Singletons

Vladimir Voevodsky

Started January 4, 2008

Contents

0.1 Introduction

Cornuet, J.M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.M., Balding, D.J., Guillemaud, T., and Estoup, A. (2008). Inferring population history with DIY ABC: A user-friendly approach to Approximate Bayesian Computation. Bioinformatics 24, 27132719.

Verdu et al., Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa, Current Biology (2009), doi:10.1016/j.cub.2008.12.049

The goal of this paper is to solve the following problem. Consider a population of identical age-less individuals (singletons) where each individual can go through one of the two possible transformations - it can die or it can divide into two. Suppose that the past history of the population was determined by the conditions that the birth (division) rate was constant and equal to 1 and the death rate was an unknown function of time $d(t)$. Suppose further that we know the ancestral tree of the present day population i.e. for each pair of singletons we know the time distance from the present to their "last common ancestor". Given this data what is the maximal likelihood reconstruction of the death rate function?

My interest in this problem originated from multiple recent papers which attempt to use the variation in the non-recombinant genetic loci to reconstruct histories of populations. While there are several standard models which the authors use to interpret the experimental data none of these models is adapted to address the most interesting question - how the population size changed in time? The singleton model outlined above is clearly the simplest possible one where the time is continuous and the population size is allowed to vary. While for the actual reconstruction problems one may need to consider more sophisticated models it seems clear that all the *negative* results obtained in the framework of singletons are likely to remain valid in more complex cases. For example, if one can show that for a given size of the present date population the uncertainty in the reconstruction of the population size T time units ago is large in the singleton model then it is likely to be even larger in more complex ones.

The precise mathematical problem which we address looks as follows. The ancestral tree of the present day population is a finite balanced weighted tree $\tilde{\Gamma}$[1]. For a given function $d(t)$ we want to compute the 'probability' of obtaining Γ in the environment determined by $d(t)$ and then find the function which maximizes this value.

We face several technical difficulties here. First of all in order to get a measure on the space of ancestral trees we have to fix the time point $T < 0$ when we start to trace the development of the population and the number N of population members at this time. These data together with the restriction of $d(t)$ to $[-T, 0]$ defines a (sub-)probability measure on the set of ancestral trees of depth $\leq T$[2].To deal with the case $T = \infty$ which we are interested in we have to find for a given $\tilde{\Gamma}$ and $T > t_1(\tilde{\Gamma})$ the most likely reconstruction of N at $-T$ and $d(t)$ on $[-T, 0]$ and then to take the limit for $T \to \infty$.

The second problem is that the space H of ancestral trees is continuous and the probability of getting any particular tree is zero. Therefore, we have to consider sufficiently small neighborhoods of $\tilde{\Gamma}$ instead of $\tilde{\Gamma}$ itself and then show that there exists a well defined limit when the neighborhoods shrink to one point.

The third problem arises from the fact that our function does not reach its maximal value on the space of actual functions $d(t)$ and in order to obtain the solution we have to allow for δ-functions. In fact, our first result (see **??**) states that for any initial $\tilde{\Gamma}$ the maximal likelihood reconstruction of $d(t)$ is a sum of δ-functions (with coefficients) concentrated at some of the time points which

[1] Recall that a weighted tree is a tree whose edges are labeled by non-negative numbers. A weighted tree is called balanced if there is a function on the vertices such that the label on an edge is the difference of the values of this function on its starting and ending vertices.

[2] We define the depth $t_1(\tilde{\Gamma})$ of $\tilde{\Gamma}$ as the time to the oldest coalescence event.

occur as vertex labels in $\tilde{\Gamma}$.

We further present an algorithm for the computation of this maximal likelihood d. This algorithm was implemented and I ran multiple reconstructions with it starting with trees obtained with a constant death rate function. In all the trials the maximal likelihood reconstruction turns out to be a series of 'tall' δ-functions separated by long time intervals. In other words we observe that the most likely reconstruction of history from the ancestral tree which formed in constant environment looks like a series of widely spaces catastrophes.

1 Singleton processes

1.1 Singleton histories

[**sec1.1**]

Definition 1.1.1 [**histdef**] *Let $s < t$ be two real numbers. A singleton history on time interval $[s, t]$ is a set of data of the form:*

$$\Gamma = (V; E \subset V \times V; \tau : V \to [s, t]; \psi : \tau^{-1}(t) \to \mathbf{N})$$

where (V, E) is a finite directed graph with the set of vertices V and the set of edges E and $\tau : V \to [s, t]$ is a function satisfying the following conditions:

1. *given an edge from v to v' one has $\tau(v) < \tau(v')$,*
2. *if $\tau(v) = s$ there is exactly 1 edge starting in v,*
3. *if $\tau(v) \neq s$ there is exactly one edge ending in v and 0 or > 1 edges starting in v.*

Intuitively, $\tau^{-1}(s)$ is the set of the population members at the initial time s. The graph, which is necessarily a union of trees in view of the condition (3), is the genealogy of these members. Its vertices correspond to the transformation events with $\tau(v)$ being the time of the corresponding event. The subset $\psi^{-1}(i)$ of the final population $V_t = \tau^{-1}(t)$ consists of members which transform into i new members at the exact moment t.

It will be convenient for us to envision a singleton history through its "geometric realization" $|\Gamma|$ which is a 1-dimensional CW-compex which is the union of intervals $[\tau(v_1), \tau(v_2)]$ for all $(v_1, v_2) \in E$ glued together at the ending points in the obvious way. This space comes together with a map $|\tau| : |\Gamma| \to [s, t]$ which is defined by the original τ in the obvious way and with a natural-valued function τ on the set $|\tau^{-1}|(t)$. The pictures used below for illustrative purposes are actually pictures of these geometric realizations.

For $u \in [s, t]$ set $V_u(\Gamma) = |\tau|^{-1}(u)$. This is the set of population members at the time infinitesimally preceding time u. Note that $V_s = \tau^{-1}(s)$ and $V_t = \tau^{-1}(t)$.

For any Γ in $H[s,t]$ the image of τ_Γ is a finite set of points in $[s,t]$ which contains $\{s\}$ and $\{t\}$. We will write $x_1(\Gamma),\ldots,x_q(\Gamma)$ for the points of this set lying in (s,t) ordered such that $x_1(\Gamma)<\cdots<x_q(\Gamma)$. The number $q=q(\Gamma)$ is an invariant of Γ which will be the basis of most of the inductive arguments below. In some cases we will write $x_0(\Gamma)$ for s and $x_{q+1}(\Gamma)$ for t.

For any $u\in(x_{i-1},x_i]$ we have a canonical identification $V_u(\Gamma)=V_{x_i}(\Gamma)$. We denote these sets by $V_i(\Gamma)$ where $i=1,\ldots,q+1$. The combinatorics of Γ defines maps

$$f_i : V_{i+1}(\Gamma)\to V_i(\Gamma)$$

The function ψ on V_{q+1} can be interpreted as a map $f_{q+1}: V_{q+2}\to V_{q+1}$ such that $\psi(v)=\#(f_{q+1}^{-1}(v))$. Therefore to each Γ we can assign a sequence of points $s<x_1<\cdots<x_q<t$ and a sequence of maps of finite sets

$$V_1\leftarrow V_2\leftarrow\ldots\leftarrow V_{q+2}$$

and one can easily see that up to an isomorphism Γ is determined by $(x_1,\ldots,x_q)$ and the isomorphism class of this sequence of maps. Moreover, such data corresponds to a history Γ if an only if the maps f_i are not isomorphisms for $i\le q$. A more detailed analysis of the structure of the space of histories based on this description is given in **??**. In our constructions it will be more convenient for us to work with ordered histories which are defined below.

Definition 1.1.2 *[**ordhist**] Let $\Gamma=(V,E,\tau,\psi)$ be a singleton history over $[s,t]$. On ordering on Γ is in ordering on $V_s=V_1$ and for each $v\in V$ an ordering on the set of edges starting at v.*

We denote the set of isomorphism classes of ordered histories over $[s,t]$ by $H[s,t]$. Note that it is a (non-commutative) monoid with respect to the obvious operation of disjoint union of ordered histories. For $s=t$ we set $H[s,s]=\mathbf{N}$ and consider it as the set of isomorphism classes of ordered finite sets.

The ordering on the set of edges starting at vertices lying over $x_i(\Gamma)$ is the same as the ordering on the fibers of the map $V_{i+1}\to V_i$. Therefore, for an ordered history all the sets $V_i(\Gamma)$ carry a natural ordering and the maps f_i preserve this ordering. Conversely, if we are given orderings on all of the sets V_i which are preserved by the maps f_i we get an ordering on Γ. Therefore, there is a bijection between $H[s,t]$ and pairs of the form $(\underline{x},\underline{f})$ where $\underline{x}=(x_1,\ldots,x_q)$, $x_i\in(s,t)$, $x_1<\cdots<x_q$ and

$$[\mathbf{uuf}]\underline{f}=(V_1\stackrel{f_1}{\leftarrow}V_2\stackrel{f_2}{\leftarrow}\cdots\stackrel{f_{q+1}}{\leftarrow}V_{q+2}) \qquad (1)$$

where V_i are ordered finite sets, f_i are order preserving maps and for $i\le q$ the map f_i is not an isomorphism.

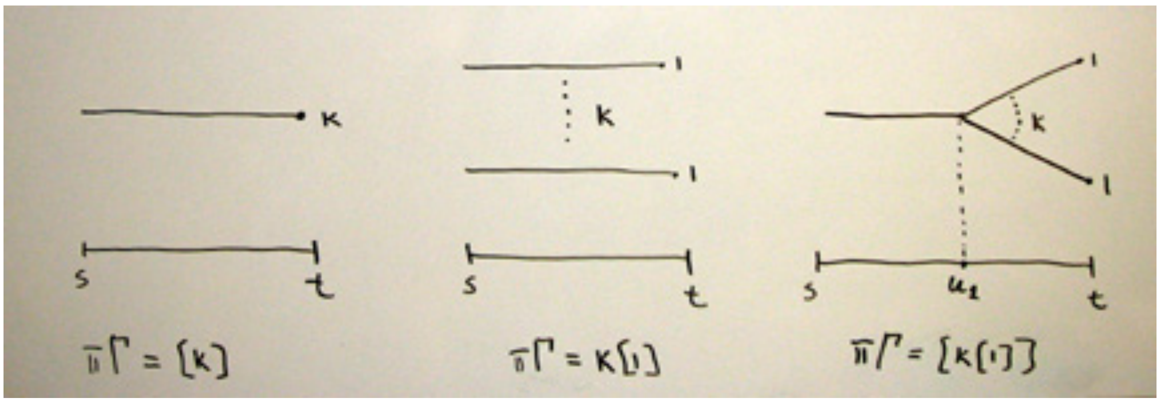
Sequences of maps of this form between the standard ordered sets $\{1,\ldots,n\}$ will be called (ordered) combinatorial types and the ones in which f_i is not an isomorphism for $i\le q$ will be called non-degenerate combinatorial types. The number $q(\pi)$ is called the level of a combinatorial type π. We will also write

$$n_i(\pi)=\#V_{i+1}(\pi)$$

Let $F_q(\mathbf{N})$ be the set of combinatorial types of level q. This set carries a natural structure of a non-commutative monoid with respect to the disjoint union of the sequences.

Denote the disjoint union of combinatorial types π_1 and π_2 by $\pi_1 + \pi_2$. For a combinatorial type π denote by $[\pi]$ the combinatorial type obtained by extending π to the right by a morphism to the point. To any natural number there corresponds a combinatorial type $[k]$ which is represented by the map $\{1, \ldots, k\} \to pt$. It is easy to see that any combinatorial type of level ≥ 0 can be obtained from types $[k]$ by iterated application of the disjoint union operation and the $[-]$ operation.

Example 1.1.3 The most important histories which we will encounter below are the ones corresponding to the combinatorial types $[k]$, $k[1]$ and $[k[1]]$ (they are of level 0, 0 and 1 respectively. The corresponding pictures look as follows:



The sets $F_q(\mathbf{N})$ form a simplicial set where the boundary operations are given compositions and the removal of f_1 and the degeneracy operations by the insertion of identities. Denote by $| - |_{[s,t]}$ the geometric realization functor which uses simplexes

$$\Delta^q_{[s,t]} = \{x_1, \ldots, x_q \in [s,t] \,|\, x_1 \leq \cdots \leq x_q\}$$

instead of the standard simplexes

$$\Delta^q = \Delta^q_{[0,1]}$$

One has the following result.

Theorem 1.1.4 *[**descr**] There is a natural bijection*

$$[\mathbf{simplreal}]H[s,t] = |F_*(\mathbf{N})|_{[s,t]} \tag{2}$$

Proof: Let $\Delta^q_{s,t}$ be the open simplex

$$\Delta^q_{s,t} = \{x_1, \ldots, x_q \in (s,t) \,|\, x_1 < \cdots < x_q\}$$

for $q > 0$ and $\Delta^0_{s,t} = \Delta^0_{[s,t]} = pt$. Then

$$|F_*(\mathbf{N})|_{[s,t]} = \amalg_{q \geq 0} \amalg_{\pi \in F_q(\mathbf{N})^{nd}} \Delta^q_{s,t}$$

where $F_q(\mathbf{N})^{nd}$ is the set of non-degenerate simplexes in $F_q(\mathbf{N})$. We assign to a history Γ the point $(x_1(\Gamma), \ldots, x_q(\Gamma))$ in the simplex corresponding to $\underline{f}(\Gamma)$.

The fact that it is a bijection follows from the comments made above. $\square$

In the following consideration we will write $\Delta^{\pi}_{s,t}$ for the open simplex in $H[s,t]$ corresponding to a (non-degenerate) combinatorial type π so that we have

$$[\textbf{simplrep}]H[s,t] = \coprod_{q\geq 0} \coprod_{\pi\in F^{nd}_q(\mathbf{N})} \Delta^{\pi}_{s,t} \tag{3}$$

We will also use the standard notation

$$sk_q H[s,t] = \amalg_{\pi,q(\pi)\leq q} \Delta^{\pi}_{s,t}$$

for the skeletons of $H[s,t]$ with respect to our triangulation.

Remark 1.1.5 [**asfree**] The simplicial set $|F_*(\mathbf{N})|$ can also be described as follows. Consider the category Mon of monoids and the pair of adjoint functors $Mon \to Sets$, $Sets \to Mon$ where the first one is the forgetting functor and the second one the free monoid functor. Their composition defines a co-triple F on the category of monoids. This co-triple defines for any monoid M a simplicial monoid $F_*(M)$. One can easily see that $F_*(\mathbf{N})$ is the simplicial monoid described above it terms of the sequences of order preserving maps.

Remark 1.1.6 [**gener**] It seems likely that if we take a set A and the simplicial monoid associated with the forgetting functor and the functor of the free monoid generated by $X \times A$ then we will obtain by the same construction the space of labelled histories which correspond to genealogies of populations with several distinguished sub-populations.

Given a singleton history Γ over $[s,t]$ and $u \in [s,t]$ one can cut Γ at u obtaining two histories $R_u(\Gamma) \in H[u,t]$ and $L_u(\Gamma) \in H[s,u]$. If there is a vertex v with $\tau(v) = u$ and n edges starting in it then it appears as one vertex v' in $L_u(\Gamma)$ with $\psi(v') = n$ and as n vertices in $R_u(\Gamma)$.

If Γ is an ordered history then $L_u(\Gamma)$ carries an obvious ordering. The right "half" $R_u(\Gamma)$ carries obvious orderings on the sets of edges starting at a given vertex and we give it an ordering on $V_u(R_u(\Gamma))$ using the identification

$$V_u(R_u(\Gamma)) = V_i(\Gamma)$$

where i is such that $u \in [x_{i-1}, x_i)$ if $u < t$ and $i = q+2$ if $u = t$.

For $s \leq u \leq v \leq t$ define the restriction maps

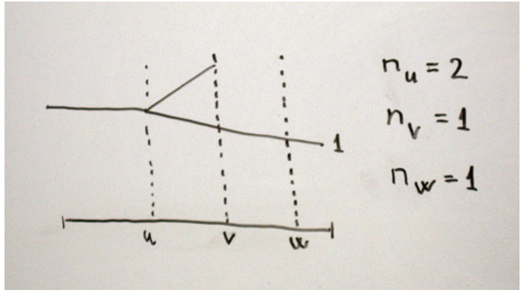
$$[\textbf{restr}]res_{u,v} : H[s,t] \to H[u,v] \tag{4}$$

as $res_{u,v} = L_v \circ R_u$. Note that for $(u',v') \subset (u,v)$ one has

$$res_{u',v'}res_{u,v} = res_{u',v'}$$

For $u = v$ we will write n_u instead of $res_{u,u}$. It is a map from $H[s,t]$ to $\mathbf{N}$ which assigns to a history the number of its members at time u. Note that in general $n_u(\Gamma) \neq \#V_u(\Gamma)$ since some vertices may have to be counted with multiplicities, instead the set V_u is equipped with a function $\psi_{\Gamma,u} : V_u \to \mathbf{N}$ such that $n_u(\Gamma) = \sum_{v\in V_u} \psi_{\Gamma,u}(v)$.

We let $H[s,t]^n_m$ denote the subset of histories Γ such that $n_s(\Gamma) = m$ and $n_t(\Gamma) = n$.



Lemma 1.1.7 *[**bij2**] For $s \leq u <\leq t$ the map*

$$res_{s,u} \times res_{u,t} : H[s,u] \to \amalg_n H[u,t]^n_* H[s,t]^*_n$$

is bijective.

Proof: We may clearly assume that $s < u < t$. Let us define an inverse c^{ord} map to $res_{s,u} \times res_{u,t}$ as follows. Let $\Gamma' \in H[u,t]^n_*$ and $\Gamma'' \in H[s,t]^*_n$ for some $n \in \mathbf{N}$ and $q = q(\Gamma')$. Then $V_{q+2}(\Gamma')$ and $V_1(\Gamma'')$ are ordered finite sets with n elements and there exists a unique order-preserving bijection between them. Using this bijection to glue the combinatorial types of Γ' and Γ'' we get a combinatorial type and since the event points are known we get a well defined Γ such that $res_{s,u}(\Gamma) = \Gamma'$ and $res_{u,t}(\Gamma) = \Gamma''$. One verifies easily that this is indeed a two-sided inverse to $res \times res$. $\square$

Lemma 1.1.8 *[**bij1**] For any $n \geq 0$ the iterated "addtion" map*

$$H[s,t]_1 \times \cdots \times H[s,t]_1 \to H[s,t]_n$$

is a bijection.

Proof: Straightforward. $\square$ In what follows we will be considering sets $H[s,t]$ as measurable spaces with respect to the Borel σ-algebra on $|F_*(\mathbf{N})|_{[s,t]}$ or, equivalently, with respect to the sum of the Borel σ-algebras on simplexes $\Delta^\pi_{s,t}$. One verifies easily that all the maps considered above are measurable. A measure on $H[s,t]$ is the same as a collection of measures on the simplexes $\Delta^\pi_{s,t}$ given for all non-degenerate ordered combinatorial types π.

For $s \leq u \leq v \leq t$ define the σ-algebra $\mathfrak{B}^v_u$ on $H[s,t]$ as forllows:

1. for $u < v$ set $\mathfrak{B}^v_u = res^{-1}_{u,v}(\mathfrak{B})$ where $\mathfrak{B}$ is the Borel σ-algebra on $H[u,v]$,
2. for $u = v$ set $\mathfrak{B}^v_u = n_u^{-1}(\mathfrak{B}_{\mathbf{N}})$ where $\mathfrak{B}_{\mathbf{N}}$ is the algebra of all subsets of $\mathbf{N}$.

We have the following result obvious result (for the definition of a path system see [6]).

Proposition 1.1.9 *[**ispath**] The collection of data $((H[s,t],\mathfrak{B}),\mathfrak{B}^v_u, n_u)$ defines a path system on $\mathbf{N}$ over T.*

We denote this path system by $\mathcal{H}[s,t]$. Up to an isomorphism in the category of probability kernels, we have

$$(H[s,t], \mathfrak{B}_u^v) = H[u,v]$$

for all $s \leq u \leq v \leq t$. We will freely use these identifications below.

1.2 Processes on $\mathcal{HD}[s,t]$

Recall (see [6]) that a process on $\mathcal{HD}[s,t]$ is a collection of probability kernels $\mu_u^v : \mathbf{N} \to HD[u,v]$ given for all $u \leq v$ such that $\mu_{u,n}^v = \mu_u^v(n)$ is supported on $HD[u,v]_n^*$. We set

$$\phi_{u,\mu}^v(n,m) = \mu_{u,n}^v(H[u,v]_n^m)$$

When no confusion is possible we will write ϕ_u^v instead of $\phi_{u,\mu}^v$ etc.

Recall (see [6]) that a process is a Markov process if it satisfies condition (M) of *loc.cit.* In the context of the path system $\mathcal{H}[s,t]$ this condition asserts that for all $s \leq u \leq v \leq w \leq t$ and all $n \geq 0$ the square

$$[\mathbf{m2diag}] \quad \begin{array}{ccc} \mathbf{1} & \xrightarrow{\mu_{u,n}^v} & HD[u,v]_n^* \\ \mu_{u,n}^w \downarrow & & \downarrow Id\otimes(\mu_{v,n}^w \circ n_v) \\ HD[u,w]_n^* & \xrightarrow{res_{u,v}\times res_{v,w}} & HD[u,v]_n^* \times HD[v,w] \end{array} \tag{5}$$

commutes. Applying [6,] and taking into account that $\mu_u^u = Id$ we get the following reformulation.

Lemma 1.2.1 *[**crit1**] A process μ_*^* on $\mathcal{HD}[s,t]$ is a Markov process if and only if for any $m,n \geq 0$, any $s \leq u < v < w \leq t$, any measurable U_1 in $HD[u,v]_n^m$ and any measurable U_2 in $HD[v,w]_m^*$ one has*

$$[\mathbf{eqcrit1}]\mu_{u,n}^w((res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2)) = \mu_{u,n}^v(U_1)\, \mu_{v,m}^w(U_2). \tag{6}$$

The re-numeration of initial vertices defines an action of the symmetric group Σ^n on $HD[u,v]_{n,*}$ and we have the following obvious result.

Lemma 1.2.2 *[**th80l1**] Let μ be an additive process on $\mathcal{HD}[s,t]$. Then for any $s \leq u < v \leq t$ and any $n \geq 0$ the measure $\mu_{u,n}^v$ is invariant under the action of Σ^n.*

For a combinatorial type $\pi = (f_{q+1}, \ldots, f_1)$ of level q define its local invariant $\underline{K}(\pi)$ as a sequence $(\underline{k}_1, \ldots, \underline{k}_{q+1})$ where $\underline{k}_i \in S^\infty(\mathbf{N})$ is the isomorphism class of the map f_i. If π is the combinatorial type of a history Γ then $\underline{k}_i(\pi)$ is the list of branching multiplicities of points of Γ over $x_i(\Gamma)$.

Theorem 1.2.3 *[**th8o**] Let μ be an additive Markov process on $HD[s,t]$ and π, π' be two ordered combinatorial types with the same local invariant $\underline{K}$ and therefore of the same level q. Then for all $s \leq u < v \leq t$ and $n \geq 0$ the co-restrictions of $\mu_{u,n}^v$ to*

$$\Delta_{u,v}^q = \Delta_{u,v}^\pi = \Delta_{u,v}^{pi'}$$

coincide.

Proof: If $q=0$ and $n_0(\pi)=n_0(\pi')=n$ then

$$\pi=[k_{11}]+\cdots+[k_{1n}]$$

$$\pi'=[k'_{11}]+\cdots+[k'_{1n}]$$

where the sequences (k_{1i}) and (k'_{1i}) differ by a permutation on $\{1,\ldots,n\}$. Then the zero simplexes $\Delta^{\pi}_{u,v}$, $\Delta^{\pi'}_{u,v}$ are transformed into each other by the action of Σ^n on $HD[u,v]_{n,*}$ and our claim follows from the fact that $\mu^v_{u,n}$ is invariant under this action.

Let $q=1$ and

$$\pi=[k_{11}[1]]+\cdots+[k_{1n}[1]]$$

$$\pi'=[k'_{11}[1]]+\cdots+[k'_{1n}[1]]$$

Then π and π' are again transformed into each other by the action of Σ^n and the claim of the theorem holds.

For a general non-degenerate combinatorial type π such that $q>0$, $u<w_1<w_2<v$ and a Borel subset B of $\Delta^{q-1}_{w_2,v}$ set

$$((w_1,w_2),B,\pi)=\{(x_1,\ldots,x_q)\in\Delta^{\pi}_{u,v}\,|\,x_1\in(w_1,w_2),\ (x_2,\ldots,x_q)\in B\}$$

Intersection of two subsets of this form is again of this form and they generate the Borel σ-algebra of $\Delta^{\pi}_{u,v}$.

For $\pi=[\pi_1]+\cdots+[\pi_n]$ the types

$$L(\pi)=[n_0(\pi_1)[1]]+\cdots+[n_0(\pi_n)[1]]$$

and

$$R(\pi)=\pi_1+\cdots+\pi_n$$

are non-degenerate and we have

$$((w_1,w_2),B,\pi)=(res_{u,w_2}\times res_{w_2,v})^{-1}(U_1\times U_2)$$

where

$$U_1=((w_1,w_2)\subset\Delta^{L(\pi)}_{u,w_2})$$

$$U_2=(B\subset\Delta^{R(\pi)}_{w_2,v})$$

and since μ is a Markov process we have

$$\mu^v_{u,n}((w_1,w_2),B,\pi)=\mu^{w_2}_{u,n}(U_1)\mu^v_{w_2,n_1}(U_2)$$

for $n_1=n_0(R(\pi))$. If $\underline{K}(\pi)=\underline{K}(\pi')$ then $\underline{K}(R(\pi))=\underline{K}(R(\pi'))$ and $\underline{K}(L(\pi))=\underline{K}(L(\pi'))$ and the claim of the theorem follows by an obvious inductive argument. $\square$

Lemma 1.2.4 *[genm] If μ_*^* is a Markov process then for $u \leq v \leq w$ in $[s,t]$ one has*

$$[\mathbf{eq001}]\phi_u^w(n,k) = \sum_{m\geq 0} \phi_u^v(n,m)\phi_v^w(m,k) \tag{7}$$

Proof: It follows from the general properties of Markov pre-processes (see [6,]). □ We set

$$h^n(u,v) = \mu_{u,1}^v(\Delta_{u,v}^{n[1]})$$

Lemma 1.2.5 *[ob1] If μ_*^* is a Markov pre-process then for $n \geq 0$ and $u \leq v \leq w$ in $[s,t]$ one has*

$$h^n(u,v)h^n(v,w) = h^n(u,w)$$

Proof: It follows from Lemma 1.2.1 applied to $U_1 = \Delta_{u,v}^{n[1]}$, $U_2 = \Delta_{v,w}^{n[1]}$. □ Note that all the maps which participate in the definition of $\mathcal{H}[s,t]$ are homomorphisms of monoids.

Definition 1.2.6 *A pre-process μ_*^* on $\mathcal{H}[s,t]$ is called an additive pre-process if $\mu_{u,0}^v(\Delta^{\emptyset}) = 1$ for all u, v and the kernels $\mu_u^v : \mathbf{N} \to H[u,v]$ are homomorphisms of monoids.*

If μ is additive then

$$h_\mu^n(u,v) = (h_\mu^1(u,v))^n$$

and

$$v_{u,k}^v = (v_{u,1}^v)^k.$$

When no confusion is possible we will write $h(u,v)$ instead of $h^1(u,v)$ so that for an additive pre-process

$$h^n(u,v) = h(u,v)^n$$

In what follows we consider almost exclusively additive (pre-)processes.

Lemma 1.2.7 *[ob00] Let μ_*^* be a Markov pre-process. Then for any $n, m \geq 0$ and any $u \leq v < w$ in $[s,t]$ the function $h^n(u, v+\epsilon)\phi_{v+\epsilon}^w(n,m)$ is monotone decreasing in ϵ and one has*

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u,v+\epsilon)\phi_{v+\epsilon}^w(n,m) = h^n(u,v)\phi_v^w(n,m)$$

Proof: Applying Lemma 1.2.1 to $U_1 = \Delta_{u,v+\epsilon}^{n[1]}$ and $U_2 = H[v+\epsilon, w]_{n,m}$ we get

$$h^n(u,v+\epsilon)\phi_{v+\epsilon}^w(n,m) = \mu_{u,n}^w(res_{u,v+\epsilon}^{-1}(\Delta_{u,v+\epsilon}^{n[1]}) \cap n_w^{-1}(m)).$$

Since for $\epsilon' \geq \epsilon$ one has

$$res_{u,v+\epsilon'}^{-1}(\Delta_{u,v+\epsilon'}^{n[1]}) \cap n_w^{-1}(m) \subset res_{u,v+\epsilon}^{-1}(\Delta_{u,v+\epsilon}^{n[1]}) \cap n_w^{-1}(m)$$

and

$$\cup_{\epsilon\to 0}(res^{-1}_{u,v+\epsilon}(\Delta^{n[1]}_{u,v+\epsilon})\cap n_w^{-1}(m)) = res^{-1}_{u,v}(\Delta^{n[1]}_{u,v})\cap n_w^{-1}(m)$$

our claims follow. □

Recall that a function f on $[s,t]$ is called monotone increasing (resp. decreasing) if for $x \le y$ one has $f(x) \le f(y)$ (resp. $f(x) \ge f(y)$). A function is called right continuous if for all $u \in [s,t)$ one has

$$\lim_{\epsilon>0,\epsilon\to 0} f(u+\epsilon) = f(u).$$

The following two lemmas give some elementary properties of such functions which will be used below.

Lemma 1.2.8 *[**rcim**] Any right continuous function f on $[s,t]$ is measurable.*

Proof: See 4.3.2. □

Lemma 1.2.9 *[**pirc**] Let f be a right continuous on $[s,t)$. If f is monotone increasing then for any $a_+ > a$ such that $f^{-1}([a,a_+)) \ne \emptyset$ there exists $b_+ > b$ such that $f^{-1}([a,a_+)) = [b,b_+)$. If f is monotone decreasing then for any $a_+ > a$ such that $f^{-1}((a,a_+]) \ne \emptyset$ there exists $b_- < b$ such that $f^{-1}((a,a_+]) = [b_-,b)$.*

Proof: Consider for example the case of an increasing f. Then if $f^{-1}([a,a_+)) \ne \emptyset$ we have

$$f^{-1}([a,\infty)) = [b,t)$$

and

$$f^{-1}((-\infty,a_+)) = [s,b_+)$$

which implies the claim of the lemma. □

As a corollary of Lemma 1.2.5 we see in particular that for a Markov pre-process the functions $h^n(u,v)$ are monotone increasing in u and monotone decreasing in v. Since $\upsilon^w_{v,m} \le 1$ and

$$[\mathbf{eq01}]\sum_{m\ge 0}\phi_{u,v}(n,m) = \upsilon^v_{u,n} \tag{8}$$

we also see that for a Markov pre-process the functions $\upsilon^v_{u,n}$ are monotone decreasing in v.

Remark 1.2.10 We will see from examples below (**??**) that there are Markov pre-processes on $\mathcal{H}[s,t]$ such that $\upsilon^v_{u,n}$ are not monotone in u.

Lemma 1.2.11 *[**ob01**] Let μ^*_* be a Markov pre-process. Then for any $m,n \ge 0$ and any $u \le v < w$ in $[s,t]$ the function $\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w)$ is monotone increasing in ϵ and one has*

$$\lim_{\epsilon>0,\epsilon\to 0}\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w) = \phi_{u,v}(m,n)h^n(v,w)$$

Proof: Applying Lemma 1.2.1 to $U_1 = H[u, v+\epsilon]_{m,n}$ and $U_2 = \Delta^{n[1]}_{v+\epsilon,w}$ we get

$$\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w) = \mu^w_{u,m}(res^{-1}_{v+\epsilon,w}(\Delta^{n[1]}_{v+\epsilon,w}))$$

and since

$$\cap_{\epsilon\to 0}(res^{-1}_{v+\epsilon,w}(\Delta^{n[1]}_{v+\epsilon,w})) = res^{-1}_{v,w}(\Delta^{n[1]}_{v,w})$$

our claim follows. $\square$

Definition 1.2.12 *[***rcont***] A pre-process μ^*_* is called non-degenerate if $v^u_{u,k} = 1$ for all u, k. It is called right continuous if for any $u \in [s,t]$ and any k, $v^v_{u,k}$ is a right continuous function in v from $[s,v]$ to $[0,1]$.*

If μ is non-degenate then $h^n(u,u) = 1$ for all n and u. Note that any process on $\mathcal{H}[s,t]$ is automatically non-degenerate and right continuous.

Remark 1.2.13 For a Markov pre-process one has $(v^u_{u,k})^2 = v^u_{u,k}$ and therefore a Markov pre-process is non-degenerate if and only if $v^u_{u,k} \neq 0$ for all u, k.

Theorem 1.2.14 *[***th1***] Let μ^*_* be a non-degenerate Markov pre-process on $\mathcal{H}[s,t]$. Then the following conditions are equivalent:*

1. *for all $n \geq 0$ functions $v^v_{u,n}$ are right continuous in u and if $u < t$ then there exits $w > u$ such that $v^w_{u,n} \neq 0$,*
2. *for all $n \geq 0$ functions $h^n(u,v)$ are right continuous in u and if $u < t$ then there exits $w > u$ such that $v^w_{u,n} \neq 0$,*
3. *for all $n \geq 0$ functions $\phi^v_u(n,m)$ are right continuous in u and if $u < t$ then there exits $w > u$ such that $v^w_{u,n} \neq 0$,*
4. *for all $n \geq 0$ functions $v^v_{u,n}$ are right continuous in v,*
5. *for all $n \geq 0$ functions $h^n(u,v)$ are right continuous in v,*
6. *for all $n \geq 0$ functions $\phi^v_u(n,m)$ are right continuous in v.*

Proof: Observe first that if for all $u < t$ then there exits $v > u$ such that $v^w_{u,n} \neq 0$ then, since $v^v_{u,n}$ are monotone decreasing in v we have $v^v_{u,n} \neq 0$ for all $u \leq v \leq w$.

Let u and w be as above. Taking the sum over m in Lemma 4.3.1 and setting $v = u$ we get

$$[\textbf{feqp}] \lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon)v^w_{u+\epsilon,n} = v^w_{u,n} \tag{9}$$

which implies that there exists $\epsilon > 0$ such that $h^n(u,u+\epsilon) \neq 0$. Without loss of generality we may assume that $u + \epsilon = w$.

$(1) \Rightarrow (2), (5)$ When $v_{u,n}^v$ is right continuous in u equation (68) implies that

$$(\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon))v_{u,n}^w = v_{u,n}^w$$

and since $v_{u,n}^w \neq 0$ we conclude that

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon) = 1$$

Together with Lemma 1.2.5 we conclude that (2) and (5) hold.

$(2) \Rightarrow (5)$ Immediate from Lemma 1.2.5 since for all u there exists w such that $h^n(u,w) \neq 0$.

$(5) \Rightarrow (3)$ Since $h^n(u,u) = 1$ condition (5) also implies that for any u there exists $w > u$ satisfying $h^n(u,w) \neq 0$. Since $v_{u,n}^w \geq h^n(u,w)$ we conclude that $v_{n,u}^w \neq 0$.

Taking in Lemma 4.3.5 $v = u$ we get

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon)\phi_{u+\epsilon}^w(n,m) = \phi_u^w(n,m)$$

for all $w > u$ and using condition (5) we get that $\phi_u^w(n,m)$ is right continuous in u.

$(2) \Rightarrow (6)$ We need to show that

$$[\textbf{seqp}] \lim_{\epsilon>0,\epsilon\to 0} \phi_u^{v+\epsilon}(m,n) = \phi_u^v(m,n) \tag{10}$$

Let w be such that $h^n(v,w) \neq 0$. Then Lemma 4.3.5 together with the right continuity of $h^n(-,-)$ in the first variable implies (69).

$(6) \Rightarrow (4)$ Immediately follows from the fact that $v_{u,n}^v = \sum_m \phi_u^v(n,m)$.

