Singletons

Vladimir Voevodsky

Started January 4, 2008

Contents

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 Singleton histories

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.

Need a picture here

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 $[s,t]$.

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 $R_u(\Gamma)$.

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

We will need topology on spaces $H[s,t]$ and $\tilde{H}[s,t]$ which we get by identifying histories and death free histories with the points of $[s,t]$-geometric realizations of two commutative simplicial monoids $F^*(\mathbf{N})$ and $\tilde{F}^*(\mathbf{N})$.

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_{nd}^i 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. [?]). If Δ_{op}^i is the open simplex for $i > 0$ and the point for $i = 0$ then there is a bijection of sets

$$|X_*| = \amalg_{i\geq 0}X_i^{nd} \times \Delta_{op}^i$$

Let $\Delta_{[s,t]}^i$ be the set of increasing sequences $u_1 < \cdots < u_i$ 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^*.

Recall further 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 two monads on the category of commutative monoids:

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

Proposition 1.1 *[**top**] There are natural isomorphisms of monoids*

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

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

where the spaces on the right hand sides are considered as sets of points.

Proof: Let us consider only the case of $H[s,t]$. The case of $\tilde{H}[s,t]$ follows by the same scheme. We are going to show that $H[s,t]$ can be identified with the set of points of the disjoint union

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

where $F_i^{nd}(X)$ is the set of on-degenerate simplicies of $F_*(X)$ and $\Delta_{(s,t)}^i$ is the point for $i = 0$ and the open simplex $s < u_1 < \cdots < u_i < s$ 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 $x_i \in X$.

Then

$$F_i(X) = S^{\circ(i+1)}(X)$$

where $S^{\circ i}$ is the i-th iteration of the functor $S(-)$. Elements of $\amalg_i F_i(X)$ which belong to $F_i(X)$ will be called elements of level i. We let $*_i$ denote the point $Symm^0(S^{\circ(i+1)}(X))$. 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 $i > 0$ is of the form $*_i$ or $\sum[\gamma_i]$ where γ_i are elements of level $i-1$.

Let us define a map $\pi : H[s,t] \to \amalg_i F_i(\mathbf{N})$ 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 $u_1 < \cdots < u_i$. The number i 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 $\gamma \in F_i(\mathbf{N})$ be an element of level i and $u_1 < \cdots < u_i$ be an increasing sequence in (s,t). Define $\Gamma = \Pi'(\gamma; u_1, \ldots, u_i) \in H[s,t]$ inductively as follows.

If $i = 0$ and $\gamma = *_0$ we set $\Gamma = \emptyset$. If $i = 0$ and $\gamma = \sum_{j=1}^m [n_j]$ 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_j.

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

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

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

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

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 $\Pi(\gamma; u_1, \ldots, u_i)$.

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

$$\Pi(\pi(\Gamma); u_1(\Gamma), \ldots, u_i(\Gamma))) = \Gamma.$$

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

$$\pi(\Pi(\gamma; u_1, \ldots, u_i)) = \gamma$$

which finishes the proof.

Corollary 1.2 *[**homot**] The spaces $H[s,t]$ 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 [**?**,] 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 to the natural homomorphisms of monads $F \to \tilde{F}$ and $\tilde{F} \to F$ which proves the following result.

Corollary 1.3 *[**rcont**] The maps*

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

are continuous.

The simplicial set $F_*(\mathbf{N})$ is not locally finite. For example, the vertex [1] is the boundary of any of the 1-simplexes of the form $n[*_0] + [1]$, as a consequence $H[s,t]$ is not a nice topological space. The simplicial set $\tilde{F}_*$ to the contrary is locally finite. Moreover, one has the following result.

Proposition 1.4 *[**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: ???

Proposition 1.5 *[**str1**] Every point of $\tilde{H}_n[s,t]$ lies in the closure of a simplex of dimension $n-1$. A point of $H_n[s,t]$ 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: ???

