-3)	.18-1(.67-3)	.19-1(.10-2)	.21-1(.10-2)	.10-1	.12-1(.36-3)	.11-1(.42-3)	.13-1(.65-3)	.14-1(.60-3)
	5	.18-1	.20-1(.81-3)	.19-1(.84-3)	.24-1(.14-2)	.26-1(.13-2)	.10-1	.11-1(.40-3)	.12-1(.45-3)	.13-1(.72-3)	.13-1(.73-3)
	6	.17-1	.18-1(.64-3)	.17-1(.64-3)	.21-1(.11-2)	.22-1(.11-2)	.10-1	.11-1(.37-3)	.10-1(.38-3)	.12-1(.67-3)	.14-1(.75-3)
	7	.16-1	.18-1(.57-3)	.17-1(.56-3)	.19-1(.10-2)	.17-1(.87-3)	.10-1	.11-1(.35-3)	.11-1(.36-3)	.12-1(.57-3)	.12-1(.64-3)
	8	.19-1	.21-1(.99-3)	.22-1(.11-2)	.29-1(.16-2)	.26-1(.18-2)	.10-1	.11-1(.55-3)	.11-1(.63-3)	.15-1(.80-3)	.14-1(.84-3)
	9	.18-1	.19-1(.11-2)	.19-1(.10-2)	.25-1(.14-2)	.23-1(.14-2)	.10-1	.11-1(.52-3)	.11-1(.43-3)	.13-1(.73-3)	.14-1(.81-3)
	10	.17-1	.19-1(.89-3)	.19-1(.76-3)	.22-1(.12-2)	.21-1(.11-2)	.10-1	.12-1(.47-3)	.11-1(.55-3)	.15-1(.79-3)	.14-1(.76-3)

Note that values are given in scientific notation; e.g., .018 is written as .18-1. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate if the regression coefficient b exists (i.e.; b≠number/zero). Otherwise, risk is estimated by the control-variate mse estimate. [See Sections 5.9 and 7.1.] Undefined regression estimate occurs in some cases when generated Dirichlet probability $\underset{\sim}{p}$ approximately equals (0,0,1).

TABLE 7A.4

RISK FOR APMR1, THE TAYLOR-SERIES APPROXIMATE POSTERIOR MEAN FOR ROBUSTNESS STUDY 1 (UNIFORM PRIOR USED IN BAYESIAN ESTIMATORS)

		SS=25					SS=50				
NU	Dir. Gen.	PID=0	PID=15		PID=40		PID=0	PID=15		PID=40	
			r1	r2	r1	r2		r1	r2	r1	r2
ν_1	1	.13-1	.14-1(.89-3)	.15-1(.83-3)	.17-1(.11-2)	.17-1(.11-2)	.58-2	.69-2(.34-3)	.63-2(.31-3)	.75-2(.48-3)	.78-2(.49-3)
	2	.84-2	.99-2(.29-3)	.10-1(.31-3)	.12-1(.47-3)	.12-1(.40-3)	.27-2	.31-2(.15-3)	.29-2(.15-3)	.38-2(.17-3)	.41-2(.20-3)
	3	.77-2	.90-2(.46-4)	.89-2(.47-4)	.12-1(.10-3)	.12-1(.92-4)	.21-2	.25-2(.95-5)	.25-2(.10-4)	.33-2(.20-4)	.33-2(.20-4)
	4	.77-2	.88-2(.44-4)	.89-2(.45-4)	.12-1(.11-3)	.12-1(.10-3)	.21-2	.25-2(.89-5)	.25-2(.94-5)	.34-2(.25-4)	.34-2(.20-4)
	5	.12-1	.12-1(.79-3)	.14-1(.79-3)	.16-1(.12-2)	.16-1(.11-2)	.51-2	.56-2(.33-3)	.60-2(.32-3)	.75-2(.49-3)	.75-2(.50-3)
	6	.93-2	.12-1(.53-3)	.10-1(.58-3)	.13-1(.72-3)	.14-1(.73-3)	.33-2	.36-2(.23-3)	.39-2(.20-3)	.45-2(.30-3)	.46-2(.32-3)
	7	.11-1	.12-1(.69-3)	.13-1(.75-3)	.17-1(.11-2)	.17-1(.10-2)	.45-2	.53-2(.30-3)	.50-2(.30-3)	.61-2(.47-3)	.69-2(.49-3)
	8	.10-1	.11-1(.65-3)	.11-1(.68-3)	.14-1(.96-3)	.15-1(.90-3)	.41-2	.43-2(.28-3)	.41-2(.29-3)	.57-2(.37-3)	.56-2(.44-3)
	9	.82-2	.93-2(.24-3)	.91-2(.21-3)	.13-1(.37-3)	.13-1(.31-3)	.25-2	.30-2(.12-3)	.29-2(.11-3)	.37-2(.15-3)	.38-2(.16-3)
	10	.77-2	.89-2(.43-4)	.89-2(.50-4)	.12-1(.10-3)	.12-1(.11-3)	.21-2	.25-2(.97-5)	.25-2(.94-5)	.33-2(.19-4)	.33-2(.20-4)
ν_2	1	.81-2	.94-2(.19-3)	.94-2(.16-3)	.12-1(.35-3)	.12-1(.28-3)	.24-2	.30-2(.81-4)	.29-2(.73-4)	.35-2(.13-3)	.40-2(.12-3)
	2	.13-1	.12-1(.79-3)	.14-1(.10-2)	.16-1(.12-2)	.18-1(.12-2)	.59-2	.64-2(.38-3)	.73-2(.38-3)	.74-2(.50-3)	.96-2(.65-3)
	3	.12-1	.14-1(.80-3)	.14-1(.82-3)	.17-1(.10-2)	.16-1(.10-2)	.54-2	.56-2(.31-3)	.58-2(.31-3)	.65-2(.39-3)	.78-2(.51-3)
	4	.14-1	.15-1(.89-3)	.16-1(.90-3)	.20-1(.13-2)	.18-1(.11-2)	.65-2	.71-2(.34-3)	.74-2(.37-3)	.80-2(.54-3)	.83-2(.55-3)
	5	.20-1	.23-1(.83-3)	.22-1(.11-2)	.27-1(.15-2)	.27-1(.16-2)	.11-1	.12-1(.43-3)	.12-1(.44-3)	.15-1(.76-3)	.15-1(.82-3)
	6	.13-1	.16-1(.93-3)	.15-1(.96-3)	.18-1(.11-2)	.20-1(.14-2)	.62-2	.62-2(.34-3)	.70-2(.34-3)	.76-2(.46-3)	.95-2(.64-3)
	7	.11-1	.12-1(.74-3)	.12-1(.71-3)	.15-1(.99-3)	.15-1(.10-2)	.45-2	.45-2(.28-3)	.48-2(.32-3)	.61-2(.40-3)	.66-2(.48-3)
	8	.11-1	.12-1(.79-3)	.12-1(.70-3)	.14-1(.96-3)	.15-1(.12-2)	.44-2	.44-2(.28-3)	.43-2(.29-3)	.55-2(.39-3)	.50-2(.34-3)
	9	.21-1	.22-1(.78-3)	.23-1(.79-3)	.28-1(.13-2)	.26-1(.14-2)	.12-1	.13-1(.40-3)	.13-1(.39-3)	.15-1(.81-3)	.16-1(.93-3)
	10	.18-1	.19-1(.92-3)	.20-1(.11-2)	.23-1(.15-2)	.25-1(.17-2)	.93-2	.11-1(.44-3)	.10-1(.46-3)	.14-1(.98-3)	.14-1(.71-3)
ν_3	1	.18-1	.21-1(.92-3)	.19-1(.74-3)	.24-1(.12-2)	.25-1(.13-2)	.95-2	.10-1(.49-3)	.11-1(.45-3)	.15-1(.80-3)	.11-1(.48-3)
	2	.20-1	.22-1(.90-3)	.24-1(.76-3)	.27-1(.14-2)	.28-1(16-2)	.11-1	.12-1(.40-3)	.13-1(.55-3)	.15-1(.70-3)	.15-1(.80-3)
	3	.16-1	.16-1(.93-3)	.19-1(.92-3)	.23-1(.17-2)	.22-1(.16-2)	.84-2	.96-2(.48-3)	.10-1(.46-3)	.12-1(.69-3)	.10-1(.65-3)
	4	.20-1	.23-1(.88-3)	.23-1(.84-3)	.27-1(.14-2)	.25-1(.13-2)	.11-1	.13-1(.39-3)	.12-1(.44-3)	.15-1(.74-3)	.16-1(.75-3)
	5	.18-1	.20-1(.73-3)	.20-1(.76-3)	.23-1(.15-2)	.25-1(.14-2)	.97-2	.11-1(.39-3)	.10-1(.38-3)	.13-1(.71-3)	.13-1(.68-3)
	6	.21-1	.23-1(.74-3)	.24-1(.83-3)	.26-1(.14-2)	.27-1(.16-2)	.12-1	.13-1(.46-3)	.12-1(.44-3)	.16-1(.69-3)	.17-1(.80-3)
	7	.14-1	.16-1(.94-3)	.17-1(.93-3)	.21-1(.15-2)	.19-1(14-2)	.68-2	.75-2(.40-3)	.77-2(.39-3)	.98-2(.60-3)	.96-2(.66-3)
	8	.20-1	.21-1(.71-3)	.22-1(.78-3)	.25-1(.13-2)	.26-1(.13-2)	.11-1	.12-1(.36-3)	.12-1(.46-3)	.15-1(.85-3)	.14-1(.75-3)
	9	.21-1	.21-1(.75-3)	.23-1(.83-3)	.27-1(.15-2)	.27-1(.14-2)	.12-1	.13-1(.37-3)	.12-1(.42-3)	.15-1(.77-3)	.16-1(.89-3)
	10	.18-1	.20-1(.83-3)	.20-1(.90-3)	.23-1(.13-2)	.24-1(.14-2)	.97-2	.11-1(.40-3)	.11-1(.41-3)	.12-1(.59-3)	.13-1(.63-3)
ν_4	1	.19-1	.22-1(.95-3)	.20-1(.72-3)	.24-1(.12-2)	.25-1(.12-2)	.10-1	.11-1(.42-3)	.12-1(.41-3)	.14-1(.76-3)	.14-1(.80-3)
	2	.21-1	.24-1(.72-3)	.22-1(.90-3)	.32-1(.16-2)	.27-1(.15-2)	.12-1	.13-1(.40-3)	.13-1(.40-3)	.16-1(.69-3)	.17-1(.94-3)
	3	.21-1	.24-1(.71-3)	.23-1(.84-3)	.26-1(.14-2)	.29-1(.15-2)	.12-1	.13-1(.40-3)	.13-1(.41-3)	.16-1(.68-3)	.15-1(.75-3)
	4	.21-1	.24-1(.87-3)	.23-1(.85-3)	.27-1(.14-2)	.29-1(.14-2)	.12-1	.14-1(.40-3)	.13-1(.49-3)	.16-1(.78-3)	.17-1(.74-3)
	5	.20-1	.22-1(.86-3)	.21-1(.80-3)	.27-1(.16-2)	.29-1(.16-2)	.11-1	.11-1(.39-3)	.12-1(.43-3)	.14-1(.76-3)	.14-1(.78-3)
	6	.21-1	.23-1(.77-3)	.23-1(.71-3)	.28-1(.14-2)	.29-1(.14-2)	.12-1	.12-1(.39-3)	.12-1(.40-3)	.15-1(.79-3)	.16-1(.84-3)
	7	.21-1	.24-1(.75-3)	.23-1(.75-3)	.27-1(.14-2)	.25-1(.12-2)	.12-1	.13-1(.41-3)	.12-1(.42-3)	.15-1(.69-3)	.15-2(.78-3)
	8	.19-1	.21-1(.78-3)	.21-1(.82-3)	.28-1(.15-2)	.25-1(.18-2)	.10-1	.11-1(.44-3)	.11-1(.51-3)	.14-1(.68-3)	.13-1(.75-3)
	9	.20-1	.22-1(.98-3)	.22-1(.89-3)	.29-1(.14-2)	.26-1(.14-2)	.11-1	.12-1(.43-3)	.12-1(.37-3)	.14-1(.72-3)	.15-1(.79-3)
	10	.21-1	.24-1(.94-3)	.23-1(.80-3)	.28-1(.14-2)	.28-1(.14-2)	.11-1	.13-1(.47-3)	.13-1(.55-3)	.17-1(.88-3)	.16-1(.88-3)

Note that values are given in scientific notation; e.g., .013 is written as .13-1. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate [see Section 5.9]

TABLE 7A.5

RISK FOR APMR2, THE TAYLOR-SERIES APPROXIMATE POSTERIOR MEAN FOR ROBUSTNESS STUDY 2 (PERTURBED PRIOR USED IN BAYESIAN ESTIMATORS)

