Singletons

Vladimir Voevodsky

January 4, 2008- September 11, 2008

Contents

0.1 Introduction

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 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] 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.

1 Singleton processes

1.1 Singleton path system

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

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

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

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

Intuitively, the set $\phi^{-1}(s)$ 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 $\phi(v)$ being the time of the corresponding event. The subsets $\psi^{-1}(i)$ of the final population $\phi^{-1}(t)$ consist of members which transform into i new members at the exact moment t. We let $H[s,t]$ denote the set of isomorphism classes of singleton histories over $[s,t]$. This set carries a natural structure of a commutative monoid given by the disjoint union of histories.

We will need topology on the space $H[s,t]$ which we get by identifying histories with the points of $[s,t]$-geometric realizations of a commutative simplicial monoid $F^*(\mathbf{N})$.

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 $u_1 \leq \cdots \leq u_i$ 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^*. We further let $\Delta^i_{(s,t)}$ denote the open analogs of $\Delta^i_{[s,t]}$.

Recall that for any monad 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 monad F on the category of commutative monoids which takes a monoid A to the free monoid generated by A as a set, e.g. $F(pt) = \mathbf{N}$,

Proposition 1.1.1 *[top] There are natural bijections of monoids*

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

Proof: We are going to show that $H[s,t]$ can be identified with the set of points of the disjoint union

$$\amalg_{i\geq 0} F_i^{nd}(\mathbf{N}) \times \Delta^i_{(s,t)}$$

where $F_q^{nd}(X)$ is the set of non-degenerate simplicies of $F_*(X)$ of dimension q and $\Delta^q_{(s,t)}$ is the point for $q=0$ and the open simplex $s < u_1 < \cdots < u_q < s$ for $i \geq 1$.

For a set X let $Symm^n(X)$ be the n-th symmetric power of X and let

$$S(X) = \amalg_{n\geq 0} Symm^n(X)$$

Any element of $S(X)$ is of one of the three types. If it belongs to $Symm^0(X) = pt$ we denote it by $*$. If it belongs to $Symm^1(X) = X$ we denote it by $[x]$ where x is the corresponding element of X. If it belongs to $Symm^n(X)$ for $n > 1$ it can be written in a unique way as a commutative sum $[x_1] + \cdots + [x_n]$ where $x_i \in X$.

By construction

$$F_q(X) = S^{\circ(q+1)}(X)$$

where $S^{\circ i}$ is the i-th iteration of the functor $S(-)$. Elements of $\amalg_q F_q(X)$ which belong to $F_q(X)$ will be called elements of level q. We let $*_q$ denote the point $Symm^0(S^{\circ(q+1)}(X))$. Then one has:

1. any element of level 0 is of the form $*_0$ or $\sum [x_i]$ where $x_i \in X$,
2. any element of level $q > 0$ is of the form $*_q$ or $\sum[\gamma_i]$ where γ_i are elements of level $q-1$.

Let us define a map $\pi : H[s,t] \to \amalg_q F_q(\mathbf{N})$ as follows. For $\Gamma \in H[s,t]$ let $Supp(\Gamma) = Im(\phi) \cap (s,t)$. It is a finite subset of (s,t) which we can write down as a unique increasing sequence $u_1 < \cdots < u_q$. The number q is called the level of Γ and will coincide with the level of $\pi(\Gamma)$.

If $l(\Gamma) = 0$ i.e. $Supp(\Gamma) = \emptyset$ then Γ is a disjoint union of intervals, one for each point of $\phi^{-1}(t)$. If Γ is empty we set $\pi(\Gamma) = *_0$. Otherwise we set

$$\pi(\Gamma) = \sum_{v \in \phi^{-1}(t)} [\psi(v)].$$

If $l(\Gamma) > 1$ and Γ is connected set

$$\pi(\Gamma) = [\pi(R_{u_1}(\Gamma))]$$

and in general set

$$\pi(\Gamma) = \pi(\Gamma_1) + \cdots + \pi(\Gamma_m)$$

where $\Gamma_1, \ldots, \Gamma_m$ are the connected components of Γ.

Let now $\gamma \in F_q(\mathbf{N})$ be an element of level q and $u_1 < \cdots < u_q$ be an increasing sequence in (s,t). Define $\Gamma = \Pi'(\gamma; u_1, \ldots, u_q) \in H[s,t]$ inductively as follows.

If $q = 0$ and $\gamma = *_0$ we set $\Gamma = \emptyset$. If $q = 0$ and $\gamma = \sum_{i=1}^{m}[n_i]$ we define Γ as the disjoint union of m intervals starting at s and ending at t with the function ψ defined by the numbers n_i.

If $q > 0$ and $\gamma = *_q$ we set $\Gamma = \emptyset$. If $\gamma = [\gamma']$, consider

$$\Gamma' = \Pi'(\gamma'; u_2, \dots, u_q) \in H[u_1, t]$$

and construct Γ by contracting all the initial vertices of Γ' to one vertex v and adding a new initial vertex v_0 and an edge from v_0 to v. If $\gamma = \sum[\gamma'_i]$ set

$$\Gamma = \sum \Pi'([\gamma'_i]; u_1, \dots, u_q)$$

A graph created by this procedure need not satisfy the definition of a population history since it may contain internal vertices where exactly one edge starts. By erasing all such vertices we get a population history which we denote by $\Pi(\gamma; u_1, \dots, u_q)$.

One verifies easily that for any $\Gamma \in H[s,t]$ one has

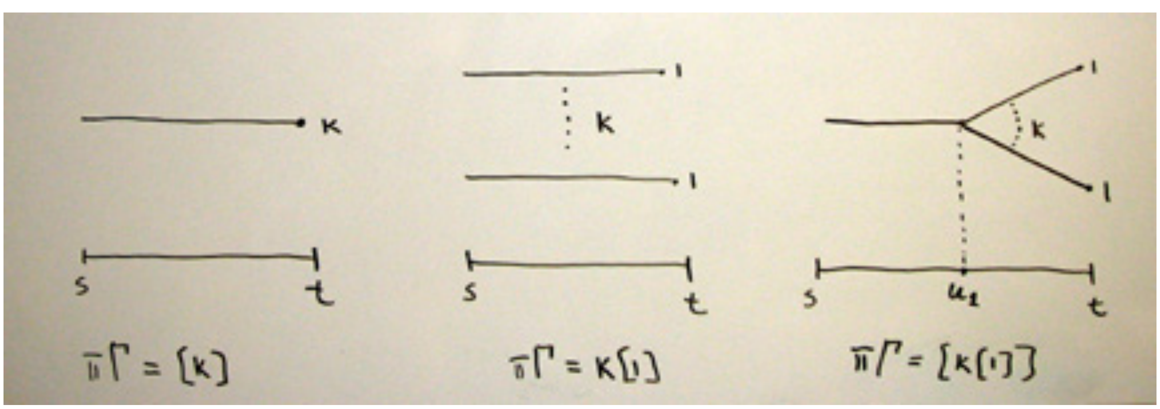
$$\Pi(\pi(\Gamma); u_1(\Gamma), \dots, u_q(\Gamma))) = \Gamma.$$

The converse is not necessarily true since erasing the extra vertices may create a situation when for one of the original u_i's there are no vertices v with $\phi(v) = u_i$. One verifies easily that this happens if and only if the corresponding element of $\amalg F_q$ is degenerate. For a non-degenerate γ and any $u_1, \dots, u_q$ one has

$$\pi(\Pi(\gamma; u_1, \dots, u_q)) = \gamma$$

which finishes the proof. $\square$

Example 1.1.2 The most important elementary 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:



Corollary 1.1.3 *[***homot***] The space $H[s,t]$ is homotopy equivalent to* **N**. *A history Γ belongs to the connected component given by the number of final vertices with multiplicities defined b ψ.*

Proof: It follows by [5,] from the fact that the monad F_* is given by the composition of the forgetful functor to sets (resp. pointed sets) with its left adjoint. $\square$

In what follows we consider $H[s,t]$ as a measurable space with respect to the Borel σ-algebra defined by our topology. The simplicial set $F_*(\mathbf{N})$ is not locally finite, for example, the vertex corresponding to [1] is the boundary of all of the 1-simplexes corresponding to $n[*_0] + [1]$, and as a consequence the topological space $H[s,t]$ is not very nice. In particular it is not locally compact. However it is a disjoint union of countable number of open simplexes which makes it equivalent to a locally compact space from the point of view of measure theory.

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 $\phi(v) = u$ 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 $R_u(\Gamma)$. For $s \leq u < v \leq t$ define the restriction maps

$$[\mathbf{restr}] res_{u,v} : H[s,t] \to H[u,v] \tag{1}$$

as follows:

1. for $s < u < v < t$ set $res_{u,v} = L_v \circ R_u$,
2. for $u = s, v < t$ set $res_{u,v} = L_v$,
3. for $s < u, v = t$ set $res_{u,v} = R_u$,
4. for $u = s, v = t$ set $res_{u,v} = Id$.