We will call histories which satisfy the conditions of Proposition 1.5 *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 $\tilde{B}_n[s,t] \to \Delta_{op}^{n-1}[s,t]$ 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.

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

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

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

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

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

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

$$[\mathbf{mcomp}]F(u,w;x) = F(u,v;F(v,w;x)). \tag{2}$$

This description provides a bijection between collections of formal power series $F(u,v;x)$ of the form (1) satisfying the conditions

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

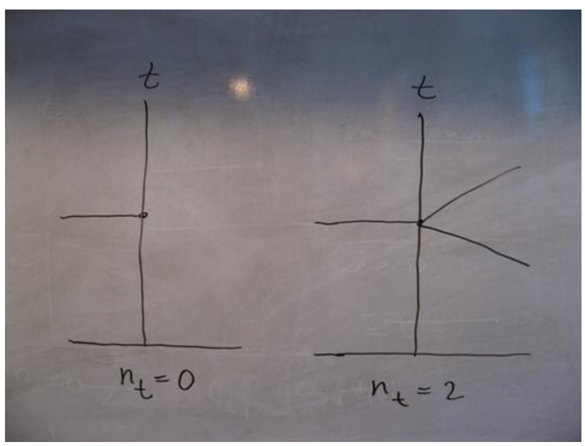
$$P_{u,v}(1,n) \geq 0$$

and (2) and the isomorphism classes of branching Markov processes on $\mathbf{N}$. We let $BM(\mathbf{N};s,t)$ denote this set of isomorphism classes.

3 Markov processes on histories

Let $u_1 \leq \cdots \leq u_q$ be a non-decreasing 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 3.1 *[**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 $s < v_1 \leq v_2 \leq t$.

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 3.2 *[**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

$$[\mathbf{eq1}] a_*(P_u(k, -) \otimes P_u(l, -)) = P_u(k + l, -) \tag{3}$$

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

Proposition 3.3 *[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: ???

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

$$[\mathbf{ca}] \lim_{\epsilon\geq 0,\epsilon\to 0} n_{u+\epsilon}(\Gamma)=n_u(\Gamma) \tag{4}$$

Remark 3.4 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 3.5 *[mes1] The functions n_u and the maps $c_{u,v}$ are measurable with respect to the Borel σ-algebras.*

Proof: ???

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 3.5, 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 [**?**, 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 $P_{u,v}=n_vP_u:\mathbf{N}\to\mathbf{N}$ 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 3.6 *[condA] For any $s\leq u\leq v\leq t$ one has*

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

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

4 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 4.1 The invariant $e(\Gamma)$ has a better behavior than a more simple invariant which assigns to Γ the function

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

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

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

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

Proposition 4.2 *[**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: ???

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

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

The property (4) 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 4.3 *[**pr1**] The collection of data $(n_t, \mathfrak{R}_t^s, P_{s,m})$ is a Markov process (in the sense of [**?**, 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 4.4 *[**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 (8):

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

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

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

$$[\mathbf{eq4}]\phi_t(\phi_t^{-1}(x)) = \phi_t^{-1}(\phi_t(x)) = Id. \tag{5}$$

Set

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

The equations (5) imply immediately that the series $\tilde{F}$ satisfy the relations (2) 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 4.5 *[**11**] 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 4.4.

Let's write $\phi_s^*(n) = \sum_k a_k \delta_k$ where δ_k is the δ-measure concentrated at k. By Corollary 4.4 we have

$$P_s(n)[S_{t_1,n_1} \cap \dots \cap S_{t_q,n_q}] = P_s(n)[r^{-1}(R_{t_1,n_1} \cap \dots \cap R_{t_q,n_q})] =$$
$$= \tilde{P}_s \phi_s^*(n)[R_{t_1,n_1} \cap \dots \cap R_{t_q,n_q}] = \sum_k a_k \tilde{P}_s(k)[R_{t_1,n_1} \cap \dots \cap R_{t_q,n_q}].$$

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

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

and therefore, again by Corollary 4.4

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

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

Using again formal power series we get the following result.