		SS=25					SS=50				
NU	Dir. Gen.	PID=0	PID=15		PID=40		PID=0	PID=15		PID=40	
			r1	r2	r1	r2		r1	r2	r1	r2
ν_1	1	.65-2	.72-2(.23-3)	.70-2(.23-3)	.79-2(.32-3)	.79-2(.37-3)	.39-2	.42-2(.15-3)	.42-2(.17-3)	.52-2(.21-3)	.48-2(.27-3)
	2	.32-2	.36-2(.14-3)	.38-2(.15-3)	.40-2(.20-3)	.39-2(.18-3)	.14-2	.15-2(.88-4)	.14-2(.88-4)	.17-2(.99-4)	.19-2(.11-3)
	3	.32-2	.36-2(.15-4)	.36-2(.16-4)	.45-2(.31-4)	.44-2(.28-4)	.11-2	.12-2(.44-5)	.12-2(.47-5)	.16-2(.86-5)	.16-2(.89-5)
	4	.32-2	.36-2(.14-4)	.36-2(.15-4)	.44-2(.30-4)	.44-2(.30-4)	.11-2	.12-2(.43-5)	.12-2(.45-5)	.16-2(.13-4)	.16-2(.81-5)
	5	.45-2	.43-2(.19-3)	.47-2(.21-3)	.49-2(.36-3)	.47-2(.29-3)	.29-2	.32-2(.13-3)	.33-2(.11-3)	.37-2(.21-3)	.37-2(.22-3)
	6	.34-2	.41-2(.22-3)	.34-2(.25-3)	.40-2(.29-3)	.41-2(.30-3)	.17-2	.18-2(.12-3)	.20-2(.10-3)	.20-2(.15-3)	.21-2(.17-3)
	7	.40-2	.42-2(.21-3)	.43-2(.25-3)	.50-2(.38-3)	.51-2(.34-3)	.25-2	.27-2(.13-3)	.26-2(.13-3)	.31-2(.22-3)	.33-2(.23-3)
	8	.42-2	.44-2(.18-3)	.44-2(.18-3)	.51.2(.27-3)	.53-2(.28-3)	.23-2	.25-2(.10-3)	.25-2(.11-3)	.31-2(.15-3)	.31-2(.19-3)
	9	.32-2	.35-2(.12-3)	.34-2(.10-3)	.44-2(.17-3)	.46-2(.13-3)	.13-2	.15-2(.73-4)	.14-2(.69-4)	.17-2(.90-4)	.18-2(.96-4)
	10	.32-2	.36-2(.15-4)	.36-2(.16-4)	.44-2(.30-4)	.45-2(.31-4)	.11-2	.12-2(.44-5)	.12-2(.44-5)	.16-2(.85-5)	.16-2(.89-5)
ν_2	1	.14-1	.16-1(.20-3)	.16-1(.17-3)	.19-1(.32-3)	.19-1(.27-3)	.48-2	.58-2(.10-3)	.56-2(.93-4)	.68-2(.15-3)	.75-2(.14-3)
	2	.99-2	.89-2(.61-3)	.10-1(.75-3)	.12-1(.82-3)	.12-1(.82-3)	.52-2	.55-2(.35-3)	.63-2(.35-3)	.60-2(.43-3)	.79-2(.55-3)
	3	.11-1	.13-1(.54-3)	.12-1(.55-3)	.15-1(.69-3)	.15-1(.63-3)	.53-2	.56-2(.26-3)	.59-2(.26-3)	.68-2(.31-3)	.77-2(.40-3)
	4	.99-2	.11-1(.50-3)	.11-1(.51-3)	.14-1(.71-3)	.13-1(.62-3)	.55-2	.59-2(.24-3)	.61-2(.27-3)	.67-2(.38-3)	.69-2(.38-3)
	5	.14-1	.15-1(.68-3)	.14-1(.69-3)	.17-1(.89-3)	.17-1(.92-3)	.90-2	.95-2(.46-3)	.91-2(.40-3)	.12-1(.63-3)	.13-1(.76-3)
	6	.99-2	.11-1(.60-3)	.11-1(.61-3)	.12-1(.65-3)	.14-1(.82-3)	.53-2	.53-2(.26-3)	.60-2(.27-3)	.64-2(.35-3)	.79-2(.47-3)
	7	.12-1	.12-1(.59-3)	.13-1(.56-3)	.15-1(.72-3)	.15-1(.75-3)	.51-2	.52-2(.27-3)	.54-2(.31-3)	.71-2(.37-3)	.74-2(.43-3)
	8	.11-1	.12-1(.68-3)	.12-1(.61-3)	.14-1(.75-3)	.14-1(.88-3)	.49-2	.49-2(.30-3)	.48-2(.31-3)	.60-2(.39-3)	.56-2(.35-3)
	9	.19-1	.21-1(.12-2)	.20-1(.10-2)	.24-1(.15-2)	.23-1(.13-2)	.11-1	.12-1(.59-3)	.11-1(.54-3)	.13-1(.81-3)	.14-1(.78-3)
	10	.10-1	.10-1(.32-3)	.11-1(.38-3)	.11-1(.55-3)	.11-1(.63-3)	.69-2	.74-2(.24-3)	.74-2(.24-3)	.93-2(.55-3)	.88-2(.40-3)
ν_3	1	.21-1	.24-1(.11-2)	.22-1(.95-3)	.30-1(.13-2)	.30-1(.13-2)	.11-1	.12-1(.65-3)	.13-1(.58-3)	.17-1(.81-3)	.14-1(.61-3)
	2	.16-1	.16-1(.92-3)	.18-1(.78-3)	.19-1(.11-2)	.20-1(.12-2)	.96-2	.10-1(.46-3)	.12-1(.59-3)	.12-1(.60-3)	.12-1(.73-3)
	3	.26-1	.26-1(.15-2)	.30-1(.15-2)	.37-1(.21-2)	.35-1(.20-2)	.12-1	.13-1(.83-3)	.14-1(.82-3)	.18-1(.10-2)	.14-1(.97-3)
	4	.15-1	.17-1(.79-3)	.17-1(.83-3)	.19-1(.10-2)	.18-1(.98-3)	.94-2	.11-1(.39-3)	.10-1(.42-3)	.12-1(.62-3)	.13-1(.64-3)
	5	.20-1	.22-1(.77-3)	.21-1(.77-3)	.27-1(.12-2)	.28-1(.11-2)	.10-1	.12-1(.43-3)	.11-1(.45-3)	.15-1(.62-3)	.14-1(.60-3)
	6	.14-1	.15-1(.52-3)	.15-1(.58-3)	.15-1(.83-3)	.16-1(.10-2)	.92-2	.99-2(.39-3)	.96-2(.38-3)	.11-1(.51-3)	.12-1(.59-3)
	7	.32-1	.36-1(.16-2)	.39-1(.16-2)	.46-1(.21-2)	.45-1(.21-2)	.13-1	.15-1(.81-3)	.15-1(.80-3)	.19-1(.11-2)	.19-1(.11-2)
	8	.15-1	.15-1(.70-3)	.17-1(.72-3)	.17-1(.10-2)	.19-1(.11-2)	.95-2	.97-2(.38-3)	.11-1(.47-3)	.12-1(.73-3)	.11-1(.66-3)
	9	.14-1	.14-1(.58-3)	.15-1(.68-3)	.16-1(.10-2)	.16-1(.89-3)	.93-2	.10-1(.34-3)	.99-2(.37-3)	.11-1(.60-3)	.12-1(.68-3)
	10	.20-1	.22-1(.12-2)	.23-1(.12-2)	.27-1(.15-2)	.27-1(.14-2)	.11-1	.12-1(.60-3)	.11-1(.63-3)	.13-1(.69-3)	.15-1(.73-3)
ν_4	1	.23-1	.27-1(.10-2)	.26-1(.87-3)	.33-1(.14-2)	.32-1(.12-2)	.12-1	.13-1(.51-3)	.14-1(.53-3)	.18-1(.86-3)	.18-1(.81-3)
	2	.15-1	.17-1(.73-3)	.16-1(.65-3)	.20-1(.11-2)	.17-1(.11-2)	.97-2	.10-1(.39-3)	.10-1(.38-3)	.12-1(.54-3)	.14-1(.75-3)
	3	.18-1	.20-1(.94-3)	.19-1(.11-2)	.23-1(.13-2)	.23-1(.14-2)	.11-1	.11-1(.53-3)	.12-1(.51-3)	.14-1(.71-3)	.14-1(.77-3)
	4	.17-1	.18-1(.95-3)	.18-1(.96-3)	.20-1(.13-2)	.21-1(.13-2)	.10-1	.12-1(.44-3)	.11-1(.57-3)	.13-1(.79-3)	.13-1(.63-3)
	5	.25-1	.30-1(.12-2)	.29-1(.13-2)	.37-1(.19-2)	.39-1(.16-2)	.13-1	.14-1(.65-3)	.15-1(.67-3)	.18-1(.99-3)	.18-1(.92-3)
	6	.15-1	.16-1(.64-3)	.15-1(.60-3)	.17-1(.89-3)	.19-1(.95-3)	.95-2	.94-2(.37-3)	.99-2(.37-3)	.12-1(.62-3)	.13-1(.65-3)
	7	.17-1	.19-1(.95-3)	.18-1(.89-3)	.22-1(.13-2)	.19-1(.12-2)	.10-1	.11-1(.49-3)	.11-1(.48-3)	.13-1(.71-3)	.13-1(.75-3)
	8	.34-1	.38-1(.17-2)	.38-1(.18-2)	.54-1(.25-2)	.49-1(.26-2)	.15-1	.17-1(.92-3)	.16-1(.11-2)	.25-1(.12-2)	.23-1(.13-2)
	9	.27-1	.31-1(.19-2)	.31-1(.18-2)	.41-1(.22-2)	.38-1(.22-2)	.14-1	.14-1(.89-3)	.16-1(.77-3)	.19-1(.11.2)	.19-1(.12-2)
	10	.16-1	.18-1(.82-3)	.17-1(.84-3)	.19-1(.10-2)	.21-1(.12-2)	.98-2	.11-1(.48-3)	.11-1(.54-3)	.13-1(.74-3)	.13-1(.70-3)

Note that values are given in scientific notation; e.g., .0065 is written as .65-2. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate [see Section 5.9]

TABLE 7A.6

RISK FOR PMDR2, THE POSTERIOR MODE FOR ROBUSTNESS STUDY R2 (PERTURBED PRIOR USED IN BAYESIAN ESTIMATORS)

		SS=25					SS=50				
NU	Gen.	PID=0	PID=15		PID=40		PID=0	PID=25		PID=40	
			r1	r2	r1	r2		r1	r2	r1	r2
ν	1	.10-1	.11-1(.70-3)	.10-1(.68-3)	.12-1(.80-3)	.12-1(.82-3)	.51-2	.48-2(.33-3)	.53-2(.37-3)	.64-2(.37-3)	.57-2(.45-3)
~1	2	.14-2	.99-3(.37-4)	.95-3(.30-4)	.95-3(.53-4)	.11-2(.60-4)	.73-3	.62-3(.22-4)	.59-3(.23-4)	.70-3(.39-4)	.70-3(.32-4)
	3	.31-3	.75-6(.6-12)	.75-6(.5-12)	.75-6(.2-11)	.75-6(.2-11)	.99-4	.37-6(.7-12)	.37-6(.6-12)	.37-6(.1-11)	.37-6(.2-11)
	4	.32-3	.88-5(.3-11)	.19-5(.3-11)	.88-5(.1-10)	.88-5(.9-11)	.10-3	.44-5(.4-11)	.13-5(.4-11)	.25-5(.32-5)	.18-5(.50-7)
	5	.70-2	.68-2(.31-3)	.64-2(.30-3)	.78-2(.38-3)	.70-2(.31-3)	.38-2	.42-2(.18-3)	.40-2(.16-3)	.46-2(.20-3)	.46-2(.24-3)
	6	.28-2	.23-2(.96-4)	.25-2(.14-3)	.25-2(.14-3)	.24-2(.17-3)	.15-2	.15-2(.50-4)	.14-2(.58-4)	.15-2(.81-4)	.17-2(.90-4)
	7	.55-2	.49-2(.20-3)	.47-2(.23-3)	.53-2(.27-3)	.51-2(.25-3)	.30-2	.26-2(.14-3)	.29-2(.13-3)	.38-2(.20-3)	.36-2(.19-3)
	8	.48-2	.44-2(.24-3)	.44-2(.25-3)	.54-2(.28-3)	.52-2(.27-3)	.25-2	.27-2(.14-3)	.29-2(.17-3)	.30-2(.21-3)	.31-2(.20-3)
	9	.11-2	.73-3(.30-4)	.75-3(.13-4)	.85-3(.32-4)	.89-3(.29-4)	.55-3	.47-3(.17-4)	.47-3(.13-4)	.48-3(.31-4)	.53-3(.33-4)
	10	.31-3	.27-6(.2-12)	.27-6(.2-12)	.27-6(.4-12)	.27-6(.4-12)	.99-4	.14-6(.2-12)	.14-6(.2-12)	.14-6(.4-12)	.14-6(.4-12)
ν	1	.32-2	.36-2(.11-3)	.36-2(.87-4)	.44-2(.17-3)	.42-2(.14-3)	.12-2	.14-2(.48-4)	.13-2(.42-4)	.15-2(.75.4)	.18-2(.67-4)
~2	2	.66-2	.62-2(.21-3)	.73-2(.35-3)	.75-2(.44-3)	.75-2(.40-3)	.41-2	.44-2(.16-3)	.47-2(.17-3)	.49-2(.24-3)	.60-2(.32-3)
	3	.66-2	.75-2(.26-3)	.74-2(.25-3)	.82-2(.41-3)	.84-2(.35-3)	.38-2	.40-2(.12-3)	.39-2(.13-3)	.46-2(.22-3)	.52-2(.25-3)
	4	.77-2	.81-2(.28-3)	.85-2(.30-3)	.89-2(.58-3)	.90-2(.51-3)	.47-2	.49-2(.17-3)	.53-2(.18-3)	.52-2(.32-3)	.56-2(.29-3)
	5	.20-1	.21-1(.11-2)	.21-1(.11-2)	.26-1(.14-2)	.25-1(.14-2)	.11-1	.12-1(.65-3)	.11-1(.56-3)	.15-1(.85-3)	.17-1(.10-2)
	6	.71-2	.81-2(.25-3)	.74-2(.30-3)	.78-2(.40-3)	.89-2(.50-3)	.44-2	.46-2(.14-3)	.48-2(.15-3)	.48-2(.29-3)	.55-2(.37-3)
	7	.53-2	.56-2(.25-3)	.59-2(.26-3)	.66-2(.35-3)	.65-2(.35-3)	.29-2	.30-2(.11-3)	.32-2(.13-3)	.39-2(.19-3)	.43-2(.24-3)
	8	.48-2	.50-2(.33-3)	.48-2(.29-3)	.52-2(.41-3)	.55-2(.51-3)	.27-2	.27-2(.13-3)	.26-2(.13-3)	.30-2(.20-3)	.27-2(.18-3)
	9	.25-1	.28-1(.16-2)	.27-1(.14-2)	.33-1(.21-2)	.33-1(.19-2)	.13-1	.14-1(.74-3)	.13-1(.67-3)	.17-1(.10-2)	.17-1(.96-3)
	10	.13-1	.14-1(.50-3)	.13-1(.58-3)	.15-1(.75-3)	.16-1(.84-3)	.80-2	.84-2(.32-3)	.87-2(.33-3)	.11-1(.61-3)	.10-1(.48-3)
ν	1	.18-1	.21-1(.10-2)	.19-1(.85-3)	.25-1(.12-2)	.25-1(.13-2)	.95-2	.10-1(.57.3)	.12-1(.51-3)	.15-1(.77-3)	.12-1(.54-3)
~3	2	.17-1	.18-1(.85-3)	.20-1(.75-3)	.21-1(.12-2)	.22-1(.12-2)	.10-1	.11-1(.42-3)	.12-1(.55-3)	.13-1(.61-3)	.13-1(.72-3)
	3	.20-1	.20-1(.13-2)	.23-1(.12-2)	.28-1(.19-2)	.27-1(.18-2)	.96-2	.11-1(.67-3)	.12-1(.66-3)	.15-1(.87-3)	.11-1(.82-3)
	4	.16-1	.18-1(.75-3)	.18-1(.76-3)	.20-1(.10-2)	.19-1(.10-2)	.98-2	.11-1(.37-3)	.11-1(.41-3)	.13-1(.64-3)	.13-1(.65-3)
	5	.17-1	.19-1(.70-3)	.19-1(.72-3)	.23-1(.12-2)	.24-1(.12-2)	.94-2	.10-1(.40-3)	.97-2(.40-3)	.13-1(.63-3)	.13-1(.61-3)
	6	.16-1	.17-1(.56-3)	.17-1(.63-3)	.18-1(.97-3)	.19-1(.12-2)	.10-1	.11-1(.41-3)	.11-1(.39-3)	.13-1(.57-3)	.14-1(.65-3)
	7	.22-1	.24-1(.14-2)	.27-1(.14-2)	.32-1(.19-2)	.31-1(.18-2)	.93-2	.10-1(.64-3)	.11-1(.63-3)	.14-1(.88-3)	.13-1(.92-3)
	8	.16-1	.17-1(.65-3)	.18-1(.69-3)	.18-1(.11-2)	.20-1(.11-2)	.99-2	.10-1(.35-3)	.11-1(.44-3)	.13-1(.73-3)	.12-1(.66-3)
	9	.16-1	.16-1(.60-3)	.18-1(.69-3)	.19-1(.11-2)	.19-1(.99-3)	.10-1	.11-1(.34-3)	.11-1(.38-3)	.13-1(.64-3)	.14-1(.73-3)
	10	.18-1	.19-1(.10-2)	.20-1(.11-2)	.23-1(.14-2)	.23-1(.13-2)	.96-2	.11-1(.50-3)	.10-1(.52-3)	.12-1(.62-3)	.13-1(.65-3)
ν	1	.22-1	.26-1(.12-2)	.24-2(.92-3)	.32-1(.15-2)	.31-1(.14-2)	.11-1	.13-1(.51-3)	.13-1(.53-3)	.17-1(.90-3)	.17-1(.87-3)
~4	2	.18-1	.21-1(.91-3)	.19-1(.81-3)	.25-1(.14-2)	.23-1(.14-2)	.11-1	.12-1(.43-3)	.12-1(.44-3)	.14-1(.62-3)	.16-1(.89-3)
	3	.21-1	.24-1(.11-2)	.22-1(.13-2)	.27-1(.15-2)	.28-1(.17-2)	.12-1	.13-1(.56-3)	.13-1(.54-3)	.15-1(.79-3)	.15-1(.85-3)
	4	.19-1	.22-1(.12-2)	.21-1(.11-2)	.24-1(.16.2)	.27-1(.16-2)	.11-1	.13-1(.48-3)	.13-1(.63-3)	.15-1(.90-3)	.15-1(.72-3)
	5	.25-1	.30-1(.13-2)	.29-1(.14-2)	.37-1(.21-2)	.39-1(.18-2)	.12-1	.14-1(.65-3)	.14-1(.66-3)	.18-1(.10-2)	.17-1(.96-3)
	6	.18-1	.19-1(.79-3)	.19-1(.75-3)	.21-1(.11-2)	.24-1(.12-2)	.11-1	.11-1(.41-3)	.11-1(.42-3)	.13-1(.72-3)	.14-1(.74-3)
	7	.20-1	.23-1(.11-2)	.22-1(.10-2)	.27-1(.16-2)	.24-1(.15-2)	.11-1	.12-1(.54-3)	.12-1(.53-3)	.15-1(.80-3)	.14-1(.85-3)
	8	.30-1	.34-1(.17-2)	.35-1(.18-2)	.50-1(.26-2)	.46-1(.28-2)	.14-1	.15-1(.85-3)	.15-1(.98-3)	.22-1(.12-2)	.21-1(.12-2)
	9	.27-1	.30-1(.19-2)	.30-1(.18-2)	.41-1(.23-2)	.38-1(.23-2)	.13-1	.14-1(.84-3)	.15-1(.72-3)	.18-1(.11-2)	.19-1(.12-2)
	10	.18-1	.21-1(.88-3)	.20-1(.94-3)	.23-1(.12-2)	.26-1(.14-2)	.11-1	.13-1(.50-3)	.12-1(.55-3)	.15-1(.82-3)	.14-1(.78-3)

Note that values are given in scientific notation; e.g., .010 is written as .10-1. Values in parenthesis are standard errors.