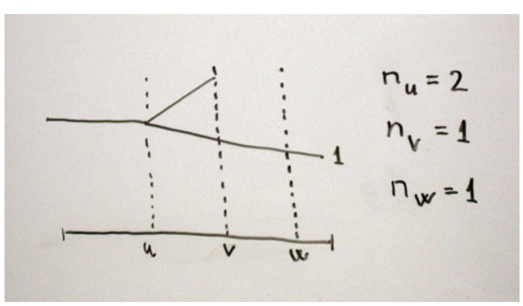
For any $u \in [s,t]$ define

$$[\mathbf{nu}] n_u : H[s,t] \to \mathbf{N} \tag{2}$$

as follows:

1. if $u \in (s,t)$ then this map takes Γ to be the number of initial vertices of $R_u(\Gamma)$ or equivalently as the number of final vertices of $L_u(\Gamma)$ counted with their multiplicities,
2. if $u = s$ then this map takes Γ to the number of initial vertices of Γ,
3. if $u = t$ then this map takes Γ to the number of final vertices of Γ (counted with multiplicities).

Intuitively, n_u takes Γ to the number of population members at time u.



One can easily see that the maps (1) and (2) are not continuous. However we have the following obvious result.

Lemma 1.1.4 *[**arem**] The maps (1) and (2) are measurable.*

For $s \leq u \leq v \leq t$ define the σ-algebra $\mathfrak{B}_u^v$ as forllows:

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

Lemma 1.1.4 immediately implies the following result (for the definition of a path system see [7]).

Proposition 1.1.5 *[**ispath**] The collection of data $((H[s,t],\mathfrak{B}),\mathfrak{B}_u^v,n_u)$ defines a path system on $\mathbf{N}$ over T.*

We denote this path system by $\mathcal{H}[s,t]$. For the uniformity of notation let us set

$$H[u,u] = \mathbf{N}$$

then, 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.

Remark 1.1.6 It seems that if we start with a monad 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.

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

For $s \leq u < v \leq t$ and a combinatorial type π let $\Delta_{u,v}^{\pi}$ be the open simplex of $H[u,v]$ corresponding to π. When no confusion is possible we will write Δ_u^{π} instead of $\Delta_{u,t}^{\pi}$. A measure on $H[u,v]$ is the same as a collection of measures on the simplexes $\Delta_{u,v}^{\pi}$ given for all combinatorial types π.

For $s \leq u < v \leq t$ and $n,q \geq 0$, let $H[u,v]_{n,q}$ be the subset of histories of level q such that $n_u(\Gamma) = n$. Note that we have $H[u,v]_{0,0} = pt$, $H[u,v]_{0,>0} = \emptyset$ and

$$H[u,v]_{1,0} = \coprod_{k \geq 0} \Delta_{u,v}^{[k]}$$

where each $\Delta_{u,v}^{[k]}$ is a point. In general

$$H[u,v]_{*,q} = sk_q(H[u,v])$$

and in particular

$$H[u,v]_{*,0} = \amalg_{\underline{k}\in S(\mathbf{N})}\Delta^{\underline{k}}_{u,v}$$

Define a map

$$(x_1, \underline{k}_1) : H[u,v] \to ((u,v] \amalg \{\infty\}) \times S(\mathbf{N})$$

as follows. It sends $\Delta^{n[1]}_{u,v}$ to $(\infty, n[1])$, $\Delta^{\underline{k}}_{u,v}$ for $\underline{k} \neq n[1]$ to $(t, \underline{k})$ and a history Γ of level $q \geq 1$ to the pair $(x_1(\Gamma), \underline{k}_1(\Gamma))$ where $x_1(\Gamma)$ is first event point in Γ and $\underline{k}_1(\Gamma)$ is the local type of this event point.

Since elements of $S(\mathbf{N})$ of the form $n[1]$ play a special role in constructions considered below we will use the notation $\tilde{S}(\mathbf{N})$ for the set of elements of $S(\mathbf{N})$ which are not of the form $n[1]$ for any n. For $\underline{k} = [k_1] + \cdots + [k_n]$ we write $n(\underline{k}) = n$ and $k = \sum_{i=1}^n k_i$.

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

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

which is determined by the conditions

$$R_v(j^w_{u,v}(\Gamma)) = \Gamma$$

$$L_v(j^w_{u,v}(\Gamma)) = \Delta^{n[1]}_{u,v}$$

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

$$j^w_{u,v}(\Delta^\pi_{v,w}) = \left\{(x_1,\ldots,x_q) \in \Delta^\pi_{u,w} \,|\, x_1 > w\right\}.$$

which implies in particular that $j^w_{u,v}$ are measurable.

For $x \in (u,v]$ and $\Gamma \in H[x,v]_{n,*}$ we let $[n] *_x \Gamma$ denote the unique history such that

$$\pi(L_x([n] *_x \Gamma)) = [n]$$

and

$$R_x([n] *_x \Gamma) = \Gamma.$$

and 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 $\underline{k} \in \tilde{S}(\mathbf{N})$, a measurable B in $(u,v]$ and a measurable $U \in H[u,v]$ let

$$(\underline{k}, B, U) = \{\Gamma \in H[u,v] |\, (x_1, \underline{k}_1)(\Gamma) \in B \times \underline{k},\; R(\Gamma) \in U\}$$

!!! Lemma is false for $n(\underline{k}) > 1$.

Lemma 1.2.1 *[**gens**] For $q > 0$ the Borel σ-algebra on $H[u,v]_{n,q}$ is generated in the strong sense by subsets of the form*

$$(\underline{k},(w_1,w_2),U) = \{\Gamma \in H[u,v]_{*,q} | \, \underline{k}_1(\Gamma) = \underline{k}, \; x_1(\Gamma) \in (w_1,w_2), \; R(\Gamma) \in U\}$$

where $n(k) = n$, $u < w_1 < w_2 < t$ and there exists $w_2 < w_3 < t$ such that U is a measurable subset of $j^v_{u,w_3}(H[w_3,v]_{tr(\underline{k}),q-1})$.

Proof: If $u = v$ then $H[u,v]_{*,>0} = \emptyset$ and the statement becomes trivial. Let $u < v$. The collection of subsets $(\underline{k},(w_1,w_2),U)$ is closed under intersections and it remains to show that for any combinatorial type π with $q(\pi) = q$ and $\underline{k}_1(\pi) = \underline{k}$ the σ-algebra generated by those of these subsets which lie in $\Delta^{\pi}_{u,v}$ coincides with the Borel σ-algebra. Since

$$\Delta^{\pi}_{u,v} = \{x_1,\dots,x_q | \, u < x_1 < \dots < x_q < v\}$$

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

$$\Delta^{R(\pi)}_{w_3,v} = \{w_3 < x_2 < \dots < x_q < v\}$$

for some $w_2 < w_3 < v$. Observe now that the image of this subset in $H[u,v]$ coincides with $(\underline{k},(w_1,w_2),U)$. $\square$

Recall (see [7]) that a process on $\mathcal{H}[s,t]$ is a collection of sub-probability kernels $\lambda^v_u : \mathbf{N} \to H[u,v]$ given for all $u \le v$ such that $\lambda^v_u(n)$ is supported on $H[u,v]_{n,*}$. We set

$$\phi^v_{u,\lambda}(n,m) = \lambda^v_{u,n}(H[u,v] \cap n_v^{-1}(m))$$

$$v^v_{u,n,\lambda} = \lambda^v_{u,n}(H[u,v]) = \sum_{m \ge 0} \phi^v_{u,\lambda}(n,m)$$

$$h^n_\lambda(u,v) = \lambda^v_{u,n}(\Delta^{n[1]}_{u,v}).$$

When no confusion is possible we will write $v^v_{u,k}$ instead of $v^v_{u,k,\lambda}$ etc.

Recall (see [7]) that a process is a sub-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 $s \le u \le v \le w \le t$ and all $n \ge 0$ the square

$$[\textbf{m2diag}] \quad \begin{array}{ccc} \mathbf{1} & \xrightarrow{\lambda^v_{u,n}} & H[u,v]_{n,*} \\ \lambda^w_{u,n} \downarrow & & \downarrow Id \otimes (\lambda^w_{v,n} \circ n_v) \\ H[u,w]_{n,*} & \xrightarrow{res_{u,v} \times res_{v,w}} & H[u,v]_{n,*} \times H[v,w] \end{array} \tag{3}$$

commutes. Applying [7,] and taking into account that $\lambda^u_u = Id$ to our case we get the following reformulation.