$(4) \Rightarrow (2)$ Since functions $v_{u,n}^v$ are right continuous in v and $v_{u,k}^u = 1$ there exists $w > u$ such that $v_{u,n}^w \neq 0$ and as explained above such that $h^n(u,w) \neq 0$. Taking in Lemma 4.3.5 $m \neq n$ and $v = u$ we get

$$[\textbf{eq020}] \lim_{\epsilon\to 0} \phi_{u,u+\epsilon}(m,n) = 0 \tag{11}$$

Therefore we have

$$[\textbf{teqp}] 1 = \lim_{\epsilon\to 0} v_{u,n}^{u+\epsilon} = \lim_{\epsilon\to 0} \sum_m \phi_{u,u+\epsilon}(n,m) = \lim_{\epsilon\to 0} \phi_{u,u+\epsilon}(n,n) \tag{12}$$

Form Lemma 4.3.5 for $m = n$ and $v = u$ we get for all $w > u$

$$\lim_{\epsilon>0,\epsilon\to 0} \phi_{u,u+\epsilon}(n,n)h^n(u+\epsilon,w) = h^n(u,w)$$

which together with (71) implies that $h^n(u,v)$ is right continuous in u.

$(3) \Rightarrow (1)$ Immediately follows from the fact that $v_{u,n}^v = \sum_m \phi_u^v(n,m)$.

Theorem is proved. $\square$

For a pre-process μ_*^* define $E_{n,\mu} \subset [s,t]$ by the rule $x \in E_{n,\mu}$ if and only if $e = s$ or for all sufficiently small $\epsilon > 0$ one has $h^n(x - \epsilon, x) = 0$. When no confusion is possible we will write E_n instead of $E_{n,\mu}$.

Lemma 1.2.15 *[**ob2**] Let μ be a non-degenerate right continuous Markov pre-process. Then for any $e \in E_n$ such that $h^n(e,t) = 0$ there exists a unique $e_{+1} > e$ in E_n such that for all $x \in [e, e_{+1})]$ one has $h^n(e,x) \neq 0$.*

Proof: By Theorem 4.3.8 the function $h^n(e,-)$ is right continuous and therefore the set of zeros of $h^n(e,-)$ is of the form $[e_{+1}, t]$ for some e_{+1} in $(e,t]$. For $\epsilon < e_{+1} - e$ we have $0 = h(e, e_{+1} + \epsilon) = h(e, e_{+1} - \epsilon)h(e_{+1} - \epsilon, e_{+1})$ and since $h(e, e_{+1} - \epsilon) \neq 0$ we conclude that $e_{+1} \in E_n$. $\square$ Note that if $E_n \neq \emptyset$ then there exists a unique $e \in E_n$ such that $h^n(e,t) \neq 0$. For this e we set $e_{+1} = t$.

Lemma 1.2.16 *[**ob3**] For a a non-degenerate right continuous Markov pre-process μ the sets E_n are countable.*

Proof: We have

$$[\mathbf{ecov}][s,t) = \amalg_{e \in E_n}[e, e_{+1}) \tag{13}$$

and since the sum of an uncountable number of non-zero numbers is infinite we conclude that E_n is countable. $\square$

Since for an additive process $E_n = E_m$ for all $m, n \neq 0$ we will write $E = E(\mu)$ for this set in the additive context.

Define a map

$$(x_1, k_1) : H[u,v]_1^* \to (u,v] \times \mathbf{N}$$

as follows. It sends $\Delta_{u,v}^{[n]}$ to (v,n) and a history Γ of level $q \geq 1$ to the pair $(x_1(\Gamma), n_1(\Gamma))$ where $x_1(\Gamma)$ is first event point in Γ and $n_1(\Gamma)$ is the branching multiplicity of this point.

For $u \leq v \leq w$ we have an embedding

$$j_{u,v}^w : H[v,w] \to H[u,w]$$

which is determined by the conditions

$$R_v(j_{u,v}^w(\Gamma)) = \Gamma$$

$$L_v(j_{u,v}^w(\Gamma)) = \Delta_{u,v}^{n[1]}$$

where $n = n_v(\Gamma)$. Note that for $v < w$

$$j_{u,v}^w(\Delta_{v,w}^{\pi}) = \left\{(x_1, \ldots, x_q) \in \Delta_{u,w}^{\pi} | \, x_1 > w\right\}.$$

which implies in particular that $j_{u,v}^w$ are measurable.

For a history Γ with $q(\Gamma) > 0$ we set

$$R(\Gamma) = j^v_{u,x_1(\Gamma)}(R_{x_1(\Gamma)}(\Gamma))$$

The combinatorial type of $R(\Gamma)$ depends only on the combinatorial type of Γ and we write $R(\pi)$ for the combinatorial type of $R(\Gamma)$ for any Γ such that $\pi(\Gamma) = \pi$. Note $q(R(\pi)) = q(\pi) - 1$.

For $u < v$, $(w_1, w_2) \subset (u, v)$ and $U \subset H[w_2, v]^*_k$ set

$$[\mathbf{gener}](k, (w_1, w_2), U) = (res_{u,w_2} \times res_{w_2,v})^{-1}(((w_1, w_2) \subset \Delta^{[k[1]]}_{u,w_2}) \times U) \tag{14}$$

or equivalently

$$(k, (w_1, w_2), U) = \{\Gamma \in H[u, v]^*_1 | \, n_1(\Gamma) = k, \; x_1(\Gamma) \in (w_1, w_2), \; res_{w_2,v}(\Gamma) \in U\}$$

Lemma 1.2.17 *[**gens**] The collection of subsets of the form (14) is closed under finite intersections and for any $q > 0$ the Borel σ-algebra on $sk_q H[u, v]^*_1$ is generated by subsets of the form $(k, (w_1, w_2), U)$ where $k \neq 1$ and $U \subset sk_{q-1} H[w_2, v]^*_k$.*

Proof: We may assume that $u < v$. Suppose that $w'_2 \geq w_2$ then

$$(k, (w_1, w_2), U) \cap (k, (w'_1, w'_2), U') = (k, (max(w_1, w'_1), w_2), U \cap j^v_{w_2,w'_2}(U'))$$

and

$$(k, (w_1, w_2), U) \cap (k', (w'_1, w'_2), U') = \emptyset$$

if $k \neq k'$. This proves the first assertion.

It remains to show that for any combinatorial type π with $q(\pi) = q$ and $n_1(\pi) = k$ the σ-algebra generated by the subsets $(k, (w_1, w_2), U)$ which lie in $\Delta^\pi_{u,v}$ coincides with the Borel σ-algebra. Since

$$\Delta^\pi_{u,v} = \{x_1, \ldots, x_q | \, u < x_1 < \cdots < x_q < v\}$$

its Borel σ algebra is generated by subsets of the form $w_1 < x_1 < w_2$, $(x_2, \ldots, x_q) \in U$ where U is a Borel measurable subset of

$$\Delta^{R(\pi)}_{w_2,v} = \{w_2 < x_2 < \cdots < x_q < v\}.$$

Observe now this subset coincides with $(k, (w_1, w_2), U)$ where U is considered as a subset of $\Delta^{R(\pi)}_{w_2,v}$.
$\square$

Proposition 1.2.18 *[**crit2**] An additive pre-process μ on $\mathcal{H}[s, t]$ is a Markov pre-process if and only if the following two conditions hold:*

1. *for any $u < v < w$ and $U \subset H[v, w]^*_1$ one has*

$$\mu^w_{u,1}(j^w_{u,v}(U)) = h(u, v)\mu^w_{v,1}(U)$$

2. for any $u < w_1 < w_2 < v$, $k \neq 1$ and $U \in H[w_2, v]^*_k$ one has

$$\mu^v_{u,1}(k, (w_1, w_2), U) = \mu^{w_2}_{u,1}((w_1, w_2) \subset \Delta^{[k[1]]}_{u,w_2})\mu^v_{w_2,k}(U)$$

Proof: The "only if" part is obvious. To prove the "if" part we need to verify the condition of Lemma 1.2.1. Using additivity one can easily see that if this condition holds for all $U_1 \subset sk_q H[v,t]_{1,*}$ then it holds for all n and all $U_1 \subset sk_q H[v,t]_{n,*}$. Therefore we may proceed by induction on q and for each q we only need to consider the case $n = 1$.

Let $q = 0$. Then we have to consider two cases.

1. Let $U_1 = \Delta^{[1]}_{u,v}$. Then
$$(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = j^w_{u,v}(U_2)$$
and the condition of Lemma 1.2.1 follows immediately our condition (1).

2. Let $U_1 = \Delta^{[n]}_{u,v}$ where $n \neq 1$. Then
$$(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = \left\{\Gamma \in sk_{>0}H[u,w]^*_1 | x_1(\Gamma) = v, n_1(\Gamma) = n, R(\Gamma) \in j^w_{u,v}(U_2)\right\}$$
We may assume without loss of generality that there exists $v < v' < w$ such that $U_2 = j^w_{v,v'}(U'_2)$. Then
$$(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = \cap_{(w_1,w_2)\in I}(n, (w_1, w_2), U'_2)$$
where I is the set of pairs $w_1 < w_2$ such that $U < w_1 < v < w_2 < v'$ and $w_1, w_2 \in \mathbf{Q}$. By our second condition and σ-additivity of μ we conclude that
$$\mu^v_{u,1}(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = \lim_I \mu^v_{u,1}((n, (w_1, w_2), U'_2)) = \lim_I \mu^{w_2}_{u,1}((w_1, w_2))$$
It follows by an obvious limit argument from Lemma (1.3.5), that the value of $\mu^w_{1,u}$ on this subset is $\theta_{u,n}(\{v\})\mu^w_{v,n}(U_2)$ which together with Lemma 1.3.4 implies the condition of Lemma 1.2.1 in this case.

Let $q > 0$. Assume by induction that the condition of Lemma 1.2.1 is known for all $U_1 \subset sk_{q-1}H[u,v]$ and all U_2 and let $U_1 \subset sk_q H[u,v]_{1,*}$. By Lemma 1.2.17 we may assume that $U_1 = (k, (w_1, w_2), U'_1)$ where $u < w_1 < w_2 < v$ and U'_1 is a measurable subset of $sk_{q-1}H[w_2, v]_{m,*}$ for some $m \neq 1$. Then

$$[\mathbf{ss}](res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = (k, (w_1, w_2), (res_{w_2,v} \times res_{v,w})^{-1}(U'_1 \times U_2)) \qquad (15)$$

By the inductive assumption

$$\mu^w_{x,k}(j^w_{x,w_2}(res_{w_2,v} \times res_{v,w})^{-1}(U'_1 \times U_2)) = \mu^w_{x,k}(j^w_{x,w_2}\, res^{-1}_{w_2,v}(U'_1))\mu^w_{v,l}(U_2)$$

where l is such that $U_2 \subset n_v^{-1}(\{l\})$. By Lemma 1.3.5, the value of $\mu^w_{u,1}$ on (29) is

$$\int_{x\in(w_1,w_2)} \mu^w_{x,k}(j^w_{x,w_2}(res_{w_2,v} \times res_{v,w})^{-1}(U'_1 \times U_2))d\theta_{u,m} =$$

$$= \left(\int_{x \in (w_1, w_2)} \mu^w_{x,k} (j^w_{x,w_2}\, res^{-1}_{w_2,v}(U'_1)) d\theta_{u,k} \right) \mu^w_{v,l}(U_2)$$

and using Lemma 1.3.5 again we get (6). Theorem is proved.

$\square$

For $u < v$ and $k \neq 1$ let $\lambda^v_{u,k}$ denote the co-restriction of $\mu^v_{u,1}$ to $\Delta^{[k[1]]}_{u,v} \amalg \Delta^{[k]}_{u,v} = (u, v]$.

Proposition 1.2.19 *[**adddet**] An additive Markov pre-process on $\mathcal{H}[s,t]$ is determined by the function $h(-,-)$ and measures $\lambda^v_{u,k}$ for $k \neq 1$ and $s \leq u < v \leq t$.*

Proof: Let μ and ν be two additive Markov pre-processes such that the corresponding functions h and measures λ coincide. Let us prove that the restrictions of μ and ν to $sk_q H[u,v]_{n,*}$ coincide for all n and q. Let

$$add_n : (H[u,v]_{1,*})^n \to H[u,v]_{n,*}$$

be the iterated addition map. Since $add_n^{-1}(sk_q H[u,v]_{n,*}) \subset (sk_q H[u,v]_{1,*})^n$ and our pre-processes are additive it is sufficient to show that they coincide on $sk_1 H[u,v]_{1,*}$ and that if they coincide on $sk_{q-1} H[u,v]_{n,1}$ for all n then they coincide on $sk_q H[u,v]_{1,*}$. That μ and ν coincide on $sk_1 H[u,v]_{1,*}$ follows from the definition of h and λ and the fact that

$$sk_1 H[u,v]_{1,*} = \Delta^{[1]}_{u,v} \amalg (\amalg_{k \neq 1} (\Delta^{[k[1]]}_{u,v} \amalg \Delta^{[k]}_{u,v}))$$

Assume that they coincide on $sq_{q-1} H[u,v]_{n,*}$ for all n and all u, v. By Lemma 1.2.17 it is sufficient to show that they coincide on subsets $(k, (w_1, w_2), U)$ in $sk_q H[u,v]_{1,*}$.

We have

$$(k, (w_1, w_2), U) = (res_{u,w_2} \times res_{w_2,v})^{-1}(((w_1, w_2) \subset \Delta^{[k[1]]}_{u,w_2}) \times U$$

and we conclude that μ and ν agree on $(k, (w_1, w_2), U)$ by the Markov property. $\square$

Let μ be an additive pre-process. For any $u \in [s,t]$ let $e(u)$ be the smallest element of E which is greater than u. If no such element exist i.e. if $h(u,t) \neq 0$ we set $e(u) = \infty$.

For $s \leq u < v \leq t$ and $k \neq 1$ define measures $\alpha^v_{u,k}$ on $(u, v]$ by the formula:

$$\alpha^v_{u,k} = ((x_1, k_1)_*(\mu^v_{u,1}))^{|(u,v] \times k}.$$

Intuitively, $\alpha^v_{u,k}(B)$, for a measurable B in $(u, v]$, is the probability that a singleton which is alive at time u will have its history traceable up to time v and the first transformation event in this history will occur at $x \in B$ and will have multiplicity k.

Theorem 1.2.20 *[**th2**] For an additive Markov process μ^*_*, any $k \neq 1$ and $s \leq u < v \leq t$ one has*

$$[\textbf{th2eq0}] \lambda^{v,k}_u = (\alpha^v_{u,k}) * h^k(-, v) \tag{16}$$

Proof: One verifies immediately using the Markov property for u, v, t that the two measures agree on $\{v\}$. Therefore it is sufficient to show that

$$[\mathbf{th2eq0a}](\lambda_u^{v,k})^{|(u,v)} = ((\alpha_{u,k}^v) * h^k(-,v))^{|(u,v)} \tag{17}$$

For convenience we will consider (16) as an equality of two measures on $[u, v)$ which are zero on $\{u\}$.

Lemma 1.2.21 *[**th2l1**] For any Markov sub-process μ_*^*, any $k \neq 1$ and any $s \leq u \leq y < y_+ \leq v \leq t$ one has*

$$[\mathbf{th2eq1}]\lambda_{u,k}^v([y, y_+)) = \lambda_{u,k}^{y_+}([y, y_+))h^k(y_+, v) \tag{18}$$

$$[\mathbf{th2eq2}] \quad \begin{gathered} \lambda_{u,k}^{y_+}([y, y_+))v_{y_+,k}^v = \\ = \mu_{u,n}^v(\{\Gamma \in H[u,v] \,|\, (x_1, k_1)(\Gamma) \in [y, y_+) \times \{k\} \text{ and } x_1(R(\Gamma)) > y_+\}) \end{gathered} \tag{19}$$

Proof: Equation (18) follow from Lemma 1.2.1 with $U_1 = \{[y, y_+) \subset \Delta_{u,y_+}^k\}$ and $U_2 = \Delta_{y_+,v}^{k[1]}$.

Equation (19) follow from Lemma 1.2.1 with $U_1 = \{[y, y_+) \subset \Delta_{u,y_+}^k\}$ and $U_2 = H[y_+, v]_{k,*}$. $\square$ It is sufficient to consider the cases $v > e(u)$ and $v < e(u)$. Suppose that $v > e(u)$. Then Markov property applied to points $u, e(u), v$ implies that the measures on both sides of (16) are supported in $e(u)$ and their values at this point agree. Assume that $v < e(u)$. Then for all $x \in [u, v)$ one has $h(x, v) \geq h(u, v) > 0$ and (16) is equivalent to the assertion that for all $w \in [u, v)$ one has

$$[\mathbf{th2eq3}]\alpha_{u,k}^v([u, w)) = \int_{x \in [u,w)} (h(x, v))^{-k} d\lambda_{u,k}^v \tag{20}$$

Let us denote the function under the integral by $f(x)$ and the measures involved by α and λ respectively.

Lemma 1.2.22 *[**th2l2**] For all $\epsilon > 0$ there exists $\delta > 0$ such that for any partition $u = x_0 < \cdots < x_n = w$ of the interval $[u, w)$ such that $|x_{i+1} - x_i| < \delta$ one has*

$$\sum |\, \alpha([x_i, x_{i+1})) - f(x_{i+1})\lambda([x_i, x_{i+1})) \,| < \epsilon$$

Proof: By Lemma 1.2.21 we have

$$\begin{gathered} f(x_{i+1})\mu_{u,k}^v([x_i, x_{i+1})) = \\ = \mu_{u,n}^v(\{\Gamma \in H[u,v] \,|\, (x_1, k_1)(\Gamma) \in [x_i, x_{i+1}) \times \{k\} \text{ and } x_1(R(\Gamma)) > x_{i+1}\}) \end{gathered}$$

Therefore

$$\begin{gathered} \alpha([x_i, x_{i+1})) - f(x_{i+1})\lambda([x_i, x_{i+1})) = \\ = \mu_{u,n}^v(\{\Gamma \in H[u,v] \,|\, (x_1, k_1)(\Gamma) \in [x_i, x_{i+1}) \times \{k\}, \text{ and } x_1(R(\Gamma)) \leq x_{i+1})\}) \end{gathered}$$

If $|x_{i+1} - x_i| < \delta$ we conclude that

$$\sum_i \alpha([x_i, x_{i+1})) - f(x_{i+1})\lambda([x_i, x_{i+1})) \leq \mu^v_{u,n}(\{\Gamma \in sk_{>1}H[u,v]_{n,*} \,|\, x_1(R(\Gamma)) - x_1(\Gamma) < \delta\})$$

Since

$$\bigcap_{\delta \to 0} \{\Gamma \in sk_{>1}H[u,v]_{n,*} \,|\, x_1(R(\Gamma)) - x_1(\Gamma) < \delta\} = \emptyset$$

we conclude by σ-additivity of $\mu^v_{u,n}$ that for each $\epsilon > 0$ there exists $\delta > 0$ such that

$$\sum_i \alpha([x_i, x_{i+1})) - f(x_{i+1})\lambda([x_i, x_{i+1})) < \epsilon$$

and all terms in this sum are non-negative. $\square$ Let $(h(u,w))^{-k} = C < \infty$. To prove the theorem it remains to verify that

$$inf \left\{ \sum_i \,|\, f(x_{i+1})\lambda([x_i, x_{i+1})) - \int_{x \in [x_i, x_{i+1})} f(x) d\lambda \,|\right\} = 0$$

where inf is taken over all partitions $u = x_0 < \cdots < x_n = w$ of $[u,w)$. Since both $\int_{x \in [x_i, x_{i+1})} f(x)d\lambda$ and $f(x_{i+1})\lambda([x_i, x_{i+1}))$ lie between $inf_{x \in [x_i, x_{i+1}]} f(x)\lambda([x_i, x_{i+1}))$ and $sup_{x \in [x_i, x_{i+1}]} f(x)\lambda([x_i, x_{i+1}))$ it is sufficient to verify that

$$[\mathbf{th2eq4}] inf \left\{ \sum_i \,|\, sup_{x \in [x_i, x_{i+1}]} f(x) - inf_{x \in [x_i, x_{i+1}]} f(x) \,|\, \lambda([x_i, x_{i+1})) \right\} = 0 \qquad (21)$$

Lemma 1.2.23 *[**th2l4**] Let f be a right continuous monotone increasing function on $[u,w]$. Then for all $\epsilon > 0$ there exists a finite set of points $a_1, \ldots a_{N(\epsilon)} \in [u,w)$ and $\delta > 0$ such that for all $(y, y_+] \subset [u,w) \backslash \{a_1, \ldots, a_N\}$ satisfying $|y_+ - y| < \delta$ one has $|sup_{x \in [y,y_+]} f(x) - inf_{x \in [y,y_+]} f(x)| < \epsilon$.*

Proof: Observe first that if the conclusion of the lemma holds for two functions then it holds for their sum. Since f is of right continuous and monotone increasing we can write it as a sum $f = f_1 + f_2$ where f_1 is continuous and f_3 is a right continuous step function with countable set of points of discontinuity (see e.g. [1]). In addition both functions are monotone increasing.

For f_1 which is continuous we may take $N = 0$ since a bounded continuous function on an interval is uniformly continuous and $sup_{x \in [y,y_+]} f(x) - inf_{x \in [y,y_+]} f(x) = f(y_+) - f(y)$.

Let A be the set of discontinuity points of f_2 and for $a \in A$ let $\Delta(f,a)$ be the jump in this point. Then $\sum_{a \in A} \Delta(f,a) < \infty$. Therefore there is a finite number of points $a_1, \ldots, a_N \in A$ such that $\sum_{a \in A'} \Delta(f,a) < \epsilon$ where $A' = A \backslash \{a_1, \ldots, a_N\}$. The conclusion of the lemma is then satisfied for these points $a_1, \ldots, a_N$ and any $\delta > 0$. If $[y, y_+] \subset [u,w) \backslash \{a_1, \ldots, a_N\}$ then obviously $|sup_{x \in [y,y_+]} f_2(x) - inf_{x \in [y,y_+]} f_2(x)| < \epsilon$. If $(y, y_+] \subset [u,w) \backslash \{a_1, \ldots, a_N\}$ but $y \in \{a_1, \ldots, a_N\}$ we still have $|sup_{x \in [y,y_+]} f_2(x) - inf_{x \in [y,y_+]} f_2(x)| < \epsilon$ due to the fact that f_2 is right continuous. $\square$ To prove (21) we have to show that for any $\epsilon > 0$ there exists a partition such that

$$[\mathbf{th2eq5}] \sum_i \,|\, sup_{x \in [x_i, x_{i+1}]} f(x) - inf_{x \in [x_i, x_{i+1}]} f(x) \,|\, \lambda([x_i, x_{i+1})) < \epsilon \qquad (22)$$

Let

$$C_1 = \lambda([u,w))$$

$$C_2 = sup_{x\in[u,w]} f(x) - inf_{x\in[u,w]} f(x)$$

Using Lemma 1.2.23 let we may find a finite subset $a_1, \ldots, a_N$ and $\delta > 0$ such that for any $(y_+, y] \in [u,w)$ satisfying $y_+ - y < \delta$ one has

$$|sup_{x\in[y,y_+]} f(x) - inf_{x\in[y,y_+]} f(x)| < \epsilon/2C_1$$

Consider partitions which contain intervals $[a_i - \delta', a_i)$, the lengths of all the intervals are less than δ and each interval contains at most one of the points from $a_1, \ldots, a_N$. By σ-additivity of λ we can choose δ' such that

$$\sum_i \lambda([a_i - \delta', a_i)) < \epsilon/2C_2$$

Elementary computation shows that for such a partition (22) is satisfied. $\square$

Embeddings $j^w_{u,v}$ allow us to consider a process μ^*_* on $\mathcal{H}[s,t]$ as a collection of measures $j^v_{s,u} \circ \mu^v_{u,n}$ on spaces $H[s,v]_{n,*}$ for $v \le t$.

Definition 1.2.24 *[***com***] A process on $\mathcal{H}[s,t]$ is called co-measurable if for all n and v the mappings*

$$u \mapsto j^v_{s,x} \circ \mu^v_{x,n}$$

are kernels from $[v,t)$ to $H[v,t]_{n,*}$.

Remark 1.2.25 The name co-measurable is chosen to avoid confusion with standard notion of a measurable process. See e.g. [3].

Lemma 1.2.26 *[***th2l8***] If μ is co-measurable then the mapping $x \mapsto \sum_{k\neq 1} \alpha^v_{x,k} \otimes \delta_{\{k\}}$ is a sub-probability kernel from $[s,v)$ to $(s,v] \times \mathbf{N}_{\neq 1}$.*

Proof: It is sufficient to show that for any $k \neq 1$ and any $s < u \le v$ the function $x \mapsto \alpha^v_{x,k}((s,u])$ is a measurable function on $[s,v)$. This function is zero for $x \ge u$ and for $x < u$ one has

$$\alpha^v_{x,k}((s,u]) = \mu^v_{x,1}((j^v_{s,x})^{-1}(x_1,k_1)^{-1}((s,u] \times \{k\}))$$

which proves the lemma. $\square$

Proposition 1.2.27 *[***arecom***] An additive Markov process μ_* on $\mathcal{H}[s,t]$ is co-measurable.*

Proof: Since for a Markov process measures $\mu^v_{u,n}$ are projections of measures $\mu^t_{u,n}$ it is sufficient to consider the case $v = t$.

Since our process is additive it is further sufficient to show that the measures $\mu^t_{x,1}$ considered as measures on $H[s,t]$ form a kernel from $[s,t)$. In view of Lemma 1.2.17 it is sufficient to verify that the functions

$$f_1 : x \mapsto \mu^t_{x,1}(\Delta^{[1]}_{x,t})$$

$$f_2 : x \mapsto \mu^t_{x,1}(\Delta^{[k]}_{x,t}) \text{ for } k \neq 1$$

and

$$f_3 : x \mapsto \mu^t_{x,1}(H[x,t] \cap (k,(w_1,w_2),U))$$

are measurable. For any Markov process functions f_1, f_2 are monotone increasing on $[s,t)$ and therefore are measurable. To show that f_3 is measurable let $I_1 = (s,w_1)$, $I_2 = (w_1,w_2)$ and $I_3 = (w_2,t)$. It is clearly sufficient to verify that the restrictions of f_3 to I_1, I_2 and I_3 are measurable. Observe first that

$$\text{for}, x \in I_1 \text{ one has } H[x,t] \cap (k,(w_1,w_2),U) = (k,(w_1,w_2),U),$$

$$\text{for}, x \in I_2 \text{ one has } H[x,t] \cap (k,(w_1,w_2),U) = (k,(x,w_2),U),$$

$$\text{for}, x \in I_3 \text{ one has } H[x,t] \cap (k,(w_1,w_2),U) = \emptyset.$$

Using Markov property we conclude that

$$f_3(x \in I_1) = h(x,w_1) f_3(w_1)$$

which is measurable since $h(-,w_1)$ is a monotone increasing function. To prove that f_3 is measurable on I_2 it is sufficient to show that it is measurable on $I_2 \cap [e,e_{+1})$ for all $e \in E$. For x in this intersection we have

$$f_3(x) = h(e,x)^{-1} f_3(e)$$

and since $h(e,-)$ is measurable and non zero on $[e,e_{+1})$ we conclude that f_3 is measurable. $\square$

Corollary 1.2.28 *[**comcor**] Let μ^*_* be an additive Markov process. Then the mapping*

$$x \mapsto \sum_k \alpha^v_{x,k} \otimes \delta_k$$

defines a sub-probability kernel from $[s,v)$ to $(s,v] \times \mathbf{N}_{\neq 1}$.

Proof: It follows from Lemma 1.2.26 and Proposition 1.2.27. $\square$

Lemma 1.2.29 *[**th2l5**] For an additive Markov process and any $s \leq u < v \leq t$ one has*

$$\alpha^v_{u,k} = (\alpha^t_{u,k})^{|(u,v]}$$

Proof: Follows from the Markov property with respect to the points u, v, t. $\square$

Lemma 1.2.30 *[**th2l6**] For an additive Markov process and any $s \leq u < v \leq t$ one has*

$$[\mathbf{maineq1}] h(u,v) = 1 - \sum_{k \neq 1} \alpha^t_{u,k}((u,v]) \tag{23}$$

Proof: For any process one has

$$h(u,v) = 1 - \sum_{k\neq 1} \alpha^v_{u,k}((u,v])$$

together with Lemma 1.2.29 it implies (23). $\square$

Lemma 1.2.31 *[th2l7] Let μ be an additive Markov process. Then for any $k \neq 1$ and any $s \leq u < v \leq t$ one has*

$$(\alpha^t_{u,k})^{|(v,t]} = h(u,v)(\alpha^t_{v,k})^{|(v,t]}$$

Proof: The condition (3) is equivalent to the condition that for $k \neq 1$ and $s \leq u < v < w \leq t$ one has

$$\alpha^t_{u,k}((v,w]) = h(u,v)\alpha^t_{v,k}((v,w])$$

which we get from immediately from Lemma 1.2.1 applied to points u, v, t and subsets

$$U_1 = \Delta^{[1]}_{u,v}$$

$$U_2 = \{\Gamma \in H[v,t]_{1,*} | x_1(\Gamma) \leq w, k_1(\Gamma) = k\}\,.$$

$\square$

Proposition 1.2.32 *[pr4] An additive Markov process on $\mathcal{H}[s,t]$ is completely determined by the collection of measures $\alpha^t_{u,k}$ for $s \leq u < t$ and $k \neq 1$ on $(u,t]$.*

Proof: It follows from Proposition 1.2.19, Lemma 1.2.30 and Theorem 1.2.20. $\square$

Summarizing some of the results of this section we see that any additive Markov process μ on $[s,t]$ defines a sub-probability kernel $[s,t) \to (s,t] \times \mathbf{N}_{\neq 1}$ of the form

$$u \mapsto \sum_{k \neq 1} \alpha^t_{u,k} \otimes \delta_{\{k\}}$$

such that for the function $h(-,-)$ defined by the formula of Lemma 1.2.30 one has:

$$h(u,v)h(v,w) = h(u,w)$$

and

$$\alpha^{|(v,t]}_{u,k} = h(u,v)\alpha^{|(v,t]}_{v,k}$$

and moreover that μ is uniquely determined by $\alpha^t_{u,k}$.

Let us say that a process μ is irreducible if for all $v < t$ one has $\mu^v_{s,1}(\Delta^{[1]}) \neq 0$. An irreducible additive Markov process is completely determined by a single sub-probability measure $\alpha^t_{s,*}$ on $(s,t] \times \mathbf{N}_{\neq 1}$ such that $\alpha^t_{s,*}((s,t) \times \mathbf{N}_{\neq 1}) < 1$.

We will see in the next section that to any measure satisfying this condition and such that in addition $\alpha_{s,k} = 0$ for sufficiently large k there corresponds a unique irreducible additive Markov process on

$\mathcal{H}[s,t]$ therefore obtaining a complete classification of irreducible additive Markov processes with restricted branching multiplicities on $\mathcal{H}[s,t]$.

A process which is not irreducible may be considered as a collection of irreducible processes on $\mathcal{H}[e,e_+]$ for $e \in E$. Conversely, for any countable subset $E \subset [s,t]$ such that for any $e \in E$ there exists $e_+ \in E \cap \{t\}$ satisfying the condition $(e,e_+) \cap E = \emptyset$ and any collection of additive Markov processes on $\mathcal{H}[e,e_+]$ such that, in addition $h(e,e_+) = 0$ for $e_+ \in E$ there exists a unique process on $\mathcal{H}[s,t]$ with this E and these restrictions to intervals $[e,e_+]$. Due to this fact we will often restrict our attention below to irreducible processes.

Proposition 1.2.33 *[**pr5**] An additive Markov process on $\mathcal{H}[s,t]$ is uniquely determined its transition kernels ϕ_u^v.*

Proof: The transition kernels determine the projections of measures $\mu_{u,n}^v$ under the map

$$n : H[u,v]_{n,*} \rightarrow \mathbf{N}^{[u,v]}$$

which sends Γ to the function $n_\Gamma : x \mapsto n_x(\Gamma)$. In view of Proposition 1.2.32 it remains to show that these projections determine the measures $\alpha_{u,k}^v$. It follows immediately from the definition of this measures and the lemma below. $\square$

Lemma 1.2.34 *[**p5l1**] Let A be a dense countable subset of (u,v). Then for $\Gamma \in H[u,v]_{1,*}$ one has:*

1. *for $k \neq 1$ and $w \leq v$, $(x_1,k_1)(\Gamma) \in (u,w) \times \{k\}$ if and only if for all $N > 0$ there exists $a_1, a_2 \in A$ such that $u < a_1 < a_2 < w$, $|a_2 - a_1| < \epsilon$, $n_\Gamma(a_2) = k$ and for all $a \in A$ such that $a \leq a_1$, $n_\Gamma(a) = 1$,*
2. *for any k, $(x_1,k_1)(\Gamma) = (v,k)$ if and only if $n_\Gamma(v) = k$ and for all $a \in A$ such that $a < v$ one has $n_\Gamma(a) = 1$.*

Proof: Straightforward, using the fact that the functions n_Γ are right continuous. $\square$

Lemma 1.2.35 *[**kcontl1**] For an additive Markov process and $k \neq 1$ such that for all $e \in E$, $\alpha_{e,k}^t(\{e_+\}) = 0$ there exists a unique measure γ_k such that for any $u \in [s,t)$ one has*

$$[\mathbf{kconteq1}]\alpha_{u,k}^t = h(u,-) * \gamma_k \tag{24}$$

Proof: Let γ_k be the unique measure on $[s,t]$ such that for any $e \in E_\mu$ one has

$$\gamma_k^{|[e,e_+)} = h(e,-)^{-1} * (\alpha_{e,k}^t)^{|[e,e_+)}$$

Let us show that it satisfies the condition of the lemma. Since $\alpha^t_{u,k}(u_+) = 0$, both sides of (**??**) are concentrated on (u, u_+) and it is sufficient to check that

$$h(u,-)^{-1} * (\alpha^t_{u,k})^{|(u,u_+)} = \gamma_k^{|(u,u_+)}$$

which follows immediately from Lemmas 1.2.31 and 1.2.5.

If γ_k and γ'_k are two measures satisfying the condition of the lemma then they are equal on each interval $[e, e_+)$ they coincide with $h(e,-)^{-1}(\alpha^t_{e,k})^{|(e,e_+)}$ and therefore they coincide with each other. $\square$ Measure γ_k is called the rate measure for events of multiplicity k. Note that these measures are bounded on closed intervals which do not contain points from E but may be unbounded around points from this set. Because of the structure of E we get the following property of γ_k's:

Lemma 1.2.36 *[**th6l3**] For any $k \neq 1$ and any $x \in [s,t)$ there exists $x' > x$ such that $\gamma_k((x,x')) < \infty$.*

For a measure γ satisfying the conclusion of Lemma 1.2.36 define $E(\gamma)$ as the set such that $x \in E(\gamma)$ if and only if $x = s$ or $x > s$ and for all $x' < x$ one has $\gamma_k((x',x)) = \infty$. The conclusions of Lemmas 4.3.9, 4.3.10 hold, with obvious modifications, for the sets $E(\gamma)$. This implies in particular that measures γ_k are σ-finite.

Lemma 1.2.37 *[**th6l4**] For any k one has $E(\gamma_k(\mu)) \subset E_\mu$.*

Proof: Follows immediately from the fact that γ_k for any k is bounded on closed intervals which do not contain points of E_μ. $\square$

Recall that a measure on $\mathbf{R}$ is called non-atomic if its value on any point is zero. A measure is non-atomic if and only if its distribution function is continuous.

Definition 1.2.38 *[**kcont**] An additive Markov process is called k-continuous if measures $\alpha^t_{u,k}$ are non-atomic for all $u \in [s,t)$. An additive Markov process is called continuous if it is k-continuous for all $k \neq 1$.*

For a measure α on an interval $[u,v]$ let $Distr(\alpha)$ be the (right-continuous) distribution function of α:

$$Distr(\alpha)(x) = \alpha[u,x]$$

Lemma 1.2.39 *[**th6l1**] Let γ be a bounded non-atomic measure on $[u,v]$ and F be a bounded non-negative measurable function on this interval. Then there exists a unique (bounded) measure α on $[u,v]$ such that*

$$[\mathbf{inteq}](F - Distr(\alpha)) * \gamma = \alpha \tag{25}$$

The function $Distr(\alpha)$ is of the form

$$[\mathbf{inteq1}]Distr(\alpha)(x) = e^{-G(x)} \int_{t\in[u,x]} F(t)e^{G(t)} d\gamma \tag{26}$$

where $G = Distr(\gamma)$. If F takes values in $[0,1]$ then α is a sub-probability measure.

