

*The University of Melbourne,
Department of Mathematics and Statistics*

Association Behaviour in Non-Linear Death Processes

David Ringer

Supervisor: Associate Professor Aihua Xia
Second Reader: Dr Owen Jones

Thesis, November 2007

Abstract

Transition probabilities within a birth and death process are difficult to find analytically. As a consequence pure birth and pure death processes are considered in separate systems and it is found that both systems can be fully specified from intensity specifications. In a linear continuous-time Markovian pure death process it is found that optional death times are independent. Faddy (1985) conjectured that concave (convex) death rates $\mu_0, \mu_1, \dots, \mu_N$ exhibit increasing (decreasing) variability compared to the linear case. Ball and Donnelly (1987) claim that the conjecture is true and their result gives circumstances in which the covariance, and thus the correlation, is increased (decreased) relative to the linear case. Brown and Donnelly (1993) find an error in Ball and Donnelly (1987) as an aspect of their proof cannot be used to relate death intensities to conditional probabilities. Lefèvre and Milhaud (1990) take advantage of the conditional independence structure and produce a general result which relates individual death rates $\mu_1, \frac{\mu_2}{2}, \dots, \frac{\mu_n}{n}$ to interparticle correlation and covariance configuration. This is proved by Brown and Donnelly (1993). The interparticle correlation results are applied to life data and it is found that male and female death times are positively correlated.

Introduction

Birth and death processes have an underlying complexity which cannot be overcome by strictly focusing on birth and death intensity rates. Kolmogorov equations show us that first derivative transition probabilities are difficult to specify since they require knowledge of three distinct transition probabilities. Parameters can be specified in the equation but the transition probabilities cannot be obtained analytically. Thus pure birth and pure death processes are considered separately. We will show that information regarding birth and death intensities, respectively, is sufficient in determining generalised models in their respective systems, regardless of population size. It is conjectured that techniques used to solve pure birth processes can also be used to solve pure death processes as we can view them as reverse representations of each other. We will consider a linear continuous-time Markovian pure death process and find the interparticle association behaviour in such a model.

Faddy [?] (1985) conjectured that the variability of a death process described by a continuous-time Markov chain is increased, relative to the linear case, if the death rates $\mu_0, \mu_1, \dots, \mu_N$ form a concave sequence, and decreased if the death rates form a convex sequence. Ball and Donnelly [?] (1987) announce that the conjecture is true and their result gives circumstances when the correlation between death times in a non-linear Markovian death process are positive or negative. Ball and Donnelly (1987) claimed that the interparticle covariance is positive (negative) for all $t > 0$ if $\frac{\mu_n}{n}$ decreases (increases) with n .

Brown and Donnelly [?] (1993) also establish circumstances in which there is positive or negative correlation between death times in a non-linear Markovian death process. Brown and Donnelly (1993) claim that the proof in Ball and Donnelly (1987) has an error, but the result is true. Lefèvre and Milhaud [?] (1990) exploit the conditional independence structure and produce a more general result. Brown and Donnelly (1993) ultimately prove Ball and Donnelly (1987) claim using *multivariate total positivity of order 2*.

This leads us to the analysis aspect of this paper. We seek to use the theory to identify the interparticle correlation structure of human males and females in a population. Life data describes human population death experience from age 0 to a limiting age ω and individual death rates are obtainable, with appropriate assumptions. Life data may exhibit increasing or decreasing individual death rates as the number of people present in the study decreases. By obtaining individual death rates, we can identify the correlation structure for the studied population. As a consequence, the life data will serve to assess the suitability of a non-linear continuous-time Markov chain and its implications.

Contents

Chapter 1

Preliminary Theory for Conditional Intensity in Pure Death Processes

Before considering conditional pure death processes, this chapter establishes a contrast between pure death processes, pure birth processes and birth & death processes. We will begin by considering a birth & death process, and we will show the difficulties involved and how we overcome these difficulties by considering pure birth and pure death processes. It will be shown that aspects of a pure birth process can be relevant in a pure death process and thus techniques used to solve pure birth process problems can be applied to pure death processes. This chapter will ultimately outline the techniques needed to investigate linear continuous-time Markov chain models.