Risk is estimated by the regression estimate if the regression coefficient b exists (i.e.; b≠number/zero). Otherwise, risk is estimated by the control-variate mse estimate. [See Sections 5.9 and 7.1.] Undefined regression estimate occurs in some cases when generated Dirichlet probability p approximately equals (0,0,1).

TABLE 7A.7

Mean Over 10 Dirichlet Simulations for Quadratic-Loss Estimated MSE.

			NU1				NU2				NU3				NU4			
			Replication 1		Replication 2		Replication 1		Replication 2		Replication 1		Replication 2		Replication 1		Replication 2	
Estimator	SS	PID	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)
APMR0	25	15	.0025	(.0009)	.0024	(.0009)	.0110	(.0029)	.0109	(.0028)	.0180	(.0014)	.0182	(.0013)	.0180	(.0011)	.0174	(.0012)
PMRD0	↓	↓	.0042	(.0016)	.0040	(.0015)	.0147	(.0044)	.0145	(.0042)	.0182	(.0013)	.0181	(.0010)	.0192	(.0004)	.0185	(.0005)
MLE			.0039	(.0013)	.0039	(.0013)	.0140	(.0029)	.0142	(.0029)	.0243	(.0016)	.0244	(.0016)	.0288	(.0008)	.0277	(.0007)
APMR1			.0107	(.0006)	.0109	(.0007)	.0154	(.0014)	.0158	(.0014)	.0204	(.0008)	.0210	(.0007)	.0230	(.0004)	.0221	(.0003)
APMR2			.0042	(.0003)	.0042	(.0003)	.0129	(.0011)	.0130	(.0010)	.0208	(.0022)	.0218	(.0024)	.0233	(.0024)	.0228	(.0025)
PMDR2		↓	.0031	(.0011)	.0030	(.0011)	.0107	(.0025)	.0106	(.0024)	.0190	(.0008)	.0198	(.0010)	.0249	(.0016)	.0242	(.0017)
APMR0		40	.0028	(.0010)	.0027	(.0010)	.0127	(.0035)	.0129	(.0035)	.0211	(.0020)	.0211	(.0018)	.0210	(.0019)	.0203	(.0017)
PMDR0		↓	.0045	(.0018)	.0045	(.0018)	.0175	(.0055)	.0177	(.0055)	.0212	(.0015)	.0212	(.0015)	.0225	(.0011)	.0219	(.0009)
MLE			.0047	(.0015)	.0046	(.0015)	.0174	(.0039)	.0175	(.0038)	.0304	(.0021)	.0306	(.0021)	.0368	(.0012)	.0366	(.0012)
APMR1			.0138	(.0007)	.0139	(.0007)	.0191	(.0017)	.0192	(.0016)	.0247	(.0007)	.0248	(.0008)	.0276	(.0006)	.0272	(.0006)
APMR2			.0049	(.0004)	.0049	(.0004)	.0152	(.0012)	.0153	(.0011)	.0254	(.0032)	.0254	(.0030)	.0284	(.0039)	.0279	(.0035)
PMDR2	↓	↓	.0034	(.0013)	.0034	(.0013)	.0123	(.0031)	.0124	(.0030)	.0228	(.0014)	.0229	(.0012)	.0308	(.0029)	.0304	(.0025)
APMR0	50	15	.0015	(.0005)	.0015	(.0005)	.0060	(.0014)	.0060	(.0013)	.0102	(.0004)	.0106	(.0005)	.0107	(.0002)	.0107	(.0004)
PMDR0	↓	↓	.0023	(.0008)	.0024	(.0008)	.0072	(.0019)	.0070	(.0018)	.0102	(.0006)	.0106	(.0006)	.0111	(.0001)	.0112	(.0002)
MLE			.0020	(.0006)	.0020	(.0006)	.0069	(.0015)	.0070	(.0014)	.0121	(.0007)	.0122	(.0008)	.0139	(.0004)	.0139	(.0002)
APMR1			.0039	(.0005)	.0039	(.0005)	.0073	(.0011)	.0075	(.0011)	.0111	(.0005)	.0113	(.0005)	.0124	(.0003)	.0124	(.0002)
APMR2			.0021	(.0003)	.0021	(.0003)	.0067	(.0007)	.0068	(.0006)	.0113	(.0005)	.0117	(.0006)	.0124	(.0007)	.0125	(.0008)
PMDR2		↓	.0017	(.0006)	.0018	(.0006)	.0059	(.0013)	.0059	(.0012)	.0106	(.0001)	.0109	(.0002)	.0128	(.0004)	.0129	(.0004)
APMR0		40	.0018	(.0006)	.0017	(.0006)	.0072	(.0018)	.0076	(.0018)	.0126	(.0007)	.0122	(.0006)	.0129	(.0006)	.0129	(.0005)
PMDR0		↓	.0027	(.0010)	.0026	(.0009)	.0089	(.0025)	.0091	(.0026)	.0127	(.0007)	.0123	(.0008)	.0135	(.0003)	.0135	(.0002)
MLE			.0024	(.0008)	.0024	(.0008)	.0085	(.0019)	.0089	(.0019)	.0152	(.0010)	.0152	(.0012)	.0176	(.0005)	.0177	(.0006)
APMR1			.0049	(.0005)	.0050	(.0006)	.0089	(.0013)	.0095	(.0013)	.0137	(.0006)	.0134	(.0008)	.0152	(.0003)	.0152	(.0004)
APMR2			.0025	(.0004)	.0025	(.0004)	.0080	(.0008)	.0086	(.0008)	.0142	(.0010)	.0136	(.0008)	.0157	(.0013)	.0156	(.0012)
PMDR2	↓	↓	.0020	(.0007)	.0020	(.0007)	.0071	(.0017)	.0075	(.0017)	.0132	(.0003)	.0127	(.0003)	.0163	(.0008)	.0162	(.0007)

TABLE 7A.8A

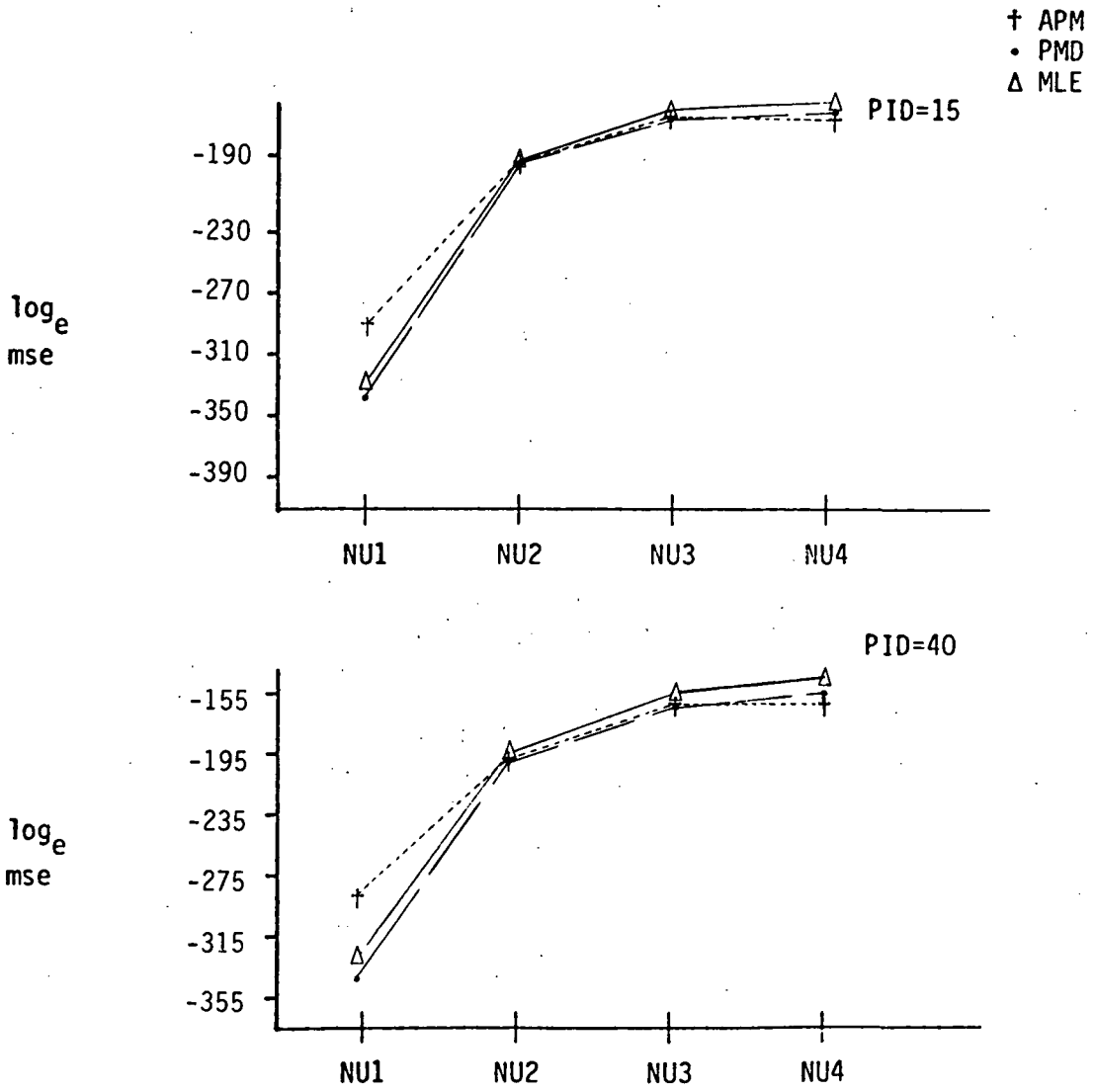
ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED QUADRATIC-LOSS MEAN SQUARED ERRORS FOR ROBUSTNESS SET 0 (ORIGINAL PRIOR IN ESTIMATORS)

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
NU	3	.251377 4	.837924 3	10.499 ***
SS	1	.896719 2	.896719 2	467.367 ***
PID	1	.644600 1	.644600 1	607.009 ***
EST	2	.156130 2	.780652 1	1.879
NU×SS	3	.366810 0	.122270 0	.637
NU×PID	3	.255790 0	.852634 -1	8.029 ***
NU×EST	6	.527583 2	.879306 1	2.116 *
SS×PID	1	.124512 -1	.124512 -1	.118
SS×EST	2	.109549 1	.547744 0	4.703 **
PID×EST	2	.156421 0	.782106 -1	14.588 ***
NU×SS×PID	3	.306224 0	.102075 0	.967
NU×SS×EST	6	.348970 0	.581616 -1	.499
NU×PID×EST	6	.921781 -1	.153630 -1	2.866 **
SS×PID×EST	2	.181913 0	.909564 -1	.901
NU×SS×PID×EST	6	.654974 0	.109162 0	1.081
Pw.NU	36	.287310 4	.798085 2	1672.306 ***
(Pw.NU)×SS	36	.690718 1	.191866 0	4.020 ***
(Pw.NU)×PID	36	.382294 0	.106193 -1	.223
(Pw.NU)×EST	72	.299159 3	.415499 1	87.064 ***
(Pw.NU)×SS×PID	36	.380123 1	.105590 0	2.213 ***
(Pw.NU)×SS×EST	72	.838485 1	.116456 0	2.440 ***
(Pw.NU)×PID×EST	72	.386013 0	.536129 -2	.112
(Pw.NU)×SS×PID×EST	72	.727218 1	.101002 0	2.116 ***
ERROR	480	.229073 2	.477236 -1	
TOTAL	959	.590404 4		

* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

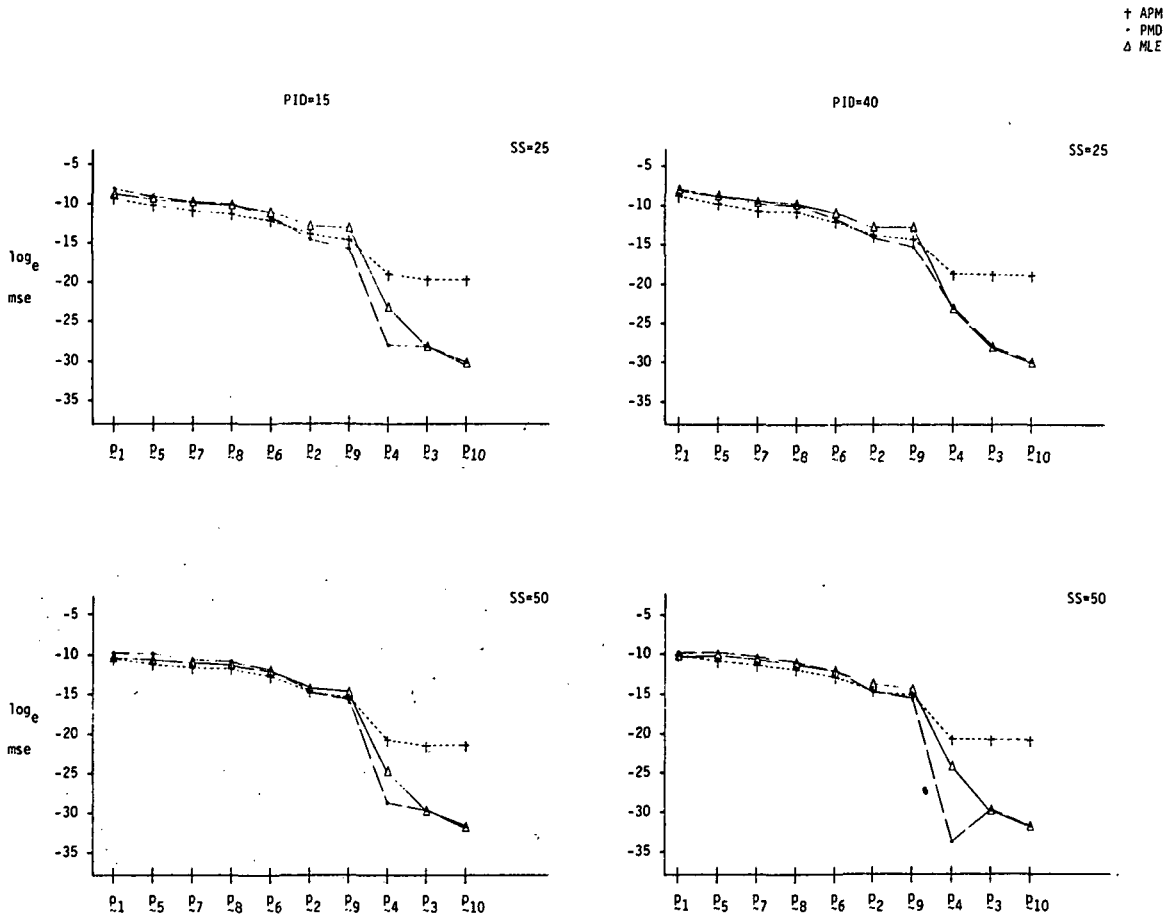
Note that the usual exponential notation is used for the third and fourth columns; for example, 5904.04 is written as .590404 4.