Proof: Consider first the case when $\gamma = dx$ is the Lebesgue measure on $[u,v]$. Then solutions of (25) are in one to one correspondence with monotone increasing non-negative functions $A = Distr(\alpha)$ such that for all $x \in [u,v]$ one has

$$\textbf{[inteq2]} \int_u^x (F(t) - A(t))dt = A(x) \tag{27}$$

which holds if and only if A is absolutely continuous, $A(u) = 0$ and

$$A'(x) = F(x) - A(x)$$

almost everywhere (see e.g. [1, p.106]). Set

$$A(x) = e^{-x} \int_u^x F(t)e^t dt$$

Then A is absolutely continuous and the equation (27) is satisfied (e.g. by [1, p.108,ex.35]). Since $e^{t-x} \leq 1$ on $[0,x]$ we have $A \leq F$ and by (27) we conclude that A is monotone increasing. If there are two solutions A_1, A_2 then their difference $A = A_1 - A_2$ is a solution of the equation

$$A(x) = -\int_u^x A(t)dt$$

A solution of this equation has a continuous derivative everywhere which implies that it is zero by the standard uniqueness for the ordinary differential equations.

Consider now the case of general non-atomic γ. We may assume that $\gamma \neq 0$. Its distribution function is a continuous mapping

$$G : [u,v] \to [0, G(v)]$$

Define a mapping $X_+ : [0, G(v)] \to [u,v]$ by the formula

$$X^+(y) = sup\{x \,|\, G(x) \leq y\}$$

Then $\gamma = (X^+)_*(dy)$ (see e.g. [7, p.34]). Since G is continuous, X^+ is an order preserving embedding with a measurable image U which a complement to the disjoint union of countably many intervals of the form $[y, y')$. Any solution α of (25) is supported in U and therefore there exists a unique measure α_1 on $[0, G(v)]$ such that $\alpha = (X^+)_*(\alpha_1)$. Using the fact that X^+ is an order preserving embedding one verifies further that

$$Distr(\alpha_1) = Distr(\alpha) \circ X^+$$

Therefore the mapping $\alpha_1 \mapsto X^+_*(\alpha_1)$ defines a bijection between the solutions of (25) and the solutions of the equation

$$(F \circ X^+ - Distr(\alpha_1)) * dy = \alpha_1$$

which implies, by the first part of the proof, that a solution to (25) exists and is unique. The explicit formula for

$$Distr(\alpha) = Distr(X^+_*(\alpha_1))$$

where

$$Distr(\alpha_1)(y) = e^{-y} \int_0^y F(X^+(z))e^z dz$$

easily follows. $\square$

Remark 1.2.40 It is plausible that the equation (25) has a unique solution for any bounded γ. For example, if γ is the delta-measure concentrated at a point $a \in [u,v]$ then there exists a unique solution and it is of the form $(F(a)/2)\delta_a$. In this case however it is not given by the formula (26) which, for $\gamma = \delta_a$ gives the distribution function of $F(a)\delta_a$.

Example 1.2.41 [**th6ex1**] If, in Lemma 1.2.39, we have $F = c$ where c is a constant then

$$Distr(\alpha)(x) = c(1 - e^{-Distr(\gamma)})$$

Lemma 1.2.42 *[***th6l2***] Let μ, μ' be two irreducible processes and I a subset of $\mathbf{N}_{\neq 1}$ such that*

1. *for $i \in I$ the processes μ and μ' are i-continuous and $\gamma_i(\mu) = \gamma_i(\mu')$,*
2. *for $j \in J = \mathbf{N}_{\neq 1} \backslash I$ one has $\alpha^t_{s,j}(\mu) = \alpha^t_{s,j}(\mu')$.*

Then $\mu = \mu'$.

Proof: Let $\gamma_I = \sum_{i\in I} \gamma_i$, $\alpha_I = \sum_{i \in I} \alpha^t_{u,i}$ and A_I be the distribution function for α_I. Then the definitions of h and γ_i imply that

$$\alpha_I = (1 - A_J - A_I) * \gamma$$

where A_J is the distribution function for $\sum_{j\in J} \alpha^t_{s,j}$.

Applying Lemma 1.2.39 to the restrictions of γ to intervals $[s,u]$ with $u < t$ we conclude that there is a unique α_I satisfying this equation. Then

$$h = 1 - (A_J + A_I)$$

and we recover measures $\alpha^t_{s,i}$ for $i \in I$ from the defining property of γ_i. $\square$

Example 1.2.43 [**th6ex2**] In the assumptions of Lemma 1.2.42 we have the following explicit formulas for $A_i = Distr(\alpha^t_{s,i})$ for $i \in I$. As in the proof set

$$A_J = Distr(\sum_{j\in J} \alpha^t_{s,j})$$

$$A_I = Distr(\sum_{i\in I} \alpha^t_{s,i})$$

and let

$$\gamma = sum_{i\in I}\gamma_i$$
$$G = Distr(\gamma)$$

Then

$$A_I(x) = e^{-G(x)} \int_{t\in[s,x]} (1 - A_J(t)) e^{G(t)} d\gamma$$

and

$$A_i(x) = \int_{t\in[s,x]} (1 - A_I(t) - A_J(t)) d\gamma_i$$

Example 1.2.44 [**th6ex3**] Combining Examples 1.2.41 and 1.2.43 we get in the case when $J = \emptyset$ the following explicit formulas for A_i:

$$A_i = \int_{t\in[s,x]} e^{-\sum_{i\in I} Distr(\gamma_i)(t)} d\gamma_i$$

In particular, if $\gamma_i = c_i dx$ where c_i's are constants we get:

$$A_i(x) = (c_i/\sum_{i\in I} c_i)(1 - e^{-(\sum c_i)(x-s)})$$

$$\alpha^t_{s,i} = c_i e^{-(\sum c_i)(x-s)} dx$$

The following proposition connects measures γ_k with the probability density functions for events of multiplicity k which form the basis of the classical theory of branching Markov processes.

Proposition 1.2.45 *[**pr7**] An additive Markov process is k-continuous if and only if for any $u \in [s,t)$ one has*

$$\lim_{\epsilon\to 0, \epsilon>0} \phi_u^{u+\epsilon}(1,k) = 0$$

In the case when $\gamma_k = g_k(x)dx$ one further has

$$\lim_{\epsilon\to 0, \epsilon>0} \phi_u^{u+\epsilon}(1,k)/\epsilon = g_k(u)$$

Proof: ??? $\square$

1.3 Construction of processes

We start with a construction of a wide class of additive processes on $\mathcal{H}[s,t]$ not all of which are Markov processes. Let

$$\theta : [s,t) \to [s,t] \times \mathbf{N}_{\neq 1}$$

be a sub-probability kernel such that for any $u \in [s,t)$ the measure $\theta(u)$ is concentrated on $(u,t]$. As usually we will write θ_u instead of $\theta(u)$ and $\theta_{u,k}$ for the co-restriction of θ_u to $[s,t] \times \{k\}$.

Then there exists a measurable space $(\Omega, \mathfrak{F})$ and a probability measure P on it together with a measurable map

$$I = I_\theta : [s,t) \times \Omega \to ((s,t] \times \mathbf{N}_{\neq 1}) \amalg pt$$

such that

$$[\mathbf{aalpha}] I(x,-)_*(P) = \theta(x). \tag{28}$$

Let $I_u : \Omega \to ((s,t] \times \mathbf{N}_{\neq 1}) \amalg pt$ be the restriction of I to $\{u\} \times \Omega$.

Remark 1.3.1 In fact we can always choose $\Omega = [0,1]$ and $P = dx$. The map I_u in this case is defined by the following explicit algorithm starting with the distribution functions $T_{u,k}$ of measures $\theta_{u,k}$.Let $c_{u,k} = \theta_{u,k}([u,t])$ and let $C_k = c_0 + \cdots + c_{k-1}$. Define, following [7, p.34], a map

$$I^+_{u,k} : [0, c_k) \to [s,t]$$

by

$$I^+_{u,k}(y) = sup\{y \,|\, T_{u,k}(x) \le y\}$$

For $y \in [0,1]$ set

$$I_u(y) = \begin{cases} I_{u,k}(y) & \text{if } C_k \le y < C_{k+1} \\ pt & \text{otherwise} \end{cases}$$

Since $\theta(u)$ is concentrated on $(u,t] \times \mathbf{N}_{\neq 1}$ we may assume that

$$I_u(\Omega) \subset ((u,t] \times \mathbf{N}_{\neq 1}) \amalg pt.$$

Let us define subsets $X^v_{u,n,N}$ of Ω^∞ inductively as follows:

$$X^v_{u,n,0} = \emptyset$$

and for $N > 0$, $X^v_{u,0,N} = \Omega^\infty$ and for $n > 0$:

$$X^v_{u,n,N} = \left\{ \underline{\omega} \in \Omega^\infty | \; \forall 1 \le i \le n \; (X_u(\omega_i) \in (([v,t] \times \mathbf{N}_{\neq 1}) \amalg pt) \, or \, (\omega_{i+n}, \omega_{i+2n}, \dots) \in X^v_{X(\omega_i),N-1}) \right\}$$

Set

$$X^v_{u,n} = \cup_{N \ge 0} X^v_{u,n,N}.$$

Lemma 1.3.2 *[**simpl7**] The subsets $X^v_{u,n,N}$ and $X^v_{u,n}$ are measurable.*

Proof: Straightforward. $\Box$

Note that $X^u_{u,n} = \Omega^\infty$.

Example 1.3.3 [**divergent**] Let $x_1, \dots, x_i, \dots$ be an increasing sequence of points of $[s,t]$ such that $\lim_n x_n = t$ and for any i one has $x_i < t$. Consider the kernel α which sends $s \le u < t$ to the measure $\delta_{x_i} \times \{2\}$ where i is the first index for which $x_i > u$. Then $X^v_{u,1} = \Omega^\infty$ for all u and $v < t$ and $X^t_{u,1} = \emptyset$ for all $u < t$.

Consider the maps $M^v_{u,n}$ from Ω^∞ to $H[u,v] \amalg pt$ defined by the following inductive construction. For $\underline{\omega} \in X^v_{u,n}$ set

$$M^v_{u,0}(\underline{\omega}) = *_0$$

$$M^v_{u,1}(\underline{\omega}) = \begin{cases} \Delta^{[1]}_{u,v} & \text{if } I(u,\omega_1) = (x,k) \text{ and } x > v \text{ or } I(u,\omega_1) = pt \\ [k] *_x M^v_{x,k}(\omega_2, \dots) & \text{if } I(u,\omega_1) = (x,k) \text{ and } x < v \\ \Delta^{[k]}_{u,v} & \text{if } I(u,\omega_1) = (v,k) \end{cases}$$

$$M^v_{u,n}(\underline{\omega}) = \textstyle\sum_{i=1}^n M^v_{1,u}(\omega_i, \omega_{i+n}, \omega_{i+2n}, \dots)$$

where $\sum$ refers to the disjoint union of histories. For $\underline{\omega} \in \Omega^{\infty} \backslash X^v_{u,n}$ set $M^v_{u,n}(\underline{\omega}) = pt$. For $u = v$ we set $M^u_{u,n} \equiv \{n\}$.

Set $\mu^v_{u,n} = (M^v_{u,n})_*(P^{\otimes\infty})^{|H[u,v]}$. Considering $\mu^v_{u,*}$ as (sub-probability) kernels from $\mathbf{N}$ to $H[u,v]$ we get a pre-process on $\mathcal{H}[s,t]$.

Let

$$\theta_{u,k} = \theta(u)^{|(s,t]\times\{k\}}$$

be the measure on $(s,t]$ which is the co-restriction of $\theta(u)$ to $(s,t] \times \{k\}$. The following three lemmas give an inductive description of measures $\mu^v_{u,n}$ directly in terms of $\theta_{u,k}$. Since $\mu^u_{u,n} = \delta_{\{n\}}$ we only consider the case $u < v$.

Lemma 1.3.4 *[q0] For any $s \leq u < v \leq t$ one has*

$$\mu^v_{u,0}(\Delta^{*0}_{u,v}) = 1$$

$$\mu^v_{u,1}(\Delta^{[n]}_{u,v}) = \begin{cases} 1 - \sum_{k\neq 1} \theta_{u,k}((u,v]) & \text{for } n = 1 \\ \\ \theta_{u,n}(\{v\}) & \text{for } n \neq 1 \end{cases}$$

Proof: We have

$$(M^v_{u,0})^{-1}(\Delta^{*0}_{u,v}) = X_{0,u} = \Omega^{\infty}$$

which proves the first equality. We have

$$(M^v_{u,1})^{-1}(\Delta^{[1]}_{u,v}) = \{\underline{\omega} \in \Omega^{\infty} |\, I_u(\omega_1) \in ((v,t] \times \mathbf{N}_{\neq 1}) \amalg pt\}\,.$$

Therefore

$$P^{\otimes\infty}((M^v_{u,1})^{-1}(\Delta^{[1]}_{u,v})) = P(I_u^{-1}((v,t] \times \mathbf{N}_{\neq 1}) \amalg pt) = 1 - \sum_{k\neq 1} \theta_{u,k}((u,v])$$

Finally

$$(M^v_{u,1})^{-1}(\Delta^{[n]}_u) = \{\underline{\omega} \in \Omega^{\infty} |\, I_u(\omega_1) = (v,n)\}$$

which proves the last equality. $\square$

Lemma 1.3.5 *[n1] For any $u < w_1 < w_2 < v$ and any measurable $U \subset H[w_2,v]_{n,*}$ one has*

$$\mu^v_{u,1}(k,(w_1,w_2),U) = \int_{x\in(w_1,w_2)} \mu^v_{x,k}(j^v_{x,w_2}(U))d\theta_{u,k}$$

Proof: It follows from the fact that

$$(M^v_{u,1})^{-1}(k,(w_1,w_2),U) =$$

$$= \left\{\underline{\omega} \in \Omega^{\infty} |\ I(\omega_1) \in (w_1,w_2) \times \{k\},\ (\omega_2,\dots) \in X^v_{I(\omega_1)},\ M^v_{I(\omega_1)}(\omega_2\dots) \in j^v_{I(\omega_1),w_2}(U)\right\}$$

where $I(-) = I_u(-)$. $\square$

Lemma 1.3.6 *[ng1] Let π be a combinatorial type with $n(\pi) > 1$. Then*

$$(\mu^v_{n,u})^{|\Delta^\pi_u} = ((\mu^v_{1,u})^{\otimes n})^{|add_n^{-1}(\Delta^\pi_u)}$$

where add_n is the addition map

$$H[u,t]_{1,*} \times \cdots \times H[u,t]_{1,*} \to H[u,t]_{n,*}.$$

Proof: Follows immediately from the definition of $M^v_{u,n}$ for $n > 1$. $\square$ As an immediate corollary from Lemma 1.3.6 we see that the pre-processes μ^*_* are additive. In view of Lemma 1.2.17, we conclude that Lemmas 1.3.4 and 1.3.5 completely determined μ in terms of θ.

Lemma 1.3.7 *[mu0] The pre-processes constructed above are right continuous.*

Proof: By Lemma 1.3.4 we have

$$h(u,v) = 1 - \sum_{k \neq 1} \theta_{u,k}((u,v])$$

which implies that $h(-,-)$ is right continuous and the claim of the lemma follows from Theorem 4.3.8. $\square$

Lemma 1.3.8 *[xinv] For any $w > v > u$ and any k one has $X^w_{u,k} \subset X^v_{u,k}$ and*

$$X^v_{u,k} = \cup_{w>v} X^w_{u,k}$$

Proof: Follows by easy induction from the construction of $X^v_{u,k,N}$. $\square$

Lemma 1.3.9 *[ta] For the process constructed from a kernel θ one has*

$$\alpha^v_{u,k} = \upsilon^v_{-,k} * \theta^{|(u,v]}_{u,k}$$

Proof: It is sufficient to compare the measures on the right and left hand sides on intervals $(u,w]$ where $w \leq v$. We have

$$(M^v_{u,1})^{-1}(x_1,k_1)^{-1}((u,w] \times \{k\}) = \left\{ \underline{\omega} \in \Omega^\infty \,|\, I(\omega_1) \in (u,w] \times \{k\} \; and \; (\omega_2, \dots) \in X^v_{I(\omega_1)} \right\}$$

(where $I(-) = I_u(-)$). Therefore

$$\alpha^v_{u,k}((u,w]) = \mu^v_{u,1}((x_1,k_1)^{-1}((u,w] \times \{k\})) = \int_{x \in (u,w]} P^{\otimes\infty}(X^v_{x,k}) d\theta_{u,k}$$

and since

$$\upsilon^v_{u,k} = P^{\otimes\infty}(X^v_{u,k})$$

the claim of the lemma follows. $\square$

Theorem 1.3.10 *[th3] Let θ be a sub-probability kernel as above such that the following conditions hold:*

1. *for all $s \leq u < v < w \leq t$ one has*
$$(1 - \sum_{k \neq 1} \theta_{u,k}((u, v]))(1 - \sum_{k \neq 1} \theta_{v,k}((v, w])) = (1 - \sum_{k \neq 1} \theta_{u,k}((u, w]))$$

2. *for all $s \leq u < v < t$ and $n \neq 1$ one has*
$$\theta_{u,n}^{|(v,t]} = (1 - \sum_{k \neq 1} \theta_{u,k}((u, v]))\, \theta_{v,n}^{|(v,t]}$$

The the corresponding pre-process μ is a Markov pre-process.

Proof:

Lemma 1.3.11 *[jmul] Suppose that θ satisfies conditions (1), (2). Then for any $s \leq u \leq v < w \leq t$ one has*
$$(\mu_{1,u}^{w})^{|j_{u,v}^{w}(H[v,w]_{1,*})} = (1 - \sum_{k \neq 1} \theta_{u,k}((u, v]))\, \mu_{1,v}^{w}$$

Proof: For $(\mu_{1,u}^{w})^{|j_{u,v}^{w}(sk_{>0}H[v,w]_{1,*})}$ it follows immediately from Lemmas 1.2.17 and 1.3.5 and condition (2). For $\mu_{1,u}^{w}(\Delta_{u,w}^{[n]})$ and $n \neq 1$ from Lemma 1.3.4 and condition (2) and finally for $\mu_{1,u}^{w}(\Delta_{u,w}^{[1]})$ from Lemma 1.3.4 and condition (1). $\square$

"If" We need to verify the condition of Lemma 1.2.1. Using additivity one can easily see that if this condition holds for all $U_1 \subset sk_q H[v, t]_{1,*}$ then it holds for all n and all $U_1 \subset sk_q H[v, t]_{n,*}$. Therefore we may proceed by induction on q and for each q we only need to consider the case $n = 1$.

Let $q = 0$. Then we have to consider two cases.

1. Let $U_1 = \Delta_{u,v}^{[1]}$. Then
$$(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = j_{u,v}^{w}(U_2)$$
and the condition of Lemma 1.2.1 follows immediately from Lemma 1.3.11 and Lemma 1.3.4 for $n = 1$.

2. Let $U_1 = \Delta_{u,v}^{[n]}$ where $n \neq 1$. Then
$$(res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = \left\{\Gamma \in sk_{>0}H[u, w]_{1,*} | x_1(\Gamma) = v, k_1(\Gamma) = n, R(\Gamma) \in j_{u,v}^{w}(U_2)\right\}$$
It follows by an obvious limit argument from Lemma (1.3.5), that the value of $\mu_{1,u}^{w}$ on this subset is $\theta_{u,n}(\{v\})\mu_{v,n}^{w}(U_2)$ which together with Lemma 1.3.4 implies the condition of Lemma 1.2.1 in this case.

Let $q > 0$. Assume by induction that the condition of Lemma 1.2.1 is known for all $U_1 \subset sk_{q-1}H[u,v]$ and all U_2 and let $U_1 \subset sk_q H[u,v]_{1,*}$. By Lemma 1.2.17 we may assume that $U_1 = (k,(w_1,w_2),U_1')$ where $u < w_1 < w_2 < v$ and U_1' is a measurable subset of $sk_{q-1}H[w_2,v]_{m,*}$ for some $m \neq 1$. Then

$$[\mathbf{ss}](res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2) = (k,(w_1,w_2),(res_{w_2,v} \times res_{v,w})^{-1}(U_1' \times U_2)) \qquad (29)$$

By the inductive assumption

$$\mu^w_{x,k}(j^w_{x,w_2}(res_{w_2,v} \times res_{v,w})^{-1}(U_1' \times U_2)) = \mu^w_{x,k}(j^w_{x,w_2}\, res^{-1}_{w_2,v}(U_1'))\mu^w_{v,l}(U_2)$$

where l is such that $U_2 \subset n_v^{-1}(\{l\})$. By Lemma 1.3.5, the value of $\mu^w_{u,1}$ on (29) is

$$\int_{x\in(w_1,w_2)} \mu^w_{x,k}(j^w_{x,w_2}(res_{w_2,v} \times res_{v,w})^{-1}(U_1' \times U_2))d\theta_{u,m} =$$

$$= \left(\int_{x\in(w_1,w_2)} \mu^w_{x,k}(j^w_{x,w_2}\, res^{-1}_{w_2,v}(U_1'))d\theta_{u,k}\right)\mu^w_{v,l}(U_2)$$

and using Lemma 1.3.5 again we get (6). Theorem is proved. $\square$

We are now going to formulate a sufficient condition for the pre-process constructed from θ to be a process.

Definition 1.3.12 *[***admiss***] The map I is called admissible if the corresponding pre-process is a process i.e. if $P^{\otimes\infty}(X^v_{u,n}) = 1$ for all u, v, n.*

One can verify easily that admissibility of I depends only on θ.

Consider the map $B^v_u : [u,v] \times \Omega \to (u,v]$ which equals $\{v\}$ on $\{v\} \times \Omega$ and whose restriction to $[u,v) \times \Omega$ is the composition of I with the projection which takes $((v,t] \times \mathbf{N}_{\neq 1}) \amalg pt$ to v and $(x,k) \in (u,v) \times \mathbf{N}_{\neq 1}$ to x. Since u and v are fixed below we will write B instead of B^v_u. Define a map

$$L : [u,v] \times \Omega^N \to (u,v]^N$$

setting

$$L(x_0,\omega_1,\omega_2,\ldots,\omega_N) = (B(x_0,\omega_1), B(B(x_0,\omega_1),\omega_2),\ldots).$$

We will write $x_i(x_0,\underline{\omega}) \in (s,t]$ for the i-component of $L(x_0,\underline{\omega})$. Let

$$Y^{<v}_{x_0,N} = \left\{\underline{\omega} \in \Omega^N \,|\, x_{N-1}(x_0,\underline{\omega}) < v\right\}$$

Proposition 1.3.13 *[***pr2***] Let $\theta_{*,k} = 0$ for $k > K$. Then*

$$P^{\otimes\infty}(\Omega^\infty \backslash X^v_{u,1,N}) \leq K^{N-1}\, P^{\otimes N}(Y^{<v}_{u,N})$$

Proof: For $n \leq K$ define subsets $Z^v_{u,n,N}$ setting $Z^v_{u,n,0} = \emptyset$ and for $N > 0$, $Z^v_{u,0,N} = \Omega^\infty$ and for $n > 0$:

$$Z^v_{u,n,N} = \left\{ \underline{\omega} \in \Omega^\infty | \; \forall 1 \leq i \leq n \; (I_u(\omega_i) \in (([v,t] \times \mathbf{N}_{\neq 1}) \amalg pt) \; or \; (\omega_{i+K}, \omega_{i+2K}, \dots) \in Z^v_{I(\omega_i),N-1}) \right\}$$

One observes easily that

$$P^\infty(Z^v_{u,n,N}) = P^\infty(X^v_{u,n,N})$$

We will write $xI_u(\omega)$ and $kI_u(\omega)$ for the first and the second component of $I_u(\omega)$ if it lies in $(s,t] \times \mathbf{N}_{\neq 1}$. If $I_u(\omega) = pt$ we will write $xI_u(\omega) = t$ and $kI_u(\omega) = 1$.

Lemma 1.3.14 *[th4l1] One has $\omega \in \Omega^\infty \backslash Z^v_{x_0,k_0,N}$ if and only if there exists a sequence $i_0, \dots, i_{N-1}$ such that for*

$$x_j = xI_{x_{j-1}}(\omega_{i_0 + i_1 K + \dots + i_{j-1} K^{j-1}})$$

$$k_j = kI_{x_{j-1}}(\omega_{i_0 + i_1 K + \dots + i_{j-1} K^{j-1}})$$

we have $1 \leq i_j \leq k_j$ and $x_j < v$ for all $0 \leq j \leq N-1$.

Proof: Follows easily by induction on N from the definition of subsets $Z^v_{x_0,k_0}$. □ Let

$$Y^{<v}_{x_0,k_0}(i_0, \dots, i_{N-1}) = \left\{ \underline{\omega} \,|\, (\omega_{i_0}, \omega_{i_0+i_1 K}, \dots) \in Y^{<v}_{x_0,N} \right\}$$

Lemma 1.3.15 *[th4l2] For any sequence $i_0, \dots, i_{N-1}$ one has*

$$P^{\otimes\infty}(Y^{<v}_{x_0,k_0}(i_0, \dots, i_{N-1})) = P^{\otimes N}(Y^{<v}_{x_0,N})$$

Proof: Consider the map $I : \Omega^\infty \to \Omega^N$ which sends $\underline{\omega}$ to $(\omega_{i_0}, \omega_{i_0+i_1 K}, \dots)$. Since it is just a partial projection the image of $P^{\otimes\infty}$ under this map coincides with $P^{\otimes N}$ which implies the claim of the lemma. □ Under the assumption of the proposition $k_j \leq K$ for all $j > 0$ and therefore

$$\Omega^\infty \backslash Z^v_{x_0,1,N} \subset \bigcup_{i_1, \dots, i_{N-1}} Y^v_{x_0,k_0}(1, i_1, \dots, i_{N-1})$$

where $1 \leq i_j \leq K$ for all j. Together with Lemma 1.3.15 this implies the claim of the proposition. □

Proposition 1.3.16 *[pr3] Let θ be a kernel satisfying condition (2) of Theorem 1.3.10 and let $\beta^v_u = (B^v_u)_*(\delta_u \otimes P)$. Then for u, v such that $h(u,v) \neq 0$ one has*

$$P^{\otimes N}(Y^{<v}_{u,N}) = \int_{u < x_1 < \dots < x_N < v} h(u,x_1)^{-1} \dots h(u,x_{N-1})^{-1} d(\beta^v_u)^{\otimes N}$$

Proof: We have

$$Y^{<v}_{u,N} = L_u^{-1}(u,v)^N$$

From the definition of L we conclude immediately that for any measurable $B_1, \ldots, B_N \in [s,t]$ one has

$$P^{\otimes N}(L_u^{-1}(B_1 \times \cdots \times B_N)) = \int_{x \in B_1} P^{\otimes(N-1)}(L_x^{-1}(B_2 \times \cdots \times B_N)) d\beta_u^v$$

If θ satisfies the condition (2) of Theorem 1.3.10 then $h(u,x)(\beta_x^v)^{|(x,v)} = (\beta_u^v)^{|(x,v)}$. Since β_x^v is concentrated on $(x,v]$ we get

$$P^{\otimes N}(L_u^{-1}((u,v)^N)) \int_{u<x_1<x_2<\cdots<x_N<v} h(u,x_1)^{-1} \ldots h(u,x_{N-1})^{-1} d(\beta_u^v)^{\otimes N}$$

□

Theorem 1.3.17 *[th4] Suppose that $\theta_{*,k} = 0$ for all but finitely many k and $lim_{v \to t} h(s,v) = c > 0$. Then θ is admissible.*

Proof: From Proposition 1.3.16 we get

$$P^{\otimes N}(Y_{u,N}^{<t}) \leq \frac{c^{1-N}}{N!} \beta_u(u,t)$$

In view of Proposition 1.3.13 it remains to show that

$$lim_{N \to \infty} \frac{(K/c)^{N-1}}{N!} = 0$$

which follows from the fact that for any finite number X one has

$$\lim_{N \to \infty} \frac{X^N}{N!} = 0.$$

□

Summarizing the results of the two previous sections we get the following theorem.

Theorem 1.3.18 *[th5] Our constructions provide a bijection between irreducible, additive Markov processes on $\mathcal{H}[s,t]$ whose branching multiplicities are bounded by $K \geq 0$ and collections of sub-probability measures $\alpha_{s,k}^t$ on $(s,t]$ given for $k \in \{0, 2, 3, \ldots, K\}$ and satisfying the condition*

$$\sum_k \alpha_{s,k}^t((s,t)) < 1$$

In the case of processes which are i-continuous for some subset of indexes $I \subset \{0, 2, \ldots, K\}$ we get the following classification.

Theorem 1.3.19 *[th6] Let $K \geq 0$ and I a subset in $\{0, 2, \ldots, K\}$. Consider the following collection of data:*

1. *for each $i \in I$, a non-atomic measure γ_i on $(s,t]$ which is bounded on $(s,u]$ for all $u < t$,*
2. *for each $j \in \{0, 2, \ldots, K\} \backslash I$ a sub-probability measure $\alpha^t_{s,j}$ on $(s,t]$ such that $\sum_j \alpha^t_{s,j}((s,t)) < 1$.*

Then there exists a unique irreducible additive Markov process μ on $\mathcal{H}[s,t]$ which is i-continuous for all $i \in I$ such that for $i \in I$ one has $\gamma_i = \gamma_i(\mu)$ and for $j \in \{0, 2, \ldots, K\} \backslash I$, one has $\alpha^t_{s,j} = \alpha^t_{s,j}(\mu)$.

Proof: The uniqueness result follows from Lemma 1.2.42. The existence follows from the same argument which is used in the proof of this lemma combined with Theorem 1.3.18. $\square$

Proposition 1.3.20 *[**pr8**] Let μ be an additive Markov process with bounded multiplicities. Then for all u, v and n one has*

$$\lim_{q \to \infty} \mu^v_{u,n}(sk_{>q} H[u,v]_{n,*}) \to 0$$

Proof: It is clearly sufficient to consider the case $u = s$ and $v = t$. Let K be a multiplicities bound. A simple inductive argument shows that

$$M^t_{s,n}(X^t_{s,n,N}) \subset sk_{K^N} H[s,t]_{n,*}$$

since. Therefore

$$\mu^t_{s,n}(sk_{>K^N} H[s,t]_{n,*}) \leq 1 - M^t_{s,n}(X^t_{s,n,N})$$

and since by Theorem 1.3.17 we have

$$lim_{N \to \infty} M^t_{s,n}(X^t_{s,n,N}) = 1$$

the claim of the proposition follows. $\square$

1.4 Birth and death processes

Definition 1.4.1 *[**bddef**] A birth and death (resp. birth, death) process on $\mathcal{H}[s,t]$ is an additive Markov process such that $\alpha^v_{u,k} = 0$ for all u, v and $k \neq 0, 2$ (resp. $k \neq 2$, $k \neq 0$).*

Definition 1.4.2 *[**nbddef**] A normalized birth and death process is a 2-continuous birth and death process such that $\gamma_2 = dx$ where dx is the Lebesgue measure on $(u,t]$.*

Let $f : [s', t'] \to [s,t]$ be a strictly increasing function. Then it defines maps $f_{u,v} : H[u,v] \to H[f(u), f(v)]$ according to an obvious rule. For a pre-process μ^*_* on $\mathcal{H}[s,t]$ define a pre-process $f^*(\mu)$ on $\mathcal{H}[s',t']$ setting $f^*(\mu)^v_{u,n} = f_*(\mu^{f(v)}_{f(u)})$. One verifies easily that this operation preserves all the properties of pre-processes considered above.

There are two cases which are of special interest. One is the case of and inclusion $i : [s', t'] \subset [s, t]$ when $i^*(\mu)$ is the restriction of μ to $[s', t']$. Another one is the case of an order preserving bijection $f : [s', t'] \to [s, t]$. In this case f^* is a bijection between pre-processes on $\mathcal{H}[s, t]$ and $\mathcal{H}[s', t']$. The α-invariants of pre-processes are transformed by this bijection by the rule

$$\alpha^v_{u,k}(f^*(\mu)) = f^*(\alpha^{f(v)}_{f(u),k})$$

Definition 1.4.3 *Two process μ and μ' are called weakly equivalent if there exists an order preserving bijection $f : [s, t] \to [s, t]$ such that $\mu' = f^*(\mu)$.*

Proposition 1.4.4 *[**pr6**] A birth and death process is weakly equivalent to a normalized birth and death process if and only if it is 2-continuous and $\gamma_2((u, v)) > 0$ for $v > u$. In this case the corresponding normalized process is unique.*

Proof: Straightforward. $\square$

Proposition 1.4.5 *[**pr11**] Let $\alpha^t_{s,0}$ be a sub-probability measure on $(s, t]$ such that $\alpha^t_{s,0}((s, t)) < 1$. Then there exists a unique irreducible normalized birth and death process on $\mathcal{H}[s, t]$ such that $\alpha^t_{s,0}(\mu) = \alpha^t_{s,0}$.*

Proof: It is a particular case of Theorem 1.3.19. $\square$

1.5 Compositions, re-gluings and related constructions

For finite sets X, B and a function $\psi : X \to \mathbf{N}$ such that

$$tr(\psi) = \sum_{x\in X} \psi(x) = \#B$$

set

$$C(B, X, \psi) = \{f : B \to X \mid \psi(f) = \psi\}$$

and let

$$c(B, X, \psi) = \#C(B, X, \psi)$$

The number $c(B, X, \psi)$ depends only on X and ψ but it will be convenient for us to keep B in the notation. We will also use the set of maps

$$C'(B, X, \psi) = \{f : B \to X \mid \psi(f) \cong \psi\}$$

where $\psi \cong \psi'$ of the exists a permutation $s : X \to X$ such that $\psi' = \psi \circ s$. Set further

$$G(\Gamma', \Gamma'') = C(\tau^{-1}_{\Gamma''}(v), \tau^{-1}_{\Gamma'}(v), \psi_{\Gamma'})$$

$$G'(\Gamma', \Gamma'') = C'(\tau^{-1}_{\Gamma''}(v), \tau^{-1}_{\Gamma'}(v), \psi_{\Gamma'})$$

Set

$$c(\Gamma', \Gamma'') = (\#G(\Gamma', \Gamma''))^{-1} \sum_{f \in G(\Gamma', \Gamma'')} \Gamma' \cup_f \Gamma''$$

$$c'(\Gamma', \Gamma'') = (\#G'(\Gamma', \Gamma''))^{-1} \sum_{f \in G'(\Gamma', \Gamma'')} \Gamma' \cup_f \Gamma''$$

where the right hand sides are considered as a measures on $H[u, w]$. This defines probability kernels

$$c : H[u, v] \times_{\mathbf{N}} H[v, w] \to H[u, w]$$

$$c' : H[u, v] \times_{\mathbf{N}} H[v, w] \to H[u, w]$$

Lemma 1.5.1 *[**cl1**] One has*

$$(res_{u,v} \times res_{v,w}) \circ c = Id$$

Proof: Let $\Gamma' = res_{u,v}(\Gamma)$ and $\Gamma'' = res_{v,w}(\Gamma)$. Then there is a well defined map

$$[\mathbf{ceq1}] g : \tau_{\Gamma''}^{-1}(v) \to \tau_{\Gamma'}^{-1}(v) \tag{30}$$

such that for any vertex $x \in \tau_{\Gamma'}^{-1}(v)$ one has

$$[\mathbf{ceq2}] \#(g^{-1}(x)) = \psi_{\Gamma'}(x) \tag{31}$$

Conversely, given $\Gamma \in H[u, v]$, $\Gamma' \in H[v, w]$ and a map g as above there exists a unique $\Gamma' \cup_g \Gamma'' \in H[u, w]$ corresponding to these data which implies the claim of the lemma. $\square$

Set

$$m_{v+}(\Gamma) = c \circ (res_{u,v} \times res_{v,w})$$

$$m_v(\Gamma) = c' \circ (res_{u,v} \times res_{v,w})$$

Lemma 1.5.2 *[**proj**] The kernels m_v and m_{v+} are projectors i.e.*

$$m_v m_v = m_v$$

$$m_{v+} m_{v+} = m_{v+}$$

Proof: For m_{v+} it follows immediately from Lemma 1.5.1. To prove that m_v observe first that for $\Gamma' \in H[u, v]_{n,*}$ and $\Gamma'' \in H[v, w]_{n,*}$ one has

$$(res_{u,v} \times res_{v,w})(c'(\Gamma', \Gamma'')) = m_v(\Gamma') \otimes \Gamma''$$

and

$$c'(\Gamma', \Gamma'') = c'(m_v(\Gamma'), \Gamma'')$$

Applying these equalities to $\Gamma' = res_{u,v}(\Gamma)$ and $\Gamma'' = res_{v,w}(\Gamma)$ we get

$$m_v m_v(\Gamma) = c' \circ (res \times res) \circ c' \circ (res \times res)(\Gamma) = c'(m_v(\Gamma') \otimes \Gamma'') = c'(\Gamma', \Gamma'') = m_v(\Gamma)$$

$\square$

Lemma 1.5.3 *[**mvpmv**] One has*

$$m_{v+}m_v = m_v$$

Proof: Straightforward. $\square$

Remark 1.5.4 Note the neither m_v nor m_{v+} is a homomorphism with respect to the disjoint union of histories.