1.1 Introduction to Continuous-Time Markov Chains

Consider a continuous-time Markov chain process $\{X_t; t \geq 0\}$ which takes values in the set of nonnegative integers. Then for all $s, t \geq 0$ and nonnegative integers i, j, x_u we have: (see Ross [?] 1996 p. 232)

$$P(X_{t+s} = j | X_s = i, X_u = x_u, 0 \leq u < s) = P(X_{t+s} = j | X_s = i)$$

This suggests that a continuous-time Markov chain is a stochastic process, subject to a Markov property, which depends only on the present state and is independent of the past states. A continuous-time Markov process is said to have stationary or homogeneous transition probabilities if $P(X_{t+s} = j | X_s = i)$ is independent of s . Let us denote τ_i as the amount of time the process spends in state i before it transits towards a different state, then:

$$P(\tau_i > s + t | \tau_i > s) = P(\tau_i > t)$$

for all $s, t \geq 0$. Therefore, the random variable τ_i is memoryless and is exponentially distributed. (This will be confirmed in the pure births section of this chapter.)

We denote ν_i as the transition rate at which the process leaves state i and P_{ij} is the probability that the process transits towards j . Let us define $q_{ij} := \nu_i P_{ij}$ for $i \neq j$, and so we have that q_{ij} is the transition rate from i to j . The probability that the process, presently in state i , transits to state j after an additional time t is: (see Ross (1996) p. 233)

$$P_{ij}(t) = P(X_{t+s} = j | X_s = i)$$

1.2 Birth and Death Processes

We will consider two examples in this section. The first example will be a two-sex population in a small colony and the second example will be one-celled organisms within a two-state environment. We illustrate the analytical complexity of birth-death processes and show that pure birth and pure death processes are analytically simpler.

(Both examples were obtained from (Ross) 1996 p. 286)

1.2.1 Two Sex Population Within a Small Colony

Consider a population of organisms which consists of both male and female members. In a small colony any particular male is likely to mate with any particular female in any time interval of length h , with probability $\lambda h + o(h)$. Each mating immediately produces one offspring, equally likely to be male or female. Let $N_1(t)$ and $N_2(t)$ denote the number of males and females in the population at t . Suppose we want to derive the parameters of the continuous-time Markov chain $\{N_1(t), N_2(t)\}$. Then we proceed as follows.

$$P(\text{Male mates with female}) = \lambda h + o(h)$$

$N_1(t)$ - The number of males in the population, (which initially equals n).

$N_2(t)$ - The number of females in the population, (which initially equals m).

We aim to find the behaviour of transitions as the joint population goes from (i_1, i_2) to (j_1, j_2) .

Generator Matrix: (see Ross (1996) p. 244.)

$$q_{[(n,m),(n+1,m)]} = 0.5\lambda nm$$

$$q_{[(n,m),(n,m+1)]} = 0.5\lambda nm$$

Then we have:

$$\begin{aligned} \nu_{(i_1, i_2)} &= \sum_{(k_1, k_2) \neq (i_1, i_2)} q_{[(i_1, i_2), (k_1, k_2)]} \\ &= 0.5\lambda i_1 i_2 + 0.5\lambda i_1 i_2 \\ &= \lambda i_1 i_2 \end{aligned}$$

Kolmogorov's Backward Equation:

(see Theorem 5.4.3 Ross (1996) p. 240.)

General Form:

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t), \text{ for all } i, j, \text{ and } t \geq 0.$$

So:

$$\begin{aligned} P'_{(i_1, i_2), (j_1, j_2)}(t) &= \sum_{(k_1, k_2) \neq (i_1, i_2)} q_{(i_1, i_2), (k_1, k_2)} P_{(k_1, k_2), (j_1, j_2)}(t) - \nu_{(i_1, i_2)} P_{(i_1, i_2), (j_1, j_2)}(t) \\ &= q_{(i_1, i_2), (i_1+1, i_2)} P_{(i_1+1, i_2), (j_1, j_2)}(t) + q_{(i_1, i_2), (i_1, i_2+1)} P_{(i_1, i_2+1), (j_1, j_2)}(t) \\ &\quad - \nu_{(i_1, i_2)} P_{(i_1, i_2), (j_1, j_2)}(t) \\ &= 0.5\lambda i_1 i_2 P_{(i_1+1, i_2), (j_1, j_2)}(t) + 0.5\lambda i_1 i_2 P_{(i_1, i_2+1), (j_1, j_2)}(t) - \lambda i_1 i_2 P_{(i_1, i_2), (j_1, j_2)}(t) \end{aligned}$$

For all i_1, i_2, j_1, j_2 and $t \geq 0$.