7A.8B PLOT OF NU×EST×PID INTERACTION*



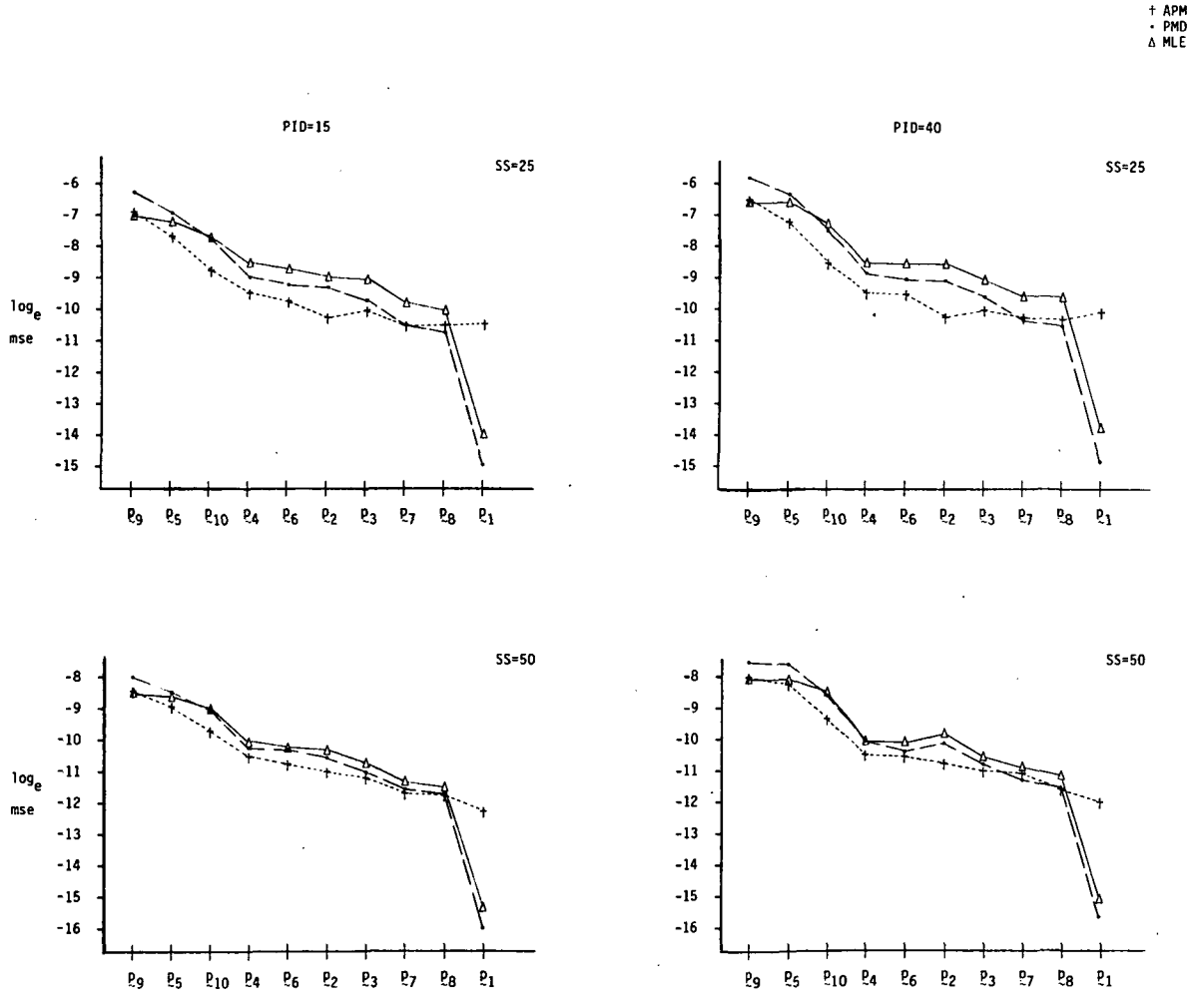
*Values are sums over P, SS, and replication.

7A.8C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(0.1,0.1,9.8).



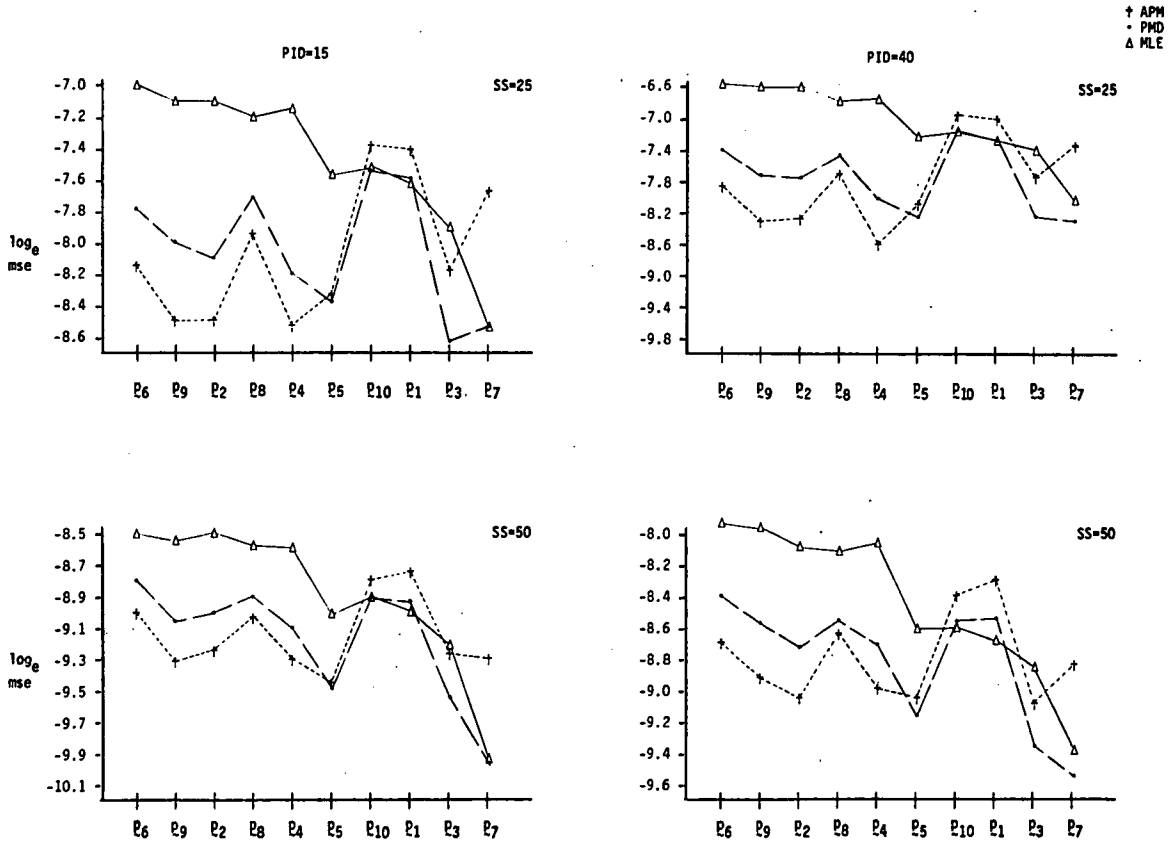
*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.8C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(1.0,1.0,8.0).



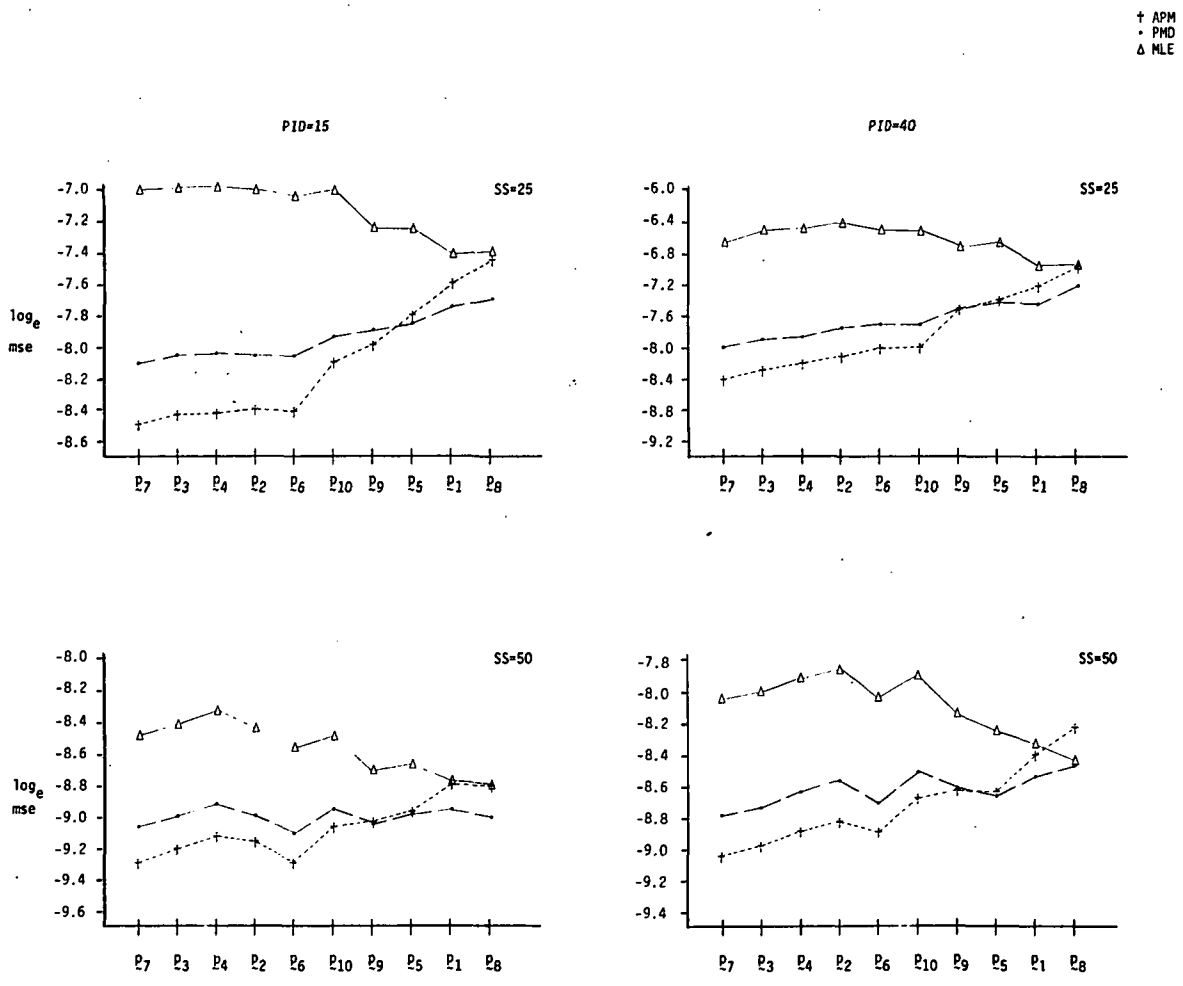
*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.8C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(2.0,3.0,5.0).



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.8C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(10/3,10/3,10/3).



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

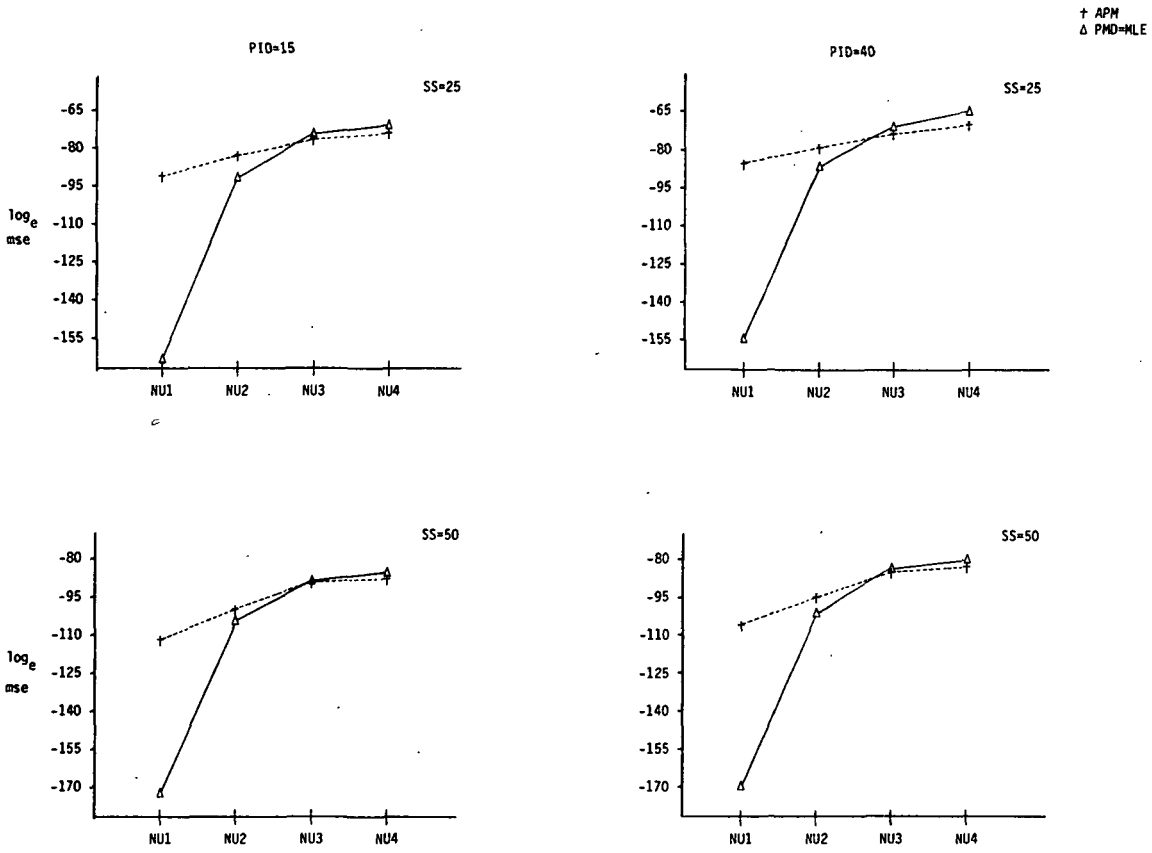
TABLE 7A.9A

ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED QUADRATIC-LOSS MEAN SQUARED ERRORS FOR ROBUSTNESS SET 1 (UNIFORM PRIOR IN ESTIMATORS)