Proposition 1.5.5 *[**pr12**] Let μ be an additive Markov process on $\mathcal{H}[s,t]$. Then its is invariant under mixings i.e. for any $u \leq y \leq v$ one has*

$$m_y \circ \mu_u^v = \mu_u^v$$

and

$$m_{y+} \circ \mu_u^v = \mu_u^v$$

Proof: The second statement follows from the first one by Lemma 1.5.3. To prove the first one let us generalize the construction of Lemma 1.2.17 as follows. Let $k \neq 1$, $u < w_1 < w_2 < v$ and let f be a measurable function on $H[w_2,v]_{k,*}$. Define a function $(k,(w_1,w_2),f)$ on $H[u,v]_{1,*}$ setting

$$(k,(w_1,w_2),f)(\Gamma) = \left\{ \begin{array}{ll} f(R(\Gamma)) & \text{if } (x_1,k_1)(\Gamma) \in (w_1,w_2) \times \{k\} \\ 0 & \text{otherwise} \end{array} \right.$$

such that if f is the indicator function for $U \subset H[w_2,v]_{k,*}$ then $(k,(w_1,w_2),f)$ is the indicator function for $(k,(w_1,w_2),U)$.

Lemma 1.5.6 *[**pr11l1**] For a Markov process μ one has*

$$\int_{H[u,v]_{1,*}} (k,(w_1,w_2),f)d\mu_{u,1}^v = \lambda_{u,k}^{w_2} \int_{H[w_2,v]_{k,*}} fd\mu_{w_2,k}^v$$

Proof: It follows immediately from the Markov property applied to the triple (u,w_2,v). $\square$ The equality $\mu_{u,k}^v \circ m_y = \mu_{u,k}^v$ holds on $sk_0H[u,v]$ for all k since for $\Gamma \in sk_0H[u,v]$ one has $m_y(\Gamma) = \Gamma$. Assume that it holds on $sk_{q-1}H[u,v]$ for all k. Let us show that the equality $\mu_{u,1}^v \circ m_y = \mu_{u,1}^v$ holds on $sk_qH[u,v]$. In view of Lemma 1.2.17 it is sufficient to show that

$$(\mu_{u,1}^v \circ m_y)(k,(w_1,w_2),U) = \mu_{u,1}^v(k,(w_1,w_2),U)$$

for all $k \neq 1$, $u < w_1 < w_2 < v$ and $U \in sk_{q-1}H[w_2,v]_{k,*}$. We have

$$(\mu_{u,1}^v \circ m_y)(k,(w_1,w_2),U) = \int_{H[u,v]} m_y^*(k,(w_1,w_2),I_U)d\mu_{u,1}^v$$

where I_U is the indicator function of U and $m_y^*(f)$ is the pull-back of this function with respect to the kernel m_y.

Lemma 1.5.7 *[pr11l2] One has*

$$m_y^*(k,(w_1,w_2),f) = \left\{ \begin{array}{ll} (k,(w_1,w_2),m_y^*(f)) & \text{for } y \geq w_2 \\ (k,(w_1,w_2),f) & \text{for } y \leq w_2 \end{array} \right.$$

Proof: Straightforward from the fact that

$$(k,(w_1,w_2),f) = (res_{u,w_2} \times res_{w_2,v})^*(I_{(w_1,w_2)} \times f)$$

where $I_{(w_1,w_2)}$ is the indicator function of (w_1,w_2) considered as a subset in $\Delta_{u,w_2}^{[k[1]]}$. $\square$ Applying Lemmas 1.5.7 and 1.5.6 we get

$$\int_{H[u,v]} m_y^*(k,(w_1,w_2),I_U)d\mu_{u,1}^v = \int_{H[u,v]} m_y^*(k,(w_1,w_2),m_y^*(I_U))d\mu_{u,1}^v =$$

$$= \lambda_{u,k}^{w_2} \int_{H[w_2,v]_{k,*}} m_y^*(I_U)d\mu_{w_2,k}^v = \lambda_{u,k}^{w_2} \int_{H[w_2,v]_{k,*}} I_U d\mu_{w_2,k}^v =$$

$$= \int_{H[u,v]} (k,(w_1,w_2),I_U)d\mu_{u,1}^v$$

where the third equality holds by the inductive assumption since I_U is supported on $sk_{q-1}H[u,v]$. To finish the inductive step it remains to verify that the equality $\mu_{u,k}^v \circ m_y = \mu_{u,k}^v$ holds on $sk_q H[u,v]_{k,*}$ for $k \neq 1$. This follows from the additivity of μ and the fact that $add_k^{-1}(sk_q H[u,v]_{k,*}) \subset (sk_q H[u,v]_{1,*})^k$. $\square$

Set

$$F(X,\psi',k) = \{\psi : X \to \mathbf{N} \,|\, \psi \leq \psi' \textit{ and } tr(\psi) = k\}$$

Let B be a finite set, $A \subset B$ its subset with k elements and $\Gamma \in H[u,v]$ a history such that $n_v(\Gamma) = tr(\psi_\gamma) = \#B$. Denote $\psi_\Gamma^{-1}(v)$ by X. Set

$$s_k(A,B,\Gamma) = \sum_{\psi \in F(X,\psi_\Gamma,k)} \frac{c(B\backslash A, X, \psi_\Gamma - \psi)c(A,X,\psi_\Gamma)}{c(B,X,\psi(\Gamma'))}\Gamma_\psi$$

which we interpret as a measure (a sum of δ-measures) on $H[u,v]$. It depends only on the isomorphism class of the pair of sets $A \subset B$ and since the number of elements of B is known only on the number of elements k of A. When possible we will denote it by $s_k(\Gamma)$.

Note that if $k = n_v(\Gamma)$ then $s_k(\Gamma) = \Gamma$ and if $k > n_v(\Gamma)$ then $s_k(\Gamma) = 0$.

The measure $s_k(\Gamma)$ is always a probability measure. Indeed consider a map

$$C(B,X,\psi_\Gamma) \to F(X,\psi',k)$$

which sends f to $\psi(f_{|A})$. One verifies easily that this map is well defined and its fiber over ψ is the product of $C(A,X,\psi)$ and $C(B\backslash A, X, \psi_\Gamma - \psi)$ which implies that

$$\sum_{\psi \in D(X,\psi_\Gamma,k)} c(B\backslash A, X, \psi_\Gamma - \psi)c(A,X,\psi_\Gamma) = c(B,X,\psi(\Gamma')).$$

1.6 Death free histories

A history $\Gamma \in H[u,v]$ is called *death free* if $\psi^{-1}(0) = \emptyset$ and for any vertex x such that $\tau(x) < v$ there exists at least one edge starting in x. We let $\tilde{H}[u,v]$ denote the set of death free histories over $[u,v]$ and $\widetilde{HD}[u,v]$ the set of ordered death free histories over $[u,v]$. The space $\widetilde{HD}[u,v]$ (resp. $\tilde{H}[u,v]$) can also be described as the $[u,v]$-geometric realization of the simplicial monoid defined as follows. Let $\tilde{F}$ (resp. $\tilde{F}^{Ab}$) be the co-triple on monoids (resp. commutative monoids) which takes a monoid A to the free monoid generated by the pointed set $(A,1)$, e.g. $\tilde{F}(pt) = pt$. Let further $\tilde{F}_*$ (resp. $\tilde{F}_*^{Ab}$) be the functor which sends a monoid to the simplicial monoid defined by this co-triple. There are obvious natural transformations of co-triples $i : \tilde{F} \to F$, $r : F \to \tilde{F}$ (resp. $i : \tilde{F}^{Ab} \to F^{Ab}$ and $r : F^{Ab} \to \tilde{F}^{Ab}$) where F and F^{Ab} are the co-triples considered in Section **??**.

Proposition 1.6.1 *[***pr9***] The space $\widetilde{HD}[u,v]$ (resp. $\tilde{H}[u,v]$) is naturally identified with $|\tilde{F}_*(\mathbf{N})|_{[u,v]}$ (resp. $|\tilde{F}_*^{Ab}(\mathbf{N})|_{[u,v]}$). The geometric realization of i gives the natural embedding of it to $HD[u,v]$ (resp. to $H[u,v]$). The geometric realization of r gives maps*

$$H[u,v] \to \tilde{H}[u,v]$$

$$HD[u,v] \to \widetilde{HD}[u,v]$$

which which we denote by r_v. These maps send a history to the ancestral history of the present day survivors. More precisely, $r_v(\Gamma)$ is the death free history obtained from Γ by the successive removal of all vertices x such that $\tau(x) = v$ and $\psi(x) = 0$ or $\tau(x) < v$ and there are no edges starting at x.

Proof: Straightforward. $\square$

The simplicial sets $\tilde{F}_*(\mathbf{N})$ and $\tilde{F}_*^{Ab}(\mathbf{N})$ are locally finite. Moreover, one has the following result.

Proposition 1.6.2 *[***topstr***] The space $\widetilde{HD}[u,v]$ (resp. $\tilde{H}[u,v]$) is the disjoint union of the form*

$$\widetilde{HD}[u,v] = \amalg_{n\geq 0}\widetilde{HD}[u,v]_{*,n}$$

$$\tilde{H}[u,v] = \amalg_{n\geq 0}\tilde{H}[u,v]_{*,n}$$

The space $\tilde{H}_{*,0}$ consists of one point corresponding to the empty history. For $n > 0$, the spaces $\widetilde{HD}[u,v]_{*,n}$ and $\tilde{H}[s,t]_{*,n}$ are finite contractible CW-complex of dimension $n-1$.

Proof: Straightforward. $\square$

Remark 1.6.3 [**cubes**] There are obvious descriptions of $\widetilde{HD}[u,v]_{*,n}$ and $\tilde{H}[u,v]_{*,n}$ along the lines of the descriptions of $HD[u,v]_{*,n}$ and $H[u,v]_{*,n}$ given in Remark **??**. In the ordered case and for $n > 0$ we get that $HD[u,v]_{*,n}$ is the $[u,v]$-nerve of the category whose objects are order preserving surjections $[n-1] \to [i]$ and morphisms are morphisms under $[n-1]$. One observes easily that this category is a partially-ordered set which is isomorphic to the set of subsets of $\{1,\ldots,n-1\}$. Correspondingly

$$\widetilde{HD}[u,v]_{*,n} = [u,v]^{n-1}$$

The geometry of $\tilde{H}[u,v]_{*,n}$ appears to be less regular. Explicit computation shows that

$$\tilde{H}[u,v]_{*,0} = \tilde{H}[u,v]_{*,1} = pt$$

$$\tilde{H}[u,v]_{*,2} = \Delta^1_{[u,v]}$$

$$\tilde{H}[u,v]_{*,3} = \Delta^2_{[u,v]}$$

and $\tilde{H}[u,v]_{*,4}$ is the union of two copies of $\Delta^3_{[u,v]}$ along a 2-dimensional face.

Remark 1.6.4 [**ultrametric**] For an element $\Gamma \in \tilde{H}[u,v]$ of level q define a distance on the set $V_{q+1}(\Gamma)$ of final vertices of Γ setting

$$d(v_1, v_2) = v - x_{i+1}(\Gamma)$$

where i is the largest index such that the images of v_1 and v_2 in V_i coincide if such an index exists and

$$d(v_1, v_2) = \infty$$

otherwise. One verifies that $V_{q+1}(\Gamma)$ with this distance function is an ultra-metric space (where 0 and ∞ are allowed as distances). Moreover, there is a natural bijection between $\tilde{H}[u,v]_{*,n}$ and the set of isomorphism classes of ultra-metric spaces with n elements such that for any v_1, v_2 one has $d(v_1, v_2) < v - u$ or $d(v_1, v_2) = \infty$.

Remark 1.6.5 [**genhist**] Previous remarks show that every point of $\tilde{H}[u,v]_{*,n}$ (resp. $\widetilde{HD}[u,v]_{*,n}$) lies in the closure of a simplex of dimension $n-1$ and it belongs to the interior of such a simplex if and only if the corresponding history Γ has the following properties:

1. there are exactly n vertices v with $\tau(v) = t$ (i.e. $\psi \equiv 1$),
2. for any v such that $\tau(v) \neq s, t$ there exists exactly two edges starting in v,
3. for any v_1, v_2 such that $\tau(v_1) = \tau(v_2) \neq t$ one has $v_1 = v_2$, in particular there is exactly one vertex v with $\tau(v) = s$.

We will call histories which satisfy these conditions *generic* and denote their space by $\tilde{B}[u,v]$ (resp. $\widetilde{BD}[u,v]$). As a space $\tilde{B}[u,v]$ (resp. $\widetilde{BD}[u,v]$) is naturally homeomorphic to the disjoint union of open $[u,v]$-simplexes and tis dense in $\tilde{H}[u,v]$. (resp. $\widetilde{HD}[u,v]$).

In what follows we will write r_v instead of r to emphasize the special role which the final time moment plays in the definition of r. We will also restrict our attention to the case of ordered histories giving the corresponding results for un-ordered histories as remarks.

We obviously have

$$(r_v)^{-1}(\widetilde{HD}[u,v]_{*,n}) = HD[u,v]_{*,n}$$

We set

$$HD[u,v]^j_{m,n} = (r_v)^{-1}(\widetilde{HD}[u,v]_{j,n}) \cap HD[u,v]_{m,n}$$

i.e. $HD[u,v]^j_{m,n}$ is the set of histories with m initial members j of which have living descendants at time v and n final members.

For a process μ on $\mathcal{HD}[s,t]$, $m, j \geq 0$ and $s \leq u \leq v \leq t$ set

$$\eta^v_{u,\mu}(m,j) = \mu^v_{u,m}(HD[u,v]^j_{m,*})$$

As always we will omit μ from our notation whenever possible. We have $\eta^v_u(m,0) = \phi^v_u(m,0)$ and in general $\eta^v_u(m,j)$ is the probability that in a population with m members at time u exactly j of them will have living descendants at time v.

Let further $\eta^v_u : \mathbf{N} \to \mathbf{N}$ be the kernels defined by the rule

$$\{m\} \mapsto \sum_{j\geq 0} \eta_{u,\mu}(m,j)\delta_{\{j\}}$$

If μ is an additive process then one has

$$[\mathbf{pr10eq1}]\eta^v_u(m,j) = C(m,j)\eta^v_u(1,1)^j \eta^v_u(1,0)^{m-j} \tag{32}$$

Proposition 1.6.6 *[**pr10**] Let $\mu : \mathbf{N} \to HD[u,v]$ be an additive kernel such that $(n_u)_*\mu = Id$ and $\mu(1, HD[u,v]^1_{1,*}) \neq 0$. Then there exists a unique kernel $\bar{\mu} : \mathbf{N} \to \widetilde{HD}[u,v]$ such that the square*

$$[\mathbf{pr10eq1}] \quad \begin{array}{ccc} \mathbf{N} & \xrightarrow{\eta} & \mathbf{N} \\ {\scriptstyle\mu}\downarrow & & \downarrow{\scriptstyle\bar{\mu}} \\ HD[u,v] & \xrightarrow{r_v} & \widetilde{HD}[u,v] \end{array} \tag{33}$$

where

$$\eta(m) = \sum_j \mu(HD[u,v]^j_{m,*})\delta_j$$

commutes and this kernel is additive.

Proof: Let r^j_m be the restriction of r_v to a $HD[u,v]^j_{m,*}$

$$r^j_m : HD[u,v]^j_{m,*} \to \widetilde{HD}[u,v]_{j,*}$$

and μ^j_m be the co-restriction of $\mu(m)$ to $HD[u,v]^j_{m,*}$.

Lemma 1.6.7 *[**pr10l1**] One has*

$$(r^j_m)_*(\mu^j_m) = C(m,j)\eta(1,0)^{m-j}(add_j)_*(((r^1_1)_*(\mu^1_1)^{\otimes j}))$$

where

$$add_{1,j} : (\widetilde{HD}[u,v]_{1,*})^{\times j} \to \widetilde{HD}[u,v]_{j,*}$$

is the restriction of the addition map.

Proof: Since r_v is a homomorphism of monoids we have a commutative diagram

$$\begin{array}{ccc} HD[u,v]^{\times m} & \xrightarrow{add_m} & HD[u,v] \\ {\scriptstyle r_v^{\times m}}\downarrow & & \downarrow{\scriptstyle r_v} \\ \widetilde{HD}[u,v]^{\times m} & \xrightarrow{add_m} & \widetilde{HD}[u,v] \end{array}$$

We have

$$X = add_m^{-1} r_v^{-1}(\widetilde{HD}[u,v]_{j,*}) = add_m^{-1}(HD[u,v]^j_{m,*}) =$$

$$= \amalg_{I\subset\{1,\ldots,m\}} \prod_{i=1}^{m} HD[u,v]^{\epsilon(i)}_{1,*}$$

where I runs through the j-element subsets of $\{1,\ldots,m\}$ and $\epsilon(i)=1$ for $i\in I$ and $\epsilon(i)=0$ otherwise. Since $\mu(m)=(add_m)_*(\mu(1))$ we have

$$(r^j_m)_*(\mu^j_m) = (add_m)_*(r^m_v)_*((\mu(1)^{\otimes m})^{|X})$$

Since X is the disjoint union of $C(m,k)$ components such that each one is obtained by permutation of factors from $(HD[u,v]^1_{1,*})^{\times j}\times(HD[u,v]^0_{1,*})^{\times(m-j)}$ and $\widetilde{HD}[u,v]_{0,*}=pt$ we get the required equality. $\Box$

To finish the proof of the proposition observe first that under our assumptions

$$\eta(j,j)=\eta(1,1)^j>0$$

Commutativity of (33) is equivalent to the assertion that for all $m,k\geq 0$ one has

$$\eta(m,j)\bar{\mu}(j) = (r^j_m)_*(\mu^j_m)$$

Set $\bar{\mu}(j)=\eta(j,j)^{-1}(r^j_j)_*(\mu^j_j)$. From the Lemma 1.6.7 we get

$$(r^j_j)_*(\mu^j_j) = (add_j)_*(((r^1_1)_*(\mu^1_1)^{\otimes j}))$$

and therefore

$$(r_{m,j})_*(\mu^j_m) = C(m,j)\eta(1,0)^{m-j}(r^j_j)_*(\mu^j_j)$$

Then for $m>j$ we have

$$\eta(m,j)\bar{\mu}(j) = \eta(m,j)\eta(j,j)^{-1}(r^j_j)_*(\mu^j_j) = C(m,j)\eta(1,0)^{m-j}(r^j_j)_*(\mu^j_j) = (r^j_m)_*(\mu^j_m)$$

The additivity of $\bar{\mu}$ follows easily by a similar argument. $\Box$

Let now μ be an additive Markov process on $\mathcal{HD}[s,t]$. Define a death-free process $\tilde{\mu}$ by the formula

$$\tilde{\mu}^v_u = (res_{u,v})_*(\bar{\mu}^t_u)$$

The process $\tilde{\mu}$ is called the ancestral process of the process μ. From the proof of Proposition 1.6.6 we have the following explicit formula for $\tilde{\mu}^t_{u,j}$:

$$\tilde{\mu}^t_{u,j} = \eta^t_u(1,1)^{-j}(r^j_{u,j})_*((\mu^v_{u,j})^{|HD[u,t]^j_{j,*}})$$

where $r^j_{u,j}$ is the restriction of r_t to $HD[u,t]^j_{j,*}$.

Theorem 1.6.8 *[th7] If μ is an additive Markov process then $\tilde{\mu}$ is an additive Markov process.*

Proof: We already know that $\tilde{\mu}$ is additive. It remains to verify that it satisfies the conditions of Lemma 1.2.1. Since the process is additive it is enough to show that for any $\tilde{n} \geq 0$, any $s \leq u < v << \leq t$, any measurable U in $\widetilde{HD}[u,v]_{1,\tilde{n}}$ and any measurable V in $\widetilde{HD}[v,t]_{\tilde{n},*}$ one has

$$[\mathbf{th7eq1}]\tilde{\mu}^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(U \times V)) = \tilde{\mu}^v_{u,1}(U) \times \tilde{\mu}^t_{v,\tilde{n}}(V) \tag{34}$$

In view of Lemma 1.1.8 we may assume that

$$V = add_{\tilde{n}}(V_1 \times \cdots \times V_{\tilde{n}})$$

where V_i's are measurable subsets in $\widetilde{HD}[v,t]_{1,*}$.

For a subset $I \subset \{1, \ldots, n\}$ and a history $\Gamma \in HD[u,v]_{n,*}$ denote by Γ_I the history given by the same data as Γ except that $V_{q+1}(\Gamma_I) = V_{q+1}(\Gamma)\backslash I$. Let further

$$r^I_v : HD[u,v] \to \widetilde{HD}[u,v]$$

the map $\Gamma \mapsto r_v(\Gamma_I)$. Intuitively it corresponds to taking the ancestral history of Γ under the assumption that the present day members from I are considered dead.

By Lemma 1.6.9 below we have

$$\tilde{\mu}^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(U \times V)) = \eta^t_u(1,1)^{-1}(\sum_{n \geq \tilde{n}} \sum_I \eta^t_v(1,1)^{\tilde{n}} \eta^t_v(1,0)^{n-\tilde{n}} \mu^v_{u,1}((r^I_v)^{-1}(U)))\, \tilde{\mu}^t_{v,\tilde{n}}(V)$$

Setting $V = \widetilde{HD}[v,t]_{\tilde{n},*}$ we get

$$\tilde{\mu}^v_{u,1}(U) = \tilde{\mu}^t_{u,1}(res^{-1}_{u,v}(U)) = \tilde{\mu}^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(U \times \widetilde{HD}[v,t]_{\tilde{n},*})) =$$

$$= \eta^t_u(1,1)^{-1}(\sum_{n \geq \tilde{n}} \sum_I \eta^t_v(1,1)^{\tilde{n}} \eta^t_v(1,0)^{n-\tilde{n}} \mu^v_{u,1}((r^I_v)^{-1}(U)))$$

and therefore (34) holds. $\square$

Lemma 1.6.9 *[th7l1] Under the assumptions made above one has:*

$$\eta^t_u(1,1)\tilde{\mu}^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(U \times V)) = (\sum_{n \geq \tilde{n}} \sum_I \eta^t_v(1,1)^{\tilde{n}} \eta^t_v(1,0)^{n-\tilde{n}} \mu^v_{u,1}((r^I_v)^{-1}(U)))\, \tilde{\mu}^t_{v,\tilde{n}}(V)$$

where I runs through subsets of $\tilde{n}$-elements in $\{1, \ldots, n\}$.

Proof: Let r^1_1 be the restriction of r_t to $HD[v,t]^1_{1,*}$. We will use the same notation for the restriction of r_t to $HD[u,t]^1_{1,*}$. By Lemma 1.1.7 the map

$$res_{u,v} \times res_{v,t} : HD[u,t]_{1,*} \to \amalg_n\ HD[u,v]_{1,n} \times HD[v,t]_{n,*}$$

is a bijection and since a history on $[u,t]$ is death free if and only if its restrictions to $[u,v]$ and $[v,t]$ are death free so is the map

$$res_{u,v} \times res_{v,t} : \widetilde{HD}_{1,*}[u,t] \to \amalg_{\tilde{n}} \ \widetilde{HD}[u,v]_{1,\tilde{n}} \times \widetilde{HD}[v,t]_{\tilde{n},*}$$

Let X be the image of $HD[u,t]^1_{1,*}$ under the first map. Then there is a unique map

$$f : X \to \widetilde{HD}[u,v]_{1,\tilde{n}} \times \widetilde{HD}[v,t]_{\tilde{n},*}$$

such that the square

$$\begin{array}{ccc} HD[u,t]^1_{1,*} & \xrightarrow{res_{u,v} \times res_{v,t}} & X \\ {\scriptstyle r^1_1}\downarrow & & \downarrow{\scriptstyle f} \\ \widetilde{HD}[u,t]_{1,*} & \xrightarrow{res_{u,v} \times res_{v,t}} & \amalg_{\tilde{n}} \ \widetilde{HD}[u,v]_{1,\tilde{n}} \times \widetilde{HD}[v,t]_{\tilde{n},*} \end{array}$$

commutes.

For a subset $I \subset \{1,\ldots,n\}$ let $HD[v,t]^I_{n,*}$ be the subset of $HD[v,t]^{\#I}_{n,*}$ which consists of histories Γ such that $p \in V_0(\Gamma) = \{1,\ldots,n\}$ has living descendants at time t if and only if it lies in I.

We have

$$f^{-1}(U \times V) = \amalg_{n \geq \tilde{n}, I} Y_{n,I}$$

where

$$Y_{n,I} = f^{-1}(U \times V) \cap HD[v,t]^I_{n,*}$$

We further have

$$f^{-1}(U \times V) \cap HD[v,t]^I_{n,*} = (r^I_v)^{-1}(U) \times add_n(Z_{I,1} \times Z_{I,n})$$

where

$$Z_{I,j} = \begin{cases} (r^1_1)^{-1}(V_{i(j,I)}) & \text{for } j \in I \\ HD[u,t]_{1,0} & \text{for } j \in \{1,\ldots,n\} \backslash I \end{cases}$$

and $i(j,I)$ is the sequential number of j as an element of I. Therefore

$$\mu^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(f^{-1}(U \times V))) = \sum_{n \geq \tilde{n}} \sum_{I, \#I = \tilde{n}} \mu^v_{u,1}((r^I_v)^{-1}(U)) \, (\mu^t_{v,1})^{\otimes n}(Z_{I,1} \times \cdots \times Z_{I,n}) =$$

$$= \sum_{n \geq \tilde{n}} \sum_{I, \#I = \tilde{n}} \mu^v_{u,1}((r^I_v)^{-1}(U)) \, \eta^t_v(1,0)^{n-\tilde{n}} \prod_{i=1}^{\tilde{n}} \mu^t_{v,1}((r^1_1)^{-1}(V_i))$$

On the other hand

$$\tilde{\mu}^t_{v,\tilde{n}}(V) = \prod_{i=1}^{\tilde{n}} (\tilde{\mu}^t_{v,1}(V_i)) = \eta^t_v(1,1)^{-\tilde{n}} \prod_{i=1}^{\tilde{n}} \mu^t_{v,1}((r^1_1)^{-1}(V_i))$$

and

$$\mu^t_{u,1}((res_{u,v} \times res_{v,t})^{-1}(f^{-1}(U \times V))) =$$

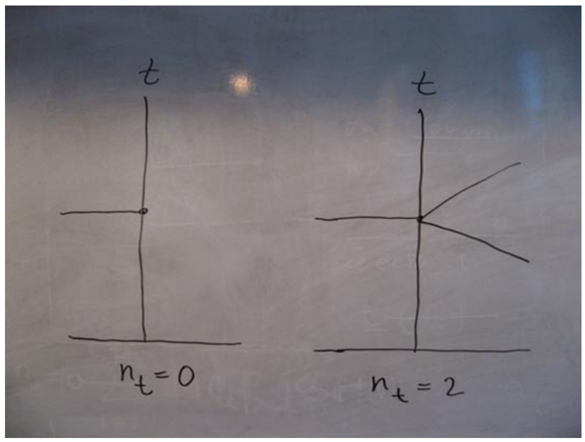
$$= \mu^t_{u,1}((r^1_1)^{-1}(res_{u,v} \times res_{v,t})^{-1}(U \times V)) = \eta^t_u(1,1) \tilde{\mu}^t_{1,u}((res_{u,v} \times res_{v,t})^{-1}(U \times V))$$

which finishes the proof. $\square$

1.7 Older stuff

Let $u_1 \leq \cdots \leq u_q$ be a monotone increasing sequence in $[s,t]$ and let Γ be a singleton history. Define $n_{u_1,\ldots,u_q}(\Gamma) \in S^{\circ(q-1)}(\mathbf{N})$ inductively as follows:

1. if $q=1$ we set $n_{u_1}(\Gamma)$ to be the number of population members at time u_1 which is defined as the number of initial vertices of $R_{u_1}(\Gamma)$ or equivalently as the number of final vertices of $L_{u_1}(\Gamma)$ counted with their multiplicities as illustrated by the picture:



2. If $q>1$ consider $R_{u_1}(\Gamma)$. If $R_{u_1}(\Gamma)=\emptyset$ we set $n_{u_1,\ldots,u_q}(\Gamma) = *_{q-2}$. Otherwise let $R_{u_1}(\Gamma) = \amalg\Gamma_i$ be the decomposition of $R_{u_1}(\Gamma)$ into the union of connected components. Then
$$n_{u_1,\ldots,u_q}(\Gamma) = \sum_i [n_{u_2,\ldots,u_q}(\Gamma_i)].$$

Proposition 1.7.1 *[borel1] The smallest σ-algebra on $H[s,t]$ which makes all the functions $n_{u_1,\ldots,u_q}$ for all $q \geq 1$ measurable coincides with the Borel σ-algebra $\mathfrak{B}$.*

Proof: For $(u_1,\ldots,u_l) \in \Delta^l$ and $\epsilon > 0$ let $U(u_1,\ldots,u_l;\epsilon)$ be the subset of $(x_1,\ldots,x_l) \in \Delta^l$ such that $|u_i - x_i| < \epsilon$. One verifies easily that subsets of the form $U = U(u_1,\ldots,u_l;\epsilon) \times \{\gamma\}$ generate $\mathfrak{B}$. It remains to show that such a subset can be defined in terms of the functions $n_{u_1,\ldots,u_q}$.

Observe that for any $\gamma \in S^{\circ(l+1)}(\mathbf{N})$ and any $0 \leq k_1 \leq \cdots \leq k_{l+1} \leq q$ there is an element $\delta_{k_1,\ldots,k_{l+1}}(\gamma) \in S^{\circ(q-1)}(\mathbf{N})$ such that
$$n_{v_1,\ldots,v_q}(u_1,\ldots,u_l;\gamma) = \delta_{k_1,\ldots,k_{l+1}}(\gamma)$$
where k_i is the number of v_i's in $[s,u_i)$ for $i \leq l$ and k_{l+1} is the number of v_i's in $[s,t)$. In particular it shows that the intersection of $n^{-1}_{v_1,\ldots,v_q}(\delta)$ with $\Delta^l \times \{\gamma\}$ is given by equations of the form $v_i < u_j$ and therefore it is Borel measurable.

Conversely, fix $\gamma \in S^{\circ(l+1)}(\mathbf{N})$ and consider the set of Γ such that for any $v_1,\ldots,v_q$ there exists $k_1 \leq \cdots \leq k_{l+1} \leq q$ such that
$$n_{v_1,\ldots,v_q}(\Gamma) = \delta_{k_1,\ldots,k_{l+1}}(\gamma).$$

Then this set coincides with $\Delta^l \times \{\gamma\} \subset H[s,t]$. Replacing all $v_1, \ldots, v_q$ in this condition by all rational ones (or all from any dense countable subset) we do not change the set. This shows that subsets of the form $\Delta^l \times \{\gamma\}$ are measurable with respect to the σ-algebra generated by functions $n_{v_1,\ldots,v_q}$.

For $(u_1, \ldots, u_l) \in \Delta^l$ and $\epsilon > 0$ let $U(u_1, \ldots, u_l; \epsilon)$ be the subset of $(x_1, \ldots, x_l) \in \Delta^l$ such that $|u_i - x_i| < \epsilon$. One verifies easily that subsets of the form $U = U(u_1, \ldots, u_l; \epsilon) \times \{\gamma\}$ generate $\mathfrak{B}$. It remains to show that such a subset can be defined in terms of the functions $n_{u-1,\ldots,u_q}$. According to the previous remark the subset $\Delta^l \times \{\gamma\}$ itself is measurable. It remains to show that $U(u_1, \ldots, u_l; \epsilon) \times \{\gamma\}$ can be defined as an intersection of $\Delta^l \times \{\gamma\}$ with a measurable subset. Such a measurable subset is easy to produce using countable combinations of functions n_{v_1,v_2} for pairs $s < v_1 \leq v_2 \leq t$. $\square$ Let $\mathfrak{S}_u^v$ be the σ-algebra on $H[s,t]$ generated by the functions $n_{w_1,\ldots,w_q}$ with $w_i \in (u,v]$. We have the following obvious result.

Lemma 1.7.2 *[**ispaths**] The collection of data $(\mathbf{N}, H[s,t], n_u, \mathfrak{S}_u^v)$ forms a path system.*

The space $H[s,t]$ has a structure of a commutative topological monoid given by the obvious map $a : H \times H \to H$ corresponding to the disjoint union of histories. One verifies easily that these maps are measurable with respect to all of the σ-algebras $\mathfrak{S}_u^v$ and that the functions n_u are homomorphisms from $H[s,t]$ to $\mathbf{N}$.

Let us say that a Markov process $P_u : \mathbf{N} \to H[s,t]$ on $H[s,t]$ is additive if the kernels P_u are homomorphisms of monoids i.e. if for $i, j \in \mathbf{N}$ one has

$$[\mathbf{eq1}] a_*(P_u(k,-) \otimes P_u(l,-)) = P_u(k+l,-) \tag{35}$$

where $P_u(n,-)$ is the measure on $\mathfrak{S}_u^t$ defined by the point n of $\mathbf{N}$.

Proposition 1.7.3 *[**ptop**] For any branching Markov process $(P_{u,v} : \mathbf{N} \to \mathbf{N})_{s \leq u \leq v \leq t}$ on $\mathbf{N}$ over $[s,t]$ there exist a unique additive Markov process P_u on $H[s,t]$ with transition kernels $P_{u,v}$.*

Proof: ??? $\square$

For a given Γ the function $u \mapsto n_u(\Gamma)$ from $[s,t]$ to $\mathbf{N}$ is continuous from the above i.e. it satisfies the condition

$$[\mathbf{ca}] \lim_{\epsilon \geq 0, \epsilon \to 0} n_{u+\epsilon}(\Gamma) = n_u(\Gamma) \tag{36}$$

Remark 1.7.4 For a given u function $\Gamma \mapsto n_u(\Gamma)$ from H to $\mathbf{N}$ need not be continuous.

Let $[u,v] \subset [s,t]$. One can easily see that there is only one reasonable way define a restriction map

$$c_{u,v} : H[s,t] \to H[u,v]$$

such that for any Γ and any $w \in [u,v]$ one has $n_w(\Gamma) = n_w(c_{u,v}(\Gamma))$.

Lemma 1.7.5 *[**mes1**] The functions n_u and the maps $c_{u,v}$ are measurable with respect to the Borel σ-algebras.*

Proof: ??? $\square$ Let $\mathfrak{S}_u^v$ be the smallest σ-algebra which makes $c_{u,v}$ measurable with respect to the Borel σ-algebra on $H[u,v]$. By Lemma 1.7.5, the system $(\mathbf{N}, H[s,t], \mathfrak{S}_u^v, n_w)$ is a 'path system' i.e. it satisfies the conditions of the definition of a Markov process (see [3, Def.1, p.40]) which do not refer to the measures. We call it the singleton path system. A Markov process on this path system is a collection of probability kernels

$$P_u : \mathbf{N} \rightarrow (H[s,t], \mathfrak{S}_u^t)$$

such that the collection $P_{u,v} = n_v P_u : \mathbf{N} \rightarrow \mathbf{N}$ has the standard Markov property

$$P_{u,u} = Id$$

$$P_{v,w} \circ P_{u,v} = P_{u,w}.$$

We will assume in addition that our processes satisfy a stronger version of the 'future depends on the past only through the present' condition.

Condition 1.7.6 *[**condA**] For any $s \leq u \leq v \leq t$ one has*

$$(P_u)_{|\mathfrak{S}_v^t} = P_v \circ P_{u,v}$$

Our first goal is to construct a class of additive Markov processes on the singleton path system which correspond to branching Markov processes on $\mathbf{N}$ satisfying certain continuity conditions.