The example shows us that we cannot explicitly derive transition probabilities by analytical means. However, we can obtain transition probabilities through numerical means.

1.2.2 One-Celled Organism Within Two States

Suppose that a one-celled organism can be in one of two states - either A or B . An individual in state A will change to state B at an exponential rate α ; an individual in state B divides into two new individuals of type A at an exponential rate β . Now suppose we want to define an appropriate continuous-time Markov chain for a population of such organisms and determine the appropriate parameters for this model.

Transition $A \rightarrow B$ at exponential rate α .

Transition $B \rightarrow 2A$ at exponential rate β .

We denote i_A as the number of A particles, and i_B as the number of B particles.

Generator Matrix:

$$q_{[(i_A, i_B), (i_A-1, i_B+1)]} = i_A \alpha$$

$$q_{[(i_A, i_B), (i_A+2, i_B-1)]} = i_B \beta$$

Then:

$$\nu_{(i_A, i_B)} = i_A \alpha + i_B \beta$$

Kolmogorov's Backward Equation:

General Form:

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t), \text{ for all } i, j, \text{ and } t \geq 0$$

So:

$$\begin{aligned} P'_{(i_A, i_B), (j_A, j_B)}(t) &= \sum_{(k_A, k_B) \neq (i_A, i_B)} q_{(i_A, i_B), (k_A, k_B)} P_{(k_A, k_B), (j_A, j_B)}(t) - \nu_{(i_A, i_B)} P_{(i_A, i_B), (j_A, j_B)}(t) \\ &= i_A \alpha P_{(i_A-1, i_B+1), (j_A, j_B)}(t) + i_B \beta P_{(i_A+2, i_B-1), (j_A, j_B)}(t) \\ &\quad - (i_A \alpha + i_B \beta) P_{(i_A, i_B), (j_A, j_B)}(t) \end{aligned}$$

For all i_A, i_B, j_A, j_B and $t \geq 0$.

We see that the differential equations are difficult to solve using analytical approaches. But as we will see, pure birth and pure death processes can be solved without resorting to numerical techniques.

1.3 Pure Birth Processes

Recall earlier we had that if τ_i represents the amount of time the process spends in state i before transition towards another state, we have $P(\tau_i > s + t | \tau_i > s) = P(\tau_i > t)$ for all $s, t \geq 0$. We produce three lemmas to confirm that this implies that the random variables τ_i 's are independent of each other and exponentially distributed with a time-varying hazard rate parameter.

1.3.1 Lemma 1

If we have a pure birth process with $U_j = \inf\{t : N_t = j\}$, $V_j = \inf\{t > U_j : N_t \neq j\}$ and we define sojourn time $V_j - U_j := \tau_j$, then $V_j - U_j$ is independent of $V_i - U_i$, $i < j$. (Sojourn time representation inspired by Kherani [?] (2006), p. 159 and Ball & Donnelly (1987) p. 758)

Proof:

$$\begin{aligned}
 P(\tau_j > t | \tau_{j-1} = z_{j-1}, \tau_{j-2} = z_{j-2}, \dots, \tau_0 = z_0) \\
 &= P(\tau_j + \tau_{j-1} + \dots + \tau_0 > t + z_{j-1} + \dots + z_0 | \tau_{j-1} = z_{j-1}, \dots, \tau_0 = z_0) \\
 &= P(N_{t+z_{j-1}+\dots+z_0} = j | N_{z_{j-1}+\dots+z_0} = j, N_{(z_{j-1}+\dots+z_0)^-} = j-1, \dots, N_{z_0} = 0) \\
 &= P(N_{t+z_{j-1}+\dots+z_0} = j | N_{z_{j-1}+\dots+z_0} = j), \text{ By Markov Property.} \\
 &= P(N_t = j | N_0 = j), \text{ By Homogeneous Property.} \\
 &= P(\tau_j > t)
 \end{aligned}$$

Thus $V_j - U_j := \tau_j$ is independent of $V_i - U_i$, $i < j$.

1.3.2 Lemma 2

If we have a continuous random variable $\tau_j := V_j - U_j$, where τ_j is independent of τ_i , for $i < j$, then we have that τ_j is memoryless.