Lemma 1.2.2 *[**crit1**] A process λ^*_* on $\mathcal{H}[s,t]$ is a sub-Markov process if and only if for any $m,n \ge 0$, any $s \le u < v < w \le t$, any measurable U_1 in $H[u,v]_{m,*} \cap n_v^{-1}(n)$ and any measurable U_2 in $H[v,w]_{n,*}$ one has*

$$[\textbf{eqcrit1}] \lambda^w_{u,m}((res_{u,v} \times res_{v,w})^{-1}(U_1 \times U_2)) = \lambda^v_{u,m}(U_1) \, \lambda^w_{v,n}(U_2). \tag{4}$$

Lemma 1.2.3 *[**genm**] If λ^*_* is a sub-Markov process then for $u \le v \le w$ in $[s,t]$ one has*

$$[\mathbf{eq001}]\phi_u^w(n,k) = \sum_{m\ge 0} \phi_u^v(n,m)\phi_v^w(m,k) \tag{5}$$

$$[\mathbf{eq00}]v_{u,n}^w = \sum_{m\ge 0} \phi_u^v(n,m)v_{v,m}^w \tag{6}$$

Proof: It follows from the general properties of sub-Markov processes (see [7,]). $\square$

Lemma 1.2.4 *[**ob1**] If λ^*_* is a sub-Markov process then for $n \ge 0$ and $u \le v \le w$ in $[s,t]$ one has*

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

Proof: It follows from Lemma 1.2.2 applied to $U_1 = \Delta_{u,v}^{n[1]}$, $U_2 = \Delta_{v,w}^{n[1]}$. $\square$

Lemma 1.2.5 *[**ob00**] Let λ^*_* be a sub-Markov process. Then for any $n, m \ge 0$ and any $u \le v < w$ 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 1.2.2 to $U_1 = \Delta_{u,v+\epsilon}^{n[1]}$ and $U_2 = H[v+\epsilon, w] \cap n_w^{-1}(m)$ we get

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

Since for $\epsilon' \ge \epsilon$ 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

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

our claims follow. $\square$

Recall that a function f on $[s,t]$ is called monotone increasing (resp. decreasing) if for $x \le y$ 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 1.2.6 *[**rcim**] Any right continuous function f on $[s,t]$ is measurable.*

Proof: It is sufficient to show that for any a the subset $U = \{x : f(x) < a\}$ is measurable. For any $y \in \mathbf{Q} \cap [s,t]$ let $y_- = inf\{w : [w,y] \subset U\}$ and let

$$U_y = \begin{cases} [x,y] & \text{if } y_- \in U \\ (x,y] & \text{otherwise} \end{cases}$$

One observes easily that $u = \cup_y U_y$ and since the set $\mathbf{Q} \cap [s,t]$ is countable it gives us a countable covering of U by measurable subsets. $\square$

Lemma 1.2.7 *[**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 1.2.4 we see in particular that for a sub-Markov process the functions $h^n(u,v)$ are monotone increasing in u and monotone decreasing in v. Since $v^w_{v,m} \leq 1$ and

$$[\mathbf{eq01}] \sum_{m\geq 0} \phi_{u,v}(n,m) = v^v_{u,n} \qquad (7)$$

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

Remark 1.2.8 We will see from examples below (**??**) that there are sub-Markov processes on $\mathcal{H}[s,t]$ such that $v^v_{u,n}$ are not monotone in u.

Lemma 1.2.9 *[**ob01**] Let λ^*_* be a sub-Markov process. Then for any $m,n \geq 0$ and any $u \leq v < w$ 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 1.2.2 to $U_1 = H[u,v+\epsilon]_{m,*} \cap n^{-1}_{v+\epsilon}(n)$ and $U_2 = \Delta^{n[1]}_{v+\epsilon,w}$ we get

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

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 1.2.10 *[**rcont**] A process λ^*_* 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 λ is non-degenate then $h^n(u,u) = 1$ for all n and u.

Remark 1.2.11 For a sub-Markov process one has $(\upsilon^u_{u,k})^2 = \upsilon^u_{u,k}$ and therefore a sub-Markov process is non-degenerate if and only if $\upsilon^u_{u,k} \neq 0$ for all u, k.

Theorem 1.2.12 *[**th1**] Let λ^*_* be a non-degenerate sub-Markov process on $\mathcal{H}[s,t]$. Then the following conditions are equivalent:*

1. *for all $n \geq 0$ 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 $n \geq 0$ 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 $\phi^v_u(n,m)$ are right continuous in u and if $u < t$ then there exits $w > u$ such that $\upsilon^w_{u,n} \neq 0$,*
4. *for all $n \geq 0$ functions $\upsilon^v_{u,n}$ 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 $\upsilon^w_{u,n} \neq 0$ then, since $\upsilon^v_{u,n}$ are monotone decreasing in v we have $\upsilon^v_{u,n} \neq 0$ for all $u \leq v \leq w$.

Let u and w be as above. Taking the sum over m in Lemma 1.2.5 and setting $v = u$ we get

$$[\mathbf{feqp}] \lim_{\epsilon>0,\epsilon\to 0} h^n(u,u+\epsilon)\upsilon^w_{u+\epsilon,n} = \upsilon^w_{u,n} \tag{8}$$

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 $\upsilon^v_{u,n}$ is right continuous in u equation (8) implies that

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

and since $\upsilon^w_{u,n} \neq 0$ we conclude that

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

Together with Lemma 1.2.4 we conclude that (2) and (5) hold.

(2) $\Rightarrow$ (5) Immediate from Lemma 1.2.4 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 1.2.9 $v = u$ 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

$$[\mathbf{seqp}] \lim_{\epsilon>0,\epsilon\to 0} \phi_u^{v+\epsilon}(m,n) = \phi_u^v(m,n) \tag{9}$$

Let w be such that $h^n(v,w) \neq 0$. Then Lemma 1.2.9 together with the right continuity of $h^n(-,-)$ in the first variable implies (9).

(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 $\upsilon_{u,n}^v$ are right continuous in v and $\upsilon_{u,k}^u = 1$ there exists $w > u$ such that $\upsilon_{u,n}^w \neq 0$ and as explained above such that $h^n(u,w) \neq 0$. Taking in Lemma 1.2.9 $m \neq n$ and $v = u$ we get

$$[\mathbf{eq020}] \lim_{\epsilon\to 0} \phi_{u,u+\epsilon}(m,n) = 0 \tag{10}$$

Therefore we have

$$[\mathbf{teqp}] 1 = \lim_{\epsilon\to 0} \upsilon_{u,n}^{u+\epsilon} = \lim_{\epsilon\to 0} \sum_m \phi_{u,u+\epsilon}(n,m) = \lim_{\epsilon\to 0} \phi_{u,u+\epsilon}(n,n) \tag{11}$$

Form Lemma 1.2.9 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 (11) implies that $h^n(u,v)$ is right continuous in u.

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

Theorem is proved. $\Box$

For a process λ_*^* define $E_{n,\lambda} \subset [s,t]$ by the rule $x \in E_{n,\lambda}$ 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 $E_{n,\lambda}$.

Lemma 1.2.13 *[**ob2**] Let λ be a non-degenerate right continuous sub-Markov 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 1.2.12 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 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 1.2.14 *[***ob3***] For a a non-degenerate right continuous sub-Markov process λ the sets E_n are countable.*

Proof: We have

$$[\mathbf{ecov}][s,t) = \amalg_{e\in E_n}[e, e_{+1}) \tag{12}$$

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

For any $u \in [s,t]$ let $e_n(u)$ be the smallest element of E_n which is greater than u. If no such element exist i.e. if $h^n(u,t) \neq 0$ we set $e_n(u) = \infty$.

For $s \le u < v \le t$ and $\underline{k} \in \tilde{S}(\mathbf{N})$ define measures $\alpha^v_{u,\underline{k}}$ on $(u,v]$ by the formula:

$$\alpha^v_{u,\underline{k}} = ((x_1, \underline{k}_1)_*(\lambda^v_{u,n}))^{|(u,v]\times\underline{k}}$$

where $n = n(\underline{k})$.

Intuitively, $\alpha^v_{u,\underline{k}}(B)$, for a measurable B in $(u,v]$, is the probability that a population with n members 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 local structure $\underline{k}$.

For a non-degenerate, right continuous sub-Markov process λ^*_* define measures $\alpha_{u,\underline{k}}$ on $(u,t]$ setting:

$$\alpha_{u,\underline{k}}(B) = \int_{x\in B\cap(u,e_n(u))} (v^v_{x,k})^{-1} d\alpha^v_{u,\underline{k}}$$

where $k = tr(\underline{k})$ and $n = n(\underline{k})$. This measure is well defined since $v^v_{x,k} \ge h^k(x,v) > 0$ for $v < e_n(u)$ and as a function of x it is measurable by Theorem 1.2.12 and Lemma 1.2.6.

??? Theoretically, it might be unbounded. The definition needs to be changed. Instead of v one should write $e_n(u)$ and one should add an extra term of the form $\alpha^v_{u,\underline{k}}(\{v\})$ with $v = e_n(u)$ if $v \in B$. Correspondingly the proof has to be adjusted.