1.8 Branching Markov processes on N

The dynamics of the population which consists identical individuals is fully described by a collection of probability kernels $P_{u,v} : \mathbf{N} \rightarrow \mathbf{N}$ given for all $u \leq v$, $u, v \in [s,t]$. The value $P_{u,v}(m, -)$ of $P_{u,v}$ on m is the measure on $\mathbf{N}$ whose value $P_{u,v}(m,n)$ on n is the probability for a population having m members at time u to have n members at time v. The assumption that the individuals are age-less is equivalent to the condition that these kernels form a Markov process i.e. that for $u \leq v \leq w$ one has

$$P_{v,w} \circ P_{u,v} = P_{u,w}.$$

We further assume that the individuals are independent (i.e. not 'aware' of each other) which is equivalent to the condition that this is a branching process i.e. that $P_{u,v}$ are homomorphisms of monoids in the category of probability kernels.

Such processes have standard description in terms of generating functions - formal power series of the form

$$[\mathbf{eform}] F(u,v;x) = \sum_{n=0}^{\infty} P_{u,v}(1,n) x^n. \tag{37}$$

The branching property implies that $P_{u,v}(m,n)$ is the n-th coefficient of the power series $F(u,v;x)^m$ and the Markovian condition becomes equivalent to the relation

$$[\mathbf{mcomp}]F(u,w;x) = F(u,v;F(v,w;x)). \tag{38}$$

This description provides a bijection between collections of formal power series $F(u,v;x)$ of the form (37) satisfying the conditions

$$F(u,v;1) = 1$$

$$P_{u,v}(1,n) \geq 0$$

and (38) and the isomorphism classes of branching Markov processes on $\mathbf{N}$. We let $BM(\mathbf{N};s,t)$ denote this set of isomorphism classes.

1.9 Branching Markov processes and E-path system

We want to construct for any such process $(F(t_1,t_2;x))_{s\leq t_1\leq t_2\leq t}$ which satisfies some continuity condition for the functions $F(t_1,t_2)(1)[n]$ an additive Markov process on the singleton path system $H[s,t]$ with the transition kernels given by $F(t_1,t_2;x)$. We will do it in two steps starting with a construction of intermediate path systems $\bar{E}[s,t]$ and $E[s,t]$.

Set:

$$\bar{E}[s,t] = \prod_{u\in[s,t]} \prod_{v\in[u,t]} (\coprod_{n\geq 0} S^n\mathbf{N})$$

where $S^n\mathbf{N}$ is the i-th symmetric power of $\mathbf{N}$. Define a map

$$e : H[s,t] \to \bar{E}[s,t]$$

by the condition that $pr_{u,v}(e(\Gamma))$ is in $S^n\mathbf{N}$ if Γ has n members $a_1,\ldots,a_n$ at time u and in this case it is given by $\{m_1\}+\cdots+\{m_n\}$ where m_i is the number of descendants of a_i at time v.

Remark 1.9.1 The invariant $e(\Gamma)$ has a better behavior than a more simple invariant which assigns to Γ the function

$$(u \mapsto n_u(\Gamma)) \in \prod_{u\in[s,t]} \mathbf{N}$$

since, as we will see below, for any $e \in \bar{E}[s,t]$ there are only finitely many Γ such that $e(\Gamma) = e$ and $n_u(\Gamma)$ does not have this property. For example consider the history Γ_w which has two members at the initial moment and the only transformation events are the death of the first one and the division of the second one into two both occurring at the same time w. Then for any $w \in (s,t]$ we have $n_u(\Gamma) \equiv 2$.

Let $\mathfrak{S}_s^t$ be the product σ-algebra of the maximal σ-algebras on the countable set $\coprod_{i\geq 0} S^n\mathbf{N}$. For any $[u,v] \subset [s,t]$ we have a projection $\bar{E}[s,t] \to \bar{E}[u,v]$ and we let $\mathfrak{S}_u^v$ denote the pull back to $\bar{E}[s,t]$ of $\mathfrak{S}_u^v$ on $\bar{E}[u,v]$.

For $u \in [s,t]$ let $n_u : \bar{E}[s,t] \to \mathbf{N}$ be the map which takes e to n such that $pr_{u,u}(e) \in S^n\mathbf{N}$. AS in the case of $H[s,t]$, one verifies immediately that the collection $(\mathbf{N}, \bar{E}[s,t], \mathfrak{S}^v_u, n_u)$ is a path system.

The monoid structure on $\amalg_{n\geq 0} S^n\mathbf{N}$ defines a monoid structure on $\bar{E}[s,t]$ and as before we call a process of this path system additive if the corresponding kernels $P_u : \mathbf{N} \to (\bar{E}[s,t], \mathfrak{S}^t_u)$ are homomorphisms of monoids.

Proposition 1.9.2 *[***ext1***] For any branching Markov process $F(t_1, t_2; x)$ on $\mathbf{N}$ over $[s,t]$ there exists a unique additive Markov process on $\bar{E}[s,t]$ with the transition kernels given by $F(t_1, t_2; x)$.*

Proof: ??? $\square$ Let $O = \{(u,v) | s \leq u \leq v \leq t\}$. Define $E[s,t]$ as the subset of $\bar{E}[s,t]$ which consists of functions $\rho : O \to S^\infty\mathbf{N}$ satisfying the following conditions:

1. ρ takes only a finite number of different values,
2. if $u < v$ then there exists $\delta > 0$ such that for all $\epsilon \leq \delta$ one has $\rho(u+\epsilon, v) = \rho(u,v)$,
3. if $v < t$ then there exists $\delta > 0$ such that for all $\epsilon \leq \delta$ one has $\rho(u, v+\epsilon) = \rho(u,v)$,

The property (36) shows that for any $\Gamma \in H[s,t]$ one has $e(\Gamma) \in E[s,t]$.

Let $\mathfrak{R}^s_t$ be the smallest σ-algebra which makes the functions n_x for $s \leq x \leq t$ measurable with respect to the obvious σ-algebra on $\mathbf{N}$. The standard construction shows that for any $m \in \mathbf{N}$, and any $s \in [-T, 0]$ there is a unique measure $P_{s,m}$ on $(V, \mathfrak{R}^s_0)$ such that for $n \in \mathbf{N}$ and $t \geq s$ one has $P_{s,m}(n_t^{-1}(n)) = P(s,t)[m,n]$ and that one has the following result.

Proposition 1.9.3 *[***pr1***] The collection of data $(n_t, \mathfrak{R}^s_t, P_{s,m})$ is a Markov process (in the sense of [3, Def.1, p.40]) with the phase space $\mathbf{N}$ and the space of elementary events $H[-T, 0]$.*

Therefore our first step is to show that the process $(n_t, \mathfrak{R}^s_t, P_{s,m})$ has a canonical extension to a process on a wider set of σ-algebras with respect to which r is measurable. Let $\mathfrak{S}^s_t = r^{-1}(\mathfrak{R}^s_t)$ be the smallest σ-algebra which makes the map r measurable with respect to the σ-algebra $\mathfrak{R}^s_t$ on $\tilde{H}$. It is generated by subsets

$$S_{x,m} = r^{-1}(R_{x,m})$$

for $s \leq x \leq t$, where

$$R_{x,m} = n_x^{-1}(m).$$

Let $\mathfrak{T}^s_t = \mathfrak{R}^s_t + \mathfrak{S}^s_t$.

Corollary 1.9.4 *[***c1***] The composition*

$$\mathbf{N} \stackrel{P'_s}{\to} H \stackrel{r}{\to} H \stackrel{n_t}{\to} \mathbf{N}$$

is a homomorphism whose value on 1 is represented by the power series $F(s,t; D(t) + (1 - D(t))x)$ where $D(t) = F(t, 0; 0)$.

Proof: We have $D(t) = F(t,0;0) = P_{t,1}(R_{0,0})$. Considering formal power series we get from (**??**):

$$\sum_{n\geq 0} P'_{s,1}(S_{t,n})x^n = \sum_{k,n\geq 0} P_{s,1}(R_{t,k}) \sum_{i_1+\cdots+i_n=n} \prod_{j=1}^{n} P'_{t,1}(S_{t,i_j})x^n =$$

$$= \sum_k P_{s,1}(R_{t,k})(\sum_i P'_{t,1}(S_{t,i})x^i)^n = \sum_k P_{s,1}(R_{t,k})(D(t) + (1-D(t))x)^k.$$

which proves the corollary. □ Let

$$\phi_t = D(t) + (1-D(t))x$$

and let

$$\phi_t^{-1} = (x - D(t))/(1-D(t))$$

such that

$$[\textbf{eq4}]\phi_t(\phi_t^{-1}(x)) = \phi_t^{-1}(\phi_t(x)) = Id. \tag{39}$$

Set

$$\tilde{F}(s,t;x) = \phi_s^{-1}(F(s,t;\phi_t(x)).$$

The equations (39) imply immediately that the series $\tilde{F}$ satisfy the relations (38) and therefore define a branching Markov process. We have:

$$\tilde{F}(s,t;0) = \phi_s^{-1}(F(s,t;D(t))) = \phi_s^{-1}(D(s)) = 0$$

i.e. this process is death free. We let $\tilde{P}_s$ denote the corresponding probability kernels $\mathbf{N} \to (\tilde{H}, \mathfrak{R}_0^s)$.

Lemma 1.9.5 *[***l1***] There are commutative diagrams of probability kernels:*

$$\begin{array}{ccc}
\mathbf{N} & \xrightarrow{\phi_s^*} & \mathbf{N} \\
{\scriptstyle P_s}\downarrow & & \downarrow{\scriptstyle \tilde{P}_s} \\
(H, \mathfrak{T}_0^s) & \xrightarrow{r} & (\tilde{H}, \mathfrak{R}_0^s) \\
{\scriptstyle n_t}\downarrow & & \downarrow{\scriptstyle n_t} \\
\mathbf{N} & \xrightarrow{\phi_t^*} & \mathbf{N}
\end{array}$$

where ϕ_s^* is the additive probability kernel $\mathbf{N} \to \mathbf{N}$ corresponding to the power series ϕ_s.

Proof: Follows immediately from Corollary 1.9.4. □ Let's write $\phi_s^*(n) = \sum_k a_k \delta_k$ where δ_k is the δ-measure concentrated at k. By Corollary 1.9.4 we have

$$P_s(n)[S_{t_1,n_1} \cap \cdots \cap S_{t_q,n_q}] = P_s(n)[r^{-1}(R_{t_1,n_1} \cap \cdots \cap R_{t_q,n_q})] =$$

$$= \tilde{P}_s\phi_s^*(n)[R_{t_1,n_1} \cap \cdots \cap R_{t_q,n_q}] = \sum_k a_k \tilde{P}_s(k)[R_{t_1,n_1} \cap \cdots \cap R_{t_q,n_q}].$$

Assume that $s \leq t_1 \leq \cdots \leq t_q$. Since $\tilde{P}_s$ for a Markov process we have

$$\tilde{P}_s(k)(R_{t_1,n_1} \cap \cdots \cap R_{t_q,n_q}) = \tilde{P}_s(k)[R_{t_1,n_1}]\tilde{P}_{t_1}(n_1)[R_{t_2,n_2}] \ldots \tilde{P}_{t_{q-1}}(n_{q-1})[R_{t_q,n_q}]$$

and therefore, again by Corollary 1.9.4

$$P_s(n)[S_{t_1,n_1}\cap\dots\cap S_{t_q,n_q}] = (\sum_k a_k\tilde{P}_s(k)[R_{t_1,n_1}])\tilde{P}_{t_1}(n_1)[R_{t_2,n_2}]\dots\tilde{P}_{t_{q-1}}(n_{q-1})[R_{t_q,n_q}] =$$

$$= n_{t_1}\tilde{P}_s\phi_s^*(n)[n_1]\tilde{P}_{t_1}(n_1)[R_{t_2,n_2}]\dots\tilde{P}_{t_{q-1}}(n_{q-1})[R_{t_q,n_q}] =$$
$$= \phi_{t_1}^* n_{t_1}P_s(n)[n_1]\tilde{P}_{t_1}(n_1)[R_{t_2,n_2}]\dots\tilde{P}_{t_{q-1}}(n_{q-1})[R_{t_q,n_q}]$$

Using again formal power series we get the following result.

Lemma 1.9.6 *[**fc1**] The value of $P_s(n)[S_{t_1,n_1}\cap\dots\cap S_{t_q,n_q}]$ is the coefficient at $x_1^{n_1}\dots x_q^{n_q}$ in the expression $(F(s,t_1;\phi_{t_1}(x_1))^n\tilde{F}(t_1,t_2;x_2)^{n_1}\dots\tilde{F}(t_{q-1},t_q;x_q)^{n_{q-1}}$.*

1.10 Reduced processes

Remark 1.10.1 The measures on $H[s,t]$ which we are going to consider in this paper vanish on the subsets of the form

$$\iota_{2,u} = \{\Gamma \text{ such that there exists a division point } v \text{ with } \phi(v)=u\}$$

but not necessarily on the subsets of the form

$$\iota_{0,u} = \{\Gamma \text{ such that there exists a death point } v \text{ with } \phi(v)=u\}$$

so we should be careful with the behavior of our constructions on the subsets of the second kind but not of the first.

Remark 1.10.2 One verifies easily that there are histories Γ,Γ' such that $n_u(\Gamma)=n_u(\Gamma)'$ for all u but $n_ur(\Gamma)\neq n_ur(\Gamma')$ for some value of u. In the most simple example of this kind the function $n_u(\Gamma)=n_u(\Gamma)'$ is the step function taking values $2,3,2$. This implies in particular that r is not measurable with respect to the minimal σ-algebras which are generated by the functions n_u.

1.11 Parameters space for singleton processes

Definition 1.11.1 *[**abar**] For $s\le t$ define the set $\bar{A}[s,t]$ as the set of functions $\sigma:[s,t]\to(0,1]$ satisfying the following conditions*

1. *σ is smooth outside of a finite number of points $\tau_i\in(s,t)$ and in all smooth points it satisfies the inequality*
$$[\textbf{maininეq}]\sigma'\ge -\sigma(1-\sigma) \quad (40)$$
2. *for any $x\in\{\tau_i\}\cup\{s\}$ the limit*
$$\sigma_+(x) = lim_{\epsilon>0,\epsilon\to 0}\sigma(x+\epsilon)$$
exists and one has $\sigma_+(x)=\sigma(x)$,

3. *for any* $x \in \{\tau_i\} \cup \{t\}$ *the limit*

$$\sigma_-(x) = lim_{\epsilon > 0, \epsilon \to 0} \sigma(x - \epsilon)$$

exists and one has $\sigma_-(x) \le \sigma(x)$

4. $\sigma(t) = 1$.

Define a topology on $\bar{A}[s,t]$ by the metric

$$dist(f,g) = |f(s) - g(s)|^2 + |f(t) - g(t)|^2 + \int_s^t |f(x) - g(x)|^2 dx$$

or by any equivalent one.

Lemma 1.11.2 *[***value***] For any* $x \in [s,t]$ *the function* $f \mapsto f(x)$ *is continuous on* $\bar{A}[s,t]$.

Proof:(Sketch) Our definition of the metric immediately implies the statement of the lemma for $x = s, t$. Therefore we may assume that $x \in (s,t)$. We need to show that for any $f \in \bar{A}, \epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that $|f(x) - g(x)| \ge \epsilon$ implies that $dist(f,g) \ge \delta(\epsilon)$. Assume for example that $g(x) > f(x)$. Then in order for g to be close to f on the interval $(x,t]$, g has to decrease as fast as possible. However, its rate of decrease is limited by the inequality (40) which allows one to find the required δ. $\square$

Proposition 1.11.3 *[***pex1***] For any* $\sigma \in \bar{A}[-T,0]$ *there exists a unique singleton process* $F(x,y;u)$ *such that for* $x \in [s,t]$ *one has:*

$$\sigma(x) = 1 - F(x,0;0).$$

Proof: Let us first consider the case when σ is smooth. Let F be a singleton process with the death rate $d(t)$. Set

$$\delta(s,t) = \int_s^t d(x)dx$$

By [5, p.47] we have:

$$F(s,t;u) = 1 - \frac{(1-u)e^{t-s-\delta(s,t)}}{1 + (1-u)\int_s^t e^{t-x-\delta(x,t)}dx}.$$

Set $F(x;u) = F(x,0;u)$ and $\delta(x) = \delta(x,0)$ then

$$F(t;u) = 1 - \frac{(1-u)e^{-(t+\delta(t))}}{1 + (1-u)\int_t^0 e^{-(x+\delta(x))}dx}$$

Set

$$\phi(t) = 1 + e^{\int_t^0 e^{-(x+\delta(x))}dx}$$

Then

$$\phi' = -e^{-(x+\delta(x))}$$

and

$$F(t;u) = 1 + \frac{(1-u)\phi'(t)}{1+(1-u)(\phi(t)-1)}$$

$$1-\sigma(t) = F(t;0) = 1 + \frac{\phi'}{\phi}$$

$$c - \int_t^0 \sigma(x)dx = ln(\phi)$$

From $\phi(0) = 2$ we get:

$$\phi(t) = 2e^{\int_t^0 \sigma(x)dx}$$

and $\phi' = -\sigma\phi$. We get:

$$F(t;u) = \frac{(\phi(t)^{-1} - 1 + \sigma(t))u + 1 - \sigma(t)}{(\phi(t)^{-1} - 1)u + 1}$$

Since this is an invertible function of u with the inverse

$$F^{\circ(-1)}(t,u) = \frac{-u + 1 - \sigma(t)}{(\phi(t)^{-1} - 1)u + 1 - \phi(t)^{-1} - \sigma(t)}$$

and from the Markovian property we get

$$F(s,t;u) = F(s;u) \circ F^{\circ(-1)}(t;u)$$

i.e.

$$F(s,t;u) = \frac{(-\sigma(s)\phi(t)^{-1} + \phi(t)^{-1} - \phi(s)^{-1})u + \phi(s)^{-1} - \phi(t)^{-1} - \sigma(t)\phi(s)^{-1} + \phi(t)^{-1}\sigma(s)}{(\phi(t)^{-1} - \phi(s)^{-1})u + \phi(s)^{-1} - \phi(t)^{-1} - \phi(s)^{-1}\sigma(t)}$$

which gives us an explicit formula for F as a function of σ when σ is smooth. Setting

$$\phi(s,t) = e^{-\int_s^t \sigma(x)dx}$$

we get

$$[\textbf{fsigma}]F(s,t;u) = 1 - \sigma(s)\frac{u-1}{(1-\phi(s,t))u + \phi(s,t) - 1 - \phi(s,t)\sigma(t)}. \tag{41}$$

Simple computation shows that such a system of functions forms a process (i.e. that all the coefficients in the Taylor series in u are non-negative) iff

$$\phi(s,t) \leq \frac{1-\sigma(s)}{1-\sigma(t)}$$

and that this condition holds for any $\sigma \in \bar{A}[-T,0]$. We denote the process (41) by F_σ.

□

2 Likelihood functional

2.1 Singleton processes

We consider here a particular class of branching Markov processes on $\mathbf{N}$ which we call singleton processes. Intuitively these processes describe the situation of a birth and death process with a constant birth rate equal 1. More precisely we consider families

$$F(s,t;u) = \sum b_k(s,t)u^k$$

such that for $\epsilon \geq 0$ one has:

$$b_k(t-\epsilon,t) = \begin{cases} o_2(\epsilon) & \text{for } k>2 \\ \epsilon + o_2(\epsilon) & \text{for } k=2 \\ o(\epsilon) & \text{for } k=0 \end{cases}$$

We assume our time interval to be $(-\infty,0]$ and write $D(t) = b_0(t,0)$ for the cumulative death rate of our process from t to 0.

We start with explicit calculation of F and $\tilde{F}$ in case when b_i's are smooth enough to use the standard differential equations describing generating functions of branching processes. Since we consider birth and death processes there are functions p_0, p_1, p_2 such that $p_0 + p_1 + p_2 = 0$ and we have:

$$[\mathbf{eq21}]\frac{\partial F(t,0;u)}{\partial t} = -f(t, F(t,0,u)) \tag{42}$$

where $f(t,x) = p_2(t)x^2 + p_1(t)x + p_0(t)$ (see e.g. [5, Th.4, p.39]). Since we assume that the birth rate is constant and equals 1 we have $p_2 = 1$ and therefore $p_1 = 1 - p_0$ where p_0 is the death rate. Then

$$f(t,x) = (x - p_0(t))(x-1)$$

We will write $d(t)$ instead of $p_0(t)$.

We further have

$$\tilde{F}(t,0;u) = \phi_t^{-1}F(t,0;u) = (F - D(t))/(1-D(t))$$

and

$$[\mathbf{eq22}]F = (1-D(t))\tilde{F} + D(t). \tag{43}$$

where $D(t) = F(t,0;0)$. Substituting (43) in (42) and using the consequence

$$\frac{\partial D(t)}{\partial t} = -f(t, D(t))$$

of (42) we get

$$\frac{\partial \tilde{F}}{\partial t} + f(t,D(t))\tilde{F} - D(t)\frac{\partial \tilde{F}}{\partial t} - f(t,D(t)) =$$

$$= -(p_0 + p_1(1-D(t))\tilde{F} + p_1 D(t) + (1-D(t))^2\tilde{F}^2 + D(t)^2 + 2D(t)(1-D(t))\tilde{F})$$

which implies for $D(t) \neq 1$:

$$(1-D(t))\tilde{F}^2 - (1-D(t))\tilde{F} = -\frac{\partial \tilde{F}}{\partial t}.$$

Since $D(t) = F(t, 0; 0)$ the (42) implies that we have

$$\frac{\partial D}{\partial t} = (D - d)(1 - D)$$

Let us denote $1 - D(t)$ by $\sigma(t)$. Then $\sigma(t)$ is the probability that one population member at time t will have at least one living descendant at time 0 and it is connected with the death rate by the equation

$$\sigma' = \sigma(\sigma + d - 1)$$

The condition $d \geq d_0$ where d_0 is a constant is then equivalent to the condition

$$1 - \sigma + (\sigma'/\sigma) \geq d_0$$

and since $\sigma \geq 0$ this is equivalent to

$$\sigma' \geq \sigma(\sigma + d_0 - 1)$$

For $d_0 = 0$ we get the inequality

$$\sigma' \geq -\sigma(1 - \sigma)$$

Since $\tilde{F}(s, t; u)$ for all s, t is determined by $\tilde{F}(t, 0; u)$ through equations 38 we see (using again [5, Th.4, p.39]) that $\tilde{F}(s, t; u)$ is the generating function of a birth process with the birth rate equal to $\sigma(t)$.

Using the explicit formula for the generating functions of such processes (see e.g. [5, Ex.9, p.46]) we get:

$$[\mathbf{m1}]\tilde{F}(s, t; u) = \frac{q(t)u}{(q(t) - q(s))u + q(s)} \tag{44}$$

where

$$q(t) = \exp(\int_t^0 \sigma(x)dx).$$

Let's write

$$[\mathbf{ared}]\tilde{F}(s, t; u) = \sum_k a_k(s, t)u^k \tag{45}$$

From (44) we get:

$$\frac{\partial \tilde{F}}{\partial u} = \frac{q(s)q(t)}{((q(t) - q(s))u + q(s))^2}$$
$$\frac{\partial^2 \tilde{F}}{\partial u^2} = 2\frac{q(s)q(t)(q(s) - q(t))}{((q(t) - q(s))u + q(s))^3}$$

and therefore

$$a_1(s, t) = \frac{q(t)}{q(s)}$$
$$a_2(s, t) = \frac{q(t)}{q(s)}(1 - \frac{q(t)}{q(s)})$$

Let us consider the sequence of $t's$ and $n's$ is of the form

$$\begin{array}{cccccccccc} t_0, & t_0, & t_1 - \epsilon, & t_1 + \epsilon, & t_2 - \epsilon, & t_2 + \epsilon, & \ldots, & t_q - \epsilon, & t_q + \epsilon, & t_{q+1} \\ N, & \tilde{n}, & \tilde{n}, & \tilde{n} + 1, & \tilde{n} + 1, & \tilde{n} + 2, & \ldots, & \tilde{n} + q - 1, & \tilde{n} + q, & \tilde{n} + q \end{array}$$

where ϵ is sufficiently small such that the sequence of $t's$ is an increasing one. We want to compute

$$F(N,\tilde{n};t_0,\dots,t_{q+1}) = P_{t_0}(N)[S_{t_0,\tilde{n}},\dots,S_{t_q,\tilde{n}+q}].$$

By Lemma 1.9.6 we get

$$F(N,\tilde{n};t_0,\dots,t_{q+1}) = \binom{N}{\tilde{n}}(1-\sigma(t_0))^{N-\tilde{n}}\sigma(t_0)^{\tilde{n}}a_1(t_0,t_1-\epsilon)^{\tilde{n}}\,\tilde{n}a_1(t_1-\epsilon,t_1+\epsilon)^{\tilde{n}-1}a_2(t_1-\epsilon,t_1+\epsilon)$$

$$a_1(t_1+\epsilon,t_2-\epsilon)^{\tilde{n}+1}(\tilde{n}+1)a_1(t_2-\epsilon,t_2+\epsilon)^{\tilde{n}}a_2(t_2-\epsilon,t_2+\epsilon)\dots$$
$$\dots(\tilde{n}+q-1)a_1(t_q-\epsilon,t_q+\epsilon)^{\tilde{n}+q-2}a_2(t_q-\epsilon,t_q+\epsilon)a_1(t_q+\epsilon,t_{q+1})^{\tilde{n}+q}$$

Set

$$[\mathbf{bi}]B_i = \begin{cases} \int_{t_0}^{t_1-\epsilon}\sigma(x)dx & \text{for } i=0 \\ \int_{t_i+\epsilon}^{t_{i+1}-\epsilon}\sigma(x)dx & \text{for } i=1,q-1 \\ \int_{t_q+\epsilon}^{t_{q+1}}\sigma(x)dx & \text{for } i=q \end{cases} \tag{46}$$

and for $i=1,\dots,q$:

$$[\mathbf{ci}]C_i = \int_{t_i-\epsilon}^{t_i+\epsilon}\sigma(x)dx \tag{47}$$

The we have:

$$F(N,\tilde{n};t_0,\dots,t_{q+1};\epsilon) = M\binom{N}{\tilde{n}}(1-\sigma(t_0))^{N-\tilde{n}}\sigma(t_0)^{\tilde{n}}e^{-\tilde{n}B_0}e^{-\tilde{n}C_1}(1-e^{-C_1})e^{-(\tilde{n}+1)B_1}e^{-(\tilde{n}+1)C_2}(1-e^{-C_2})\dots$$

$$\dots e^{-(\tilde{n}+q-1)C_q}(1-e^{-C_q})e^{-(\tilde{n}+q)B_q}$$

where

$$M = \tilde{n}(\tilde{n}+1)\dots(\tilde{n}+q-1).$$

2.2 Computation A

???This lemma has to be reproved for functions in $\bar{A}$.

Lemma 2.2.1 *[***cp1***] Let $t_0 < t_1$ and $\sigma_0,\sigma_1 \in (0,1]$. A smooth function $\sigma : [t_0,t_1] \to \mathbf{R}$ such that $\sigma(t_0)=\sigma_0$, $\sigma(t_1)=\sigma_1$ and*

$$[\mathbf{cond1}]\sigma' \geq -\sigma(1-\sigma) \tag{48}$$

exists if and only if

$$[\mathbf{asser1}]\sigma_1 \geq \frac{\sigma_0}{\sigma_0+(1-\sigma_0)e^{t_1-t_0}} \tag{49}$$

or equivalently

$$[\mathbf{asser2}]\sigma_0 \leq \frac{\sigma_1}{\sigma_1+(1-\sigma_1)e^{t_0-t_1}} \tag{50}$$

and the equalities are achieved for a unique function

$$[\mathbf{s01}]\sigma(u) = \frac{\sigma_0}{\sigma_0+(1-\sigma_0)e^{u-t_0}} \tag{51}$$

Proof: The equivalence of (49) and (50) is obvious. Let σ be a function satisfying the conditions of the proposition. Let us show that (49) holds. If $\sigma_1 = 1$ then (49) is obvious. Therefore, we may assume that $\sigma_1 < 1$. Assume that for all x, $\sigma(x) > 0$. Set

$$[\mathbf{cp1eq2}]\phi(x) = -\frac{\sigma'}{\sigma(1-\sigma)}. \tag{52}$$

Then (48) implies that $\phi(x) \leq 1$. Solving (52) with the initial condition $\sigma(t_0) = \sigma_0$ we get:

$$\sigma(u) = \frac{\sigma_0}{\sigma_0 + (1-\sigma_0)e^{\Phi(u)}}$$

where

$$\Phi(u) = \int_{t_0}^{u} \phi(x)dx \leq t_1 - t_0$$

which implies (49). This computation also implies that the condition which we have started with (that $\sigma > 0$) is superfluous and that the only smooth function for which (49) is an equality is (51).

Suppose now that $\sigma_1 \in [0,1]$ satisfies the strong version of (49). Let $\epsilon > 0$ be a sufficiently small number. Consider the function of the form (51) on the interval $[t_0, t_1 - \epsilon]$ and extend it to a smooth function on $[t_0, t_1]$ with $\sigma(t_1) = \sigma_1$ such that on the segment $[t_1 - \epsilon, t_1]$ we have $\sigma' >> 0$. Clearly, such σ satisfies (48). $\square$

??? The following lemma also has to be reproved for $\sigma \in \bar{A}$. Change the definition of $\bar{A}$ removing the normalization $\sigma(t) = 1$.

Lemma 2.2.2 *[**bcomp**] Let σ be a function satisfying the conditions of Lemma 2.2.1. Then*

$$[\mathbf{asser3}](1 + \sigma_1(e^{t_1 - t_0} - 1))^{-1} \leq e^{-\int_{t_0}^{t_1}\sigma(x)dx} \leq 1 + \sigma_0(e^{t_0 - t_1} - 1) \tag{53}$$

The equality is achieved in the class of smooth functions only if the equality holds in (49). In this case the only function which achieves the equality in any of the inequalities of (53) is (51) which makes both inequalities to be equalities.

Proof: Lemma 2.2.1 shows that

$$\sigma(u) \geq \frac{\sigma_0}{\sigma_0 + (1-\sigma_0)e^{u - t_0}}$$

and

$$\sigma(u) \leq \frac{\sigma_1}{\sigma_1 + (1-\sigma_1)e^{u - t_1}}$$

Computing the integrals we get (53). $\square$

2.3 Computation B

Set

$$[\mathbf{fofsigma}]F(t_1, \ldots, t_{q+1}; \epsilon) = e^{-C_1}(1 - e^{-C_1})e^{-2B_1}e^{-2C_2}(1 - e^{-C_2})\ldots e^{-qC_q}(1 - e^{-C_q})e^{-(q+1)B_q} \tag{54}$$

and

$$G(N, t_0; \epsilon) = N(1 - \sigma(t_0))^{N-1}\sigma(t_0)e^{-B_0}$$

such that

$$F(N, 1; t_0, \ldots, t_{q+1}; \epsilon) = q!G(N, t_0; \epsilon)F(t_1, \ldots, t_{q+1}).$$

Proposition 2.3.1 *[**redf1**] For any $\sigma \in \bar{A}[t_1, t_{q+1}]$ which maximizes $F(t_1, \ldots, t_{q+1})$ there exists $T < t_1$ such that for any $t_0 \leq T$ there is an extension of σ to an element of $\bar{A}[t_0, t_{q+1}]$ which maximizes $F(N, 1; t_0, \ldots, t_{q+1}; \epsilon)$.*

Proof: We will show that for any $y > 0$ there exists T such that for $t_0 < T$ a function $f \in \bar{A}[t_0, t_1]$ which maximizes $G(N, t_0; \epsilon)$ exists and for any such function one has $f(t_1) < y$. Applying this result to $y = \sigma(t_1)$ we get a function f which, when 'concatenated' with σ will lie in $\bar{A}[t_0, t_{q+1}]$ and maximizes both $F(t_1, \ldots, t_{q+1})$ and $G(N, t_0; \epsilon)$.

□

Proposition 2.3.2 *[**redf2**] Let ϵ be admissible with respect to $t_1, \ldots, t_{q+1}$. Then there exists $T <<$ t_1 such that for any $t_0 \leq T$ and any function $\sigma \in \bar{A}[t_0, t_{q+1}]$ which maximizes $F(N, 1; t_0, \ldots, t_{q+1}; \epsilon)$ the restriction $\sigma_{|[t_0,t_1]}$ maximizes $max_{N\geq 1}G(N, t_0; \epsilon)$ and the restriction $\sigma_{|t_1,t_{q+1}}$ maximizes $F(t_1, \ldots, t_{q+1})$.*

Proof: ??? □

Lemma 2.3.3 *[**redf3**] For any $t_1, \ldots, t_{q+1}$ and any sufficiently small ϵ there exists a function $\sigma \in \bar{A}[t_1, t_{q+1}]$ which maximizes $F(t_1, \ldots, t_{q+1})$.*

Proof: ??? □

2.4 Computation C

Here we consider the problem of maximizing $F(t_1, \ldots, t_{q+1}; \epsilon)$ as a functional on $\bar{A}[t_1 - \epsilon, t_{q+1}]$. For σ in $\bar{A}[t_1 - \epsilon, t_{q+1}]$ and $1 \leq i \leq q$ set:

$$y_i(\sigma) = \sigma(t_i + \epsilon)$$

Definition 2.4.1 *A number $\epsilon > 0$ is called admissible relative to $t_1, \ldots, t_{q+1}$ if $\epsilon < -(1/2)ln(q/(q+1))$ and $\epsilon < (t_{i+1} - t_i)/2$ for all $i = 1, \ldots, q$.*

Note that the conditions imposed on ϵ imply that the sequence $t_1 - \epsilon, t_1 + \epsilon, t_2 - \epsilon, \ldots, t_q + \epsilon, t_{q+1}$ is an increasing one and that $e^{-C_i} > i/(i+1)$ for $i = 1, \ldots, q$ which in turn implies that the functions $e^{-iC_i}(1 - e^{-C_i})$ are increasing functions of C_i.

In what follows we consider $t_1, \ldots, t_{q+1}$ to be fixed.