Proof:

$$P(\tau_j > s + t | \tau_j > s) = \frac{P(\tau_j > s + t)}{P(\tau_j > s)}$$

Now we consider:

$$\begin{aligned}
 P(\tau_j > u) &= \int_0^\infty P(\tau_j > u | \tau_{j-1} = w_{j-1}, \tau_{j-2} = w_{j-2}, \dots, \tau_0 = w_0) f_{\tau_{j-1}, \tau_{j-2}, \dots, \tau_0}(w_{j-1}, w_{j-2}, \dots, w_0) dw_{j-1} dw_{j-2} \dots dw_0 \\
 &\text{Let } W = \sum_{k=0}^{j-1} w_k, \text{ then:} \\
 &= \int_0^\infty P(N_{W+u} = j | N_W = j, N_{W^-} = j-1, N_{W-w_j} = j-1, \dots, N_{w_0} = 0) \\
 &\quad \times f_{\tau_{j-1}, \tau_{j-2}, \dots, \tau_0}(w_{j-1}, w_{j-2}, \dots, w_0) dw_{j-1} dw_{j-2} \dots dw_0
 \end{aligned}$$

Then:

$$P(\tau_j > s+t) = \int_0^\infty P(N_{W+s+t} = j | N_W = j, N_{W^-} = j-1, \dots, N_{w_0} = 0) \\ \times f_{\tau_{j-1}, \tau_{j-2}, \dots, \tau_0}(w_{j-1}, w_{j-2}, \dots, w_0) dw_{j-1} dw_{j-2} \dots dw_0$$

By the Markov Property we have:

$$= \int_0^\infty P(N_{W+s+t} = j | N_W = j) f_{\tau_{j-1}, \tau_{j-2}, \dots, \tau_0}(w_{j-1}, w_{j-2}, \dots, w_0) dw_{j-1} dw_{j-2} \dots dw_0 \\ = \int_0^\infty \frac{P(N_{W+s+t} = j, N_{W+t} = j, N_W = j)}{P(N_{W+t} = j, N_W = j)} \frac{P(N_{W+t} = j, N_W = j)}{P(N_W = j)} \\ \times f_{\tau_{j-1}, \dots, \tau_0}(w_{j-1}, \dots, w_0) dw_{j-1} \dots dw_0 \\ = \int_0^\infty P(N_{W+s+t} = j | N_{W+t} = j, N_W = j) P(N_{W+t} = j | N_W = j) \\ \times f_{\tau_{j-1}, \dots, \tau_0}(w_{j-1}, \dots, w_0) dw_{j-1} \dots dw_0$$

By the Markov Property we have that:

$$P(N_{W+s+t} = j | N_{W+t} = j, N_W = j) = \underbrace{P(N_{W+s+t} = j | N_{W+t} = j)}_{=P(\tau_j > s), \text{ By Homogeneous Property.}}$$

This gives us:

$$P(\tau_j > s+t) = P(\tau_j > s) \int_0^\infty \underbrace{P(N_{W+t} = j | N_W = j)}_{=P(\tau_j > t), \text{ By Homogeneous Property.}} f_{\tau_{j-1}, \dots, \tau_0}(w_{j-1}, \dots, w_0) dw_{j-1} \dots dw_0 \\ = P(\tau_j > s) P(\tau_j > t), \text{ Since } \int_0^\infty f_{\tau_{j-1}, \dots, \tau_0}(w_{j-1}, \dots, w_0) dw_{j-1} \dots dw_0 = 1$$

This implies that:

$$P(\tau_j > s+t | \tau_j > s) = \frac{P(\tau_j > s) P(\tau_j > t)}{P(\tau_j > s)} = P(\tau_j > t)$$

Thus $V_j - U_j := \tau_j$ is memoryless.

1.3.3 Lemma 3

If we have a pure birth process, and a memoryless continuous random variable, $\tau_j := V_j - U_j$, then τ_j has an exponential distribution.

Proof:

Let $r_{\tau_j}(t)$ be the hazard function for τ_j :

$$\begin{aligned} r_{\tau_j}(t) &:= \lim_{\Delta t \rightarrow 0} \frac{P(\tau_j \in (t, t + \Delta t) | \tau_j > t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(\tau_j \in (t, t + \Delta t))}{P(\tau_j > t) \Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{f(t) \Delta t}{\bar{F}(t) \Delta t}, \text{ where } \bar{F}(t) = 1 - F(t) = P(\tau_j > t). \\ &= \frac{f(t)}{\bar{F}(t)} \end{aligned}$$