Let $n = n(\underline{k})$ and let $\lambda^v_{u,\underline{k}}$ denote the co-restriction of $\lambda^v_{u,n}$ to $\Delta^{\underline{k}}_{u,v} = (u,v)$.

Theorem 1.2.15 *[***th2***] For λ^*_* as above and any $\underline{k} \in \tilde{S}(\mathbf{N})$ with $n(\underline{k}) = n$ and $tr(\underline{k}) = k$ one has*

$$[\mathbf{th2eq0}]\lambda^{v,\underline{k}}_u = (\alpha_{u,\underline{k}})^{|(u,v)} * h^k(-,v) \tag{13}$$

Proof: For convenience we will consider (13) as an equality of two measures on $[u, v)$ which are zero on $\{u\}$.

Lemma 1.2.16 *[**th2l1**] For any sub-Markov process λ_*^*, any $\underline{k} \in \tilde{S}(\mathbf{N})$ and any $s \le u < y < y_+ < v \le t$ one has*

$$[\mathbf{th2eq1}]\lambda_{u,\underline{k}}^v([y, y_+)) = \lambda_{u,\underline{k}}^{y_+}([y, y_+))h^k(y_+, v) \tag{14}$$

$$[\mathbf{th2eq2}] \quad \begin{aligned} &\lambda_{u,\underline{k}}^{y_+}([y, y_+))\upsilon_{y_+,k}^v = \\ &= \lambda_{u,n}^v(\{\Gamma \in H[u, v] \,|\, (x_1, \underline{k}_1)(\Gamma) \in [y, y_+) \times \{\underline{k}\} \text{ and } x_1(R(\Gamma)) > y_+\}) \end{aligned} \tag{15}$$

Proof: Equation (14) follow from Lemma 1.2.2 with $U_1 = \{[y, y_+) \subset \Delta_{u,y_+}^{\underline{k}}\}$ and $U_2 = \Delta_{y_+,v}^{k[1]}$.

Equation (15) follow from Lemma 1.2.2 with $U_1 = \{[y, y_+) \subset \Delta_{u,y_+}^{\underline{k}}\}$ and $U_2 = H[y_+, v]_{k,*}$. $\square$

Suppose that $v \ge e_k(u)$. Then the right hand side of (13) is zero and (14) implies that the left hand side of (13) is zero. Assume that $v < e_k(u)$. Then for all $x \in [u, v)$ one has $\upsilon_{x,k}^v \ge h^k(x, v) \ge h^k(u, v) > 0$ and (13) is equivalent to the assertion that for all $w \in [u, v)$ one has

$$[\mathbf{th2eq3}]\alpha_{u,\underline{k}}^v([u, w)) = \int_{x \in [u,w)} (h^k(x, v))^{-1} \upsilon_{x,k}^v \, d\lambda_{u,\underline{k}}^v \tag{16}$$

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

Lemma 1.2.17 *[**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 1.2.16 we have

$$\begin{gathered} f(x_{i+1})\lambda_{u,\underline{k}}^v([x_i, x_{i+1})) = \\ = \lambda_{u,n}^v(\{\Gamma \in H[u, v] \,|\, (x_1, \underline{k}_1)(\Gamma) \in [x_i, x_{i+1}) \times \{\underline{k}\} \text{ and } x_1(R(\Gamma)) > x_{i+1}\}) \end{gathered}$$

Therefore

$$\begin{gathered} \alpha([x_i, x_{i+1})) - f(x_{i+1})\lambda([x_i, x_{i+1})) = \\ = \lambda_{u,n}^v(\{\Gamma \in H[u, v] \,|\, (x_1, \underline{k}_1)(\Gamma) \in [x_i, x_{i+1}) \times \{\underline{k}\}, \text{ and } x_1(R(\Gamma)) \le x_{i+1})\}) \end{gathered}$$

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

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

Since

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

we conclude by σ-additivity of $\lambda^v_{u,n}$ 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. $\Box$

Let $(h^k(u,w))^{-1}=C<\infty$.

Lemma 1.2.18 *[***th2l3***] The function $f(x)$ is a right continuous function of bounded variation with values in the interval $[1,C]$.*

Proof: The range of values of f is inside $[1,C]$ in view of our definition of C and the fact that $\upsilon^v_{x,k}\leq 1$. By Theorem 1.2.12 we know that f is the ration of two right continuous functions and therefore is right continuous. The ration of two functions of bounded variation such that the function in the denominator is separated from zero is of bounded variation. The function $h^k(x,v)$ is increasing in x and therefore is of bounded variation. It remains to prove that $\upsilon^v_{x,k}$ is of bounded variation in x on $[u,w]$. Let $u=x_0<\ldots x_n=w$ be a partition of $[u,w]$. We have

$$\sum_{i=0}^{n-1}|\upsilon^v_{x_i,k}-\upsilon^v_{x_{i+1},k}|\leq\sum_{i=0}^{n-1}|\upsilon^v_{x_i,k}-h^k(x_i,x_{i+1})\upsilon^v_{x_{i+1},k}|+\sum_{i=0}^{n-1}|(1-h^k(x_i,x_{i+1}))\upsilon^v_{x_{i+1},k}|$$

Using the sub-markov property we can write the first sum as

$$\sum_{i=0}^{n-1}|\upsilon^v_{x_i,k}-h^k(x_i,x_{i+1})\upsilon^v_{x_{i+1},k}|=\sum_i h^k(u,x_i)^{-1}\sum_{tr(\underline{k})=k}\alpha^v_{u,\underline{k}}(x_i,x_{i+1}]\leq C\sum_{tr(\underline{k})=k}\alpha^v_{u,\underline{k}}([u,w])\leq C$$

and estimate the second sum as

$$\sum_{i=0}^{n-1}|(1-h^k(x_i,x_{i+1}))\upsilon^v_{x_{i+1},k}|\leq\sum_i(1-h^k(x_i,x_{i+1}))=$$

$$=\sum_i(h^k(u,x_i)-h^k(u,x_{i+1}))h^k(u,x_i)^{-1}\leq C-1$$

$\Box$

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 $inf_{x\in[x_i,x_{i+1}]} f(x)\alpha([x_i, x_{i+1}))$ and $sup_{x\in[x_i,x_{i+1}]} f(x)\alpha([x_i, x_{i+1}))$ it is sufficient to verify that

$$[\mathbf{th2eq4}] inf \left\{ \sum_i \mid sup_{x\in[x_i,x_{i+1}]} f(x) - inf_{x\in[x_i,x_{i+1}]} f(x) \mid \alpha([x_i, x_{i+1})) \right\} = 0 \qquad (17)$$

Lemma 1.2.19 *[th2l4] Let f be a right continuous function of bounded variation 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 of bounded variation we can write it as a sum $f = f_1 + f_2 + f_3 + f_4$ where f_1, f_2 are continuous and f_3, f_4 are right continuous step functions with countable sets of points of discontinuity (see e.g. [1]). In addition f_1 and f_3 are monotone increasing and f_2 and f_4 are monotone decreasing. Let us prove the lemma for f_1 and f_3. The decreasing case is strictly parallel.

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_3 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 $|sup_{x\in[y,y_+]} f(x) - inf_{x\in[y,y_+]} f(x)| < \epsilon$. If $(y, y_+] \subset [u, w)\backslash\{a_1, \ldots, a_N\}$ but $y \in \{a_1, \ldots, a_N\}$ we still have $|sup_{x\in[y,y_+]} f(x) - inf_{x\in[y,y_+]} f(x)| < \epsilon$ due to the fact that f is right continuous. □

To prove (17) we have to show that for any $\epsilon > 0$ there exists a partition such that

$$[\mathbf{th2eq5}] \sum_i \mid sup_{x\in[x_i,x_{i+1}]} f(x) - inf_{x\in[x_i,x_{i+1}]} f(x) \mid \alpha([x_i, x_{i+1})) < \epsilon \qquad (18)$$

Let

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

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

Using Lemma 1.2.19 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 α we can choose δ' such that

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

Elementary computation shows that for such a partition (18) is satisfied. □

Remark 1.2.20 We will show below (see Example **??**) that there are right continuous non-degenerate sub-Markov processes for which the functions $v^v_{x,k}$ are not of bounded variation in x on $[u, v]$.

Proposition 1.2.21 *[***pr4***] Let λ^*_* be a right continuous non-degenerate sub-Markov process. Then the mapping*

$$u \mapsto \sum_{\underline{k}} \alpha_{u,\underline{k}} \otimes \delta_{\underline{k}}$$

defines a sub-probability kernel from $[s, t)$ to $(s, t] \times \tilde{S}(\mathbf{N})$.