Lemma 2.4.2 *[***ccl1***] For a given collection $0 \le y_1, \ldots, y_q \le 1$ the set $C(y_1, \ldots, y_q; \epsilon)$ of functions $\sigma \in \bar{A}[t_1 - \epsilon, t_{q+1}]$ such that $y_i(\sigma) = y_i$ for $i = 1, \ldots q-1$ is non-empty if and only if*

$$[\mathbf{conc}]\frac{y_i}{y_i + (1-y_i)e^{t_{i+1}-t_i-2\epsilon}} \le \frac{y_{i+1}}{y_{i+1} + (1-y_{i+1})e^{-2\epsilon}} \tag{55}$$

Proof: It follows easily from Lemma 2.2.1. □

Lemma 2.4.3 *[***ccl2***] If $C(y_1, \ldots, y_q; \epsilon)$ is non-empty then there exists a unique element σ there which maximizes $F(t_1, \ldots, t_q; \epsilon)$ and one has*

$$\sigma(t_i - \epsilon) = \frac{y_i}{y_i + (1-y_i)e^{-2\epsilon}}$$

$$\sigma_-(t_{i+1} - \epsilon) = \frac{y_i}{y_i + (1-y_i)e^{t_{i+1}-t_i-2\epsilon}}$$

$$\sigma_-(t_{q+1}) = \frac{y_q}{y_q + (1-y_q)e^{t_{q+1}-t_q-\epsilon}}$$

$$e^{-C_i} = (1 + y_i(e^{2\epsilon} - 1))^{-1}$$

$$e^{-B_i} = \begin{cases} 1 + y_i(e^{2\epsilon-(t_{i+1}-t_i)} - 1) & \text{for } i < q \\ 1 + y_q(e^{\epsilon-(t_{q+1}-t_q)} - 1) & \text{for } i = q \end{cases}$$

Proof: By definition F is given by (54) where B_i and C_i are defined by (46) and (47) respectively. The terms of the product depending on B_i's are decreasing in B_i's and in view of the fact that ϵ is admissible the terms depending on C_i are increasing in C_i. For a given y_i, Lemma 2.2.2 shows that there exists a unique function $\sigma \in \bar{A}[t_i - \epsilon, t_i + \epsilon]$ (resp. $\sigma \in \bar{A}[t_i + \epsilon, t_{i+1} - \epsilon]$ for $i < q$ and $\sigma \in \bar{A}[q_i + \epsilon, t_{q+1}]$ for $i = q$) such that $\sigma(t_i + \epsilon) = y_i$ which maximizes C_i (resp. minimizes B_i). The inequalities (55) show that we can concatenate these functions and get a function σ in $\bar{A}(t_1 - \epsilon, t_{q+1})$ which maximizes the product. One can easily see now that any other function which maximizes the product also should maximize each of the term and therefore it coincides with the σ which we have constructed. □ Set

$$\delta = e^{2\epsilon} - 1$$

$$r_i = \begin{cases} e^{2\epsilon-(t_{i+1}-t_i)} & \text{for } i < q \\ e^{\epsilon-(t_{q+1}-t_q)} & \text{for } i = q \end{cases}$$

Re-writing the formulas of Lemma 2.4.3 we get:

$$\sigma(t_i - \epsilon) = (1+\delta)y_i(\delta y_i + 1)^{-1}$$

$$\sigma_-(t_{i+1} - \epsilon) = r_i y_i((r_i - 1)y_i + 1)^{-1}$$

$$e^{-C_i} = (\delta y_i + 1)^{-1}$$

$$1 - e^{-C_i} = \delta y_i(\delta y_i + 1)^{-1}$$

$$e^{-B_i} = (r_i - 1)y_i + 1$$

and we get for our function $F(t_1,\dots,t_{q+1};\epsilon)$ the expression:

$$F=\delta^q\prod_{i=1}^{q} y_i((r_i-1)y_i+1)^{i+1}(\delta y_i+1)^{-(i+1)}$$

which we have to maximize on the set of $y_1,\dots,y_q$ satisfying

$$y_1\geq 0$$
$$y_{i+1}\geq(1+\delta)y_i((1+\delta-r_{i+1})y_i+r_{i+1})^{-1}\ \text{ for i=1,...,q}$$
$$1\geq y_{q+1}$$

Note that all the expressions involve Moebius (linear fractional) functions of y_i which we may describe in terms of 2x2 matrices considered up to a scalar multiple:

$$M_i=\begin{pmatrix} r_i-1 & 1\\ \delta & 1\end{pmatrix}$$

$$E_i=\begin{pmatrix} 1+\delta & 0\\ 1+\delta-r_i & r_i\end{pmatrix}^{-1}=\begin{pmatrix} r_i & 0\\ r_i-(1+\delta) & 1+\delta\end{pmatrix}$$

Then our function becomes

$$F=\delta^q\prod_{i=1}^{q} y_iM_i(y_i)^{i+1}$$

and the conditions

$$y_1\geq 0$$
$$y_{i+1}\geq E_{i+1}^{-1}(y_i)\ \text{ for i=1,...,q}$$
$$1\geq y_{q+1}$$

we have

$$det(E_i)=r_i(1+\delta)>0$$

which implies that $E_i(y)$ are increasing functions. Set

$$A_i=E_{i+1}\dots E_q$$

and introduce new variables:

$$u_i=A_i^{-1}(y_i)$$

Then the function becomes

$$[\textbf{ufun}]F=\delta^q\prod_{i=1}^{q} A_i(u_i)M_i(A_i(u_i))^{i+1} \tag{56}$$

and the inequalities become

$$[\textbf{uineq}]0\leq u_1\leq\dots\leq u_q\leq 1 \tag{57}$$

i.e. we have to find maximums of (56) on the simplex (57). We have:

$$E_jE_{j+1}=\begin{pmatrix} r_j & 0\\ r_j-(1+\delta) & 1+\delta\end{pmatrix}\begin{pmatrix} r_{j+1} & 0\\ r_{j+1}-(1+\delta) & 1+\delta\end{pmatrix}=\begin{pmatrix} r_jr_{j+1} & 0\\ r_jr_{j+1}-(1+\delta)^2 & (1+\delta)^2\end{pmatrix}$$

which implies that

$$A_i = \begin{pmatrix} r_{i+1}\dots r_q & 0 \\ r_{i+1}\dots r_q - (1+\delta)^{q-i} & (1+\delta)^{q-i} \end{pmatrix}$$

and

$$M_i A_i = (1+\delta)^{-1} \begin{pmatrix} r_i \dots r_q - (1+\delta)^{q-i} & (1+\delta)^{q-i} \\ r_{i+1}\dots r_q - (1+\delta)^{q-i-1} & (1+\delta)^{q-i-1} \end{pmatrix}$$

Proposition 2.4.4 *[umax] There exists $\rho > 0$ such that for any $0 < \epsilon < \rho$, any $i = 1, \dots q$ and any $k = 1, \dots, q+1-i$ the function*

$$\prod_{j=0}^{k-1} A_{i+j}(u) M_{i+j}(A_{i+j}(u))^{i+j+1}$$

has a unique maximum for $u \in (0, 1]$.

Proof: ??? $\square$

2.5 Computation for $\delta = 0$

Set $s_i = 1 - r_i \dots r_q$ since $r_j \leq 1$ we have $1 > s_i \geq s_{i+1} \geq 0$ and any monotone decreasing sequence of s_i's may arise from a combinations of the event times $t_1 \leq \dots \leq t_q$. For $\delta = 0$ our formulas become:

$$A_i = \begin{pmatrix} 1 - s_{i+1} & 0 \\ -s_{i+1} & 1 \end{pmatrix} \qquad M_i A_i = \begin{pmatrix} -s_i & 1 \\ -s_{i+1} & 1 \end{pmatrix}$$

$$f_i(x) = A_i(x) M_i(A_i(x))^{i+1} = (1 - s_{i+1}) x (-s_i x + 1)^{i+1} (-s_{i+1} x + 1)^{-(i+2)}$$

$$f_{i,k} = \prod_{j=0}^{k-1} f_{i+j}(x) = (\prod_{j=0}^{k-1} (1 - s_{i+j+1})) x^k (-s_i x + 1)^{i+1} (-s_{i+k} x + 1)^{-(i+k+1)}$$

Lemma 2.5.1 *[**maxfik**] For $k > 0$ the function $f_{i,k}(x)$ has a unique maximum on $[0, 1]$ at the point*

$$x_{i,k} = \frac{k}{(i+k+1)s_i - (i+1)s_{i+k}}$$

Proof: Elementary computation. $\square$

3 Algorithms

4 Appendix. Some basic notions of probability

The main notion which we need is that of a probability kernel. Consider two measurable spaces (X, A), (Y, B) where X and Y are sets and A, B are σ-algebras of subsets of X and Y respectively.

A probability kernel $P:(X,A)\rightarrow(Y,B)$ is a function $X\times B\rightarrow\mathbf{R}_{\geq 0}$ such that for any $x\in X$ the function $P(x,-)$ is a probability measure on B and for any $U\in B$ the function $P(-,U)$ is a measurable function on (X,A). Probability kernels can be composed in a natural way. The category whose objects are measurable spaces and morphisms are probability kernels was first considered in [4] and we will call it the Giry category. Any measurable map $f:(X,A)\rightarrow(Y,B)$ may be considered as a probability kernel which takes a point x of X to the δ-measure $\delta_{f(x)}$.

The Giry category has a monoidal structure given on the level of spaces by the direct product. The monoidal category axioms are essentially equivalent to the Fubbini theorems.

The definition of a Markov process which we use is similar to but slightly different from the one adopted in [].

Definition 4.0.2 *[***pathsystem***] A path system over the interval $[s,t]$ is the following collection of data:*

1. *A measurable space (X,A) which is called the phase space of the system,*
2. *A set Ω which is called the path space of the system,*
3. *A family of maps $\xi_u:\Omega\rightarrow X$ given for all $u\in[s,t]$,*
4. *A family of σ-algebras $\mathfrak{S}_u^v$ on Ω given for all $u\leq v$ in $[s,t]$.*

These data should satisfy the following conditions:

1. *For $[u,v]\subset[a,b]$ one has $\mathfrak{S}_u^v\subset\mathfrak{S}_a^b$,*
2. *For $u\in[s,t]$ the map $\xi_u:(\Omega,\mathfrak{S}_u^u)\rightarrow(X,A)$ is measurable.*

For simplicity of notation we will sometimes abbreviate the notation for a path system omitting some of its components e.g. we may write $(\Omega,\mathfrak{S}_u^v)$ instead of $(X,A,\Omega,\xi_u,\mathfrak{S}_u^v)$.

We define the standard path system $St(X,A)$ associated with (X,A) setting $\Omega=X^{[s,t]}$, ξ_u to be the projections and $\mathfrak{S}_u^v$ to be the smallest σ-algebra which makes ξ_w for $w\in[u,v]$ measurable.

Definition 4.0.3 *[***mprocess***] A Markov process on a path system $((X,A),\Omega,\xi_u,\mathfrak{S}_u^v)$ is a collection of probability kernels*

$$P_u:(X,A)\rightarrow(\Omega,\mathfrak{S}_u^t)$$

such that $\xi_u\circ P_u=Id$ and for $u\leq v$ the square

$$\begin{tikzcd}(X,A) \arrow[r, "P_u"] \arrow[d, "P_{u,v}"'] & (\Omega,\mathfrak{S}_u^t) \arrow[d] \\ (X,A) \arrow[r, "P_v"] & (\Omega,\mathfrak{S}_v^t)\end{tikzcd} \tag{58}$$

where

$$P_{u,v} = \xi_v \circ P_u,$$

commutes.

One verifies easily that for a Markov process P and for $u \leq v \leq w$ one has

$$[\mathbf{comp0}]P_{u,u} = Id \tag{59}$$

$$[\mathbf{comp1}]P_{v,w} \circ P_{u,v} = P_{u,w} \tag{60}$$

Conversely, suppose that we are given a family of probability kernels $P_{u,v} : (X, A) \to (X, A)$ for all $[u, v] \subset [s, t]$ which satisfy the conditions (59) and (60). Then it is easy to define a Markov process on the standard path system associated with (X, A) with these transition kernels. We will say that a Markov process on (X, A) is such a collection of kernels or equivalently a Markov process on the standard path system associated with (X, A).

Definition 4.0.4 *[**mps**] Let $(X, A, \Omega, \xi_u, \mathfrak{S}_u^v)$ and $(X, A, \Omega', \xi'_u, \mathfrak{R}_u^v)$ be two path systems over $[s, t]$ with the same phase space. A morphism from the first to the second is a map $f : \Omega \to \Omega'$ such that:*

1. *for any $u \in [s, t]$ one has $\xi'_u \circ f = \xi'_u$,*
2. *for any $u \leq v$ in $[s, t]$ the map f is measurable with respect to $\mathfrak{S}_u^v$ and $\mathfrak{R}_u^v$.*

For any path system on (X, A) there is a unique morphism from it to the standard path system $St(X, A)$ on (X, A).

Lemma 4.0.5 *[**mpm**] Let f be a morphism of path systems as in Definition 4.0.4 and $(P_u)_{u \in [s,t]}$ a Markov process on the first one. Then the kernels fP_u form a Markov process on the second system.*

Proof: Elementary verification. $\square$ Note that for a morphism f of paths systems and a process P on the first one the transition kernels $P_{u,v}$ for P and fP coincide.

Definition 4.0.6 *[**lkh**] Let (Y, B) be a measurable space and $y \in Y$. Suppose that Y also carries a topology. The we define a partial order $\geq_y$ on the set of measures on (Y, B) setting $\mu \geq_y \mu'$ if there exists an open neighborhood U of y such that for any measurable Z in U one has $\mu(U) \geq mu'(U)$.*

Lemma 4.0.7 *[**contcase**] Let (Y, B) be a measure space which also carries a topology and $y \in Y$. Let further μ be a measure on Y and f, g two continuous non-negative functions on Y. If $f(y) > g(y)$ then $f\mu \geq_y g\mu$.*

Proof: ??? $\square$

Example 4.0.8 *[add1] Note that if under the assumptions of Lemma 4.0.7 we have $f(y) = g(y)$ then one may have $f\mu \geq_y g\mu$, $g\mu \geq_y f\mu$ or $f\mu$ and $g\mu$ may be incomparable relative to $\leq_y$.*

Definition 4.0.9 *[likelihood] Let $P : (X, A) \to (Y, B)$ be a probability kernel, y a point of Y and assume that Y has a topology.*

A maximal likelihood reconstruction of y relative to P is a point x of X such that for any x' one has $P(x, -) \geq_y P(x', -)$.

Lemma 4.0.10 *[existence] Let $P : (X, A) \to (Y, B)$ be a probability kernel of the form $x \mapsto f_x\mu$ where μ is a measure on (Y, B) and $(f_x)_{x\in X}$ is a collection of continuous functions on Y. Let $y \in Y$ and suppose that there exists a point $x \in X$ such that for any $x' \neq x$ one has $f_x(y) > f_{x'}(y)$. Then x is the maximal likelihood reconstruction of y relative to P.*

Proof: It follows immediately from Lemma 4.0.7. $\square$

4.1 Leftovers

General definitions The bijection (3) can be extended to a bijection between $H[s, t]$ and the geometric realization of a simplicial monoid. Recall that for a simplicial set $X_* = (X_i, \sigma_i^j, \partial_i^j)_{i\geq 0}$ its geometric realization $|X_*|$ is the topological space of the form

$$|X_*| = \amalg_{i\geq 0}(X_i^{nd} \times \Delta^i)/ \approx$$

where X^i_{nd} is the subset of non-degenerate simplexes in X^i and $\approx$ is an equivalence relation defined in the standard way by the boundary maps ∂_i^j (see e.g. [2]). If Δ^i_{op} is the open simplex for $i > 0$ and the point for $i = 0$ then there is a bijection of sets

$$|X_*| = \amalg_{i\geq 0} X_i^{nd} \times \Delta^i_{op}$$

Let $\Delta^i_{[s,t]}$ be the set of non decreasing increasing sequences $x_1 \leq \cdots \leq x_i$ in $[s, t]$ for $i > 0$ and the point for $i = 0$. These spaces are canonically homeomorphic to the standard simplexes and we may consider the topological realization functor $| - |_{[s,t]}$ based on $\Delta^*_{[s,t]}$ instead of Δ^*. The simplexes $\Delta^i_{(s,t)}$ considered above are the open analogs of $\Delta^i_{[s,t]}$.

Recall that for any co-triple M on a category C and any object X of C we have a simplical object $M_*(C)$ whose i-simplicies are given by $M_i(X) = M^{\circ(i+1)}(X)$. Consider the co-triple FAb on the category of commutative monoids which takes a monoid A to the free commutative monoid generated by A as a set, e.g. $FAb(pt) = \mathbf{N}$, One verifies easily that the set of combinatorial types of level q is naturally isomorphic as a monoid to $FAb_q(\mathbf{N})$ and that a combinatorial type is non-degenerate in our sense if and only if it corresponds to a non-degenerate simplex of $FAb_*(\mathbf{N})$. Together with (3) this observation implies immediately that

$$[\mathbf{simplrep2}]H[s,t] = |FAb_*(\mathbf{N})|_{[s,t]}. \tag{61}$$

Remark 4.1.1 *[**htype**] Since the co-triple FAb_* is given by the composition of the forgetful functor to sets with its left adjoint we conclude that $H[s,t]$ with the topology defined by (61) is homotopy equivalent to $\mathbf{N}$. A history Γ belongs to the connected component given by the number of final vertices with multiplicities defined by ψ.*

Remark 4.1.2 It seems that if we start with a co-triple which takes a commutative monoid A to the free commutative monoid generated by the set $A \times X$ where X is a set and apply the same constructions we will get a path system for branching processes with X-types.

For a combinatorial type we let $n_0(\pi)$ denote the number of elements in the last set of π or equivalently the number of "connected components" of π.

The commutative monoid structure on $H[s,t]$ provided by its realization as $|FAb_*(\mathbf{N})|_{[s,t]}$ corresponds to the disjoint union of histories. To distinguish it below from the addition of points of $H[s,t]$ considered as δ-measures we will denote this operation by $(\Gamma_1, \Gamma_2) \mapsto \Gamma_1 \amalg \Gamma_2$.

One of the important consequences of (61) is that there is a natural triangulation on $H[s,t]^{\times n}$ with respect to which the disjoint union map

$$\amalg_n : H[s,t]^{\times n} \to H[s,t]$$

is simplicial. The q-dimesnional simplexes of this triangulation are of the form $\pi_1 \times \cdots \times \pi_n$ where π_i are combinatorial types of level q (length $q+1$) such that $\pi_1 + \cdots + \pi_n$ is non-degenerate. We denote the simplex corresponding to $\pi_1 \times \cdots \times \pi_n$ by $\Delta^{\pi_1 \times \cdots \times \pi_n}_{(s,t)}$. Let $\pi_1, \ldots \pi_n$ be combinatorial types of the same level $q-1 \geq 0$ such that $[\pi_1] + \cdots + [\pi_n]$ is non-degenerate. Let further $(u,v) \subset (s,t)$ and $B \subset \Delta^{q-1}_{(v,t)}$. Denote by $((u,v), B, [\pi_1] \times \cdots \times [\pi_n])$ the subset of $\Delta^{[\pi_1] \times \cdots \times [\pi_n]}_{(s,t)}$ which consists of points $(x_1, \ldots, x_q)$ such that $x_1 \in (u,v)$ and $(x_2, \ldots, x_q) \in B$.

Lemma 4.1.3 *[**tech1**] One has*

$$((u,v), B, [\pi_1] \times \cdots \times [\pi_n]) =$$

$$= ((res_{s,v} \times res_{v,t})^{\times n})^{-1}(\{(\Gamma'_i, \Gamma''_i)_{i=1}^n \,|\, (\Gamma'_1, \ldots, \Gamma'_n) \in (u,v) \, and \, (\Gamma''_1, \ldots, \Gamma''_n) \in B\,;\, i = 1, \ldots, n\})$$

where (u,v) is considered as the subset of

$$\Delta^{[n_0(\pi_1)[1]] \times \cdots \times [n_0(\pi_n)[1]]}_{(s,v)} \subset H[s,v]^{\times n}_{1,*}$$

and B is considered as a subset of

$$\Delta^{\pi_1 \times \cdots \times \pi_n}_{(v,t)} \subset H[v,t]_{n_0(\pi_1),*} \times \cdots \times H[v,t]_{n_0(\pi_n),*}$$

Proof: Straightforward. □

There are two main ways to construct singleton histories inductively. For two singleton histories Γ_1, Γ_2 on $[s,t]$ their disjoint union $\Gamma_1 \amalg \Gamma_2$ is a singleton history. This operation makes $H[s,t]$ into a commutative monoid whose initial element is the empty history. The restriction maps are homomorphisms with respect to this monoid structure.

For $u \in (s,t]$ and $\Gamma \in H[u,t]_{k,*}$ we let $[k] *_u \Gamma$ denote the unique history such that

$$\pi(L_u([k] *_u \Gamma)) = [k]$$

and

$$R_u([k] *_u \Gamma) = \Gamma.$$

One observes easily that any history can be obtained by a combination of these two operations from the history $1 \in H[t,t]$.

Let $\Gamma \in H[s,t]$ and let $\psi : V_{\Gamma,t} \to \mathbf{N}$ be a function. We let Γ_ψ denote the history which is identical to Γ except that $\psi_{\Gamma_\psi} = \psi$.

For a map of finite sets $f : V_2 \to V_1$ denote by $\psi(f)$ the function

$$\psi(f)(x) = \#(f^{-1}(x))$$

Let $\Gamma' \in H[s,u]$, $\Gamma'' \in H[u,t]$ and $f : V_{\Gamma'',u} \to V_{\Gamma',u}$ be a map. Then we can glue Γ' and Γ'' in the obvious way obtaining a history $\Gamma' \cup_f \Gamma'' \in H[s,t]$ such that

$$res_{s,u}(\Gamma' \cup_f \Gamma'') = \Gamma'_{\psi(f)}$$

and

$$res_{u,t}(\Gamma' \cup_f \Gamma'') = \Gamma''$$

For $s = t$ we set $H[s,t]^{ord} = H[s,t] = \mathbf{N}$ and interpret it as the set of isomorphism classes of linearly ordered finite sets.

Given two ordered histories Γ_1, Γ_2 on $[s,t]$ there is a unique ordering on $\Gamma_1 \amalg \Gamma_2$ in which all elements of V_{s,Γ_1} precede all elements of V_{s,Γ_2} such that Γ_1 and Γ_2 are ordered sub-histories of $\Gamma_1 \amalg \Gamma_2$. This construction makes $H[s,t]^{ord}$ into a non-commutative monoid whose initial element is again the empty history. Observe, that given an ordered set X and a map $Y \to X$ with the orderings on each of its fibers we may equip Y with a "lexicographical" ordering in an obvious way. Conversely, given orderings on sets X and Y and a function $\psi : X \to \mathbf{N}$ such that $tr(\psi) = \#Y$ there are a unique map $f : Y \to X$ and orderings on its fibers such that $\psi(f) = \psi$ and the corresponding lexicographical order on Y coincides with the original one. If we denote the elements of X and Y by natural numbers than this map sends $1, \ldots, \psi(1)$ to 1, $\psi(1)+1, \ldots, \psi(1)+\psi(2)$ to 2 etc.

For an ordered history Γ and $u \in [s,t]$ we may define an ordering on V_u starting with the ordering on V_s and extending it lexicographically using the local orderings on edges starting at a given vertex.

Furthermore, $L_u(\Gamma)$ carries an obvious ordering and $R_u(\Gamma)$ carries the ordering which is defined by the ordering on $V_u(R_u(\Gamma)) = V_{u+\epsilon}(\Gamma)$ and the same local orderings as before.

This construction defines the restriction maps

$$res_{u,v} : H[s,t]^{ord} \to H[u,v]^{ord}$$

which satisfy the condition that for $(u', v') \subset (u, v)$ one has

$$res_{u',v'} res_{u,v} = res_{u',v'}$$

For an ordered Γ the sets $V_{x_i(\Gamma),\Gamma}$ are ordered and therefore we get a sequence of maps of ordered sets

$$V_t(\Gamma) \stackrel{f_q}{\to} V_{x_q}(\Gamma) \stackrel{f_{q-1}}{\to} \dots \stackrel{f_1}{\to} V_{x_1}(\Gamma)$$

which together with the function ψ_Γ on $V_t(\Gamma)$ and points $x_1, \dots, x_q$ defines Γ uniquely up to an isomorphism. This lets us to assign to each ordered Γ an invariant of the form

$$\pi^{ord}(\Gamma) = (n_0; k_{1,1}, \dots, k_{1,n_0}; k_{2,1}, \dots, k_{2,n_1}; \dots; k_{q+1,1}, \dots, k_{q+1,n_q})$$

where for $i > 0$ one has $n_i = k_{i,1} + \dots + k_{i,n_{i-1}}$ defined as follows:

1. $n_0 = \#V_{x_1}(\Gamma) = n_s(\Gamma)$
2. $k_{i,j} = \#f_i^{-1}(j)$ where j is the element number j in $V_{x_i}(\Gamma)$ for $i \leq q$
3. $k_{q+1,j} = \psi_\Gamma(j)$ where j is the element number j in $V_t(\Gamma)$.

One observes easily that Γ is completely determined by the collection $(\pi^{ord}, x_1, \dots, x_q)$ and that such a collection corresponds to an ordered history if and only if $s < x_1 < \dots < x_q < t$ and for all $i \leq q$ there exists j such that $k_{i,j} \neq 1$. Therefore there is a bijection

$$[\textbf{simplrep3}] H[s,t]^{ord} = \amalg_{\pi^{ord}} \Delta^{q(\pi)}_{(s,t)} \tag{62}$$

where $\Delta^q_{(s,t)} = \{s < x_1 < \dots < x_q < t\}$ and π^{ord} runs through the isomorphism classes of sequences of maps of ordered finite sets $V_{q+2} \stackrel{f_{q+1}}{\to} V_{q+1} \stackrel{f_q}{\to} \dots \stackrel{f_1}{\to} V_1$ such that for $i = 1, \dots, q$ the map f_i is not an isomorphism or equivalently through the set of sequences as above.

Remark 4.1.4 [**simplord**] There is a direct analog of (61) for ordered histories. One has

$$[\textbf{simplrep4}] H[s,t]^{ord} = |F_*(\mathbf{N})|_{[s,t]} \tag{63}$$

where F is the co-triple on the category of non-commutative monoids which takes a monoid to the free monoid generated by its underlying set.

The combinatorial types $[k]$, $k[1]$ and $[k[1]]$ correspond uniquely to the ordered combinatorial types $(1;k)$, $(k;1,\ldots,1)$ and $(1;k;1,\ldots,1)$ respectively and we will sometimes use the shorter notations $[k]$, $k[1]$ and $[k[1]]$ in the context of the ordered histories. Note that

$$[\mathbf{sk01}] sk_0H[u,v]^{ord}_{1,*} = sk_0H[u,v]_{1,*} = \coprod_{k\geq 0} \Delta^{[k]}_{(u,v)} \tag{64}$$

If $\Gamma \in H[u,t]^{ord}_{k,*}$ where $u \in (s,t]$ then $[k] *_u \Gamma$ has a unique ordering such that $R_u([k] *_u \Gamma) = \Gamma$ which allows us to extend the construction $\Gamma \to [k] *_u \Gamma$ to ordered histories.

4.2 Appendix: Processes on un-ordered histories

For a combinatorial type π define $d(\pi)$ inductively as follows:

1. $d([k]) = 1$ for all $k \in \mathbf{N}$,
2. if $\pi = \sum n_i[\sigma_i]$ where $\sigma_i \neq \sigma_j$ for $i \neq j$ then

$$d(\pi) = \frac{(\sum_i n_i)!}{\prod_i n_i!} \prod_i d(\sigma_i)^{n_i}$$

(in particular$d([\pi]) = d(\pi)$).

For a pre-process μ define $\nu^{v,\pi}_{\mu,u}$ as the measure on $\Delta^{q(\pi)}_{(u,v)}$ which is the co-restriction of $d(\pi)^{-1}\mu^v_{u,n_0(\pi)}$ to $\Delta^{\pi}_{(u,v)}$. As usual we will omit μ from the notation when possible.

Theorem 4.2.1 *[**th8**] Let μ be an additive Markov pre-process on $\mathcal{H}[s,t]$. Let π, π' be two combinatorial types with the same local invariant $\underline{K}$. Then for any $s \leq u < v \leq t$ one has $\nu^{v,\pi}_u = \nu^{v,\pi'}_u$.*

Proof: Let us first generalize the notations introduced above to simplexes $\Delta^{\pi_1\times\cdots\times\pi_n}_{(u,v)}$ of $H[u,v]^{\times n}$. Set $n_{0,i} = n_0(\pi_i)$. This simplex has dimension $q = q(\pi_i)$ and lies in $H[u,v]_{n_{0,1},*} \times \cdots \times H[u,v]_{n_{0,n},*}$. We let $\nu^{v,\pi_1\times\cdots\times\pi_n}_u$ denote the co-restriction of

$$d(\pi_1)^{-1}\ldots d(\pi_n)^{-1}\,\mu^v_{u,n_{0,1}} \otimes \ldots \otimes \mu^v_{u,n_{0,n}}$$

to this simplex. We are going to prove that for $(\pi_i)^n_{i=1}$, $(\pi'_i)^n_{i=1}$ which are of the same level q and which are locally equivalent i.e.

$$[\mathbf{th8eq0}]\underline{K}(\sum_i \pi_i) = \underline{K}(\sum_i \pi'_i) \tag{65}$$

and both sums are non-degenerate we have

$$[\mathbf{th8eq1}]\nu^{v,\pi_1\times\cdots\times\pi_n}_u = \nu^{v,\pi'_1\times\cdots\times\pi'_n}_u \tag{66}$$

It follows by induction from Lemmas 4.2.2,4.2.3 and 4.2.4 below.

Lemma 4.2.2 *[th8l1] The equality (66) holds for all locally equivalent* $([\sigma_j])$, $([\sigma'_j])$ *with* $q([\sigma_j]) = q([\sigma'_j]) = 0$.

Proof: We have $[\sigma_j] = [k_j]$ and $[\sigma'_j] = [k'_j]$ for some $k_j, k'_j \in \mathbf{N}$. Since they are locally equivalent there is a permutation of factors of $H[u,v]_{1,*}^{\times n}$ which takes $\Delta_{(u,v)}^{\times_j[\sigma_j]}$ to $\Delta_{(u,v)}^{\times_j[\sigma'_j]}$. Since $(\mu_{u,1}^v)^{\otimes n}$ is invariant under such permutations we conclude that

$$(\mu_{u,1}^v)^{\otimes n}(\Delta_{(u,v)}^{\times_j[\sigma_j]}) = (\mu_{u,1}^v)^{\otimes n}(\Delta_{(u,v)}^{\times_j[\sigma'_j]})$$

On the other hand $d([\sigma_j]) = d([\sigma'_j]) = 1$ and we conclude that (66) holds. $\square$

Lemma 4.2.3 *[th8l2] Let* $q \geq 0$. *If (66) holds for all locally equivalent* $([\sigma_j])$, $([\sigma'_j])$ *with* $q([\sigma_j]) = q([\sigma'_j]) \leq q$ *then it holds for all locally equivalent* (π_i), (π'_i) *with* $q(\pi_i) = q(\pi'_i) \leq q$.

Proof: Consider the disjoint union map

$$a = \times_{i=1}^n \amalg_{n_{0,i}} : H[u,v]_{1,*}^{\times n_{0,1}} \times \cdots \times H[u,v]_{1,*}^{\times n_{0,n}} \to H[u,v]_{n_{0,1},*} \times \cdots \times H[u,v]_{n_{0,n},*}$$

and let $n_0 = \sum_{i=1}^n n_{0,i}$. Since this map is the geometric realization of the map of simplicial sets given by addition of combinatorial types we have

$$a^{-1}(\Delta_{(u,v)}^{\pi_1 \times \cdots \times \pi_n}) = \coprod_{(\sigma_j) \in \Sigma} \Delta_{(u,v)}^{\times_j[\sigma_j]}$$

Where Σ is the set of sequences $(\sigma_1, \ldots, \sigma_{n_1})$ such that for each $i = 1, \ldots, n$

$$\sum_{j=1+n_{0,i-1}+\cdots+n_{0,1}}^{n_{0,i}+n_{0,i-1}+\cdots+n_{0,1}} [\sigma_j] = \pi_i$$

Since μ is additive we have

$$\mu_{u,n_{0,1}}^v \otimes \ldots \otimes \mu_{u,n_{0,n}}^v = a_*((\mu_{u,1}^v)^{\otimes n_0})$$

Since the decomposition of a combinatorial type into a sum of types of the form $[-]$ is unique up to the permutation of factors and the co-restriction of $(\mu_{u,1}^v)^{\otimes n_0}$ to $\Delta_{(u,v)}^{\times_j[\sigma_j]}$ does not depend on the choice of (σ_j) in Σ. We conclude that

$$\nu_u^{v,\pi_1 \times \cdots \times \pi_n} = \#\Sigma \prod_{i=1}^n d(\pi_i)^{-1} \prod_{j=1}^{n_0} d(\sigma_j) \nu_u^{v,\times_j[\sigma_j]}$$

for any $(\sigma_j) \in \Sigma$. Recalling our definition of $d(\pi_i)$ we conclude that

$$\prod_{i=1}^n d(\pi_i) = \#\Sigma \prod_{j=1}^{n_0} d(\sigma_j)$$

and therefore

$$\nu_u^{v,\pi_1 \times \cdots \times \pi_n} = \nu_u^{v,\times_j[\sigma_j]}$$

a similar equality holds for (π'_i) and for (σ'_j) in the corresponding set Σ'. On the other hand

$$\underline{K}(\sum_j [\sigma_j]) = \underline{K}(\sum_i \pi_i) = \underline{K}(\sum_i \pi'_i) = \underline{K}(\sum_j [\sigma'_j])$$

and by the assumption of the lemma we conclude that

$$\nu_u^{v,\pi_1\times\cdots\times\pi_n} = \nu_u^{v,\pi'_1\times\cdots\times\pi'_n}$$

$\square$

Lemma 4.2.4 *[**th8l3**] Let $q > 0$. If (66) holds for all locally equivalent (π_i), (π'_i) with $q(\pi_i) = q(\pi'_i) < q$ then (66) holds for all locally equivalent $([\sigma_j])_{j=1}^n$, $([\sigma'_j])_{j=1}^n$ with $q([\sigma_j]) = q([\sigma'_j]) = q$.*

Proof: The Borel σ-algebra of $\Delta_{(u,v)}^{\times_j[\sigma_j]} = \Delta_{(u,v)}^q$ is generated in the strong sense by "rectangles" of the form $((w_1, w_2), B, \times_j[\sigma_j])$ where $(w_1, w_2) \subset (u, v)$ and $B \subset \Delta_{(w_2,v)}^{q-1}$. By Lemma 4.1.3 we have

$$((w_1, w_2), B, \times_j[\sigma_j]) = (res_{u,w_2}^{\times n})^{-1}((w_1, w_2) \subset \Delta_{(u,w_2)}^{\times_j[n_0(\sigma_j)[1]]}) \times (res_{w_2,v}^{\times n})^{-1}(B \subset \Delta_{(w_2,v)}^{\times_j\sigma_j})$$

Since μ is a Markov pre-process we conclude that

$$\nu_u^{v,\times_j[\sigma_j]}((w_1, w_2), B, \times_j[\sigma_j]) = \prod_j d(\sigma_j)^{-1}\ (\mu_{u,1}^v)^{\otimes n}((w_1, w_2), B, \times_j[\sigma_j]) =$$

$$= \prod_j d(\sigma_j)^{-1}\ (\mu_{u,1}^{w_2})^{\otimes n}((w_1, w_2) \subset \Delta_{(u,w_2)}^{\times_j[n_0(\sigma_j)[1]]})\,(\otimes_j \mu_{w_2,n_0(\sigma_j)}^v)(B \subset \Delta_{(w_2,v)}^{\times_j\sigma_j}) =$$

$$= (\mu_{u,1}^{w_2})^{\otimes n}((w_1, w_2) \subset \Delta_{(u,w_2)}^{\times_j[n_0(\sigma_j)[1]]})\,\nu_{w_2}^{v,\times_j\sigma_j}(B \subset \Delta_{(w_2,v)}^{\times_j\sigma_j})$$

and a similar equality holds for $([\sigma'_j])$. We have $q(\sigma_j) = q(\sigma'_j) = q - 1$. By the assumption of the lemma we conclude that

$$\nu_{w_2}^{v,\times_j\sigma_j}(B) = \nu_{w_2}^{v,\times_j\sigma'_j}(B)$$

On the other hand there is a permutation of factors on $H[u, w_2]^{\times n}$ which takes $\Delta_{(u,w_2)}^{\times_j[n_0(\sigma_j)[1]]}$ to $\Delta_{(u,w_2)}^{\times_j[n_0(\sigma'_j)[1]]}$ and since $(\mu_{u,1}^{w_2})^{\otimes n}$ is invariant under such permutations we conclude that

$$(\mu_{u,1}^{w_2})^{\otimes n}((w_1, w_2) \subset \Delta_{(u,w_2)}^{\times_j[n_0(\sigma_j)[1]]}) = (\mu_{u,1}^{w_2})^{\otimes n}((w_1, w_2) \subset \Delta_{(u,w_2)}^{\times_j[n_0(\sigma'_j)[1]]})$$

$\square$ $\square$

4.3 Appendix: some remarks on Markov pre-processes

Lemma 4.3.1 *[**ob00**] Let μ_*^* be a Markov pre-process. Then for any $n, m \geq 0$ and any $u \leq v < w$ in $[s, t]$ the function $h^n(u, v + \epsilon)\phi_{v+\epsilon}^w(n, m)$ is monotone decreasing in ϵ and one has*

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u, v + \epsilon)\phi_{v+\epsilon}^w(n, m) = h^n(u, v)\phi_v^w(n, m)$$

Proof: Applying Lemma 1.2.1 to $U_1 = \Delta^{n[1]}_{u,v+\epsilon}$ and $U_2 = H[v+\epsilon, w]_{n,m}$ we get

$$h^n(u, v+\epsilon)\phi^w_{v+\epsilon}(n,m) = \mu^w_{u,n}(res^{-1}_{u,v+\epsilon}(\Delta^{n[1]}_{u,v+\epsilon}) \cap n_w^{-1}(m)).$$

Since for $\epsilon' \geq \epsilon$ one has

$$res^{-1}_{u,v+\epsilon'}(\Delta^{n[1]}_{u,v+\epsilon'}) \cap n_w^{-1}(m) \subset res^{-1}_{u,v+\epsilon}(\Delta^{n[1]}_{u,v+\epsilon}) \cap n_w^{-1}(m)$$

and

$$\cup_{\epsilon\to 0}(res^{-1}_{u,v+\epsilon}(\Delta^{n[1]}_{u,v+\epsilon}) \cap n_w^{-1}(m)) = res^{-1}_{u,v}(\Delta^{n[1]}_{u,v}) \cap n_w^{-1}(m)$$

our claims follow. $\square$

Recall that a function f on $[s,t]$ is called monotone increasing (resp. decreasing) if for $x \leq y$ one has $f(x) \leq f(y)$ (resp. $f(x) \geq f(y)$). A function is called right continuous if for all $u \in [s,t)$ one has

$$\lim_{\epsilon>0,\epsilon\to 0} f(u+\epsilon) = f(u).$$

The following two lemmas give some elementary properties of such functions which will be used below.