We see that $\frac{f(t)}{\bar{F}(t)} = -\frac{d}{dt} \frac{\bar{F}(t)}{\bar{F}(t)}$, then:

$$\begin{aligned} -\frac{d}{dt} \log(\bar{F}(t)) &= r_{\tau_j}(t) \\ \log(\bar{F}(t)) &= -\int_0^t r_{\tau_j}(s) ds + k \\ \bar{F}(t) &= e^k e^{-\int_0^t r_{\tau_j}(s) ds} \end{aligned}$$

If we let $t = 0$ then we obtain a value of $k = 0$, and so:

$$\begin{aligned} \bar{F}(t) &= e^{-\int_0^t r_{\tau_j}(s) ds} \\ F(t) &= 1 - e^{-\int_0^t r_{\tau_j}(s) ds}, \text{ (see Ross (1996) pp. 246–247)} \end{aligned}$$

We see that this resembles the exponential cumulative distribution function. However if we inspect $r_{\tau_j}(s)$ we see that:

$$\begin{aligned} r_{\tau_j}(s) &= \lim_{h \rightarrow 0} \frac{P(\tau_j \in (s, s + h) | \tau_j > s)}{h} \\ &= \lim_{h \rightarrow 0} \frac{P(\tau_j > s | \tau_j > s) - P(\tau_j > s + h | \tau_j > s)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1 - P(\tau_j > h)}{h}, \text{ By Memoryless Property.} \\ &= \lim_{h \rightarrow 0} \frac{\int_0^\infty f_{\tau_j}(s) ds - \int_h^\infty f_{\tau_j}(s) ds}{h} \\ &= \lim_{h \rightarrow 0} \frac{\int_0^h f_{\tau_j}(s) ds}{h} \\ &= \lim_{h \rightarrow 0} \frac{F_{\tau_j}(h) - F_{\tau_j}(0)}{h} \\ &= \left. \frac{dF(s)}{ds} \right|_{s=0} \\ &= f_{\tau_j}(0) \end{aligned}$$

Which implies that if τ_j is memoryless then $r_{\tau_j}(s) = r_{\tau_j}(t) = f_{\tau_j}(0)$, which is constant. So $\tau_j \sim \exp(f_{\tau_j}(t))$.

Lemmas 1,2,3 will be applied to two pure birth examples, in this section. The first example, which we refer to as a *0 to N birth process*, will consider the moments of the time it takes for the population, which begins with 0 lives, to reach a population of N. The second example is a conditional Yule process which begins with i lives and is conditional upon having $i + k$ lives after elapsed time t . (Both examples were obtained from Ross (1996) p. 287.)

1.3.4 0 to N Birth Process

For a pure birth process with birth parameters $\lambda_n, n \geq 0$, we intend to compute the mean, variance, and moment generating function of the time it takes the population to go from size 0 to size N . We proceed as follows.

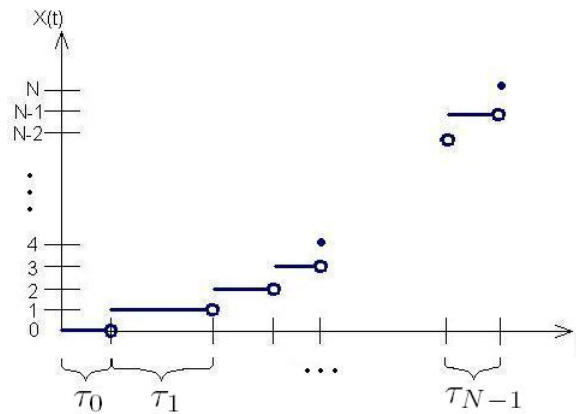
Birth parameter: $\lambda_n, n \geq 0$.

Let τ_i denote the time it takes for the population to go from size i to $i+1$. In this case we have time intervals $\tau_0, \tau_1, \dots, \tau_{N-1}$ with conditional intensities $\lambda_0, \lambda_1, \dots, \lambda_{N-1}$ respectively. Then we define:

$$T = \sum_{i=0}^{N-1} \tau_i$$

Where T represents the total time it takes for the population to reach size N , given it starts with 0.

Graphical Representation:



We see that τ_i can be defined as $\tau_i := V_i - U_i$, where $U_i = \inf\{t : N_t = i\}$ and $V_i = \inf\{t > U_i : N_t \neq i\}$. By *Lemma 1* we can say that τ_i is independent of $\tau_k, k < i$. By *Lemma 2* we can say that τ_i is memoryless for all i . This implies that, by *Lemma 3*, $\tau_i \sim \exp(\lambda_i)$ for all i .