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F.
NU	3	.723959 3	.241320 3	11.557 ***
SS	1	.865921 2	.865921 2	3702.584 ***
PID	1	.678313 1	.678313 1	1111.425 ***
EST	1	.103706 3	.103706 3	6.492 **
NU×SS	3	.129282 1	.430939 0	18.426 ***
NU×PID	3	.205667 -1	.685557 -2	1.123
NU×EST	3	.317635 3	.105878 3	6.628 ***
SS×PID	1	.105919 -2	.105919 -2	.344
SS×EST	1	.185452 0	.185452 0	9.239 ***
PID×EST	1	.818075 -2	.818075 -2	1.769
NU×SS×PID	3	.222064 -2	.740214 -3	.241
NU×SS×EST	3	.160116 1	.533721 0	26.588 ***
NU×PID×EST	3	.172753 0	.575842 -1	12.449 ***
SS×PID×EST	1	.169200 -3	.169200 -3	.280
NU×SS×PID×EST	3	.480949 -2	.160316 -2	2.654 *
Pw.NU	36	.751737 3	.208816 2	6623.127 ***
(Pw.NU)×SS	36	.841929 0	.233869 -1	7.418 ***
(Pw.NU)×PID	36	.219711 0	.610310 -2	1.936 **
(Pw.NU)×EST	36	.575080 3	.159744 2	5066.700 ***
(Pw.NU)×SS×PID	36	.110764 0	.307677 -2	.976
(Pw.NU)×SS×EST	36	.722642 0	.200734 -1	6.367 ***
(Pw.NU)×PID×EST	36	.166521 0	.462560 -2	1.467 *
(Pw.NU)×SS×PID×EST	36	.217426 -1	.603960 -3	.192
ERROR	320	.100891 1	.315283 -2	
TOTAL	639	.257187 4		

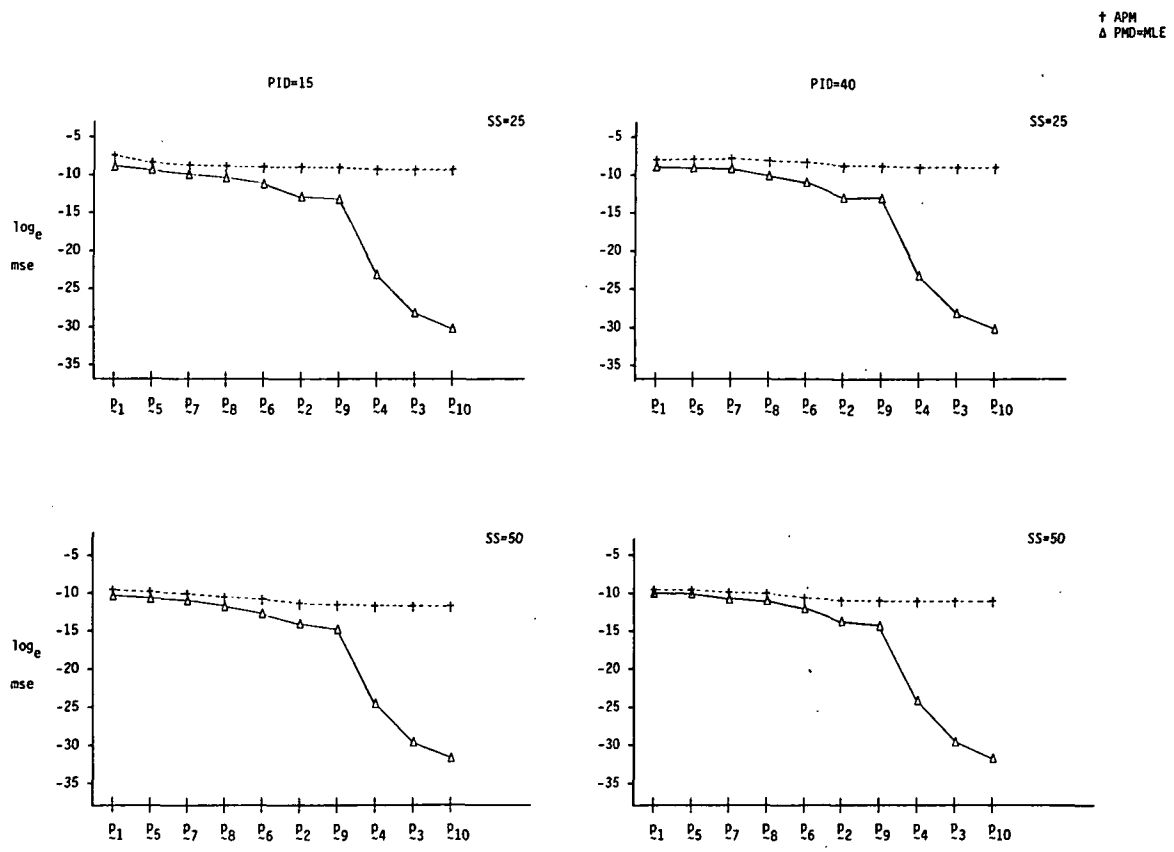
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

7A.98 PLOT OF NU×EST×SS×PID INTERACTION*



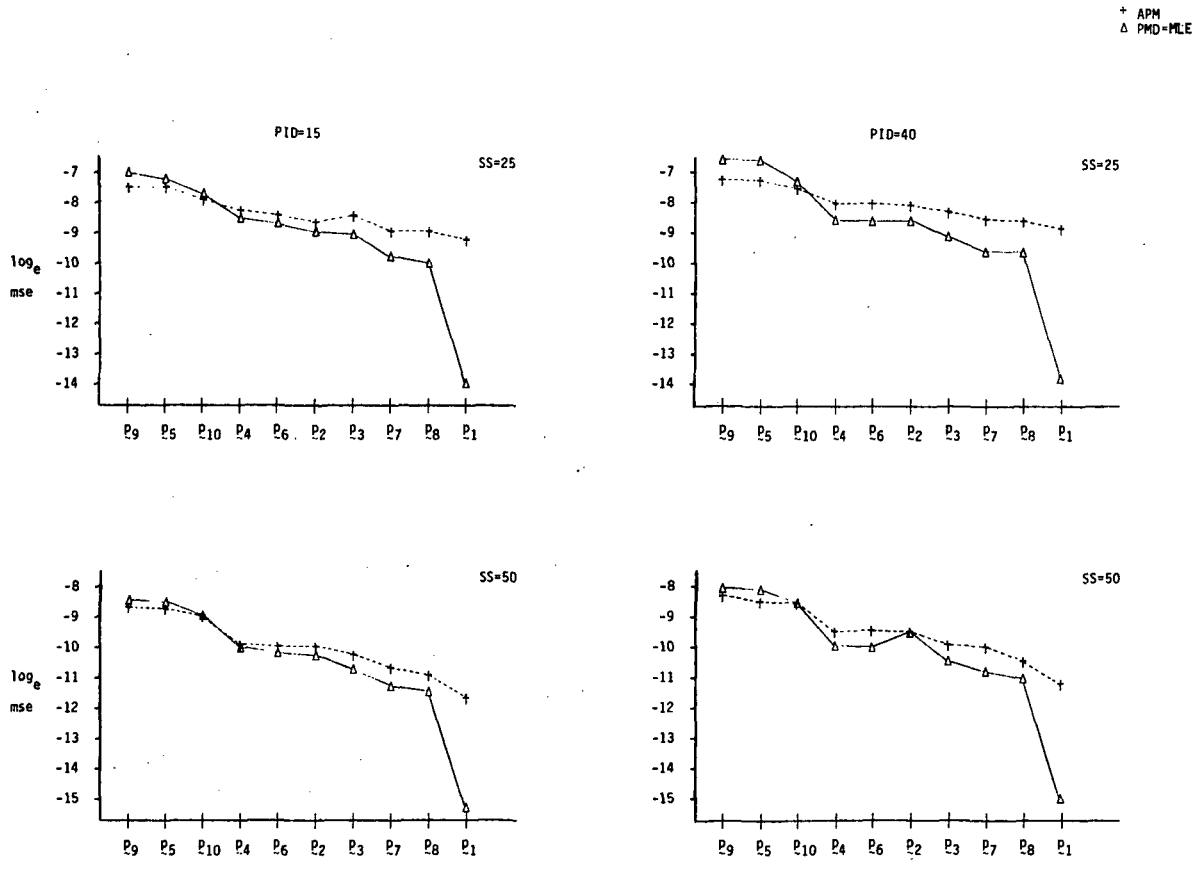
*Values are sums over P and replication.

7A.9C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(0.1,0.1,9.8).



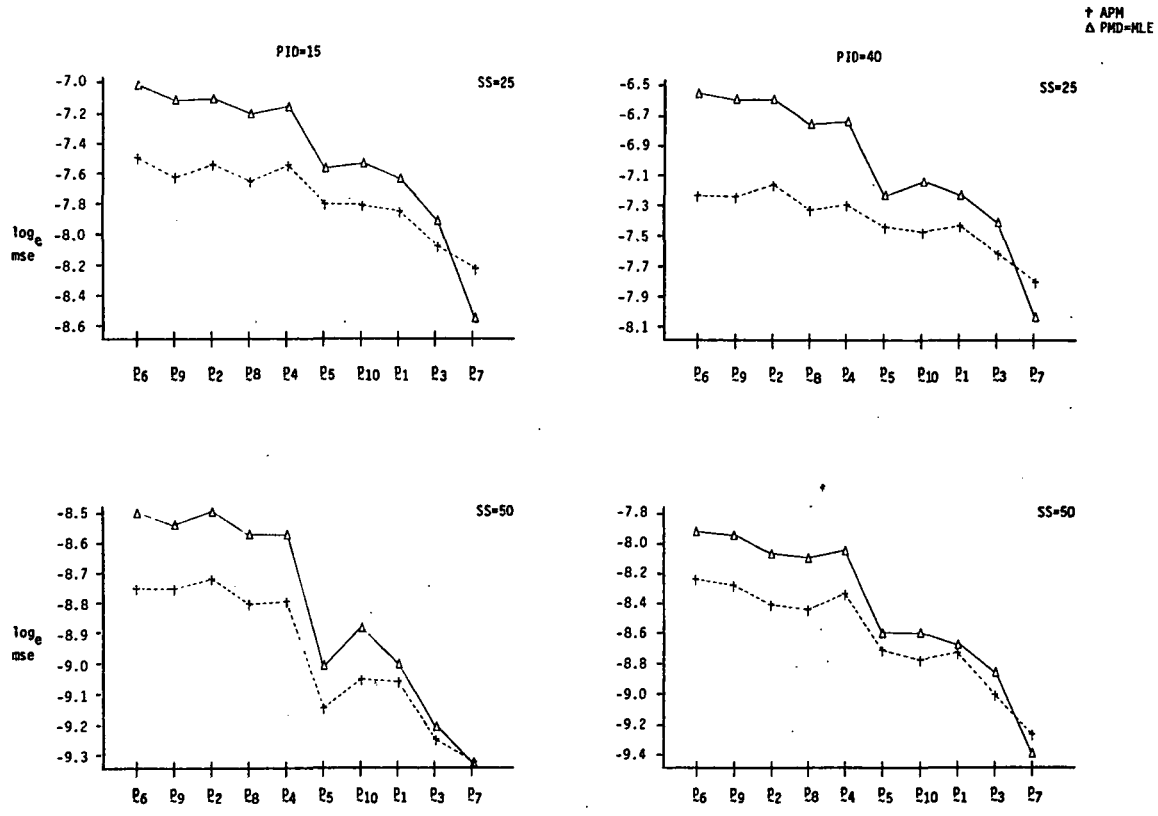
*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.9C PLOT OF (Pw,NU)×EST×SS×PID INTERACTION.* NU=(1.0,1.0,8.0).



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

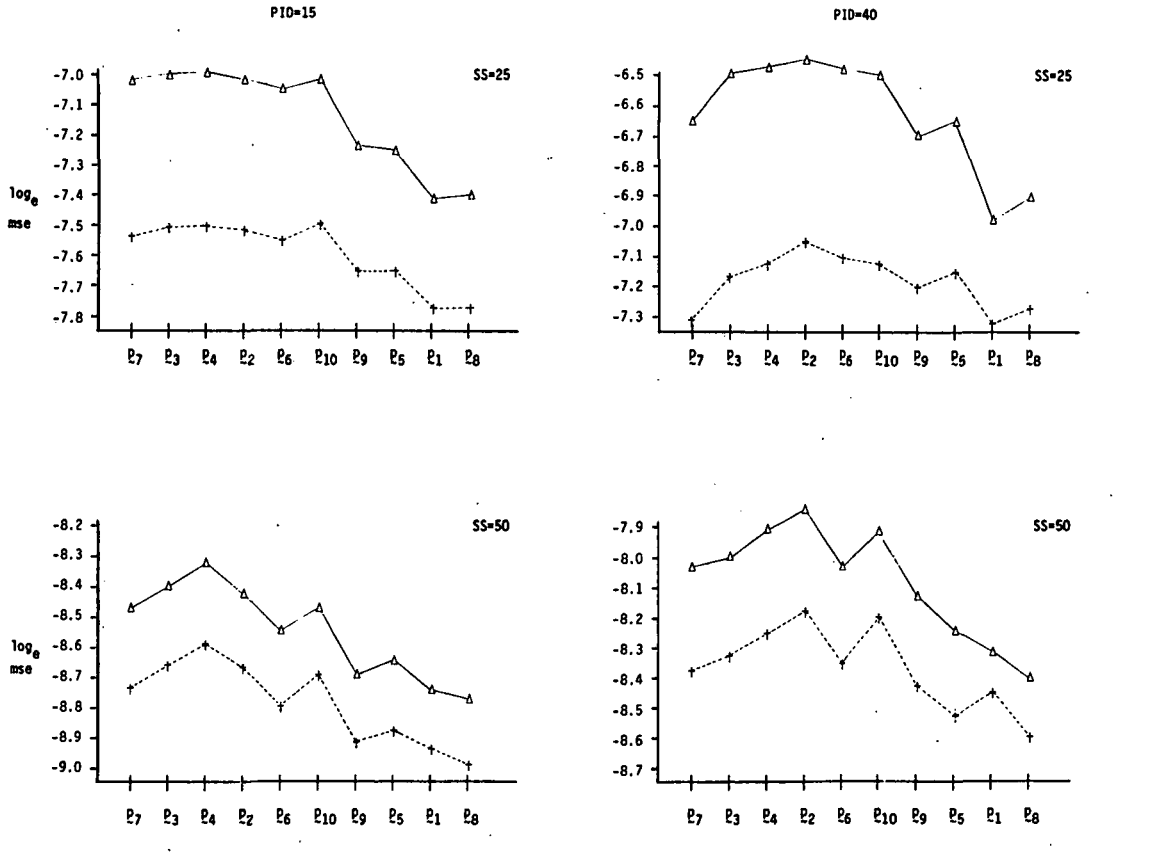
7A.9C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(2.0,3.0,5.0).



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.9C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(10/3,10/3,10/3).