Proof: ??? $\square$

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

Definition 1.2.22 *[***com***] A process on $\mathcal{H}[s, t]$ is called co-measurable if for all $s \leq v \leq t$ the mappings*

$$u \mapsto j^v_{s,u} \circ \lambda^v_u$$

are kernels from $[s, v]$ to $H[s, v]$.

Remark 1.2.23 The name co-measurable is chosen to avoid confusion with standard notion of a measurable process. See e.g. [3].

Lemma 1.2.24 *[***arecom***] Any additive Markov process λ_* on $\mathcal{H}[s, t]$ is co-measurable.*

Proof: For an additive process to be co-measurable it is sufficient for the measures $\lambda_{1,u}$ considered as measures on $H[s, t]$ to form a kernel from $[s, t)$. In view of Lemma 1.2.1 it is sufficient to verify that the functions

$$f_1 : u \mapsto \lambda_{1,u}(\Delta^{[1]}_u)$$

$$f_2 : u \mapsto \lambda_{1,u}(\Delta^{[n]}_u) \text{ for } n \neq 1$$

and

$$f_3 : u \mapsto \lambda_{1,u}(H[u, t] \cap ((w_1, w_2), w_3, U))$$

are measurable. For any Markov process which satisfies (M2) 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

$$\text{for}, u \in I_1 \text{ one has } H[u, t] \cap ((w_1, w_2), w_3, U) = ((w_1, w_2), w_3, U),$$

$$\text{for}, u \in I_2 \text{ one has } H[u, t] \cap ((w_1, w_2), w_3, U) = ((u, w_2), w_3, U),$$

$$\text{for}, u \in I_3 \text{ one has } H[u,t] \cap ((w_1, w_2), w_3, U) = \emptyset.$$

Using Markov property we conclude that

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

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 u in this intersection we have

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

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

Note that all the maps which participate in the definition of $\mathcal{H}[s,t]$ are homomorphisms of monoids. A process λ_*^* on $\mathcal{H}[s,t]$ is called an additive process if $\lambda_u^v(\{0\}, \Delta^{*0}) = 1$ for all u, v and the kernels $\lambda_u^v : \mathbf{N} \to H[u,v]$ are homomorphisms of monoids. If λ is additive then

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

and

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

Example 1.2.25 [**nonrc**] Consider a process λ on $\mathcal{H}[s,t]$ such that the measures $\lambda_{u,k}^v$ are concentrated on $\Delta_{u,v}^{k[1]}$. Such a process is simply a collection of functions $\upsilon_{u,*}^v$ on $\mathbf{N}$. It is additive if and only if $\upsilon_{u,k}^v = (\upsilon_{u,1}^v)^k$ and it is sub-Markov if and only if $\upsilon_{u,k}^v \upsilon_{v,k}^w = \upsilon_{u,k}^w$.

Set $\upsilon_{u,n}^u = 1$, $\upsilon_{u,0}^v = 1$ and $\upsilon_{u,n}^v = 0$ for $v > u$ and $n > 0$. This gives us an example of a non-degenerate, additive sub-Markov process such that the functions $\upsilon_{u,n}^v$ are right continuous in u but not in v.

Let $x \in (s,t)$ and set $\upsilon_{u,0}^v = 1$ and for $n > 0$, $\upsilon_{u,n}^v = 0$ if $u \leq x$ and $\upsilon_{u,n}^v = 1$ if $u > x$. This defines a degenerate additive, sub-Markov process for which functions $\upsilon_{u,n}^v$ are right continuous in v but not in u.

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

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

be a sub-probability kernel such that for any $u \in [s,t)$ the measure $\alpha(u)$ is concentrated on $(u,t]$.

By [7,] there exists a measurable space $(\Omega, \mathfrak{F})$ and a probability measure P on it together with a measurable map

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

such that

$$[\mathbf{aalpha}]A(x,-)_*(P)=\alpha(x). \qquad (19)$$

Let $A_u : \Omega \to ((s,t]\times \mathbf{N}_{\neq 1}) \amalg pt$ be the restriction of A to $\{u\}\times\Omega$. Since $\alpha(x)$ is concentrated on $(x,t]\times \mathbf{N}_{\neq 1}$ we may assume that

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

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

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

and for $N>0$

$$X^v_{u,n,N}=\left\{\underline{\omega}\in\Omega^\infty | \forall 1\le i\le n\ (A_u(\omega_1)\in([v,t]\times\mathbf{N}_{\neq 1})\amalg pt)\ or\ ((\omega_{i+n},\omega_{i+2n},\dots)\in X^v_{A(\omega_i),N_1})\right\}$$

Set

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

Lemma 1.3.1 *[**simpl7**] The subsets $X^v_{u,n,N}(A)$ and $X^v_{u,n}$ are measurable.*

Proof: Straightforward. □

We set $X^u_{u,n}=\Omega^\infty$ and $X^v_{u,0}=\Omega^\infty$.

Example 1.3.2 [**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

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

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

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

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 process on $\mathcal{H}[s,t]$.

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

Lemma 1.3.3 *[q0] For any $s \le u < v \le t$ one has*

$$\mu^v_{u,0}(\Delta^{*_0}_{u,v}) = 1$$

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

Proof: We have

$$(M^v_{u,0})^{-1}(\Delta^{*_0}_{u,v}) = X_{0,u} = \Omega^\infty$$

which proves the first equality. We have

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

Therefore

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

Finally

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

which proves the last equality. $\square$

Lemma 1.3.4 *[n1] For any $u < w_1 < w_2 < w_3 < v$ and any measurable $U \subset j^v_{u,w_3}(H[w_3,v]_{n,*})$ one has*

$$\mu^v_{u,1}((w_1,w_2),w_3,U) = \int_{x \in (w_1,w_2)} \mu^v_{x,n}(U) d\alpha_n(u)$$

Proof: It follows from the fact that

$$(M^v_{u,1})^{-1}((w_1,w_2),w_3,U) =$$

$$= \left\{\underline{\omega} \in \Omega^\infty |\; A_u(\omega_1) \in (w_1,w_2) \times \{n\},\, (\omega_2,\dots) \in X^v_{A_u(\omega_1)},\, M^v_{A(\omega_1)}(\omega_2\dots) \in U\right\}$$

$\square$

Lemma 1.3.5 *[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

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

Proof: Follows immediately from the definition of $M^v_{u,n}$ for $n > 1$. $\square$

As an immediate corollary from Lemma 1.3.5 we see that the processes μ^*_* are additive.

Lemma 1.3.6 *[***mu0***] Let μ^*_* be one of the processes constructed above. Then for any $u \in [s,t]$, $h(u,u) = 1$ and $h(u,v)$ is a monotone decreasing, right continuous function in v from $[u,t]$ to $[0,1]$ i.e.*

$$h(u, v_1) \geq h(u, v_2) \text{ for } v_1 \leq v_2$$

and

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

Proof: It follows immediately from Lemma 1.3.3. $\square$

Lemma 1.3.7 *[***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 1.3.8 *[***upprop***] Let μ^*_* be one of the processes constructed above. Then it is right continuous.*

Proof: The first condition is obvious from the construction. The second follows from the fact that

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

and Lemma 1.3.7 by σ-additivity of $P^{\otimes\infty}$. $\square$

Our next goal is to show that our construction gives all non-degenerate additive sub-Markov processes on $\mathcal{H}[s,t]$.

Theorem 1.3.9 *[***th2***] Let α be a sub-probability kernel as above. Then the process μ^*_* is a sub-Markov process in the sense of [7,] if and only if the following conditions hold:*

1. *for all $s \leq u < v < t$ one has*

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

2. for all $s \leq u < v < t$ and $n \neq 1$ one has

$$\alpha_n(u)^{|(v,t]} = (1 - \sum_{k\neq 1} \alpha_k(u,(u,v]))\,\alpha_n(v)$$

Proof: "Only if" Suppose that $(\mu_u)_{u\in[s,t]}$ is a Markov process. As was noted above the condition (M1) is equivalent to the condition that α is admissible. Let us show that (2) and (3) hold. To get (2) consider the condition of Lemma 1.2.2 for $U_1 = \Delta^{[1]}_{u,v}$ and $U_2 = \Delta^{[1]}_{v}$ and apply Lemma 1.3.3 for $n = 1$.

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

$$\alpha_n(u,(v,w]) = (1 - \sum_{k\neq 1} \alpha_k(u,(u,v]))\,\alpha_n(v,(v,w])$$

which we get from immediately from Lemma 1.2.2 applied to

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

$$U_2 = \left\{\Gamma \in H[u,t]_{1,*} | \Gamma \neq \Delta^{[1]}_{u},\, w \geq x_1(\Gamma) > v, k_1(\Gamma) = n\right\}$$

and Lemma **??** and Lemma 1.3.3 for $n = 1$. The "only if" part is proved.

"If" We need to verify the condition of Lemma 1.2.2. Using additivity one can easily see now that if (4) holds for all $U_1 \subset H[v,t]_{1,\leq q}$ then it holds for all m and all $U_1 \subset H[v,t]_{m,\leq q}$. Therefore we may proceed by induction on q and for each q we only need to consider the case $m = 1$.