Lemma 4.3.2 *[**rcim2**] Any right continuous function f on $[s,t]$ is measurable.*

Proof: It is sufficient to show that for any a the subset $U_a = \{x : f(x) < a\}$ is measurable. For $y \in (\mathbf{Q} \cup \{t\}) \cap U_a$ consider the set $V_{a,y} = \{w \in [s,t] \,|\, [w,y] \subset U\}$. This set is of the form $[y_-, y]$ or $(y_-, y]$ where $y_- = Inf(V_{a,y})$ and in particular it is measurable. Let us show that

$$U_a = \cup_{y\in(\mathbf{Q}\cup\{t\})\cap U_a} V_{a,y}$$

which would imply that U is measurable. The inclusion "$\supset$" is obvious from the definition of $V_{a,y}$. If $t \in U_a$ then $t \in V_{a,t}$. Let $u \in U_a \cap [s,t)$. Since f is right continuous there exists $\epsilon > 0$ such that $[u, u+\epsilon] \subset U_a$. Let y be any element of $[u, u+\epsilon] \cap \mathbf{Q}$. Then $u \in V_{a,y}$ which proves the inclusion "$\subset$". $\square$

Lemma 4.3.3 *[**pirc**] Let f be a right continuous on $[s,t)$. If f is monotone increasing then for any $a_+ > a$ such that $f^{-1}([a,a_+)) \neq \emptyset$ there exists $b_+ > b$ such that $f^{-1}([a,a_+)) = [b,b_+)$. If f is monotone decreasing then for any $a_+ > a$ such that $f^{-1}((a,a_+]) \neq \emptyset$ there exists $b_- < b$ such that $f^{-1}((a,a_+]) = [b_-, b)$.*

Proof: Consider for example the case of an increasing f. Then if $f^{-1}([a,a_+)) \neq \emptyset$ we have

$$f^{-1}([a,\infty)) = [b,t)$$

and

$$f^{-1}((-\infty,a_+)) = [s,b_+)$$

which implies the claim of the lemma. $\square$

As a corollary of Lemma 1.2.5 we see in particular that for a Markov pre-process the functions $h^n(u,v)$ are monotone increasing in u and monotone decreasing in v. Since $\upsilon^w_{v,m}\le 1$ and

$$[\mathbf{eq01}]\sum_{m\ge 0}\phi_{u,v}(n,m)=\upsilon^v_{u,n} \tag{67}$$

we also see that for a Markov pre-process the functions $\upsilon^v_{u,n}$ are monotone decreasing in v.

Remark 4.3.4 We will see from examples below (**??**) that there are Markov pre-processes on $\mathcal{H}[s,t]$ such that $\upsilon^v_{u,n}$ are not monotone in u.

Lemma 4.3.5 *[**ob01**] Let μ^*_* be a Markov pre-process. Then for any $m,n\ge 0$ and any $u\le v<w$ in $[s,t]$ the function $\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w)$ is monotone increasing in ϵ and one has*

$$\lim_{\epsilon>0,\epsilon\to 0}\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w)=\phi_{u,v}(m,n)h^n(v,w)$$

Proof: Applying Lemma 1.2.1 to $U_1=H[u,v+\epsilon]_{m,n}$ and $U_2=\Delta^{n[1]}_{v+\epsilon,w}$ we get

$$\phi_{u,v+\epsilon}(m,n)h^n(v+\epsilon,w)=\mu^w_{u,m}(res^{-1}_{v+\epsilon,w}(\Delta^{n[1]}_{v+\epsilon,w}))$$

and since

$$\cap_{\epsilon\to 0}(res^{-1}_{v+\epsilon,w}(\Delta^{n[1]}_{v+\epsilon,w}))=res^{-1}_{v,w}(\Delta^{n[1]}_{v,w})$$

our claim follows. $\square$

Definition 4.3.6 *[**rcont**] A pre-process μ^*_* is called non-degenerate if $\upsilon^u_{u,k}=1$ for all u,k. It is called right continuous if for any $u\in[s,t]$ and any k, $\upsilon^v_{u,k}$ is a right continuous function in v from $[s,v]$ to $[0,1]$.*

If μ is non-degenate then $h^n(u,u)=1$ for all n and u. Note that any process on $\mathcal{H}[s,t]$ is automatically non-degenerate and right continuous.

Remark 4.3.7 For a Markov pre-process one has $(\upsilon^u_{u,k})^2=\upsilon^u_{u,k}$ and therefore a Markov pre-process is non-degenerate if and only if $\upsilon^u_{u,k}\neq 0$ for all u,k.

Theorem 4.3.8 *[**th1**] Let μ^*_* be a non-degenerate Markov pre-process on $\mathcal{H}[s,t]$. Then the following conditions are equivalent:*

1. *for all $n\ge 0$ functions $\upsilon^v_{u,n}$ are right continuous in u and if $u<t$ then there exits $w>u$ such that $\upsilon^w_{u,n}\neq 0$,*
2. *for all $n\ge 0$ functions $h^n(u,v)$ are right continuous in u and if $u<t$ then there exits $w>u$ such that $\upsilon^w_{u,n}\neq 0$,*

3. *for all $n \geq 0$ functions $\phi_u^v(n,m)$ are right continuous in u and if $u < t$ then there exits $w > u$ such that $\upsilon_{u,n}^w \neq 0$,*

4. *for all $n \geq 0$ functions $\upsilon_{u,n}^v$ are right continuous in v,*

5. *for all $n \geq 0$ functions $h^n(u,v)$ are right continuous in v,*

6. *for all $n \geq 0$ functions $\phi_u^v(n,m)$ are right continuous in v.*

Proof: Observe first that if for all $u < t$ then there exits $v > u$ such that $\upsilon_{u,n}^w \neq 0$ then, since $\upsilon_{u,n}^v$ are monotone decreasing in v we have $\upsilon_{u,n}^v \neq 0$ for all $u \leq v \leq w$.

Let u and w be as above. Taking the sum over m in Lemma 4.3.1 and setting $v = u$ we get

$$[\mathbf{feqp}] \lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon)\upsilon_{u+\epsilon,n}^w = \upsilon_{u,n}^w \tag{68}$$

which implies that there exists $\epsilon > 0$ such that $h^n(u,u+\epsilon) \neq 0$. Without loss of generality we may assume that $u + \epsilon = w$.

$(1) \Rightarrow (2),(5)$ When $\upsilon_{u,n}^v$ is right continuous in u equation (68) implies that

$$(\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon))\upsilon_{u,n}^w = \upsilon_{u,n}^w$$

and since $\upsilon_{u,n}^w \neq 0$ we conclude that

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon) = 1$$

Together with Lemma 1.2.5 we conclude that (2) and (5) hold.

$(2) \Rightarrow (5)$ Immediate from Lemma 1.2.5 since for all u there exists w such that $h^n(u,w) \neq 0$.

$(5) \Rightarrow (3)$ Since $h^n(u,u) = 1$ condition (5) also implies that for any u there exists $w > u$ satisfying $h^n(u,w) \neq 0$. Since $\upsilon_{u,n}^w \geq h^n(u,w)$ we conclude that $\upsilon_{n,u}^w \neq 0$.

Taking in Lemma 4.3.5 $v = u$ we get

$$\lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon)\phi_{u+\epsilon}^w(n,m) = \phi_u^w(n,m)$$

for all $w > u$ and using condition (5) we get that $\phi_u^w(n,m)$ is right continuous in u.

$(2) \Rightarrow (6)$ We need to show that

$$[\mathbf{seqp}] \lim_{\epsilon>0,\epsilon\to 0} \phi_u^{v+\epsilon}(m,n) = \phi_u^v(m,n) \tag{69}$$

Let w be such that $h^n(v,w) \neq 0$. Then Lemma 4.3.5 together with the right continuity of $h^n(-,-)$ in the first variable implies (69).

$(6) \Rightarrow (4)$ Immediately follows from the fact that $\upsilon_{u,n}^v = \sum_m \phi_u^v(n,m)$.

(4) $\Rightarrow$ (2) Since functions $v^v_{u,n}$ are right continuous in v and $v^u_{u,k}=1$ there exists $w>u$ such that $v^w_{u,n}\neq 0$ and as explained above such that $h^n(u,w)\neq 0$. Taking in Lemma 4.3.5 $m\neq n$ and $v=u$ we get

$$[\mathbf{eq020}]\lim_{\epsilon\to 0}\phi_{u,u+\epsilon}(m,n)=0 \tag{70}$$

Therefore we have

$$[\mathbf{teqp}]1=\lim_{\epsilon\to 0}v^{u+\epsilon}_{u,n}=\lim_{\epsilon\to 0}\sum_m\phi_{u,u+\epsilon}(n,m)=\lim_{\epsilon\to 0}\phi_{u,u+\epsilon}(n,n) \tag{71}$$

Form Lemma 4.3.5 for $m=n$ and $v=u$ we get for all $w>u$

$$\lim_{\epsilon>0,\epsilon\to 0}\phi_{u,u+\epsilon}(n,n)h^n(u+\epsilon,w)=h^n(u,w)$$

which together with (71) implies that $h^n(u,v)$ is right continuous in u.

(3) $\Rightarrow$ (1) Immediately follows from the fact that $v^v_{u,n}=\sum_m\phi^v_u(n,m)$.

Theorem is proved. $\Box$

For a pre-process μ^*_* define $E_{n,\mu}\subset[s,t]$ by the rule $x\in E_{n,\mu}$ if and only if $e=s$ or for all sufficiently small $\epsilon>0$ one has $h^n(x-\epsilon,x)=0$. When no confusion is possible we will write E_n instead of $E_{n,\mu}$.

Lemma 4.3.9 *[***ob2***] Let μ be a non-degenerate right continuous Markov pre-process. Then for any $e\in E_n$ such that $h^n(e,t)=0$ there exists a unique $e_{+1}>e$ in E_n such that for all $x\in[e,e_{+1})]$ one has $h^n(e,x)\neq 0$.*

Proof: By Theorem 4.3.8 the function $h^n(e,-)$ is right continuous and therefore the set of zeros of $h^n(e,-)$ is of the form $[e_{+1},t]$ for some e_{+1} in $(e,t]$. For $\epsilon<e_{+1}-e$ we have $0=h(e,e_{+1}+\epsilon)=h(e,e_{+1}-\epsilon)h(e_{+1}-\epsilon,e_{+1})$ and since $h(e,e_{+1}-\epsilon)\neq 0$ we conclude that $e_{+1}\in E_n$. $\Box$ Note that if $E_n\neq\emptyset$ then there exists a unique $e\in E_n$ such that $h^n(e,t)\neq 0$. For this e we set $e_{+1}=t$.

Lemma 4.3.10 *[***ob3***] For a a non-degenerate right continuous Markov pre-process μ the sets E_n are countable.*

Proof: We have

$$[\mathbf{ecov}][s,t)=\amalg_{e\in E_n}[e,e_{+1}) \tag{72}$$

and since the sum of an uncountable number of non-zero numbers is infinite we conclude that E_n is countable. $\Box$

Since for an additive process $E_n=E_m$ for all $m,n\neq 0$ we will write $E=E(\mu)$ for this set in the additive context.

Example 4.3.11 [**nonrc**] Consider a pre-process μ on $\mathcal{H}[s,t]$ such that the measures $\mu^v_{u,k}$ are concentrated on $\Delta^{k[1]}_{u,v}$. Such a pre-process is simply a collection of functions $\upsilon^v_{u,*}$ on $\mathbf{N}$. It is additive if and only if $\upsilon^v_{u,k} = (\upsilon^v_{u,1})^k$ and it is Markov if and only if $\upsilon^v_{u,k}\upsilon^w_{v,k} = \upsilon^w_{u,k}$.

Set $\upsilon^u_{u,n} = 1$, $\upsilon^v_{u,0} = 1$ and $\upsilon^v_{u,n} = 0$ for $v > u$ and $n > 0$. This gives us an example of a non-degenerate, additive Markov pre-process such that the functions $\upsilon^v_{u,n}$ are right continuous in u but not in v.

Let $x \in (s,t)$ and set $\upsilon^v_{u,0} = 1$ and for $n > 0$, $\upsilon^v_{u,n} = 0$ if $u \leq x$ and $\upsilon^v_{u,n} = 1$ if $u > x$. This defines a degenerate additive, Markov pre-process for which functions $\upsilon^v_{u,n}$ are right continuous in v but not in u.

5 Summary

The general set-up for a problem of historical inference can be described as follows. First we should specify the time period $[t_0, t_1]$ which we are interested in. It must include both the times which we want to learn about and the time or times at which the measurements are made. Next we need to define the set $H[t_0, t_1]$ of possible histories among which we will be looking for the best fit to our data and a mapping from $H[t_0, t_1]$ to the set D of possible data points which will be used as the starting point of our inference. We also have to choose a function or functions from $H[t_0, t_1]$ to some set X whose values are the properties of histories which we are actually interested in, such as the population size at some point $t \in [t_0, t_1]$. The next choice we have to make is that of a forward historical model which describes how, for given values of historical parameters, an initial state of a population will develop into a history. Finally, we should choose the sets I and P of reasonable values of initial states and historical parameters:

$$\begin{array}{ccccc} I \times P & \longrightarrow & H[t_0, t_1] & \longrightarrow & D \\ & & \downarrow & & \\ & & X & & \end{array}$$

Let us outline now the form of different components of this set-up in the case of one of the most basic inference problems - to infer the size of a haploid population at time $t = -T$ based on the genetic distances between its members at the present time $t = 0$. The time period to choose is obviously $[-T, 0]$. The set $H[-T, 0]$ or more generally the set $H[u, v]$ of histories of non-structured haploid populations is the set of ... Check what is the biologists terminology for this.

Let us consider the case of a singleton population history on time interval $[t_0, t_1]$ under the assumption that the population size $N(t)$ remains "very large" for all $t \in [t_0, t_1]$. Suppose that the population develops under some birth-and-death dynamical model which is determined by birth rate function $b(t)$ and death rate function $d(t)$ i.e. during the time interval $[t, t + \Delta t]$ there occur approximately $d(t)N(t)\Delta(t)$ death events and approximately $b(t)N(t)\Delta(t)$ birth or division events and the probability for a given individual to undergo a death or division event does not depend on its previous history.

Let us say that an individual which is alive at time t is "lucky" if it has a living descendant at time t_1. Let $\tilde{N}(t)$ be the number of lucky individuals at time t. The ratio

$$\sigma(t) = \tilde{N}(t)/N(t)$$

will be called the survival probability (from t to t_1). Consider the sub-population of "lucky" individuals. The only events which occur in this sub-population are division events. The rate at which these events occur may be computed as follows. During the interval $[t, t+\Delta t]$ there is a total of (approximately) $C = b(t)N(t)\Delta(t)$ division events. Let C_0 be the number of division events in which both "children" are unlucky, C_1 the number of division events in which one child is "lucky" and another one "unlucky" and C_2 the number of events in which both children are "lucky". We have $C = C_0 + C_1 + C_2$ and, approximately,

$$C_0 = \left(\frac{N(t+\Delta t) - \tilde{N}(t+\Delta t)}{N(t+\Delta t)}\right)^2 \cdot C, \qquad C_2 = \left(\frac{\tilde{N}(t+\Delta t)}{N(t+\Delta t)}\right)^2 \cdot C$$

The number of division events in the lucky sub-population is C_2 and assuming that $(\tilde{N}(t+\Delta t) - \tilde{N}(t))\Delta t$ is small and the same applies to $N(-)$ we conclude that the probability for a lucky individual to divide into two lucky ones during the time interval $[t, t+\Delta t]$ equals

$$C_2/\tilde{N}(t) = b(t)\sigma(t)\Delta t$$

i.e. the sub-population of lucky individuals develops under the birth dynamic model with the birth rate $\tilde{b}(t) = b(t)\sigma(t)$. In particular we should approximately have:

$$[\mathbf{nlarge1}]\tilde{N}(t) = \tilde{N}(t_0)\, e^{\int_{t_0}^{t} \sigma(x)b(x)dx}, \quad \sigma(t) = \frac{\tilde{N}'(t)}{\tilde{N}(t)b(t)} \tag{73}$$

and therefore

$$[\mathbf{nlarge2}]N(t) = \frac{b(t)\tilde{N}(t)^2}{\tilde{N}'(t)} \tag{74}$$

which gives us a direct formula for the reconstruction of the total population size in terms of the size of the lucky sub-population when the total number of individuals is very large and the birth rate is known. The size of lucky-subpopulation at time t may be further interpreted as the number of equivalence classes in the population at time t_1 with respect to the equivalence relation in which two individuals are "equivalent" if their most recent common ancestor lived no earlier than $t_1 - t$ time units ago.

Below we construct an algorithm which gives the best possible reconstruction of the survival probability function from the ancestral genealogy of individuals alive at time t_1 in the case of finite population size. Since the birth rate in the original population plays no role other than time scaling parameter we assume $b(t) = 1$.

Let us consider a particular generic (see **??**) ancestral genealogy $\tilde{\Gamma}$ on the time interval $[t_0, t_{q+1}]$ with $\tilde{n}$ members at time t_0 and times of division events being $t_1, \ldots, t_q$. Suppose first that there are so many events that we may try to approach the problem by taking a C_1-approximation $\tilde{N}(t)$ of the function $\tilde{n}$ defined by the condition that for $t \in [t_i, t_{i+1})$ one has $\tilde{n}(t) = \tilde{n} + i$ and then applying to it formulas (73), (74) to recover $\sigma(t)$ and $N(t)$.

In the simplest approach one may take a pice-wise linear approximation of $\tilde{n}$ defined by the condition $\tilde{N}(t_i) = \tilde{n} + i - 1$. Then $\tilde{N}'$ is a pice-wise constant function defined outside of points t_i such that

$$\tilde{N}'(t) = 1/(t_{i+1} - t_i)$$

for $t \in (t_i, t_{i+1})$. Therefore we would get

$$\sigma(t) = 1/((\tilde{n} + i)(t_{i+1} - t_i))$$

which for almost all $\tilde{\Gamma}$ will have many values greater than 1.

Note: In the most simple demographic models the dynamics of a population is described by the "grows rate" g which is actually the difference between the birth rate and the death rate $g = b - d$. This difference fully determines the change in average population size according to the formula $N'/N = g$. Note however that g does not determine the dynamics of the "lucky" sub-population i.e. for (b, d), (b', d') such that $b - d = b' - d'$ the dynamics of the lucky sub-populations may be different. The most elementary example is given by the pairs $b = 0, d = 0$ and $b = 1, d = 1$. In the first case $N = \tilde{N}$ while in the second $\tilde{N} = n/(t + 1)$ where t is the time to the present.

Note: It might be better to explain the previous arguments starting with the idea of approximating the continuous time by a given sequence of discrete time moments ("generations") τ_j.

Note: The formulas derived above for any birth rate $b(t)$ and death rate $d(t)$ lead to the following:

$$[\textbf{master}]\sigma' = \sigma(b\sigma - b + d), \qquad \sigma(t_{now}) = 1 \tag{75}$$

and

$$[\textbf{relationseq}]\sigma = \tilde{N}/N, \qquad \tilde{N}'/\tilde{N} = b\sigma, \qquad N'/N = b - d \tag{76}$$

From (76) we get

$$\tilde{N}(t) = \tilde{N}(t_{now})e^{-\int_t^{t_{now}} b(\tau)\sigma(\tau)d\tau}$$

$$N(t) = N(t_{now})e^{-\int_t^{t_{now}} (b(\tau)-d(\tau))d\tau}$$

where $N(t_{now}) = \tilde{N}(t_{now})$ is the number of present day members.

If σ is a solution of (75) then one has:

$$\sigma = \frac{e^{\int_t^{t_{now}} (b(\tau)-d(\tau))d\tau}}{1 + \int_t^{t_{now}} b(x)e^{\int_x^{t_{now}} (b(\tau)-d(\tau))d\tau} dx}$$

and

$$1/\tilde{N}(t) = \frac{1}{\tilde{N}(t_{now})} \cdot e^{\int_t^{t_{now}} b(\tau)\sigma(\tau)d\tau} = \frac{1}{\tilde{N}(t_{now})} \cdot (1 + \int_t^{t_{now}} b(x)e^{\int_x^{t_{now}} (b(\tau)-d(\tau))d\tau} dx)$$

Consider $(b_1, d_1, N_{1,now})$ and $(b_2, d_2, N_{2,now})$. Then $\tilde{N}_1(t) = \tilde{N}_2(t)$ if and only if $N_{1,now} = N_{2,now}$ and

$$b_1(t)e^{\int_t^{t_{now}} (b_1(\tau)-d_1(\tau))d\tau} = b_2(t)e^{\int_t^{t_{now}} (b_2(\tau)-d_2(\tau))d\tau} \; (mod \;\; dt)$$

or equivalently

$$b_1(t)/N_1(t) = b_2(t)/N_2(t) \ (mod \ \ dt)$$

If we are given $\tilde{N}(t)$ and $d(t)$ then we can reconstruct $b(t)$ as follows. We have

$$\tilde{b} = \tilde{N}'/\tilde{N} = b\sigma$$

From (75) we get

$$(\tilde{b}/b)' = (\tilde{b}/b)(\tilde{b} - b + d)$$

or

$$[\mathbf{relation2}]\tilde{b}'b - \tilde{b}b' + \tilde{b}b^2 - \tilde{b}^2 b - \tilde{b}db = 0 \tag{77}$$

Solving it in b we get

$$b' = b(b + \tilde{b}'/\tilde{b} - \tilde{b} - d)$$

or

$$b(t) = b(t_{now})\frac{e^{\int_t^{t_{now}}(d+\tilde{b}-\tilde{b}'/\tilde{b})d\tau}}{1 + \int_t^{t_{now}} e^{\int_x^{t_{now}}(d+\tilde{b}-\tilde{b}'/\tilde{b})d\tau}dx}$$

Since $\tilde{b} = dLog(\tilde{N})/dt$ and $\tilde{b}'/\tilde{b} = dLog(\tilde{b})/dt$ we further have

$$e^{\int_t^{t_{now}}(d+\tilde{b}-\tilde{b}'/\tilde{b})d\tau} = \frac{\tilde{N}(t_{now})}{\tilde{N}(t)}\frac{\tilde{b}(t)}{\tilde{b}(t_{now})}e^{\int_t^{t_{now}} d(\tau)d\tau} = \frac{(1/\tilde{N})'(t)}{(1/\tilde{N})'(t_{now})} \cdot e^{\int_t^{t_{now}} d(\tau)d\tau}$$

If we are given $\tilde{N}(t)$ and $b(t)$ then we can reconstruct $d(t)$ from (77) as

$$d = \tilde{b}'/\tilde{b} - \tilde{b} + b - b'/b = \tilde{N}''/\tilde{N}' - 2\tilde{N}'/\tilde{N} + b - b'/b$$

We see that the reconstruction of d involves the second derivative of $\tilde{N}$ while reconstruction of b only the first derivative.

In addition we get:

$$b - d = \frac{dlog}{dt}(b + 2\tilde{N} - \tilde{N}')$$

and correspondingly

$$N(t) = N_{now}e^{-\int_t^{t_{now}}(b-d)d\tau} = b(t)\frac{\tilde{N}^2(t)}{\tilde{N}'(t)}$$

which can also be seen directly from (76).

Let $\mu_{n,t_0}^{t_1}(b,d)$ be the probability measure on $H[t_0,t_1]_{n,*}$ corresponding to the strictly continuous birth and death process with n initial population members, birth rate b and death rate d. Let $r : H[t_0,t_1] \to \tilde{H}[t_0,t_1]$ be the map which assigns to a descending genealogy the ancestral genealogy of members alive at time t_1.

Proposition 1. Let

$$\sigma(t) = \mu_{1,t}^{t_1}(b,d)(H[t,t_1]_{1,1})$$

be the non-extinction probability function associated with a strictly continuous birth and death process with parameters b and d. Then σ' exists everywhere and satisfies the equation

$$\sigma' = \sigma(b\sigma - b + d)$$

Theorem 2. One has

$$r_*(\mu^{t_1}_{n,t_0}(b,d))_{|\tilde{H}[t_0,t_1]_{\tilde{n},*}} = C(n,\tilde{n})a_0^{\tilde{n}}(1-a_0)^{n-\tilde{n}}\mu^{t_1}_{\tilde{n},t_0}(b\sigma,0)$$

For an ancestral genealogy $\tilde{\Gamma}$ let $n(\tilde{\Gamma})$ be the number of connected components of $\tilde{\Gamma}$ and $q(\Gamma)$ the number of coalescent events in $\tilde{\Gamma}$. Let further $x_1,\dots,x_q \in [t_0,t_1]$ be the times of coalescent events in the increasing order. An ancestral genealogy $\tilde{\Gamma}$ is called semi-generic if all its coalescent events are of multiplicity two and the times of distinct coalescent events are distinct.

Let $J_i = [\tau_i^-,\tau_i^+]$, $i = 1,\dots,q$ be a sequence of subintervals of the interval $[t_0,t_1]$ such that $\tau_i^+ < \tau_{i+1}^-$. Denote by $U(\tilde{n},q,J)$ the set of all ancestral genealogies $\tilde{\Gamma}$ such that $n(\tilde{\Gamma}) = \tilde{n}$, $q(\tilde{\Gamma}) = q$ and $x_i(\tilde{\Gamma}) \in J_i$.

Let further $I_0 = [t_0,\tau_1^-]$; $I_i = [\tau_i^+,\tau_{i+1}^-]$, $i = 1,\dots,q-1$; $I_q = [\tau_q^+,t_1]$; be the components of the complement $[t_0,t_1]\backslash J$.

Proposition 3. One has

$$\mu^{t_1}_{\tilde{n},t_0}(\tilde{b},0)(U(\tilde{n},q,J)) = \prod_{i=0}^{q} e^{-(\tilde{n}+i)B_i} \cdot \prod_{i=1}^{q}(\tilde{n}+i-1)e^{-(\tilde{n}+i)C_i}(1-e^{-C_i})$$

where $B_i = \int_{x\in I_i}\tilde{b}(x)dx$ and $C_i = \int_{x\in J_i}\tilde{b}(x)dx$.

A genealogy is completely determined by its combinatorial type π and the times of events $x_1,\dots,x_q$. The set of all genealogies with a given combinatorial type is the open simplex

$$\Delta^{q(\pi)}_{(t_0,t_1)} = \{x_1,\dots,x_q \,|\, t_0 < x_1 < \dots < x_q < t_1\}$$

For Γ with times of events $x_i(\Gamma)$ and

$$\epsilon < \frac{1}{2}min\{\Delta x_i(\Gamma),\ i = 0,\dots,q+1\}$$

where

$$\Delta x_0 = x_1 - t_0;\ \Delta x_i = x_{i+1} - x_i,\ i = 1,\dots,q-1;\ \Delta x_{q+1} = t_1 - x_q;$$

Let

$$U_\epsilon(\Gamma) = \{\Gamma' \,|\, \pi(\Gamma') = \pi(\Gamma) \text{ and } |x_i(\Gamma)' - x_i(\Gamma)| < \epsilon \text{ for } i = 1,\dots,q\}$$

Proposition 4 Let $\tilde{\Gamma}$ a semi-generic ancestral genealogy with $n(\tilde{\Gamma}) = \tilde{n}$, $q(\tilde{\Gamma}) = q$ and $x_i(\tilde{\Gamma}) = x_i$. Let further $\tilde{b}$ be the birth rate of a continuous birth process. Then

$$\mu^{t_1}_{\tilde{n},t_0}(\tilde{b},0)(U_\epsilon) = c(\pi(\Gamma)) \cdot \mu^{t_1}_{\tilde{n},t_0}(\tilde{b},0)(U(\tilde{n},q,J(\{x_i\},\epsilon)))$$

where $J(\{x_i\},\epsilon)_i = \{x \,|\, |x - x_i| < \epsilon\}$ and c is a constant which does not depend on $\tilde{b}$.

Note: We can modify the "master equation" introducing bias in the form of "having a lucky sibling makes one more (less) likely to be lucky" which will change the formulas for C_0 and C_2 given above. However, it is unclear whether or not our main theorem which asserts that all of the information about b and d contained in $\tilde{\Gamma}$ is contained in $\tilde{n}$ and the sequence $t_1, \ldots, t_q$. will still apply in this more general situation.

It might be very interesting to see how one can guess the value of such a bias coefficient from $\tilde{\Gamma}$ and then apply it to a real population (e.g. human).

1. [**sum0**] A normalized birth and death process $\mu^*_{*,*}$ over a time interval $[t_0, t_1]$ is determined by its survival probability function

$$\sigma(t) = \left\{ \begin{array}{c} \text{the probability that a population member alive at time } t \in [t_0, t_1] \\ \text{will have a living descendant at time } t_1. \end{array} \right\}$$

   The complement $1 - \sigma(t)$ to $\sigma(t)$ is the extinction probability function given by the transition probability $1 \mapsto 0$ over $[t, t_1]$ for the process μ.

2. [**sum1**] A function $[t_0, t_1] \to [0, 1]$ is called a normalized survival function if it is the survival probability function for a normalized birth and death process over $[t_0, t_1]$. We let $\bar{A}[t_0, t_1]$ denote the set of all normalized survival functions.

3. [**sum2**]Normalized survival functions have the following properties:

   (a) [**sum2a**] Normalized survival functions are right continuous.

   (b) [**sum2b**] Let σ_0 be a normalized survival function on $[t_0, t_1]$ and σ_1 a normalized survival function on $[t_1, t_2]$ such that $\sigma_0(t_1) = \sigma_1(t_1)$. Then the function σ which equals σ_0 on $[t_0, t_1]$ and σ_1 on $[t_1, t_2]$ is a normalized survival function.

   (c) [**sum2c**] Let $a_0, a_1 \in [0, 1]$. A normalized survival function on $[t_0, t_1]$ satisfying $a_0 = \sigma(t_0)$ and $a_1 = \sigma(t_1)$ exists if and only if the following two equivalent inequalities hold:

$$[\textbf{sum2ceq}] a_1 \geq \frac{a_0}{(1 - e^{t_1 - t_0})a_0 + e^{t_1 - t_0}} \qquad a_0 \leq \frac{a_1}{(1 - e^{t_0 - t_1})a_1 + e^{t_0 - t_1}} \tag{78}$$

   If a_1, a_0 satisfy the corresponding equalities then such a function σ is unique and is given by

$$\sigma(t) = \frac{a_0}{(1 - e^{t - t_0})a_0 + e^{t - t_0}} = \frac{a_1}{(1 - e^{t - t_1})a_1 + e^{t - t_1}}$$

   (d) [**sum2d**] Let $[t_0, t_1]$ be a time interval and a_0, a_1 be two numbers such that inequalities (78) are satisfied. Then a normalized survival function σ satisfying $\sigma(t_0) = a_0$, $\sigma(t_1) = a_1$ and

$$\int_{t_0}^{t_1} \sigma(t) dt = A$$

   exists if and only if

$$(1 + a_1(e^{t_1 - t_0} - 1))^{-1} < e^{-A} \leq 1 + a_0(e^{t_0 - t_1} - 1)$$

   or $(1 + a_1(e^{t_1 - t_0} - 1))^{-1} = e^{-A}$ and equalities hold in (78).

4. [**sum3**] Let

$$\tilde{\Gamma} = \{q, V_0 \leftarrow \cdots \leftarrow V_{q+1}; t_1, \ldots, t_q\}$$

be a generic ancestral (i.e. death free) genealogy over $[-T,0]$. Let $\epsilon > 0$ be a number such that $(-T, t_1-\epsilon, t_1+\epsilon, t_2-\epsilon, \dots, t_q-\epsilon, t_q+\epsilon, 0)$ is an increasing sequence i.e. such that

$$\epsilon < t_1 - (-T); \quad \epsilon < t_{i+1} - t_i, \;\; i = 1, \dots, q; \quad \epsilon < 0 - t_q;$$

Denote by $U_\epsilon(\tilde{\Gamma})$ the ϵ-neighborhood of $\tilde{\Gamma}$ in $\tilde{H}[-T,0]$ i.e. the set of genealogies of the form $\{q, V_0 \leftarrow \dots \leftarrow V_{q+1}; x_1, \dots, x_q\}$ where $x_i \in [t_i - \epsilon, t_i + \epsilon]$.

5. [**sum4**] Let $r : H[-T,0] \to \tilde{H}[-T,0]$ be the map which sends a singleton population history into the ancestral genealogy of members which are alive at time 0.

6. [**sum5**] Let μ be a normalized birth and death process over $[-T,0]$ with the survival probability function σ. Let $\tilde{\Gamma}$ and ϵ be as above. Denote by I_i and J_i the intervals

$$I_0 = [-T, t_1 - \epsilon); \quad I_i = [t_i + \epsilon, t_{i+1} - \epsilon), \;\; i = 1, \dots, q-1; \quad I_q = [t_q + \epsilon, 0);$$

and

$$J_i = [t_i - \epsilon, t_i + \epsilon)$$

Let $\tilde{n} = \#V_0$ be the number of population members in $\tilde{\Gamma}$ at time $-T$ and n be any integer $\geq \tilde{n}$. Then one has:

$$\mu^0_{-T,n}(r^{-1}(U_\epsilon(\tilde{\Gamma}))) =$$

$$[\mathbf{sum5eq}] = c(\tilde{\Gamma})\, C(n,\tilde{n})\, (1-\sigma(-T))^{n-\tilde{n}}\, \sigma(-T)^{\tilde{n}} \prod_{i=0}^{q} e^{-(\tilde{n}+i)B_i} \prod_{i=1}^{q} e^{-(\tilde{n}+i-1)C_i}(1-e^{-C_i}) \quad (79)$$

where $c(\tilde{\Gamma})$ is a coefficient which does not depend on σ, T or n, $C(n,\tilde{n})$ is the binomial coefficient and B_i, C_i are given by:

$$B_i = \int_{I_i} \sigma(t)dt \quad C_i = \int_{J_i} \sigma(t)dt$$

If we sum up over all $\tilde{\Gamma}$ with given $\tilde{n}$ and q then the coefficient $c(\tilde{\Gamma})$ becomes

$$c(\tilde{n}, q) = \tilde{n}(\tilde{n}+1)\dots(\tilde{n}+q-1)$$

7. [**sum5.1**] Let $S_{[t_0,t_1]}$ be the set of all normalized survival probability functions on $[t_0,t_1]$. Our "optimization" problem can be formulated as following. Consider (79) as a function $P(\tilde{\Gamma}, \epsilon)$ on the space

$$X = \mathbf{N} \times \amalg_{t_0 \in [-\infty, 0]} S_{[t_0, 0]}$$

which we set to be 0 if $t_0 \geq t_1 - \epsilon$ or $n < \tilde{n}$. Let

$$p(\tilde{\Gamma}, \epsilon) = sup_{(n,t_0,\sigma) \in X}\{P(\tilde{\Gamma}, \epsilon)(n, t_0, \sigma)\}$$

be the maximal possible value of this function and for $\delta > 0$ let

$$V_\delta(\tilde{\Gamma}, \epsilon) = P(\tilde{\Gamma}, \epsilon)^{-1}([p(\tilde{\Gamma}, \epsilon) - \delta, p(\tilde{\Gamma}, \epsilon)]) \subset X$$

We are interested in the subset

$$V(\tilde{\Gamma}) = \{x \in X \,|\, \forall \delta > 0 \exists \epsilon > 0 \,:\, \forall \epsilon' < \epsilon (x \in V_\delta(\tilde{\Gamma}, \epsilon)\}$$

8. [**sum6**] Let $S(a_0,\ldots,a_q)$ be the set of normalized survival functions σ on $[-T,0]$ such that $a_0=\sigma(-T)$ and $a_i=\sigma(t_i-\epsilon)$ for $i=1,\ldots,q$. We further denote $t_0=-T$ and $t_{q+1}=0$. Then one has:

   (a) [**sum6a**] It follows immediately from 3b and 3c that the set $S(a_0,\ldots,a_q)$ is non-empty if and only if

   $$[\textbf{sum6aeq}]a_1 \geq \frac{a_0}{(1-e^{(t_1-\epsilon-t_0)})a_0+e^{(t_1-\epsilon-t_0)}} \tag{80}$$

   and

   $$[\textbf{sum6aeq}]a_{i+1} \geq \frac{a_i}{(1-e^{(t_{i+1}-t_i)})a_i+e^{(t_{i+1}-t_i)}} \tag{81}$$

   for $i=1,\ldots,q-1$.