† APM
Δ PMD=MLE



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

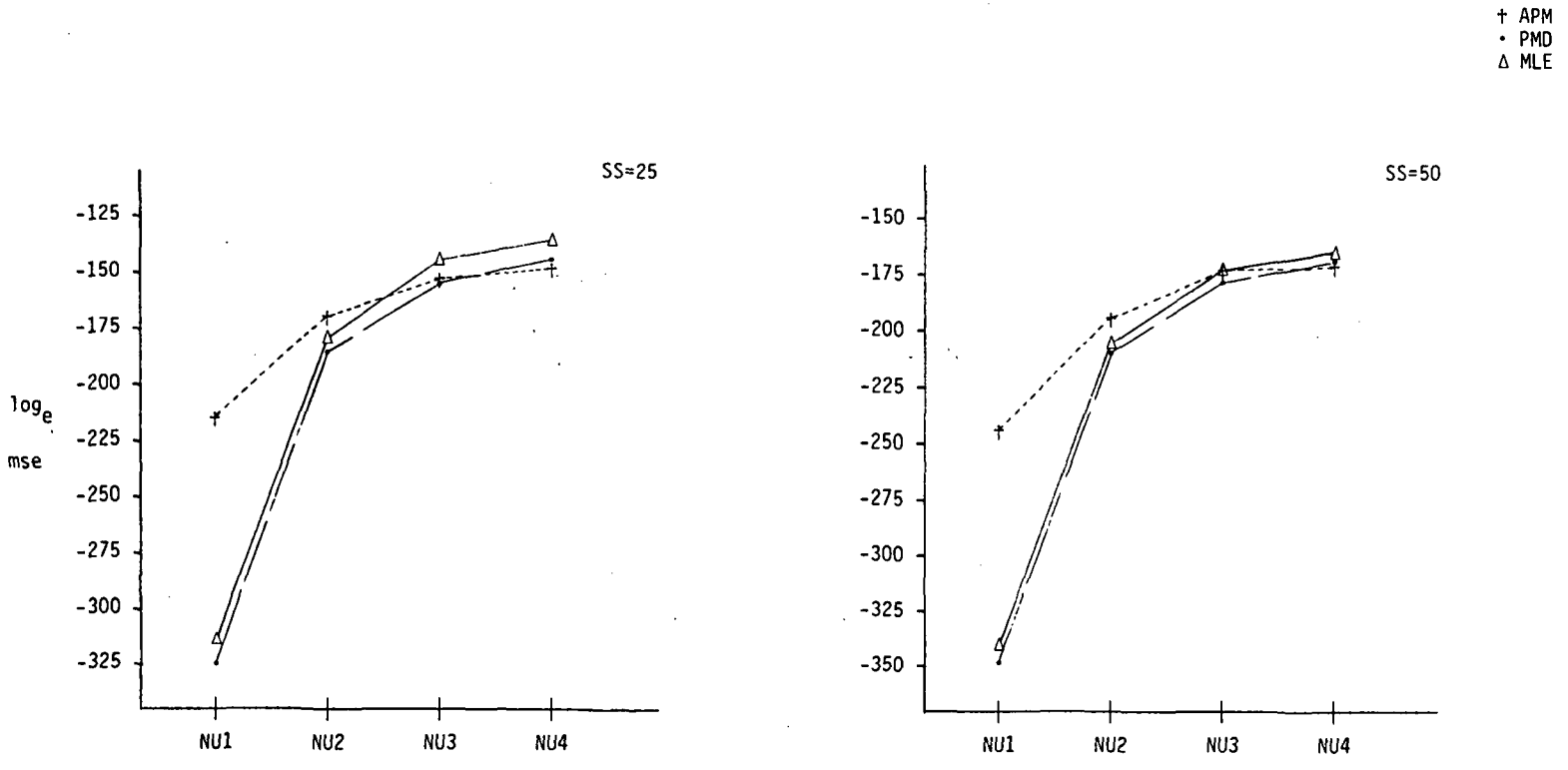
TABLE 7A.10A

ANALYSIS OF VARIANCE FOR NATURAL LOGARITHMS OF ESTIMATED QUADRATIC-LOSS MEAN SQUARED ERRORS FOR ROBUSTNESS SET 2 (PERTURBED PRIOR IN ESTIMATORS)

SOURCE	D.F.	SUM OF SQ.	MEAN SQ.	F
NU	3	.195478 4	.651592 3	13.159 ***
SS	1	.980437 2	.980437 2	1250.280 ***
PID	1	.797488 1	.797488 1	774.940 ***
EST	2	.964996 2	.482498 2	4.445 **
NU×SS	3	.102821 0	.342735 -1	.437
NU×PID	3	.184583 0	.615277 -1	5.979 ***
NU×EST	6	.251424 3	.419040 2	3.860 ***
SS×PID	1	.143585 -1	.143585 -1	1.827
SS×EST	2	.540025 0	.270013 0	8.854 ***
PID×EST	2	.453269 -1	.226634 -1	3.441 **
NU×SS×PID	3	.153708 -1	.512360 -2	.652
NU×SS×EST	6	.353239 0	.588731 -1	1.931 *
NU×PID×EST	6	.725311 -1	.120885 -1	1.835
SS×PID×EST	2	.127573 -1	.637866 -2	1.511
NU×SS×PID×EST	6	.333265 -1	.555442 -2	1.315
Pw.NU	36	.178257 4	.495159 2	7138.288 ***
(Pw.NU)×SS	36	.282303 1	.784174 -1	11.305 ***
(Pw.NU)×PID	36	.370475 0	.102910 -1	1.484 **
(Pw.NU)×EST	72	.781618 3	.108558 2	1564.988 ***
(Pw.NU)×SS×PID	36	.282862 0	.785727 -2	1.133
(Pw.NU)×SS×EST	72	.219563 1	.304948 -1	4.396 ***
(Pw.NU)×PID×EST	72	.474228 0	.658649 -2	.950
(Pw.NU)×SS×PID×EST	72	.304022 0	.422252 -2	.609
ERROR	480	.332960 1	.693667 -2	
TOTAL	959	.498407 4		

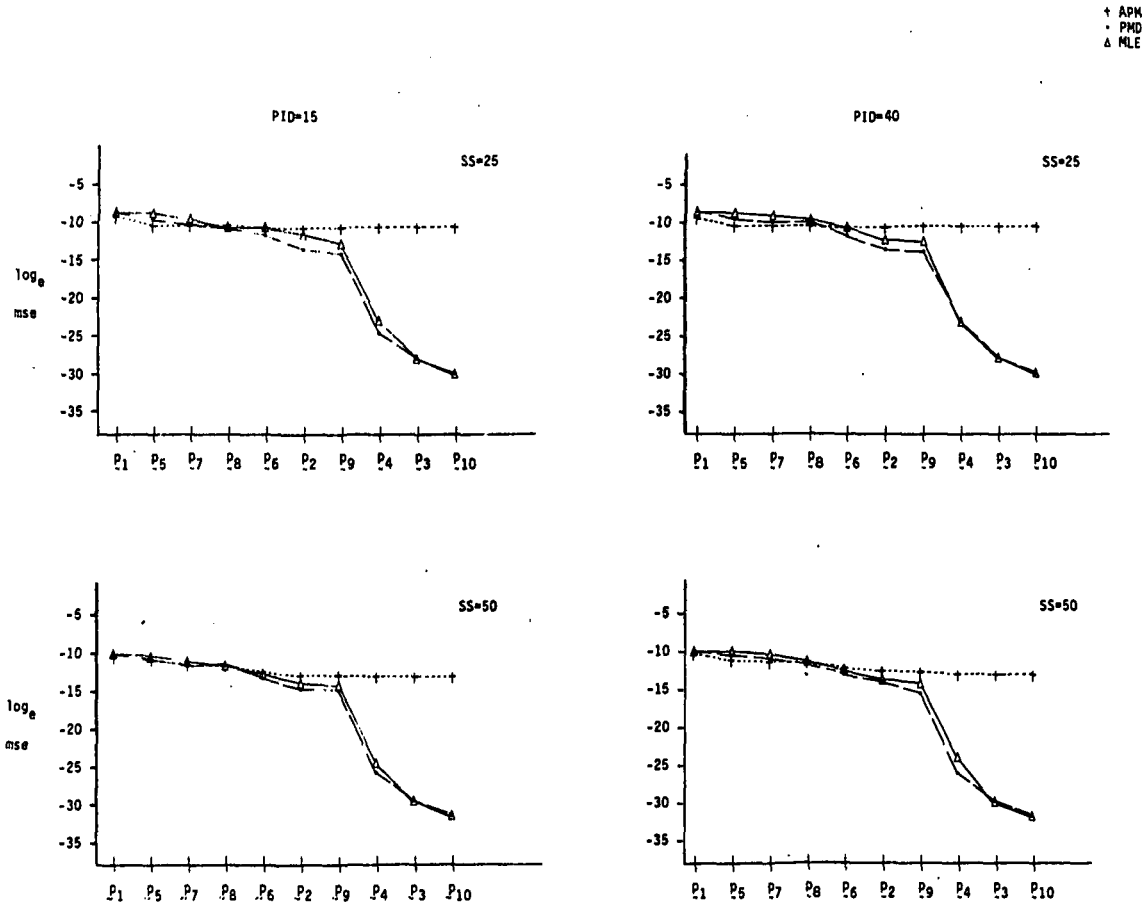
* Significant at 10% level.
** Significant at 5% level.
*** Significant at 1% level.

7A.10B PLOT OF NU×EST×SS INTERACTION*



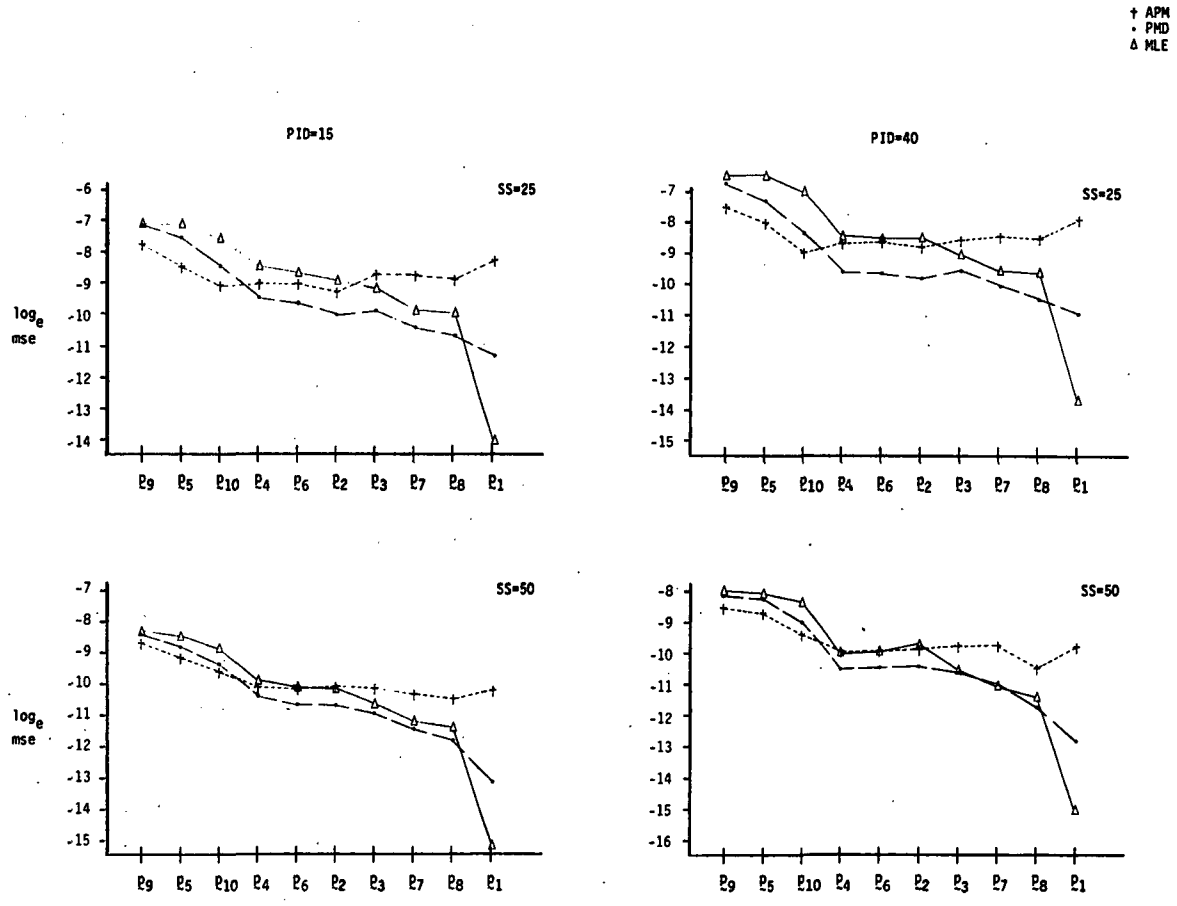
*Values are sums over P, PID, and replication.

7A.10C PLOT OF (Pw.NU)*EST*SS*PID INTERACTION.* NU=(0.1,0.1,9.8).



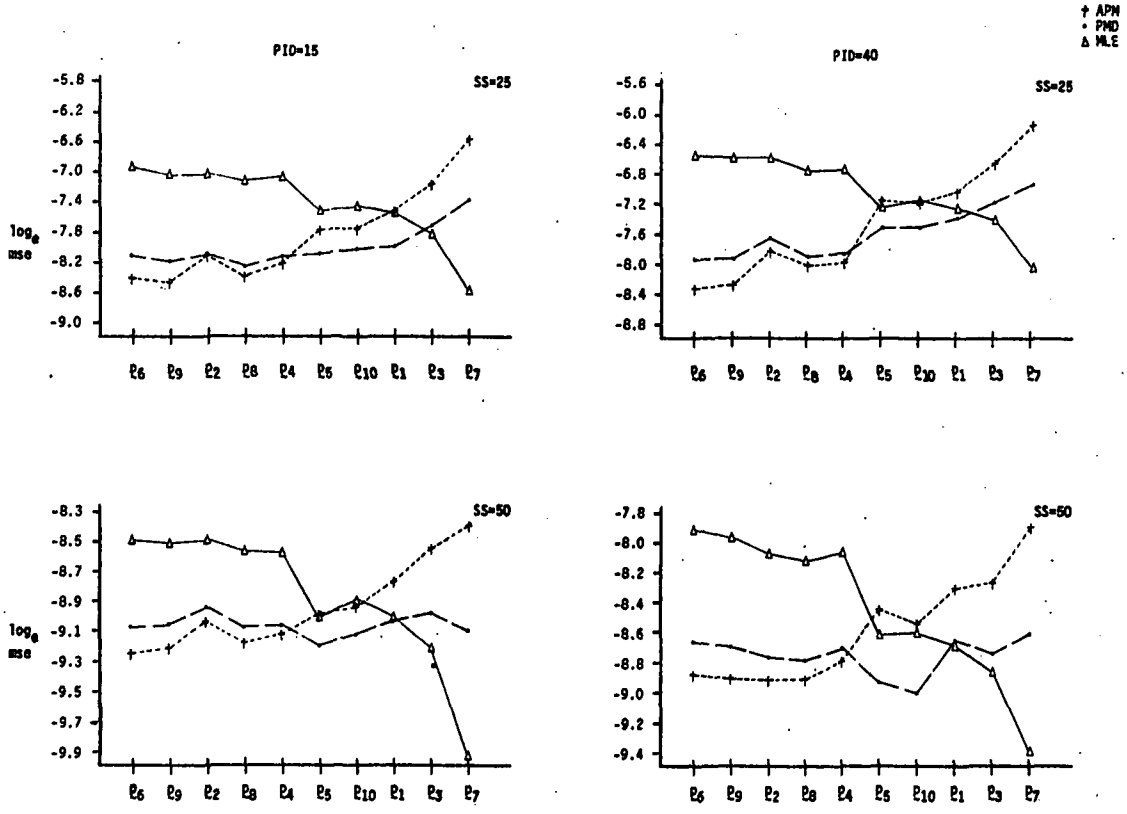
*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.10C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(1.0,1.0,8.0).