Lemma 1.3.10 *[**jmul**] Suppose that α satisfies conditions (2), (3). Then for any $s \leq u \leq v < t$ one has*

$$\mu_{1,u}^{|H[v,t]_{1,*}} = (1 - \sum_{k\neq 1} \alpha_k(u,(u,v]))\,\mu_{1,v}$$

Proof: For $\mu_{1,u}^{|H[w,t]_{1,>0}}$ it follows immediately from Lemmas 1.2.1 and 1.3.4 and condition (3). For $\mu_{1,u}(\Delta^{[n]}_u)$ and $n \neq 1$ from Lemma 1.3.3 and condition (3) and finally for $\mu_{1,u}(\Delta^{[1]}_u)$ from Lemma 1.3.3 and condition (2). $\square$

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

$$(res_{u,v} \times res_{v,t})^{-1}(U_1 \times U_2) = U_2$$

with respect to the identification $j_{u,v,t}$ of $H[v,t]$ with a subspace of $H[u,t]$. Condition (4) follows now immediately from Lemma 1.3.10 and Lemma 1.3.3 for $n = 1$.

Let $U_1 = \Delta^{[n]}_{u,v}$ where $n \neq 1$. Then

$$(res_{u,v} \times res_{v,t})^{-1}(U_1 \times U_2) = \{\Gamma \in H[u,t]_{1,>0} | x_1(\Gamma) = v, k_1(\Gamma) = n, R(\Gamma) \in U_2\}$$

It follows by an obvious limit argument from Lemma (1.3.4) that the value of $\mu_{1,u}$ on this subset is $\alpha_n(u, \{v\})\mu_{n,v}(U_2)$ and (4) follows from Lemma 1.3.3.

Let $q > 0$. Assume by induction that (4) is known for all $U_1 \subset H[v,t]_{*,<q}$ and all U_2 and let $U_1 \subset H[v,t]_{1,q}$. By Lemma 1.2.1 we may assume that $U_1 = ((w_1, w_2), w_3, U_1')$ where $u < w_1 < w_2 < w_3 < v$ and U_1' is a measurable subset of $H[w_3, v]_{m,q-1}$ for some $m \neq 1$. Then

$$[\mathbf{ss}](res_{u,v} \times res_{v,t})^{-1}(U_1 \times U_2) = ((w_1, w_2), w_3, (res_{u,v} \times res_{v,t})^{-1}(U_1' \times U_2)) \tag{20}$$

By the inductive assumption

$$\mu_{m,x}((res_{u,v} \times res_{v,t})^{-1}(U_1' \times U_2)) = \mu_{m,x}(res_{u,v}^{-1}(U_1'))\mu_{l,v}(U_2)$$

where l is such that $U_2 \subset n_v^{-1}(\{l\})$. By Lemma 1.3.4, the value of $\mu_{1,u}$ on (20) is

$$\int_{x\in(w_1,w_2)} \mu_{m,x}((res_{u,v} \times res_{v,t})^{-1}(U_1' \times U_2))d\alpha_m(u) =$$

$$= \left(\int_{x\in(w_1,w_2)} \mu_{m,x}(res_{u,v}^{-1}(U_1'))d\alpha_m(u)\right)\mu_{l,v}(U_2)$$

and using Lemma 1.3.4 again we get (4). Theorem is proved. □

Lemma 1.3.11 *[***d3uv***] Let α and μ be as above. Then*

$$[\mathbf{avtoa}]\alpha_{u,k}^v = \upsilon_{*,k}^v * \alpha_k(u) = (\upsilon_{*,1}^v)^k * \alpha_k(u) \tag{21}$$

Proof: Equation (21) is equivalent to the assertion that for a measurable B in $(u, v]$ one has

$$\alpha_{u,k}^v(B) = \int_{x\in B} \upsilon_{x,k}^v d\alpha_k(u)$$

We have

$$\alpha_{u,k}^v(B) = \mu_{u,1}^v(\{\Gamma \in H[u,v]_{1,*} | \, x_1(\Gamma) \in B \, k_1(\Gamma) = k\}) =$$

$$= P^{\otimes\infty}(\underline{\omega} \in X_{u,1}^v, A(\omega_1) \in B \times \{k\}) = P^{\otimes\infty}(\underline{\omega} \in \Omega^\infty, A(\omega_1) \in B \times \{k\}, (\omega_2, \dots) \in X_{A(\omega_1)}{}^v)$$

since $P^{\otimes\infty}(X_{x,k}^v) = \upsilon_{x,k}^v$ the claim of the lemma follows from the usual formula for the value of products measures. □

Definition 1.3.12 *[***admiss***] The map A is called admissible if $P^{\otimes\infty}(X_{u,n}^v) = 1$ for all u, v, n.*

We will say that the original kernel α is admissible if there exists A satisfying (19) which is admissible. Note that α is admissible if and only if the corresponding process μ_*^* satisfies the condition (P) of [7, ...].

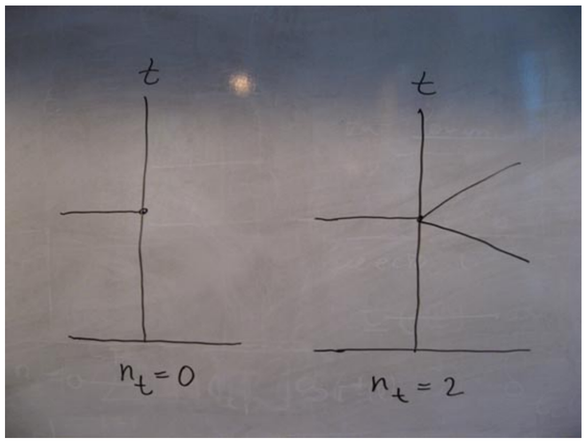
Proposition 1.3.13 *[***adm1***] Suppose that $\alpha_k = 0$ for all but finitely many k. Then α is admissible.*

Proof: ??? □

1.4 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 1.4.1 *[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 1.4.2 *[**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

$$[\mathbf{eq1}] a_*(P_u(k,-) \otimes P_u(l,-)) = P_u(k+l,-) \tag{22}$$

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

Proposition 1.4.3 *[**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

$$[\mathbf{ca}] \lim_{\epsilon \geq 0, \epsilon \to 0} n_{u+\epsilon}(\Gamma) = n_u(\Gamma) \tag{23}$$

Remark 1.4.4 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 1.4.5 *[**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 1.4.5, 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

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

such that the collection $P_{u,v} = n_v P_u : \mathbf{N} \to \mathbf{N}$ 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 1.4.6 *[**condA**] For any $s \le u \le v \le t$ 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.

1.5 Branching Markov processes on N

The dynamics of the population which consists identical individuals is fully described by a collection of probability kernels $P_{u,v} : \mathbf{N} \to \mathbf{N}$ given for all $u \le v$, $u, v \in [s, t]$. 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 $u \le v \le w$ 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

$$[\mathbf{eform}]F(u,v;x) = \sum_{n=0}^{\infty} P_{u,v}(1,n)x^n. \tag{24}$$

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

$$[\mathbf{mcomp}]F(u,w;x) = F(u,v;F(v,w;x)). \tag{25}$$

This description provides a bijection between collections of formal power series $F(u,v;x)$ of the form (24) satisfying the conditions

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

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

and (25) and the isomorphism classes of branching Markov processes on $\mathbf{N}$. We let $BM(\mathbf{N};s,t)$ denote this set of isomorphism classes.

1.6 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 1.6.1 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 $\bar{E}[s,t]\rightarrow\bar{E}[u,v]$ 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 $n_u:\bar{E}[s,t]\rightarrow\mathbf{N}$ 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 $(\mathbf{N},\bar{E}[s,t],\mathfrak{S}_u^v,n_u)$ 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 $P_u:\mathbf{N}\rightarrow(\bar{E}[s,t],\mathfrak{S}_u^t)$ are homomorphisms of monoids.

Proposition 1.6.2 *[***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 $\rho:O\rightarrow S^{\infty}\mathbf{N}$ 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 (23) shows that for any $\Gamma\in H[s,t]$ one has $e(\Gamma)\in E[s,t]$.

Let $\mathfrak{R}_t^s$ 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 $(V,\mathfrak{R}_0^s)$ 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 1.6.3 *[***pr1***] The collection of data $(n_t,\mathfrak{R}_t^s,P_{s,m})$ 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 $(n_t,\mathfrak{R}_t^s,P_{s,m})$ has a canonical extension to a process on a wider set of σ-algebras with respect to which r is measurable. Let $\mathfrak{S}_t^s=r^{-1}(\mathfrak{R}_t^s)$ be the smallest σ-algebra which makes the map r measurable with respect to the σ-algebra $\mathfrak{R}_t^s$ 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 $\mathfrak{T}_t^s=\mathfrak{R}_t^s+\mathfrak{S}_t^s$.