Lemma 4.6 *[**fc1**] The value of $P_s(n)[S_{t_1,n_1} \cap \cdots \cap S_{t_q,n_q}]$ is the coefficient at $x_1^{n_1} \dots x_q^{n_q}$ in the expression $(F(s,t_1;\phi_{t_1}(x_1))^n \tilde{F}(t_1,t_2;x_2)^{n_1} \dots \tilde{F}(t_{q-1},t_q;x_q)^{n_{q-1}}$.*

5 Reduced processes

Proposition 5.1 *[**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*

$$[\mathbf{eq3}]P'_{s,1}(R_{t,k} \cap S_{t,n}) = P'_{s,1}(R_{t,k})P'_{t,k}(S_{t,n}). \qquad (6)$$

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

$$[\mathbf{eq2}]a^{-1}(S_{t,n}) = \amalg_{i+j=n} S_{t,i} \times S_{t,j} \qquad (7)$$

The condition (3) implies that

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

and together with (7) we get

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

We further have

$$[\mathbf{eq6}]P'_{s,1}(S_{t,n}) = \sum_{k\geq 0} P'_{s,1}(R_{t,k} \cap S_{t,n}) = \sum_{k\geq 0} P'_{s,1}(R_{t,k})P'_{t,k}(S_{t,n}) = \qquad (8)$$

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

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 5.2 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 5.3 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.

6 Parameters space for singleton processes

Definition 6.1 *[**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*
$$[\textbf{maineq}]\sigma' \geq -\sigma(1-\sigma) \tag{9}$$
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 $\sigma_-(x) \leq \sigma(x)$
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 6.2 *[**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 $|f(x) - g(x)| \geq \epsilon$ implies that $dist(f,g) \geq \delta(\epsilon)$. 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 (9) which allows one to find the required δ.

Proposition 6.3 *[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 [**?**, 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

$$[\mathbf{fsigma}]F(s,t;u) = 1 - \sigma(s)\frac{u-1}{(1-\phi(s,t))u + \phi(s,t) - 1 - \phi(s,t)\sigma(t)}. \tag{10}$$

Simple computation shows that such a system of functions forms a process (i.e. that all the coefficients in the Taylor series in u are non-negative) iff

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

and that this condition holds for any $\sigma \in \bar{A}[-T,0]$. We denote the process (10) by F_σ.

2 Likelihood functional

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:

$$[\mathbf{eq21}]\frac{\partial F(t,0;u)}{\partial t} = -f(t, F(t,0,u)) \tag{11}$$

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

$$[\mathbf{eq22}] F = (1 - D(t))\tilde{F} + D(t). \tag{12}$$

where $D(t) = F(t,0;0)$. Substituting (12) in (11) and using the consequence

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

of (11) 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 (11) 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 2 we see (using again [**?**, 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. [**?**, Ex.9, p.46]) we get:

$$[\mathbf{m1}]\tilde{F}(s,t;u) = \frac{q(t)u}{(q(t) - q(s))u + q(s)} \tag{13}$$

where

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

Let's write

$$[\mathbf{ared}]\tilde{F}(s,t;u) = \sum_k a_k(s,t)u^k \tag{14}$$

From (13) 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

$$\begin{array}{cccccccccc} t_0, & t_0, & t_1-\epsilon, & t_1+\epsilon, & t_2-\epsilon, & t_2+\epsilon, & \dots, & t_q-\epsilon, & t_q+\epsilon, & t_{q+1} \\ N, & \tilde{n}, & \tilde{n}, & \tilde{n}+1, & \tilde{n}+1, & \tilde{n}+2, & \dots, & \tilde{n}+q-1, & \tilde{n}+q, & \tilde{n}+q \end{array}$$

where ϵ is sufficiently small such that the sequence of $t's$ is an increasing one. We want to compute