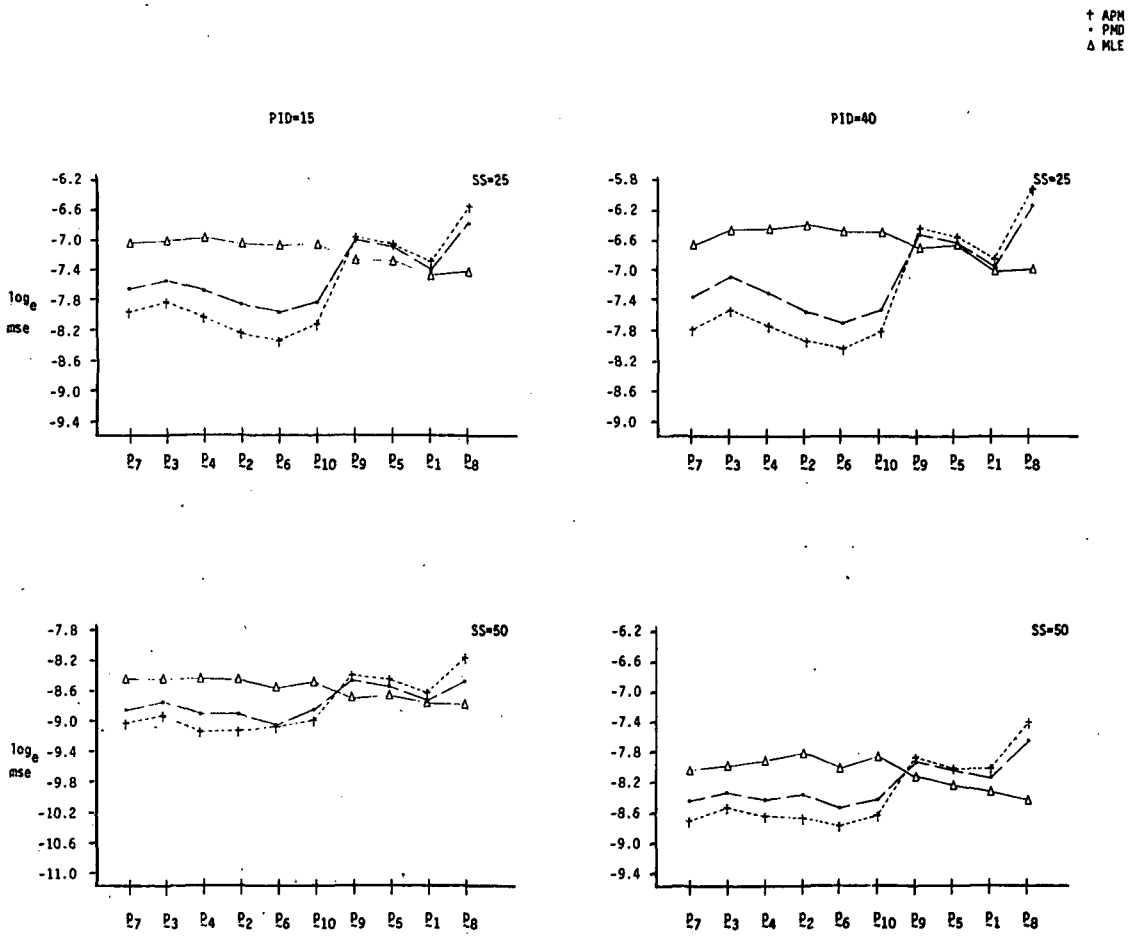
*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.10C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(2.0,3.0,5.0).



*Values are sums over replication. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

7A.10C PLOT OF (Pw.NU)×EST×SS×PID INTERACTION.* NU=(10/3,10/3,10/3).



*Values are sums over replications. Note that p's are ranked in terms of increasing C(p); recall Table 7.1.

CHAPTER 8

SUMMARY AND CONCLUSIONS

In this thesis we considered simultaneous estimation of the vector of multinomial cell probabilities $\underset{\sim}{p}$ from incomplete data, incomplete in that it contains partially classified observations. Each such partially classified observation is observed to fall in one of two or more selected categories but is not classified further. The estimation criterion was minimization of risk $E[L(\underset{\sim}{p},\dot{\underset{\sim}{p}})]$ for quadratic loss $L(\underset{\sim}{p},\dot{\underset{\sim}{p}})=(\underset{\sim}{p}-\dot{\underset{\sim}{p}})'(\underset{\sim}{p}-\dot{\underset{\sim}{p}})$ for the estimator $\dot{\underset{\sim}{p}}$ of $\underset{\sim}{p}$.

The estimators considered were the classical maximum likelihood estimate $\hat{\underset{\sim}{p}}$ and the Bayesian posterior mean $\tilde{\underset{\sim}{p}}$ and posterior mode $\hat{\tilde{\underset{\sim}{p}}}$. We chose the maximum likelihood estimate because it is frequently used in practice. In particular, the maximum likelihood estimate is often used when one has no prior information. Further, Johnson (1971) proved that the complete-data maximum likelihood estimate is admissible; that is, no other estimator can have smaller risk everywhere. The complete-data maximum likelihood estimate is admissible because it has very small risk at the corners of the P_k simplex. We chose the posterior mean because it minimizes expected risk; hence, it must be best for at least some values of $\underset{\sim}{p}$. We chose the posterior mode because it is an in-between estimator. Like the maximum likelihood estimate, it is a mode and can have zero components for a non-zero prior. Like the posterior mean, it can incorporate prior information.

A final reason for choosing these three estimators was that the maximum likelihood estimate $\hat{\underset{\sim}{p}}$, posterior mode $\hat{\tilde{\underset{\sim}{p}}}$, and a Taylor-series approximation $\overset{*}{\underset{\sim}{p}}$ of the posterior mean (discussed below) can all be evaluated by the EM algorithm of Dempster, Laird, and Rubin (1977). This was im-

portant because these three estimators each constitute a nonlinear system of k equations in k unknowns, for which the number of solutions may range from zero to infinity. Further, as illustrated in Section 4D.5, any roots that do exist need not be in P_k. Finally, when roots do exist in P_k there can be difficulty in finding that one for which the likelihood is a maximum. However, Dempster, Laird, and Rubin (1977) proved that if the eigenvalues of the covariance matrix of the complete-data sufficient statistics are bounded above zero, then the EM iterative algorithm converges in P_k to a local maximum. A global maximum is then found by choosing that root in P_k that maximizes the likelihood function

$$\prod_{i=1}^{k+1} \dot{p}_i^{z_i+\alpha_i} \prod_{D} \dot{p}_D^{z_D}$$

where $\dot{p}_i$ denotes one of the three estimators $\hat{p}$, $\tilde{p}$, and $\dot{p}$ and where $\alpha_i=0$ for the maximum likelihood estimate and $\alpha_i=\nu_i-1$ for the posterior mode and Taylor-series approximate posterior mean.

We showed these three estimators to be approximately equal in large samples. To compare these estimators in small- and medium-size samples, we used two Monte-Carlo simulation studies restricted, because of cost constraints, to samples from the trinomial distribution. In the studies, samples were of size 25 and 50, percentages of incomplete data varied around 15 and 40, and probabilities ranged from the center of the P_2 simplex to one of its corners. In the first simulation study, we chose the mean of the prior distribution, given one of four prior parameters, as the probability to be estimated. In the second study we randomly generated ten probabilities from the Dirichlet distribution given each of the

four prior parameters. For each probability, in both studies, we then generated 200 sets of complete and incomplete trinomial data from which an estimate of risk was calculated. Because the prior is not known in practice, we also explored how robust results were to use of the correct prior in calculating the Bayesian estimators. Besides the correct prior, we also used the uniform prior and a perturbed prior in the calculations.

Results indicated that an important factor in determining which estimator was best was the position of $\underset{\sim}{p}$ in the P_2 simplex; in particular, whether $\underset{\sim}{p}$ was at a corner or in the center of P_2. Another important factor was the relationship between the probability $\underset{\sim}{p}$ being estimated and the prior parameters $\underset{\sim}{\beta}$ used in the Bayesian estimators. We studied this relationship in terms of the difference between $\underset{\sim}{p}$ and the mean $\bar{\underset{\sim}{p}}$ of the prior distribution given $\underset{\sim}{\beta}$. The most satisfactory measure of this difference was the difference in the linear centrality measures $C(\underset{\sim}{p})$ and $C(\bar{\underset{\sim}{p}})$ of $\underset{\sim}{p}$ and $\bar{\underset{\sim}{p}}$, respectively, where $C(\underset{\sim}{p})=\sum_{i=1}^{2}\sum_{j>i}^{3}(p_i-p_j)^2$. Results indicated that, except at a corner $\underset{\sim}{p}=(0,0,1)$, when the centrality measure $C(\bar{\underset{\sim}{p}})$ was within a fairly wide range of $C(\underset{\sim}{p})$, then the posterior mean was best. If the difference between the two centrality measures was very large, then the maximum likelihood estimate was best. If the difference was between moderate and very large, the posterior mode was often best when the probability being estimated was toward a corner of P_2. At the $\underset{\sim}{p}=(0,0,1)$ corner, the posterior mode or maximum likelihood estimate was always far better than the posterior mean.

Based on these results, in Section 7.3 we recommended rough operating procedures to guide a practioner in choosing which estimator to use for his data and estimated prior parameters.

Risk was usually reduced by one-third to one-fourth when the best estimator was used instead of the next best estimator and by one-half to one-third when the best estimator was used instead of the worst estimator. However, the reduction was sometimes substantial. Further, the reduction in risk at the corner probability $\underset{\sim}{p}=(0,0,1)$ was huge; the risk of the posterior mean was as much as 33,000 times larger than the risk for the posterior mode or maximum likelihood estimate. [The risk of the maximum likelihood estimate and posterior mode were equal at $\underset{\sim}{p}=(0,0,1)$.] As soon as one moved even slightly away from the corner, however, the risk difference dropped sharply.

As noted, the posterior mean was the best estimator most of the time. In these cases, the posterior mode was usually next best. Other than cross-over probabilities, the smallest difference between the posterior mode and mean was at the center of the P_2 simplex. There, the risk of the posterior mean was reduced only 14% to 23% from that of the posterior mode; whereas the reduction in risk from that of the maximum likelihood estimate ranged from 22% to 42%.

As the percentage of incomplete data increased from 0 to near 40, the risk of the three estimators did not greatly increase and the relationship among the estimators changed little. As sample size increased, risk and the difference between estimators usually decreased.

Because numerical evaluation of the exact posterior central moments

is generally unfeasible, we also developed approximations for elements of the posterior mean and covariance matrices. The best of three approximations considered for the posterior mean was based on a first-order Taylor-series expansion of the exact posterior mean, which we accordingly called the Taylor-series approximate posterior mean $\dot{\tilde{p}}$. Approximations used for elements of the posterior covariance matrix were also based on first-order Taylor-series expansions. An important property of the Taylor-series approximations is that, as the percentage of incomplete data goes to zero, they go to the exact posterior moments. In addition, the relationship between the Taylor-series approximate posterior mean and the posterior mode parallels their complete-data relationship. That is, the Taylor-series approximate posterior mean for a Dirichlet density with prior parameters $(\nu_1,\cdots,\nu_k;\nu_{k+1})$ equals the posterior mode for a Dirichlet density with prior parameters $(\nu_1+1,\cdots,\nu_k+1;\nu_{k+1}+1)$.

To determine the accuracy of the Taylor-series approximate posterior mean, we first found that the Taylor-series expansion of the exact posterior mean had accuracy of magnitude $O(n^{-1})$. Because terms in the expansion were then approximated, the final approximation was not necessarily accurate to order $O(n^{-1})$. However, we showed that this approximation asymptotically equals the exact posterior mean. Further, we gave two conditions which guarantee that the error between the exact posterior mean and an iterative solution of the Taylor-series approximate posterior mean is of magnitude $O(n^{-1})$. The two conditions, given by Lemma 4E.1, concern the region in which the initial iterative estimate is chosen and a bound on the partial derivatives of the Taylor-series approximation.

If a neighborhood $||\dot{\tilde{p}}-\tilde{\tilde{p}}||_\infty<\rho$, for $\rho>0$, of the exact posterior mean $\tilde{\tilde{p}}$ can be found such that for all probabilities $\dot{\tilde{p}}$ in this neighborhood

$$\max_{1\leq i\leq k} \sum_{j=1}^{k} |\partial g_i(\dot{\tilde{p}})/\partial \dot{p}_j| \leq \lambda < 1,$$

for

$$g_i(\dot{\tilde{p}})=(z_i+\nu_i+\sum_{D\ni i} z_D\dot{p}_i/\dot{p}_D)/(n+\sum_{h=1}^{k+1}\nu_h),$$

and if an initial iterative estimate $\dot{p}_i^{(0)}$ is chosen within the inner neighborhood $||\dot{\tilde{p}}-\tilde{\tilde{p}}||_\infty<\rho_0\leq\rho-\delta/(1-\lambda)$ where δ is a bound on the error in approximating the exact posterior mean by a first-order Taylor series, then the iterative solution to the defining equations of the Taylor-series approximate posterior mean $\dot{\tilde{p}}$ will converge to within $O(n^{-1})$ of the exact posterior mean.

If a neighborhood of the exact posterior mean can be found in which the λ bound is satisfied, then for large enough sample sizes, the second condition can be satisfied by choosing an initial iterative estimate within the first neighborhood. Even for medium-size samples, the inner neighborhood is almost as large as the outer neighborhood if the percentage of incomplete data is moderate. In Appendix 4E, we showed how to determine, in practice, whether the second condition can be expected to hold.

As for the condition for the EM algorithm, the conditions of Lemma 4E.1 need not be met. In fact, there may not even exist any neighborhood of the exact posterior mean in which the λ bound holds, as we illustrated for an 11-dimensional multinomial problem. However, Appendix 4E showed that this was not the case for incomplete trinomial data; there does exist a root in P_2 of the Taylor-series approximate posterior mean that differs

from the exact posterior mean by magnitude $O(n^{-1})$. However, this root need not be unique in P_2; hence, finding it can be difficult. In these cases in P_2 and in higher dimensions, because the complete-data relationship between the posterior mode and posterior mean was paralleled by the relationship between the posterior mode and the Taylor-series approximate posterior mean for incomplete data (i.e., the Taylor-series approximate posterior mean can be written as a posterior mode), we intuitively expect that that root that is in the guaranteed-convergence region of the exact posterior mean, or at least the closest root to $\tilde{p}$, is given by whichever root in P_k maximizes the likelihood function $\prod_{i=1}^{k+1} \tilde{p}_i^{z_i+\nu_i-1} \cdot \tilde{p}_D^{z_D}$.

Finally, we gave examples showing that Lemma 4E.1 gives extremely conservative bounds on the error between the exact posterior mean and the converged iterative estimate and on the region in which an initial iterative estimate can be chosen so that successive iterates converge to within a small error of $\tilde{p}$.

Approximations used for elements of the posterior covariance matrix were based on Taylor-series expansions that were accurate to order $O(n^{-3/2})$. When the iterative solution for the Taylor-series approximate posterior mean has accuracy of magnitude $O(n^{-1})$, then the Taylor-series approximate posterior variance and covariance can be evaluated noniteratively to have accuracy of magnitude $O(n^{-3/2})$. These approximations can also be evaluated iteratively. However, insurance of accuracy of magnitude $O(n^{-3/2})$ then depends on satisfaction of the two conditions of Lemma 4E.1, where $g(\tilde{\tilde{p}})$ is replaced by the proper function.

In the same Monte-Carlo simulation used for the risk study, the

Taylor-series approximation for the posterior mean was usually accurate to at least four significant figures; that for the posterior variance, to at least three significant figures; and that for the posterior covariance, to at least two significant figures. In practice, the Taylor-series approximations will generally be more accurate than numerical evaluation of the corresponding exact posterior moments.