Corollary 1.6.4 *[**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 (29):

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

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

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

$$[\mathbf{eq4}]\phi_t(\phi_t^{-1}(x)) = \phi_t^{-1}(\phi_t(x)) = Id. \tag{26}$$

Set

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

The equations (26) imply immediately that the series $\tilde{F}$ satisfy the relations (25) 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 1.6.5 *[**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 1.6.4. □

Let's write $\phi_s^*(n) = \sum_k a_k\delta_k$ where δ_k is the δ-measure concentrated at k. By Corollary 1.6.4 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 $s \le t_1 \le \cdots \le t_q$. 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 1.6.4

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

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

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

Using again formal power series we get the following result.

Lemma 1.6.6 *[***fc1***] The value of $P_s(n)[S_{t_1,n_1} \cap \cdots \cap S_{t_q,n_q}]$ is the coefficient at $x_1^{n_1}\ldots x_q^{n_q}$ in the expression $(F(s,t_1;\phi_{t_1}(x_1))^n\tilde{F}(t_1,t_2;x_2)^{n_1}\ldots\tilde{F}(t_{q-1},t_q;x_q)^{n_{q-1}}$.*

1.7 Death free histories

A singleton history is called *death free* if $\psi^{-1}(0) = \emptyset$ and for any v such that $\phi(v) < t$ there exists at least one edge starting in v. We let $\tilde{H}[s,t]$ denote the set of death free histories over $[s,t]$. Given a general singleton history we can "reduce" it to a death free history by removing all v with $\phi(v) = t$ and $\psi(v) = 0$ and with $\phi(v) < t$ and no edges starting at v. This gives us a projection

$$r : H[s,t] \to \tilde{H}[s,t]$$

for which the natural inclusion $\tilde{H}[s,t] \to H[s,t]$ is a section. From the population point of view r corresponds to the passage from the full genealogy of a population to the ancestral genealogy of the present day survivors.

Note that $\tilde{H}[s,t]$ is also a commutative monoid with respect to the disjoint union of histories and that both the inclusion and the reduction map are homomorphisms of monoids.

and $\tilde{F}^*(\mathbf{N})$.

$\tilde{F}$ takes a monoid A to the free monoid generated by $(A,1)$ as a pointed set, e.g. $\tilde{F}(pt) = pt$.

$$\tilde{H}[s,t] = |\tilde{F}_*(\mathbf{N})|_{[s,t]}$$

Let us consider only the case of $H[s,t]$. The case of $\tilde{H}[s,t]$ follows by the same scheme.

Corollary 1.7.1 *[**homot**] The space $H[s,t]$ and $\tilde{H}[s,t]$ are are homotopy equivalent to $\mathbf{N}$. A history Γ belongs to the connected component given by the number of final vertices with multiplicities defined b ψ.*

Proof: It follows by [5,] from the fact that the monad F_* (resp. $\tilde{F}$) is given by the composition of the forgetful functor to sets (resp. pointed sets) with its left adjoint. □

The maps $\tilde{H}[s,t] \to H[s,t]$ and $H[s,t] \to \tilde{H}[s,t]$ correspond with respect to the identifications of Proposition 1.1.1 to the natural homomorphisms of monads $F \to \tilde{F}$ and $\tilde{F} \to F$ which proves the following result.

Corollary 1.7.2 *[**rcont**] The maps*

$$\tilde{H}[s,t] \to H[s,t]$$
$$H[s,t] \to \tilde{H}[s,t]$$

are continuous.

The simplicial set $\tilde{F}_*$ to the contrary is locally finite. Moreover, one has the following result.

Proposition 1.7.3 *[**topstr**] The space $\tilde{H}[s,t]$ is the disjoint union of the form*

$$\tilde{H}[s,t] = \amalg_{n\geq 0} \tilde{H}_n[s,t]$$

where $\tilde{H}_n[s,t]$ is the subset of histories with the n survivors. The space $\tilde{H}_0$ consists of one point corresponding to the empty history. For $n > 0$, the space $\tilde{H}_n[s,t]$ is a finite contractible CW-complex of dimension $n-1$.

Proof: ??? □

Picture of small parts here?

Proposition 1.7.4 *[**str1**] Every point of $\tilde{H}_n[s,t]$ lies in the closure of a simplex of dimension $n-1$. A point of $\tilde{H}_n[s,t]$ belongs to the interior of a simplex of dimension $n-1$ (i.e. has level $n-1$) if and only if the corresponding history Γ has the following properties:*

1. *there are exactly n vertices v with $\phi(v) = t$ (i.e. $\psi \equiv 1$),*

2. *for any v such that $\phi(v) \neq s, t$ there exists exactly two edges starting in v,*

3. *for any v_1, v_2 such that $\phi(v_1) = \phi(v_2) \neq t$ one has $v_1 = v_2$, in particular there is exactly one vertex v with $\phi(v) = s$.*

Proof: ??? $\square$

We will call histories which satisfy the conditions of Proposition 1.7.4 *generic histories* and denote their space by $\tilde{B}[s,t]$ since they are the ones with only binary ramification points. The proposition shows that $\tilde{B}[s,t]$ is naturally homeomorphic to the disjoint union of open $[s,t]$-simplexes and that it is dense in $\tilde{H}[s,t]$.

The natural map $\tilde{B}_n[s,t] \rightarrow \Delta_{op}^{n-1}[s,t]$ assigns to each history Γ its sequential invariant - the sequence $u_1, \ldots, u_{n-1}$ of the times of the division events in Γ.

1. structure of $\tilde{H}_n$ for small n,

2. for a generic Γ the neighborhood $U_\epsilon(\Gamma)$ isomorphic to I_ϵ^{n-1} where $I_\epsilon = [-\epsilon, \epsilon]$,

3. death free histories can be equivalently described as finite ultra-metric spaces whose metrics are allowed to take value $+\infty$ and to be degenerate (i.e. one may have $d(x,y) = 0$ for $x \neq y$). The level of such a space is the number of values in $(0, \infty)$ which the metric takes and the sequential invariant is the set of these values in the increasing order.

1.8 Reduced processes

Proposition 1.8.1 *[**p1**] For any additive Markov process $(n_t, \mathfrak{R}_t^s, P_{s,m})$ there exists a unique additive Markov process $(n_t, \mathfrak{T}_t^s, P'_{s,m})$ such that the restriction of $P'_{s,m}$ to $\mathfrak{R}_t^s$ equals $P_{s,m}$ and for $t \geq s$ one has*

$$[\mathbf{eq3}] P'_{s,1}(R_{t,k} \cap S_{t,n}) = P'_{s,1}(R_{t,k}) P'_{t,k}(S_{t,n}). \tag{27}$$

Proof: We will only prove uniqueness i.e. we will show how to express $P'_{s,m}(S_{t,n})$ through $P_{s,m}$. Note first that

$$[\mathbf{eq2}] a^{-1}(S_{t,n}) = \amalg_{i+j=n} S_{t,i} \times S_{t,j} \tag{28}$$

The condition (22) implies that

$$P'_{s,k} = a_*(\otimes_{j=1}^k P'_{s,1})$$

and together with (28) we get

$$P'_{s,k}(S_{t,n}) = \sum_{i_1+\cdots+i_k=n} \prod_{j=1}^k P'_{s,1}(S_{t,i_j}).$$

We further have

$$[\mathbf{eq6}]P'_{s,1}(S_{t,n}) = \sum_{k\geq 0} P'_{s,1}(R_{t,k}\cap S_{t,n}) = \sum_{k\geq 0} P'_{s,1}(R_{t,k})P'_{t,k}(S_{t,n}) = \tag{29}$$

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

Observe now that $P'_{t,1}(S_{t,i})$ can be non-zero only for $i = 0, 1$ and that

$$P'_{t,1}(S_{t,0}) = P_{t,1}(R_{0,0})$$

$$P'_{t,1}(S_{t,1}) = 1 - P_{t,1}(R_{0,0})$$

which finishes the proof of the proposition. □

Remark 1.8.2 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 1.8.3 One verifies easily that there are histories Γ, Γ' such that $n_u(\Gamma) = n_u(\Gamma)'$ for all u but $n_u r(\Gamma) \neq n_u r(\Gamma')$ 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.