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

By Lemma 4.6 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

$$[\mathbf{bi}]B_i = \begin{cases} \int_{t_0}^{t_1-\epsilon}\sigma(x)dx & \text{for } i=0 \\ \int_{t_i+\epsilon}^{t_{i+1}-\epsilon}\sigma(x)dx & \text{for } i=1,q-1 \\ \int_{t_q+\epsilon}^{t_{q+1}}\sigma(x)dx & \text{for } i=q \end{cases} \tag{15}$$

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

$$[\mathbf{ci}]C_i = \int_{t_i-\epsilon}^{t_i+\epsilon}\sigma(x)dx \tag{16}$$

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 Computation A

???This lemma has to be reproved for functions in $\bar{A}$.

Lemma 2.1 *[**cp1**] Let $t_0 < t_1$ and $\sigma_0, \sigma_1 \in (0,1]$. A smooth function $\sigma : [t_0, t_1] \rightarrow \mathbf{R}$ such that $\sigma(t_0) = \sigma_0$, $\sigma(t_1) = \sigma_1$ and*

$$[\mathbf{cond1}]\sigma' \leq -\sigma(1-\sigma) \tag{17}$$

exists if and only if

$$[\mathbf{asser1}]\sigma_1 \geq \frac{\sigma_0}{\sigma_0 + (1-\sigma_0)e^{t_1 - t_0}} \tag{18}$$

or equivalently

$$[\mathbf{asser2}]\sigma_0 \leq \frac{\sigma_1}{\sigma_1 + (1-\sigma_1)e^{t_0 - t_1}} \tag{19}$$

and the equalities are achieved for a unique function

$$[\mathbf{s01}]\sigma(u) = \frac{\sigma_0}{\sigma_0 + (1-\sigma_0)e^{u - t_0}} \tag{20}$$

Proof: The equivalence of (18) and (19) is obvious. Let σ be a function satisfying the conditions of the proposition. Let us show that (18) holds. If $\sigma_1 = 1$ then (18) is obvious. Therefore, we may assume that $\sigma_1 < 1$. Assume that for all x, $\sigma(x) > 0$. Set

$$[\mathbf{cp1eq2}]\phi(x) = -\frac{\sigma'}{\sigma(1-\sigma)}. \tag{21}$$

Then (17) implies that $\phi(x) \leq 1$. Solving (21) 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 (18). 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 (18) is an equality is (20).

Suppose now that $\sigma_1 \in [0,1]$ satisfies the strong version of (18). Let $\epsilon > 0$ be a sufficiently small number. Consider the function of the form (20) 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 (17).

??? 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 *[**bcomp**] Let σ be a function satisfying the conditions of Lemma 2.1. Then*

$$[\mathbf{asser3}](1 + \sigma_1(e^{t_1 - t_0} - 1))^{-1} \leq e^{-\int_{t_0}^{t_1} \sigma(x)dx} \leq 1 + \sigma_0(e^{t_0 - t_1} - 1) \tag{22}$$

The equality is achieved in the class of smooth functions only if the equality holds in (18). In this case the only function which achieves the equality in any of the inequalities of (22) is (20) which makes both inequalities to be equalities.

Proof: Lemma 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 (22).

3 Computation B

Set

$$[\textbf{fofsigma}]F(t_1,\ldots,t_{q+1};\epsilon) = e^{-C_1}(1-e^{-C_1})e^{-2B_1}e^{-2C_2}(1-e^{-C_2})\ldots e^{-qC_q}(1-e^{-C_q})e^{-(q+1)B_q} \tag{23}$$

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 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 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 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: ???

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 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 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*

$$[\textbf{conc}]\frac{y_i}{y_i + (1 - y_i)e^{t_{i+1} - t_i - 2\epsilon}} \leq \frac{y_{i+1}}{y_{i+1} + (1 - y_{i+1})e^{-2\epsilon}} \tag{24}$$

Proof: It follows easily from Lemma 2.1.