Note that, although the maximum likelihood estimate and posterior mode asymptotically equal the exact posterior mean (and, hence, the Taylor-series approximate posterior mean), neither was a good approximation of the exact posterior mean in the small- and medium-size samples studied in the simulation. Further, as the percentage of incomplete data goes to zero, neither go to the exact posterior mean. Finally, neither relate to the posterior mode in the same manner that the complete-data posterior mean relates to the complete-data posterior mode.

Among areas for future work are extensions of the simulation study to (1) more priors for the distribution of the data and for use in the Bayesian estimators, (2) investigation of the use of the linear centrality measure $C(\underset{\sim}{p})$, and (3) higher dimensions on P_k.

Between Design 1 and Design 2, nearly all types (corner, noncorner boundary, center, and in-between) probabilities were covered in the simulation studies. We do not expect different results for different values of the same type of probability. For example, we expect results for the probability (1,0,0) to be similar to those for the corner probability (0,0,1). *One type of probability not covered was the middle of a side*; e.g., (.00,.51,.49). However, this probability is further from a corner

than were the side probabilities (.00,.15,.85), (.04,.00,.96), (.00,.07, .93), and (.18,.00,.82) that were included in Design 2. Therefore, we expect the posterior mean to be the best estimator for a middle-of-a-side probability for even more values of the prior parameter $\underset{\sim}{\beta}$ used in the Bayesian estimators than were for these four. The effect of the size of the "prior-sample size" $\Sigma\nu_j$ relative to the size n of the current data sample was also thought to be adequately addressed. If the ratio $\Sigma\nu_j/n$ is much smaller, then the prior will have little effect on results. If the ratio is much larger, then the data will have little effect. It might, however, be valuable to look at more types of priors. For example, why were the results for the posterior mode when $C(\underset{\sim}{p})=.09$ in Design 2 [see Figure 7.1 and $\underset{\sim}{\nu}_4$ plot in Table 7.6] inconsistent with results for the posterior mode for neighboring values of $C(\underset{\sim}{p})$? Was this inconsistency because probabilities near the center of P_2 were more sensitive to use of wrong priors than probabilities elsewhere in P_2? [Recall the tightness of the prior distribution of $\underset{\sim}{p}$ given $\underset{\sim}{\nu}_4\equiv(10/3,10/3,10/3)$.]

To examine risk as a function of individual values of $\underset{\sim}{p}$, we used the linear centrality measure $C(\underset{\sim}{p})$. This measure reduces a probability in essentially two-dimensional space to one dimension. Thus, there are many probabilities $\underset{\sim}{p}$ that map into one value of $C(\underset{\sim}{p})$. It could be that the values of risk for these many probabilities differ greatly. If so, then $C(\underset{\sim}{p})$ would not be useful for measuring risk as a function of $\underset{\sim}{p}$; in particular, for describing the relationship between risk, the value of the probability being estimated, and the prior used in the Bayesian estimators. For those probabilities that were studied in P_2, however, $C(\underset{\sim}{p})$ was

a very good measure, as evidenced by plots in Tables 7.4 - 7.6. Risk was a smooth function of $C(\underset{\sim}{p})$ and for nearly all values of $\underset{\sim}{p}$ that had the same $C(\underset{\sim}{p})$, the risk, for a given estimator and prior, was approximately the same. A slight exception did occur, however, for the posterior mode and posterior mean at $\underset{\sim}{\nu}_3$ for the probabilities $\underset{\sim}{p}$=(.23,.42,.35) and $\underset{\sim}{p}$=(.08, .61,.31), both having $C(\underset{\sim}{p})$=.44, when the correct prior was used in the Bayesian estimators. [See $\underset{\sim}{\nu}_3$ plots in Table 7.4 and $\underset{\sim}{p}_5$ and $\underset{\sim}{p}_{10}$ in the $\underset{\sim}{\nu}_3$ plot in Table 7A.9; however, note that the risk was the same for these two probabilities when the perturbed and uniform priors were used. Hence, the unequal results when the correct prior was used could be due to a poor estimate of risk for one of these probabilties.] Thus, there might be other problems in using $C(\underset{\sim}{p})$ in P_2 that were not encountered in this study. Would there be any problems in using $C(\underset{\sim}{p})$ in higher dimensions? A good linear measure of $\underset{\sim}{p}$ is even more important in higher dimensions, where risk could otherwise be much more difficult to relate to $\underset{\sim}{p}$ in a simple manner. Note that, in P_2, $C(\underset{\sim}{p})$ was a much better measure of $\underset{\sim}{p}$ for use in analyzing risk than was the maximum, minimum, component differences, absolute component differences, or component-squared sums. Either the relationship between risk and these other measures was less smooth than that with $C(\underset{\sim}{p})$ [recall plots in Tables 7.4 - 7.6] or, unlike with $C(\underset{\sim}{p})$, usually more than one value of risk corresponded to one value of these measures.

We are especially interested in how results from the simulation study carry over to higher dimensions. However, note that several numerical problems found in this study are likely to be even worse in higher

dimensions. There will almost surely be more multiple roots of the defining equations for the estimators. If there are more in P_k, then there will be greater difficulty locating the global maximum. More initial iterative estimates will have to be tried to insure that all local maximum are found and then each of these local maximum will have to be checked to see if it is the root that maximizes the likelihood. Since P_k becomes increasingly large as k increases, the search for all local maximum could be long. Hence, study is needed to examine the roots found by the EM algorithm. Are there many in P_k or are all but one outside of P_k? For incomplete trinomial data in Appendix 4D, there was one and only one root in P_2 out of three to five roots for the maximum likelihood estimate (asymptotic posterior mean), excluding the root (0,0,1) which was eliminated upon consideration of the data.

Since there are more components to a probability in P_k, convergence problems may increase. Finding an initial iterative estimate that has each component close to the corresponding component of $\underset{\sim}{p}$ is more difficult in higher dimensions; e.g., trying to approximate 11 components entails more error than trying to approximate only two components. Under what conditions is $\hat{v}_i / \sum_{j=1}^{k+1} \hat{v}_j$ from the estimated prior or, in many cases, $z_i + \sum_{D \ni i} z_D (z_i / \sum_{j \in D} z_D)$ a good initial iterative estimate? Thus, how sensitive to the initial iterative estimate is convergence of the EM algorithm in higher dimensions? How does the number of iterations increase with an increase in the number k of dimensions? Are there more problems in higher dimensions satisfying the conditions guaranteeing that the EM algorithm will converge to a local maximum in P_k?

Similarly, there may be more problems in approximating the exact posterior mean in higher dimensions. We showed by example in Chapter 4 that in higher dimensions it will be increasingly difficult to find a region of the exact posterior mean in which an initial iterative estimate picked guarantees convergence of the EM algorithm to within a small error of the exact posterior mean. However, we also showed by examples in Appendix 4E that this lemma gives extremely conservative bounds on the guaranteed-convergence region. Initial iterative estimates were picked far outside the guaranteed-convergence sphere and the EM algorithm still converged to the exact posterior mean within the same small error. How much does the conservatism of the guaranteed-convergence region carry over to higher dimensions? In particular, when there does not exist a guaranteed-convergence region, are there any initial iterative estimates for which the EM algorithm will converge to the exact posterior mean within a small error? If the Taylor-series approximate posterior mean is a poor approximation in higher dimensions, can a good approximation be found? As illustrated in Section 2.2.4, as the number of dimensions increases, the exact posterior moments become increasingly expensive to evaluate. Thus, good approximations become increasingly important. Finally, when multiple roots of the defining equation of the Taylor-series approximate posterior mean exist in P_k, is, as speculated, the root that is closest to the exact posterior mean that root that maximizes the likelihood function?

Finally, we assumed in this work (recall Section 1.2) that all incomplete data was incomplete at random. Another area of study, therefore, concerns incomplete data where the incompleteness of an observation is not random but instead depends on the value that would have been observed.

REFERENCES

Abramowitz, M. and Stegun, I. A., ed. (1970): Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables. U.S. Dept. of Commerce, National Bureau of Standards.

Ahrens, J. H. and Dieter, U. (1974): Computer Methods for Sampling From Gamma, Beta, Poisson, and Binomial Distributions, Computing, 12, 223-246.

Anderson, T. W. (1958): An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Bartle, R. G. (1966): The Elements of Integration. Wiley, New York.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975): Discrete Multivariate Analysis: Theory and Practice. M.I.T. Press, Cambridge, Mass.

Blumenthal, S. (1968): Multinomial Sampling With Partially Categorized Data, J. Amer. Statist. Assoc., 63, 542-551.

Buck, R. C. (1965): Advanced Calculus. McGraw-Hill, New York.

Cochran, W. G. (1967): Sampling Techniques. Wiley, New York.

Computer Programing Manual (1975): Programing Efficiency. vol. 1, Sec. 8.1, NASA, Langley Reseach Center, Hampton, Va.

Coveyou, R. R. and MacPherson, R. D. (1967): Fourier Analysis of Uniform Random Number Generators, J. Assoc. Computing Machinery, 14, 100-119.

Cox, D. R. and Hinkley, D. V. (1974): Theoretical Statistics. Chapman and Hall, London.

Cramer, H. (1951): Mathematical Methods of Statistics. Princeton Univ. Press, Princeton.

CRC Standard Mathematical Tables (1962): Hodgman, C. D. ed., Chemical Rubber Publishing Co., Cleveland.

Credeur, K. R. (1978): Computer Programs for Estimation From Incomplete Multinomial Data. NASA TM-X ...., Langley Research Center, Hampton Va., in preparation.

DeGroot, M. H. (1970): Optimal Statistical Decisions. McGraw-Hill, New York.

Dempster, A. P. (1969): Elements of Continuous Multivariate Analysis. Addison-Wesley, Reading, Mass.

———, Laird, N. M., and Rubin, D. B. (1977): Maximum Likelihood From Incomplete Data Via the EM Algorithm, J. Royal Statist. Soc., Ser. B, 39, 1-22.

———, Schatzoff, M., and Wermuth, N. (1977): A Simulation Study of Alternatives to Ordinary Least Squares, J. Amer. Statist. Assoc., 72, 77-91.

Efron, B. (1977): Discussion on the paper by Dempster, Laird, and Rubin, J. Royal Statist. Soc, Ser. B, 39, 29.

Feller, W. (1968): An Introduction to Probability Theory and Its Applications, Vol. 1. Wiley, New York.

Golder, E. R. (1976): The Spectral Test for the Evaluation of Congruential Pseudo-Random Generators, Applied Statist., 25, 173-180.

Gradshteyn, I. S. and Ryzhik, I. M. (1967): Tables of Integrals, Series, and Products. Academic, New York.

Graybill, F. A. (1969): Introduction to Matrices With Applications in Statistics. Wadsworth, Belmont, Calif.

Gunst, R. F. and Mason, R. L. (1977): Biased Estimation in Regression: An Evaluation Using Mean Squared Error, J. Amer. Statist. Assoc., 77, 616-628.

Hartley, H. O. (1958): Maximum Likelihood Estimation From Incomplete Data, Biometrics, 14, 174-194.

Hoaglin, D. C. (1977): Direct Approximations for Chi-Squared Percentage Points, J. Amer. Statist. Assoc., 77, 508-515.

——— and King, M. L. (1978): A Remark on Algorithm AS 98: "The Spectral Test for the Evaluation of Congruential Pseudo-Random Generators", Applied Statist. 27, No. 3, to appear.

Hocking, R. R. and Oxspring, H. H. (1971): Maximum Likelihood Estimation With Incomplete Multinomial Data, J. Amer. Statist. Assoc., 66, 65-70.

Isaacson, E. and Keller, H. B. (1966): Analysis of Numerical Methods. Wiley, New York.

Jeffreys, H. (1939): Theory of Probability. University Press, Oxford.

Johnson, B. McK. (1971): On the Admissible Estimators for Certain Fixed Sample Binomial Problems, Annals of Mathematical Statistics, 42, 1579-1587.

Kendall, M. G. and Stuart, A. (1969): The Advanced Theory of Statistics, Vol. I. Griffin, London.

Kleinjnen, J. P. C. (1975): Statistical Techniques in Simulation, Parts I and II. Marcel Dekker, New York.

Knuth, D. E. (1969): The Art of Computer Programming, Vol. 2 (Semi-Numerical Algorithms). Addison-Wesley, Reading, Mass.

Korn, G. A. and Korn, T. M. (1968): Mathematical Handbook for Scientists and Engineers. McGraw-Hill, New York.

Lindley, D. V. (1965): Introduction to Probability and Statistics From a Bayesian Viewpoint, Part 1, Probability. Cambridge University Press, Cambridge.

MACSYMA Reference Manual (1975): The Mathlab Group, Project MAC - M.I.T., Cambridge, Mass.

MACSYMA Primer (1975): The Mathlab Group. Project MAC - M.I.T., Cambridge, Mass.

MACSYMA manual (1976): An Introduction to ITS for the MACSYMA User. Project MAC - M.I.T., Cambridge, Mass.

Mood, A. M. and Graybill, F. A. (1963): Introduction to the Theory of Statistics. McGraw-Hill, New York.

Ortega, J. M. and Rheinboldt, W. C. (1970): Iterative Solution of Non-Linear Equations in Several Variables. Academic, New York.

Parzen, E. (1962): Stochastic Processes. Holden-Day, San Francisco.

Rao, C. R. (1968): Linear Statistical Inference and Its Applications. Wiley, New York.

Rubin, D. B. (1976): Inference and Missing Data, Biometrika, 63, 581-592.

Scheffé, H. (1959): The Analysis of Variance. Wiley, New York.

Scheid, F. (1968): Theory and Problems of Numerical Analysis. McGraw-Hill, New York.

Schmetterer, L. (1974): Introduction to Mathematical Statistics. Springer-Verlag, New York.

Snedecor, G. W. and Cochran, W. G. (1968): Statistical Methods, Iowa State University Press, Ames, Iowa.

Steel, R. G. D. and Torrie, J. H. (1960): Principles and Procedures of Statistics. McGraw-Hill, New York.

Sundberg, R. (1974): Maximum Likelihood Theory for Incomplete Data From an Exponential Family, Scand. J. Statist., 1, 49-58.

Wilks, S. S. (1962): Mathematical Statistics. Wiley, New York.

Zellner, A. (1971): An Introduction to Bayesian Inference in Econometrics. Wiley, New York.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM 70703		

4. Title and Subtitle	5. Report Date
ESTIMATION FROM INCOMPLETE MULTINOMIAL DATA	4/10/78
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Karen Rackley Credeur	
	10. Work Unit No.
9. Performing Organization Name and Address	
NASA Langley Research Center Hampton, Va. 23665	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Technical Memorandum
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

dissertation in partial fulfillment of the Ph.D. requirements of Harvard University

16. Abstract

Estimators are derived and compared for the vector of multinomial cell probabilities $\underset{\sim}{p}$ from incomplete data, incomplete in that it contains partially classified observations. The estimation criterion is minimization of risk for quadratic loss. The estimators are the classical maximum likelihood estimate, the Bayesian posterior mode, and the posterior mean. The Dirichlet, the conjugate prior for the multinomial distribution, is assumed for the prior distribution. In addition, approximations are developed for elements of the posterior mean and covariance matrices.

The estimators and approximations are compared in large samples by theory and in small- and medium-size samples through Monte-Carlo simulation studies for the trinomial distribution. Sample size, percentage of incomplete data, and probability are varied. Probabilities equal the means of the prior distributions or are randomly generated from these distributions. The correct prior, a uniform prior, and a perturbed prior are used in the Bayesian estimators.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
estimation, incomplete data, multinomial, Bayesian, simulation	unclassified unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
unclassified	unclassified	343	$12.00

*For sale by the National Technical Information Service, Springfield, Virginia 22161