1.9 Parameters space for singleton processes

Definition 1.9.1 *[**abar**] For $s \leq t$ 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*

$$[\mathbf{ainineq}]\sigma' \geq -\sigma(1-\sigma) \tag{30}$$

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 1.9.2 *[***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 (30) which allows one to find the required δ. □

Proposition 1.9.3 *[***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 [6, 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

$$[\textbf{fsigma}]F(s,t;u) = 1 - \sigma(s)\frac{u-1}{(1-\phi(s,t))u + \phi(s,t) - 1 - \phi(s,t)\sigma(t)}. \tag{31}$$

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

$$\phi(s,t) \le \frac{1-\sigma(s)}{1-\sigma(t)}$$

and that this condition holds for any $\sigma \in \bar{A}[-T, 0]$. We denote the process (31) 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:

$$[\mathbf{eq21}]\frac{\partial F(t,0;u)}{\partial t} = -f(t, F(t,0,u)) \tag{32}$$

where $f(t,x) = p_2(t)x^2 + p_1(t)x + p_0(t)$ (see e.g. [6, 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

$$[\mathbf{eq22}]F = (1 - D(t))\tilde{F} + D(t). \tag{33}$$

where $D(t) = F(t,0;0)$. Substituting (33) in (32) and using the consequence

$$\frac{\partial D(t)}{\partial t} = -f(t, D(t))$$

of (32) 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 (32) 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

$$\sigma' = \sigma(\sigma + d - 1)$$

We can express d through σ and σ' using this equation and since $d \geq 0$ we conclude that σ must satisfy 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 25 we see (using again [6, 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. [6, Ex.9, p.46]) we get:

$$[\mathbf{m1}]\tilde{F}(s, t; u) = \frac{q(t)u}{(q(t) - q(s))u + q(s)} \tag{34}$$

where

$$q(t) = \exp(\int_t^0 \sigma(x)dx).$$

Let's write

$$[\mathbf{ared}]\tilde{F}(s, t; u) = \sum_k a_k(s, t)u^k \tag{35}$$

From (34) 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

$$F(N, \tilde{n}; t_0, \ldots, t_{q+1}) = P_{t_0}(N)[S_{t_0,\tilde{n}}, \ldots, S_{t_q,\tilde{n}+q}].$$

By Lemma 1.6.6 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

$$[\mathbf{bi}]B_i = \begin{cases} \int_{t_0}^{t_1-\epsilon}\sigma(x)dx & \text{for } i=0 \\ \int_{t_i+\epsilon}^{t_{i+1}-\epsilon}\sigma(x)dx & \text{for } i=1,q-1 \\ \int_{t_q+\epsilon}^{t_{q+1}}\sigma(x)dx & \text{for } i=q \end{cases} \tag{36}$$

and for $i=1,\dots,q$:

$$[\mathbf{ci}]C_i = \int_{t_i-\epsilon}^{t_i+\epsilon}\sigma(x)dx \tag{37}$$

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*

$$[\mathbf{cond1}]\sigma' \le -\sigma(1-\sigma) \tag{38}$$

exists if and only if

$$[\mathbf{asser1}]\sigma_1 \ge \frac{\sigma_0}{\sigma_0+(1-\sigma_0)e^{t_1-t_0}} \tag{39}$$

or equivalently

$$[\mathbf{asser2}]\sigma_0 \le \frac{\sigma_1}{\sigma_1+(1-\sigma_1)e^{t_0-t_1}} \tag{40}$$

and the equalities are achieved for a unique function

$$[\mathbf{s01}]\sigma(u) = \frac{\sigma_0}{\sigma_0+(1-\sigma_0)e^{u-t_0}} \tag{41}$$

Proof: The equivalence of (39) and (40) is obvious. Let σ be a function satisfying the conditions of the proposition. Let us show that (39) holds. If $\sigma_1 = 1$ then (39) is obvious. Therefore, we may assume that $\sigma_1 < 1$. Assume that for all x, $\sigma(x) > 0$. Set

$$[\mathbf{cp1eq2}]\phi(x) = -\frac{\sigma'}{\sigma(1-\sigma)}. \tag{42}$$

Then (38) implies that $\phi(x) \leq 1$. Solving (42) 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 (39). 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 (39) is an equality is (41).

Suppose now that $\sigma_1 \in [0,1]$ satisfies the strong version of (39). Let $\epsilon > 0$ be a sufficiently small number. Consider the function of the form (41) 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 (38). $\square$

??? 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*

$$[\mathbf{asser3}](1 + \sigma_1(e^{t_1 - t_0} - 1))^{-1} \leq e^{-\int_{t_0}^{t_1} \sigma(x)dx} \leq 1 + \sigma_0(e^{t_0 - t_1} - 1) \tag{43}$$

The equality is achieved in the class of smooth functions only if the equality holds in (39). In this case the only function which achieves the equality in any of the inequalities of (43) is (41) 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 (43). $\square$

2.3 Computation B

Set

$$[\textbf{fofsigma}]F(t_1,\ldots,t_{q+1};\epsilon)=e^{-C_1}(1-e^{-C_1})e^{-2B_1}e^{-2C_2}(1-e^{-C_2})\ldots e^{-qC_q}(1-e^{-C_q})e^{-(q+1)B_q} \quad (44)$$

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 $0 \leq y_1, \ldots, y_q \leq 1$ 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*

$$[\mathbf{conc}]\frac{y_i}{y_i + (1 - y_i)e^{t_{i+1} - t_i - 2\epsilon}} \leq \frac{y_{i+1}}{y_{i+1} + (1 - y_{i+1})e^{-2\epsilon}} \tag{45}$$

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 (44) where B_i and C_i are defined by (36) and (37) 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 (45) 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 $F(t_1, \ldots, t_{q+1}; \epsilon)$ 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 $y_1, \ldots, y_q$ satisfying

$$y_1 \geq 0$$

$$y_{i+1} \geq (1+\delta)y_i((1+\delta - r_{i+1})y_i + r_{i+1})^{-1} \text{ for i=1,\ldots,q}$$

$$1 \geq y_{q+1}$$

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

$$F = \delta^q \prod_{i=1}^{q} y_i M_i(y_i)^{i+1}$$

and the conditions

$$y_1 \geq 0$$

$$y_{i+1} \geq E_{i+1}^{-1}(y_i) \text{ for i=1,\ldots,q}$$

$$1 \geq y_{q+1}$$

we have

$$det(E_i) = r_i(1+\delta) > 0$$

which implies that $E_i(y)$ are increasing functions. Set

$$A_i = E_{i+1} \ldots E_q$$

and introduce new variables:

$$u_i = A_i^{-1}(y_i)$$

Then the function becomes

$$[\mathbf{ufun}]F = \delta^q \prod_{i=1}^{q} A_i(u_i) M_i(A_i(u_i))^{i+1} \tag{46}$$

and the inequalities become

$$[\mathbf{uineq}] 0 \le u_1 \le \cdots \le u_q \le 1 \tag{47}$$

i.e. we have to find maximums of (46) on the simplex (47). We have:

$$E_j E_{j+1} = \begin{pmatrix} r_j & 0 \\ r_j - (1+\delta) & 1+\delta \end{pmatrix} \begin{pmatrix} r_{j+1} & 0 \\ r_{j+1} - (1+\delta) & 1+\delta \end{pmatrix} = \begin{pmatrix} r_j r_{j+1} & 0 \\ r_j r_{j+1} - (1+\delta)^2 & (1+\delta)^2 \end{pmatrix}$$

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: ??? □

2.5 Computation for $\delta = 0$

Set $s_i = 1 - r_i \dots r_q$ since $r_j \le 1$ we have $1 > s_i \ge s_{i+1} \ge 0$ and any monotone decreasing sequence of s_i's may arise from a combinations of the event times $t_1 \le \cdots \le t_q$. 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. □

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 $P : (X, A) \to (Y, B)$ is a function $X \times B \to \mathbf{R}_{\geq 0}$ 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 $f : (X, A) \to (Y, B)$ 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 $\xi_u : \Omega \to X$ 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 $\xi_u : (\Omega, \mathfrak{S}_u^u) \to (X, A)$ 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*

$$P_u : (X,A) \to (\Omega, \mathfrak{S}_u^t)$$

such that $\xi_u \circ P_u = Id$ and for $u \le v$ the square

$$\begin{array}{ccc} (X,A) & \xrightarrow{P_u} & (\Omega,\mathfrak{S}_u^t) \\ {\scriptstyle P_{u,v}}\downarrow & & \downarrow \\ (X,A) & \xrightarrow{P_v} & (\Omega,\mathfrak{S}_v^t) \end{array} \tag{48}$$

where

$$P_{u,v} = \xi_v \circ P_u,$$

commutes.

One verifies easily that for a Markov process P and for $u \le v \le w$ one has

$$[\mathbf{comp0}] P_{u,u} = Id \tag{49}$$

$$[\mathbf{comp1}] P_{v,w} \circ P_{u,v} = P_{u,w} \tag{50}$$

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 (49) and (50). 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 $u \le v$ 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. □

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: ??? □

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. □

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] S. MacLane. *Homology.* Springer-Verlag, 1963.

[6] B. A. Sevastyanov. *Vetvyashchiesya protsessy.* Izdat. "Nauka", Moscow, 1971.

[7] Vladimir Voevodsky. Notes on categorical probability. *In preparation*, 2008.