Lemma 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 (23) where B_i and C_i are defined by (15) and (16) 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 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 (24) 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 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

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

and introduce new variables:

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

Then the function becomes

$$[\mathbf{ufun}]F = \delta^q \prod_{i=1}^{q} A_i(u_i) M_i(A_i(u_i))^{i+1} \tag{25}$$

and the inequalities become

$$[\mathbf{uineq}]0 \le u_1 \le \dots \le u_q \le 1 \tag{26}$$

i.e. we have to find maximums of (25) on the simplex (26). We have:

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

which implies that

$$A_i = \begin{pmatrix} r_{i+1}\dots r_q & 0 \\ r_{i+1}\dots r_q - (1+\delta)^{q-i} & (1+\delta)^{q-i} \end{pmatrix}$$

and

$$M_iA_i = (1+\delta)^{-1} \begin{pmatrix} r_i\dots r_q - (1+\delta)^{q-i} & (1+\delta)^{q-i} \\ r_{i+1}\dots r_q - (1+\delta)^{q-i-1} & (1+\delta)^{q-i-1} \end{pmatrix}$$

Proposition 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: ???

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 non-increasing sequence of s_i's may arise from a combinations of the event times $t_1 \le \dots \le t_q$. For $\delta = 0$ our formulas become:

$$A_i = \begin{pmatrix} 1 - s_{i+1} & 0 \\ -s_{i+1} & 1 \end{pmatrix} \qquad M_iA_i = \begin{pmatrix} -s_i & 1 \\ -s_{i+1} & 1 \end{pmatrix}$$

$$f_i(x) = A_i(x)M_i(A_i(x))^{i+1} = (1 - s_{i+1})x(-s_i x + 1)^{i+1}(-s_{i+1}x + 1)^{-(i+2)}$$

$$f_{i,k} = \prod_{j=0}^{k-1} f_{i+j}(x) = (\prod_{j=0}^{k-1}(1 - s_{i+j+1}))x^k(-s_i x + 1)^{i+1}(-s_{i+k}x + 1)^{-(i+k+1)}$$

Lemma 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 [**?**] 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 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 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* $u\leq v$ *the square*

$$\begin{array}{ccc} (X,A) & \xrightarrow{P_u} & (\Omega,\mathfrak{S}_u^t) \\ {\scriptstyle P_{u,v}}\downarrow & & \downarrow \\ (X,A) & \xrightarrow{P_v} & (\Omega,\mathfrak{S}_v^t) \end{array} \tag{27}$$

where

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

commutes.

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

$$[\mathbf{comp0}]P_{u,u}=Id \tag{28}$$

$$[\mathbf{comp1}]P_{v,w}\circ P_{u,v}=P_{u,w} \tag{29}$$

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 (28) and (29). 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 0.4 *[**mps**] Let* $(X,A,\Omega,\xi_u,\mathfrak{S}_u^v)$ *and* $(X,A,\Omega',\xi'_u,\mathfrak{R}_u^v)$ *be two path systems over* $[s,t]$ *with the same phase space. A morphism from the first to the second is a map* $f:\Omega\to\Omega'$ *such that:*

1. *for any* $u\in[s,t]$ *one has* $\xi'_u\circ f=\xi'_u$,

2. *for any* $u\leq v$ *in* $[s,t]$ *the map* f *is measurable with respect to* $\mathfrak{S}_u^v$ *and* $\mathfrak{R}_u^v$.

For any path system on (X, A) there is a unique morphism from it to the standard path system $St(X, A)$ on (X, A).

Lemma 0.5 *[**mpm**] Let f be a morphism of path systems as in Definition 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 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 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 0.8 *[**add1**] Note that if under the assumptions of Lemma 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 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 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 0.7.