   (b) [**sum6b**] If (81) hold and

   $$\epsilon < \frac{1}{2}\log(1+\frac{1}{\tilde{n}+q-1})$$

   then there exists a unique function σ in $S(a_0,\ldots,a_q)$ which maximizes the expression (79). This function is given by

   $$[\textbf{sum6beq}]\sigma(t)=\begin{cases}\frac{a_0}{(1-e^{t-t_0})a_0+e^{t-t_0}} & \text{for } t\in[t_0,t_1-\epsilon)\\ \\ \frac{a_i}{(1-e^{t-(t_i-\epsilon)})a_i+e^{t-(t_i-\epsilon)}} & \text{for } t\in[t_i-\epsilon,t_{i+1}-\epsilon) \quad i=1,\ldots,q-1\\ \\ \frac{a_q}{(1-e^{t-(t_q-\epsilon)})a_q+e^{t-(t_q-\epsilon)}} & \text{for } t\in[t_q-\epsilon,t_{q+1})\end{cases} \tag{82}$$

   (c) [**sum6c**] The values of B_i and C_i for (82) are

   $$e^{-B_0}=(e^{(t_0-(t_1-\epsilon))}-1)a_0+1$$

   for $i=1,\ldots,q-1$

   $$e^{-B_i}=\frac{(e^{(t_i-t_{i+1})}-1)a_i+1}{(e^{-2\epsilon}-1)a_i+1}$$

   and

   $$e^{-B_q}=\frac{(e^{(t_q-t_{q+1}-\epsilon)}-1)a_q+1}{(e^{-2\epsilon}-1)a_q+1}$$

   and

   $$e^{-C_i}=(e^{-2\epsilon}-1)a_i+1$$

   and the value of (79) is

   $$[\textbf{sum6ceq}]c(\tilde{\Gamma})\,C(n,\tilde{n})\prod_{i=0}^{q}f_i(a_i) \tag{83}$$

   where

   $$f_0(a_0)=a_0^{\tilde{n}}\,((e^{t_0-(t_1-\epsilon)}-1)a_0+1)^{\tilde{n}}\,(-a_0+1)^{n-\tilde{n}}$$

   for $i=1,\ldots,q-1$,

   $$f_i(a_i)=(1-e^{-2\epsilon})a_i\,((e^{(t_i-t_{i+1})}-1)a_i+1)^{\tilde{n}+i}\,((e^{-2\epsilon}-1)a_i+1)^{-1}$$

   and

   $$f_q(a_q)=(1-e^{-2\epsilon})a_q\,((e^{(t_q-t_{q+1}-\epsilon)}-1)a_q+1)^{\tilde{n}+q}\,((e^{-2\epsilon}-1)a_q+1)^{-1}$$

9. Let $M(x)$ be the matrix

$$M_x = \begin{pmatrix} 1 & 0 \\ 1-e^x & e^x \end{pmatrix}$$

Then $det(M_x) = e^x > 0$, $M_x M_y = M_{x+y}$ and $M_x^{-1} = M_{-x}$. Below we consider M_x as a Moebius transformation.

10. [**sum7**] The previous computations show that we need to find the maximum of (83) on the domain defined by the inequalities $0 \leq a_0$, $a_q \leq 1$ and inequalities (81) which are of the form

$$a_1 \geq M_{t_1-\epsilon-t_0}(a_0)$$

and for $i = 1, \ldots, q-1$

$$a_{i+1} \geq M_{t_{i+1}-t_i}(a_i)$$

Set

$$M_0 = M_{t_1-\epsilon-t_0}, \qquad M_i = M_{t_{i+1}-t_i}, \quad i = 1, \ldots, q-1;$$

Consider new variables

$$z_0 = a_0$$

and for $i = 1, \ldots, q$

$$z_i = M_0^{-1} M_1^{-1} \ldots M_{i-1}^{-1}(a_i) = M_{t_0+\epsilon-t_i}(a_i)$$

or equivalently

$$a_i = M_{t_i-t_0-\epsilon}(z_i)$$

Then the inequalities take the form

$$0 \leq z_0 \leq \cdots \leq z_q \leq 1$$

and the expression (83) takes the form

$$c(\tilde{\Gamma})\, C(n, \tilde{n})\, (1-e^{-2\epsilon})^q \prod_{i=0}^{q} g_i(z_i)$$

where

$$g_0(z_0) = z_0^{\tilde{n}}((e^{t_0-(t_1-\epsilon)}-1)z_0+1)^{\tilde{n}}(-z_0+1)^{n-\tilde{n}}$$

for $i = 1, \ldots, q-1$

$$g_i(z_i) = e^{t_0+\epsilon-t_i} z_i((e^{t_0+\epsilon-t_{i+1}}-1)z_i+1)^{\tilde{n}+i}((e^{t_0+\epsilon-t_i}-1)z_i+1)^{-(\tilde{n}+i+1)}\cdot$$
$$\cdot\frac{(e^{t_0+\epsilon-t_i}-1)z_i+1}{(e^{t_0-\epsilon-t_i}-1)z_i+1}$$

and

$$g_q(z_q) = e^{t_0+\epsilon-t_q} z_q((e^{t_0-t_{q+1}}-1)z_q+1)^{\tilde{n}+q}((e^{t_0+\epsilon-t_q}-1)z_q+1)^{-(\tilde{n}+q+1)}\cdot$$
$$\cdot\frac{(e^{t_0+\epsilon-t_q}-1)z_q+1}{(e^{t_0-\epsilon-t_q}-1)z_q+1}$$

11. [**sum8**] Let C be a compact space and $f : C \to \mathbf{R}$ be a continuous function which takes its maximal value in a unique point $p \in C$. Let further h be any other continuous function on C. Consider the "deformation" of f of the form $f + \epsilon\, h$. Then for any neighborhood U of p there exists δ such that for all $\epsilon < \delta$ the points where $f + \epsilon\, h$ takes its maximal value are in U. It follows immediately from the fact that h is bounded on C and there is positive number which separates the value of f at p from its values on $C \backslash U$.

12. [**sum9**] Comment (11) shows that we need to consider the case $\epsilon = 0$ and then among the points of maximum of the corresponding normalized function we should choose the ones which have maxima of the functional for small ϵ's in any neighborhood. Set

$$c_i = 1 - e^{t_0 - t_i},\ \ i = 1, \dots, q; \quad\ c_{q+1} = 1 - e^{-T}$$

We have $0 \leq c_1 \leq c_2 \leq \dots \leq c_{q+1} \leq 1$ and

$$g_0(z_0) = z_0^{\tilde{n}}(-z_0 + 1)^{n-\tilde{n}}(-c_1 z_0 + 1)^{\tilde{n}}$$

$$g_i(z_i) = e^{t_0+\epsilon-t_i} z_i(-c_{i+1}z_i + 1)^{\tilde{n}+i}(-c_i z_i + 1)^{-(\tilde{n}+i+1)}$$

and we need to maximize $C(n, \tilde{n}) \prod_{i=0}^{q} g_i(z_i)$ on $\{n \in \mathbf{N} \,|\, n \geq \tilde{n}\} \times \Delta^{q+1}$ where Δ^{q+1} is the standard simplex with coordinates $0 \leq z_0 \leq \dots \leq z_q \leq 1$.

13. [**sum10**] We have:

$$\frac{\partial g_0}{\partial z_0} = ...$$

For $i = 1, \dots, q$ we have:

$$\frac{\partial g_i}{\partial z_i} = e^{t_0+\epsilon-t_i}(-c_{i+1}z_i + 1)^{\tilde{n}+i-1}(-c_i z_i + 1)^{-(\tilde{n}+i+2)}((c_i(\tilde{n} + i) - c_{i+1}(\tilde{n} + i + 1))z_i + 1)$$

In particular the function g_i on $[0, 1]$ is increasing up to its maximum at the point

$$u_i = \min\{1, \frac{1}{c_{i+1}(\tilde{n} + i + 1) - c_i(\tilde{n} + i)}\}$$

and is decreasing after this point.

14. [**sum11**] Let $g_0, \dots, g_q$ be non-negative continuous functions on $[0, 1]$ such that for any $0 \leq i \leq q$, $0 \leq k \leq q - i$ the function

$$g_{i,i+k} = \prod_{j=0}^{k} g_{i+j}$$

is quasi-concave on $[0, 1]$ i.e. there is a point $u_{i,i+k} \in [0, 1]$ such that $g_{i,i+k}$ increases on $[0, u_{i,i+k}]$, reaches maximum at $u_{i,i+k}$ and decreases on $[u_{i,i+k}, 0]$. Then the function $\prod_{i=0}^{q} g_i(z_i)$ has a unique global maximum on the simplex $\{0 \leq z_0 \leq \dots \leq z_q \leq 1\}$.

Let us proceed by induction on q. For $q = 0$ the statement is obvious. Suppose $q > 0$. Let $(z_1, \dots, z_q)$ be the point of maximum for $g_1(z_1) \dots g_q(z_q)$. Consider two cases. If $u_{0,0} < z_1$ then $(u_{0,0}, z_1, \dots, z_q)$ is a point of global maximum for $g_0(z_0) \dots g_q(z_q)$ and since the value at this point is the product of the maximal values of $g_0(z_0)$ and $g_1(z_1) \dots g_q(z_q)$ it is the only point where maximum is reached. If $u_{0,0} \geq z_1$ then for any point of local maximum of $g_0(z_0) \dots g_q(z_q)$ we will have $z_0 = z_1$ and therefore any point of maximum for $g_0(z_0) \dots g_q(z_q)$ will be the point of maximum for $g_{0,1}(z_1) \dots g_q(z_q)$. By induction such a point is unique.

15. [**sum12**] Let us show that the functions g_i of (12) satisfy the conditions of (14).

(a) [**sum12a**] For $i \geq 1$ we have

$$g_{i,i+k}(z) =$$

$$= \left(\prod_{j=i}^{i+k} e^{t_0+\epsilon-t_j}\right)(-c_i z+1)^{-(\tilde{n}+i+1)}\left(\prod_{j=1}^{k}(-c_{i+j}z+1)^{-2}\right)(-c_{i+k+1}z+1)^{\tilde{n}+i+k}z^{k+1}$$

The derivative of this function is of the form

$$g'_{i,i+k}(z) =$$

$$= \left(\prod_{j=i}^{i+k} e^{t_0+\epsilon-t_j}\right)(-c_i z+1)^{-(\tilde{n}+i+2)}\left(\prod_{j=1}^{k}(-c_{i+j}z+1)^{-3}\right)(-c_{i+k+1}z+1)^{\tilde{n}+i+k-1}z^{k}\cdot$$

$$\cdot(A_0 z^{k+1}+\cdots+A_{k+1})$$

and therefore it has at most $k+1$ critical points which are not zeroes or poles. On the other hand

$$g_{i,i+k}(0) = g_{i,i+k}(1/c_{i+k+1}) = 0$$

$$g_{i,i+k}(1/c_{i+k}) = \cdots = g_{i,i+k}(1/c_i) = \infty$$

and points $0, 1/c_{i+k+1}, 1/c_{i+k}, \ldots, 1/c_i$ form an increasing sequence. Since a function must have a critical point between each two adjacent zeroes or poles and the total number of critical points which are not at zeroes or poles is $k+1$ we conclude that there is exactly one critical point $u_{i,i+k}$ of $g_{i,i+k}$ between 0 and $1/c_{i+k+1}$ and since our function is non-negative on this interval it must be a maximum.

(b) [**sum12b**] For $i = 0$ we have:

$$g_{0,k}(z) = \left(\prod_{j=1}^{k} e^{t_0+\epsilon-t_j}\right)\left(\prod_{j=1}^{k}(-c_j z+1)^{-2}\right)(-c_{k+1}z+1)^{\tilde{n}+k}(-z+1)^{n-\tilde{n}}z^{\tilde{n}+k}$$

Suppose that $n = \tilde{n}$. Our function has at most $k+1$ critical point outside of poles and zeroes. The sequence $0, 1/c_{k+1}, \ldots, 1/c_1$ is an increasing one and our function has zeroes in the first two points of this sequence and poles in other points. If $\tilde{n}+k$ is even then our function is non-negative for $z > 1/c_{k+1}$ and therefore must have critical points between points $1/c_k, 1/c_{k-1}, \ldots, 1/c_1, +\infty$. Therefore it has exactly one critical point between 0 and $1/c_{k+1}$. A similar argument leads to the same conclusion for $\tilde{n}+k$ being odd.

For $n > \tilde{n}$ similar reasoning shows that our function must have critical points on the intervals $(0,1)$, $(1, 1/c_{k+1})$, $(1/c_k, 1/c_{k-1})$,..., $(1/c_1, +\infty)$ and since the total number of critical points outside poles and zeroes is no more than $k+2$ there is exactly one critical point on each interval and in particular on $(0,1)$. This proves that functions of (12) satisfy the conditions of (14) and therefore the product $g_0(z_0)\ldots g_q(z_q)$ has a unique maximum on the simplex $0 \leq z_0 \leq \cdots \leq z_q \leq 0$.

16. [**sum13**] More formulas for $\epsilon > 0$. For $1 \leq i \leq j \leq q-1$ we have:

$$g_{i,j}(z,\epsilon) =$$

$$= \left(\prod_{k=i}^{j} e^{t_0+\epsilon-t_k}\right) z^{j-i+1}((e^{t_0-t_i+\epsilon}-1)z+1)^{-(\tilde{n}+i+1)} \cdot \left(\prod_{k=i+1}^{j}((e^{t_0-t_k+\epsilon}-1)z+1)^{-2}\right) \cdot$$

$$\cdot((e^{t_0-t_{j+1}+\epsilon}-1)z+1)^{\tilde{n}+j} \cdot \prod_{k=i}^{j} \frac{(e^{t_0-t_k+\epsilon}-1)z+1}{(e^{t_0-t_k-\epsilon}-1)z+1}$$

and

$$g_{0,j}(z,\epsilon) =$$

$$= \left(\prod_{k=1}^{j} e^{t_0+\epsilon-t_k}\right) z^{\tilde{n}+j}(-z+1)^{n-\tilde{n}}((e^{t_0-t_1+\epsilon}-1)z+1)^{-2} \cdot \left(\prod_{k=2}^{j}((e^{t_0-t_k+\epsilon}-1)z+1)^{-2}\right) \cdot$$

$$\cdot((e^{t_0-t_{j+1}+\epsilon}-1)z+1)^{\tilde{n}+j} \cdot \prod_{k=1}^{j} \frac{(e^{t_0-t_k+\epsilon}-1)z+1}{(e^{t_0-t_k-\epsilon}-1)z+1}$$

For $j = q$ there is a slight modification dues to the fact that we have to write $t_0 - t_{q+1}$ instead of $t_0 - t_{q+1} + \epsilon$.

Let further $U_\epsilon(t_1, \ldots, t_q)$ be the set of generic ancestral genealogies with event points $x_i \in (t_i - \epsilon, t_i + \epsilon)$. Then

$$\mu_{t_0}^{t_{q+1}}(n, r^{-1}(U_\epsilon(t_1, \ldots, t_q))) =$$

$$= \tilde{n}(\tilde{n}+1)\ldots(\tilde{n}+q-1)C(n,\tilde{n}) \cdot \prod_{i=0}^{q}(1-e^{-2\epsilon})g_i(z_i,\epsilon)$$

where $g_i = g_{i,i}$ are given in (10).

17. Let us specify explicitly which normalized birth and death process corresponds to the "optimal" survival probability function lying in $S(a_0, \ldots, a_q)$ i.e. what is the $\alpha_{0,t_0}^{t_{q+1}}$ measure for this process. It is clear that this measure is a sum of δ-measures concentrated in points $t_1, \ldots, t_q, t_{q+1}$ with some coefficients. As the first step towards the description of this measure we will compute the values

$$d_i = lim_{\epsilon\to 0}\alpha_{0,t_i-\epsilon}^{t_{q+1}}(\{t_i\})$$

i.e. the probability that a population member alive right before the time point t_i will die at this time point. Almost from definitions we have

$$d_i = 1 - \frac{a_{i-1}a_i^{-1}}{(1-e^{t_i-t_{i-1}})a_{i-1} + e^{t_i-t_{i-1}}}$$

for $i = 1, \ldots, q+1$ where $a_{q+1} = 1$.

18. Let us again consider the case $\epsilon = 0$ and now assume also that $q = 1$. Then situation is completely described by $\tilde{n}$ and

$$c_1 = 1 - e^{t_0-t_1}, \quad c_2 = 1 - e^{t_0-t_2}$$

where c_1, c_2 are arbitrary numbers satisfying $1 \geq c_2 \geq c_1 \geq 0$. The probability function normalized by the division by $(1-e^{-2\epsilon})$ and $(1-c_1)$ is

$$G(n, z_0, z_1) = \tilde{n}C(n,\tilde{n})z_0^{\tilde{n}}(-z_0+1)^{n-\tilde{n}}(-c_1z_0+1)^{\tilde{n}}z_1(-c_2z_1+1)^{\tilde{n}+1}(-c_1z_1+1)^{-(\tilde{n}+2)}$$

which we have to maximize on $\{n \in \mathbf{N} \,|\, n \geq \tilde{n}\} \times \{0 \leq z_0 \leq z_1 \leq 1\}$.

19. [**sum19**]Consider now the case when the death rate of the normalized birth and death process which we consider is bounded from the below by a non-genative number d (more formally, we may consider processes for which death rate is defined and is $\geq d$ on the complement to a subset of Lebesgue measure zero). For the time being let us call the survival functions of such processes "d-normalized survival functions". If σ is a d-normalized survival function on $[t_0, t_1)$ and $a_0 = \sigma(t_0)$ then one has

$$\sigma \geq \frac{(1-d)a_0}{(1-e^{(1-d)(t-t_0)})a_0 + (1-d)e^{(1-d)(t-t_0)}} = \frac{1-d}{1-ce^{(1-d)t}}$$

where $c = e^{(d-1)t_0}(a_0 + d - 1)/a_0$. One further has:

$$e^{-\int_{t_0}^{t_1} \frac{1-d}{1-ce^{(1-d)t}} dt} = \frac{e^{(d-1)t_1} - c}{e^{(d-1)t_0} - c} =$$

$$= \frac{a_0(e^{(d-1)(t_1-t_0)} - 1) + 1 - d}{1-d}$$

Note that

$$\frac{1-d}{1-ce^{(1-d)t_1}} = \frac{(1-d)a_0}{(1-e^{(1-d)(t_1-t_0)})a_0 + (1-d)e^{(1-d)(t_1-t_0)}} = M_0(a_0)$$

where

$$M_i = \begin{pmatrix} 1 & 0 \\ \frac{1-e^{(1-d)(t_{i+1}-t_i)}}{1-d} & e^{(1-d)(t_{i+1}-t_i)} \end{pmatrix}$$

For $d = 1$ we have special formulas. See **??** below.

20. Let $S(a_0, a_1, \ldots, a_q; d)$ be the set of normalized survival functions σ such that $d(\sigma) \geq d$, $\sigma(t_i) = a_i$. From the previous discussion we see that this set is non-empty if and only if one has $a_{i+1} \geq M_i(a_i)$ for $i = 0, \ldots, q-1$ and $a_0 \geq 0$, $1 \geq M_q(a_q)$.

    If $d \neq 1$ then the same reasoning as in the case of $d = 0$ shows that if $S(a_0, a_1, \ldots, a_q; d)$ is non-empty then for any $n \geq \tilde{n}$ and any sufficiently small $\epsilon > 0$ there exists a unique function σ there such that the corresponding process maximizes $\mu_{n,t_0}^{t_{q+1}}(r^{-1}(U_\epsilon(t_1, \ldots, t_q)))$ and the problem reduces to the maximization of the function (79) on the set of $(a_0, \ldots, a_q)$ satisfying the inequalities $a_{i+1} \geq M_i(a_i)$ for $i = 0, \ldots, q-1$ and $a_0 \geq 0$, $1 \geq M_q(a_q)$.

21. Doing the same computations as in (8c) we get:

    $$e^{-C_i} = a_i \frac{e^{2\epsilon(d-1)} - 1}{1-d} + 1$$

    and for $\epsilon = 0$ we get

    $$e^{-B_i} = a_i \frac{e^{(d-1)(t_{i+1}-t_i)} - 1}{1-d} + 1$$

    Modulo the factor $(2\epsilon)^q$ we may re-write the function which we need to maximize as

    $$\tilde{n} \ldots (\tilde{n} + q - 1) C(n, \tilde{n}) \prod_{i=0}^{q} f_i(a_i)$$

where

$$f_0(a_0) = (1-a_0)^{n-\tilde{n}} a_0^{\tilde{n}} (a_0 \frac{e^{(d-1)(t_1-t_0)}-1}{1-d} + 1)^{\tilde{n}}$$

and for $i = 1, \ldots, q$

$$f_i(a_i) = a_i (a_i \frac{e^{(d-1)(t_{i+1}-t_i)}-1}{1-d} + 1)^{\tilde{n}+i}$$

22. Set:

$$z_i = M_0^{-1} \ldots M_{i-1}^{-1}(a_i)$$

for $i = 0, \ldots, q$ and

$$z_{q+1} = M_0^{-1} \ldots M_q^{-1}(1)$$

Transformations M_i are order-preserving bijections and therefore the system of inequalities

$$0 \leq a_0; \quad M_i(a_i) \leq a_{i+1}, \ \ i = 0, \ldots, q; \quad M_q(a_q) \leq 1;$$

is equivalent to the system

$$0 \leq z_0 \leq \cdots \leq z_q \leq z_{q+1}$$

23. Let

$$L_i = \begin{pmatrix} \frac{e^{(d-1)(t_{i+1}-t_i)}-1}{1-d} & 1 \\ 0 & 1 \end{pmatrix}$$

Since $z_0 = a_0$ we have

$$f_0(a_0(z_0)) = (1-z_0)^{n-\tilde{n}} z_0^{\tilde{n}} L_0(z_0)^{\tilde{n}}$$

and since for $i = 1, \ldots, q$

$$a_i = M_{i-1} \ldots M_0(z_i)$$

we have:

$$f_i(a_i(z_i)) = M_{i-1} \ldots M_0(z_i)(L_i M_{i-1} \ldots M_0(z_i))^{\tilde{n}+i}$$

We further have

$$M_{i-1} \ldots M_0 = \begin{pmatrix} 1 & 0 \\ p_i & q_i \end{pmatrix}$$

$$L_i M_{i-1} \ldots M_0 = \begin{pmatrix} e^{(d-1)(t_{i+1}-t_i)} p_{i+1} & e^{(d-1)(t_{i+1}-t_i)} q_{i+1} \\ p_i & q_i \end{pmatrix}$$

where p_i and q_i are given by the recursive formulas:

$$p_0 = 0; \quad p_{i+1} = e^{(1-d)(t_{i+1}-t_i)} \left(\frac{e^{(d-1)(t_{i+1}-t_i)}-1}{1-d} + p_i \right);$$

$$q_0 = 1; \quad q_{i+1} = e^{(1-d)(t_{i+1}-t_i)} q_i = e^{(1-d)(t_{i+1}-t_0)};$$

Therefore

$$f_0(a_0(z_0)) = e^{\tilde{n}(d-1)(t_1-t_0)} (1-z_0)^{n-\tilde{n}} z_0^{\tilde{n}} (p_1 z_0 + q_1)^{\tilde{n}} =$$
$$= (1-z_0)^{n-\tilde{n}} z_0^{\tilde{n}} (z_0 \frac{p_1}{q_1} + 1)^{\tilde{n}}$$

and for $i = 1, \ldots, q$

$$f_i(a_i(z_i)) = e^{(\tilde{n}+i)(d-1)(t_{i+1}-t_i)} z_i (p_i z_i + q_i)^{-(\tilde{n}+i+1)} (p_{i+1} z_i + q_{i+1})^{\tilde{n}+i} =$$

$$= e^{(d-1)(t_i-t_0)} z_i (z_i \frac{p_i}{q_i}+1)^{-(\tilde{n}+i+1)} (z_i \frac{p_{i+1}}{q_{i+1}}+1)^{\tilde{n}+i}$$

Set $c_i = -p_i/q_i$. Then for $1 \leq i \leq j \leq q$ one has

$$g_{i,j}(z) = e^{(d-1)(t_i-t_0+\cdots+t_j-t_0)} z^{j-i+1} (-c_i z+1)^{-(\tilde{n}+i+1)} \left(\prod_{k=i+1}^{j} (-c_k z+1)^{-2} \right) (-c_{j+1} z+1)^{\tilde{n}+j}$$

and

$$g_{0,j} = e^{(d-1)(t_1-t_0+\cdots+t_j-t_0)} z^{\tilde{n}+j} (1-z)^{n-\tilde{n}} \left(\prod_{k=1}^{j} (-c_k z+1)^{-2} \right) (-c_{j+1} z+1)^{\tilde{n}+j}$$

We have

$$c_0 = 0; \quad c_{i+1} = -\frac{p_{i+1}}{q_{i+1}} = c_i + \frac{e^{(d-1)(t_{i+1}-t_0)} - e^{(d-1)(t_i-t_0)}}{d-1} = \frac{e^{(d-1)(t_{i+1}-t_0)} - 1}{d-1}$$

In particular $0 \leq c_1 \leq \cdots \leq c_{q+1}$.

We have

$$\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -p/q & 1/q \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ p' & q' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ p+qp' & qq' \end{pmatrix}$$

Our domain of definition for $z_1, \ldots, z_q$ is bounded from the above by

$$z_{q+1} = (M_q \ldots M_0)^{-1}(1) = (c_{q+1} + q_{q+1}^{-1})^{-1} = \frac{d-1}{de^{(d-1)(t_{q+1}-t_0)} - 1} \leq 1$$

In particular $1/c_i \geq z_{q+1}$ for all i. This shows that our previous reasoning applies without change to show that $g_1(z_1) \ldots g_q(z_q)$ has a unique maximum on $0 \leq z_1 \leq \cdots \leq z_q \leq z_{q+1}$.

Note: for $d = 1$ we get $z_{q+1} = 1/(1 + t_{q+1} - t_0)$.

Note: We have

$$M_i^{-1} = \begin{pmatrix} 1 & 0 \\ \frac{1-e^{(d-1)(t_{i+1}-t_i)}}{1-d} & e^{(1-d)(t_{i+1}-t_i)} \end{pmatrix}$$

and

$$M_{i^{-1}}(1) = \frac{d-1}{de^{(d-1)(t_{i+1}-t_i)} - 1}$$

24. [**sum24**] In order to model or construct a birth process with the birth rate function σ on $[t_0, t_1]$ we need to describe for any m and $u \in [t_0, t_1)$ the measure $\alpha_u^{t_1}(m \mapsto m+1)$ on $(u, t_1]$ as the image of the Lebesgue measure on $[0, 1]$ under a function $F : [0, 1] \to (u, t_1] \amalg pt$.

    From the general comments made above we know that we may take F to be the supremum or infinum inverse of the probability distribution function $A : (u, t_1] \amalg pt \to [0, 1]$ of $\alpha_u^{t_1}(m \mapsto m+1)$. In our case this function is of the form

$$A_{u,m}(t) = \alpha_u^{t_1}(m \mapsto m+1)(u, t] =$$

$$= 1 - \phi_u^t(1,1)^m = 1 - (e^{-\int_u^t \sigma(x)dx})^m$$

The main equation tells us that

$$\sigma' = \sigma(\sigma + d - 1)$$

In the case of constant death rate d we get

$$\sigma = \frac{c(d-1)e^{(d-1)t}}{1 - ce^{(d-1)t}}$$

Finding c from the condition $\sigma(\tau) = a$ we get:

$$c = e^{(1-d)\tau} \frac{a}{a + d - 1}$$

and

$$\sigma(t) = \frac{(d-1)ae^{(d-1)(t-\tau)}}{a + d - 1 - ae^{(d-1)(t-\tau)}}$$

which after a simple transformation agrees with the formula of 19. We further have:

$$e^{-\int_{t_0}^{t_1} \sigma(t)dt} = \frac{e^{(d-1)t_1} - c^{-1}}{e^{(d-1)t_0} - c^{-1}} =$$

$$= \frac{ae^{(d-1)(t_1-\tau)} - a - d + 1}{ae^{(d-1)(t_0-\tau)} - a - d + 1}$$

which for $\tau = t_0$ again agrees with (19). For $\tau = t_1$ and $a = 1$ we get

$$e^{-\int_u^t \sigma(x)dx} = \frac{e^{(d-1)(t-t_1)} - d}{e^{(d-1)(u-t_1)} - d}$$

Now we need to solve in t the equation

$$z = 1 - (e^{-\int_u^t \sigma(x)dx})^m = 1 - (e^{(d-1)(t-t_1)} - d)^m (e^{(d-1)(u-t_1)} - d)^{-m}$$

We get:

$$e^{(d-1)(t-t_1)} - d = (1-z)^{1/m}(e^{(d-1)(u-t_1)} - d)$$

$$t = t_1 + (d-1)^{-1} ln(d + (1-z)^{1/m}(e^{(d-1)(u-t_1)} - d))$$

25. Let us generalize the formulas further. Let d, b be two constants. The general form of the solution for the equation

$$\sigma' = \sigma(b\sigma - b + d)$$

is

$$\sigma = \frac{c(d-b)e^{(d-b)t}}{b - bce^{(d-b)t}}$$

where c is a constant. From the condition $\sigma(\tau) = a$ we find

$$c = e^{(b-d)\tau} \frac{ab}{ab + d - b}$$

and

$$\sigma = \frac{(d-b)ae^{(d-b)(t-\tau)}}{ab + d - b - abe^{(d-b)(t-\tau)}}$$

Then

$$\int_{t_0}^{t_1} \frac{c(d-b)e^{(d-b)t}}{b(1-ce^{(d-b)t})}dt = \int_{t\in[t_0,t_1]} \frac{d\, ce^{(d-b)t}}{b(1-ce^{(d-b)t})} =$$

$$= \frac{1}{b} ln(\frac{1-ce^{(d-b)t_0}}{1-ce^{(d-b)t_1}})$$

And for $\sigma(\tau) = a$ we get

$$e^{-\int_{t_0}^{t_1} b\sigma(t)dt} = \frac{1-ce^{(d-b)t_1}}{1-ce^{(d-b)t_0}} =$$

$$= \frac{abe^{(d-b)(t_1-\tau)} - ab - d + b}{abe^{(d-b)(t_0-\tau)} - ab - d + b}$$

For $\tau = t_0$ and $a = a_0$ we get

$$e^{-\int_{t_0}^{t_1} b\sigma(t)dt} = ab\frac{1-e^{(d-b)(t_1-t_0)}}{d-b} + 1$$

26. Let $d_{min}, b_{max} \geq 0$, $t_0 < t_1$ and $a_0, a_1 \in [0,1]$ Consider the set S of triples of continuous on the right functions b, d, σ on $[t_0, t_1]$ such that

   (a) $0 \leq b(t) \leq b_{max}$ and $d(t) \geq d_{min}$,

   (b) σ' exists everywhere except possibly for a finite number of points x_i,

   (c) $\sigma(x_i) - lim_{t\uparrow x_i}\sigma(t) \geq 0$,

   (d) for $t \neq x_i$

   $$[\textbf{maineq}]\sigma' = \sigma(b\sigma - b + d) \tag{84}$$

   (e) $\sigma(t_0) = a_0$ and $\sigma(t_1) = a_1$.

   Consider the functional $e^{-\int_{t_0}^{t_1} b\sigma d\tau}$ on S. We need to find its maximal and minimal values and the points where these values are achieved.

27. Lemma. For any $\sigma \in S$ and $t \in [t_0, t_1]$ one has $\sigma(t) \in [0,1]$.

28. Lemma S is non-empty if and only if

   $$a_n \geq e^{(d_{min}-b_{max})(t_1-t_0)}a_0$$

   Proof:

29. Proposition. If σ is a continuous function satisfying the equation (84) and $\sigma(t_0) = a_0$ then

   $$g(t) = e^{-\int_{t_0}^{t} b(\tau)\sigma(\tau)d\tau} = 1 - a_0 \int_{t_0}^{t} b(x)e^{\int_{t_0}^{x}(d(\tau)-b(\tau))d\tau}dx$$

   Proof: Direct verification that

   $$\sigma = -(1/b)g'/g$$

   satisfies (84).

30. Proposition. For given b, d, a_0, a_1 the maximal value of $e^{-\int_{t_0}^{t_1} b\sigma d\tau}$ is achieved by σ which is continuous everywhere except possibly t_0.

   Proof: (see notes for March 23, 24, 2009).

April 15, 2009 Let $D = (V, E, E \to V \times V)$ be a directed (multi-)graph with the set of vertices V and the set of edges E. Let $E_i(D)$ be the set of paths of length i in D. For $u \le v$ set

$$P_D[u,v] = V \amalg (\amalg_{q\ge 0} \amalg_{\pi\in E_{q+1}(D)} (\Delta^{q+1}_{(u,v)} \amalg \Delta^q_{(u,v)}))$$

For $u = v$ we have $\Delta^q_{(u,v)} = \emptyset$ for all $q \ge 0$ and therefore

$$P_D[u,u] = V$$

We can also describe $P_D[u,v]$ as the set of triples of the form

1. a right-continuous map $g : [u,v] \to V$,
2. a finite subset $x_1, \ldots, x_n$ of $(u,v]$ which includes all points of discontinuity of g,
3. for each $i = 1, \ldots, n$ a choice of an edge connecting $g_-(x_i) = \lim_{x\uparrow x_i} g(x)$ to $g(x_i)$.

The points of V correspond to constant functions. If $\pi = (e_1, \ldots, e_{q+1})$ is a path then for $(x_1, \ldots, x_q) \in \Delta^q_{(u,v)}$ the corresponding function is of the form

$$g([u,x_1)) = \partial_0(e_1);\;\; g([x_i,x_{i+1})) = \partial_0(e_{i+1}),\; i = 1,\ldots,q-1;\;\;\; g([x_q,v)) = \partial_0(e_{q+1});$$

and

$$g([v]) = \partial_1(e_{q+1});$$

with distinguished points $(x_1, \ldots, x_q, v)$ and for $(x_1, \ldots, x_{q+1}) \in \Delta^{q+1}_{(u,v)}$ the corresponding function is of the form

$$g([u,x_1)) = \partial_0(e_1);\;\; g([x_i,x_{i+1})) = \partial_0(e_{i+1}),\; i = 1,\ldots,q;\;\;\; g([x_q,v]) = \partial_1(e_{q+1});$$

with distinguished points $(x_1, \ldots, x_{q+1})$.

For $[u',v'] \subset [u,v]$ the restriction defines a map

$$P_D[u,v] \to P_D[u',v']$$

These maps clearly satisfy the conditions of Definition 4.0.2 and we obtain a path system $\mathcal{P}_D$.

We can modify the first description of $P_D[u,v]$ using the bijection

$$\Delta^{q+1}_{(u,v)} \amalg \Delta^q_{(u,v)} = \{u < x_1 < \cdots < x_{q+1} \le v\}$$

References

[1] Gerald B. Folland. *Real analysis.* Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[2] P. Gabriel and M. Zisman. *Calculus of fractions and homotopy theory.* Springer-Verlag, Berlin, 1967.

[3] Iosif I. Gikhman and Anatoli V. Skorokhod. *The theory of stochastic processes. II.* Classics in Mathematics. Springer-Verlag, Berlin, 2004. Translated from the Russian by S. Kotz, Reprint of the 1975 edition.

[4] Michèle Giry. A categorical approach to probability theory. In *Categorical aspects of topology and analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 68–85. Springer, Berlin, 1982.

[5] B. A. Sevastyanov. *Vetvyashchiesya protsessy.* Izdat. “Nauka”, Moscow, 1971.

[6] Vladimir Voevodsky. Notes on categorical probability. *In preparation*, 2008.

[7] David Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.