Now to find the moment generating function of $T = \sum_{i=0}^{N-1} \tau_i$, we first consider the moment generating function of τ_0 .

$$\begin{aligned}
M_{\tau_0}(t) &= E(e^{t\tau_0}) \\
&= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\
&= \int_0^t e^{tx} \lambda_0 e^{-\lambda_0 x} dx \\
&= \int_0^{\infty} \lambda_0 e^{x(t-\lambda_0)} dx \\
&= \lim_{k \rightarrow \infty} \left[\frac{\lambda_0}{t - \lambda_0} e^{x(t-\lambda_0)} \right]_0^k
\end{aligned}$$

If $t \geq \lambda_0$ then the integral is undefined, so function does not exist.
So we only consider the function when $t < \lambda_0$:
Hence we see that:

$$\lim_{k \rightarrow \infty} e^{k(t-\lambda_0)} = 0, \text{ since } t - \lambda_0 < 0$$

Then:

$$\begin{aligned}
M_{\tau_0}(t) &= \lim_{k \rightarrow \infty} \left[\frac{\lambda_0}{t - \lambda_0} e^{x(t-\lambda_0)} \right]_0^k \\
&= \frac{\lambda_0}{t - \lambda_0} [0 - e^0] \\
&= \frac{\lambda_0}{\lambda_0 - t}
\end{aligned}$$

If $t < \lambda_i$, $i = 1, 2, \dots, N-1$, then:

$$\begin{aligned}
M_{\tau_1}(t) &= \frac{\lambda_1}{\lambda_1 - t} \\
M_{\tau_2}(t) &= \frac{\lambda_2}{\lambda_2 - t} \\
&\vdots \\
M_{\tau_{N-1}}(t) &= \frac{\lambda_{N-1}}{\lambda_{N-1} - t}
\end{aligned}$$

Now let us consider $T = \sum_{i=0}^{N-1} \tau_i$:

$$\begin{aligned}
M_{\tau_0+\tau_1+\dots+\tau_{N-1}}(t) &= E(e^{t(\tau_0+\tau_1+\dots+\tau_{N-1})}) \\
&= E(e^{t\tau_0})E(e^{t\tau_1})\dots E(e^{t\tau_{N-1}}), \text{ By Independence} \\
&= \frac{\lambda_0}{\lambda_0 - t} \frac{\lambda_1}{\lambda_1 - t} \dots \frac{\lambda_{N-1}}{\lambda_{N-1} - t}, \quad (t < \lambda_i, i = 0, 1, \dots, N-1) \\
&= e^{\ln\left(\frac{\lambda_0}{\lambda_0-t}\right)} e^{\ln\left(\frac{\lambda_1}{\lambda_1-t}\right)} \dots e^{\ln\left(\frac{\lambda_{N-1}}{\lambda_{N-1}-t}\right)} \\
&= e^{\sum_{i=0}^{N-1} \ln\left(\frac{\lambda_i}{\lambda_i-t}\right)} \\
&= e^{\sum_{i=0}^{N-1} \ln(\lambda_i) - \sum_{i=0}^{N-1} \ln(\lambda_i - t)}, \text{ We shall denote this as } \textit{Result A}.
\end{aligned}$$

Now to find $E(T) = E(\tau_0 + \tau_1 + \dots + \tau_{N-1})$ we take the first derivative of *Result A* and evaluate at $t = 0$:

$$\begin{aligned}
E(T) &= \frac{d}{dt} \left[e^{\sum_{i=0}^{N-1} \ln(\lambda_i) - \sum_{i=0}^{N-1} \ln(\lambda_i - t)} \right]_{|t=0} \\
&= \sum_{j=0}^{N-1} \left(-1 \times \frac{-1}{\lambda_j - t} \right) \left(e^{\sum_{i=0}^{N-1} (\ln(\lambda_i) - \ln(\lambda_i - t))} \right)_{|t=0} \\
&= \sum_{j=0}^{N-1} \frac{1}{\lambda_j}
\end{aligned}$$

Now to find $E(T^2)$ we take the second derivative of *Result A* and evaluate at $t = 0$:

$$\begin{aligned}
E(T^2) &= \frac{d^2}{dt^2} \left[e^{\sum_{i=0}^{N-1} (\ln(\lambda_i) - \ln(\lambda_i - t))} \right]_{|t=0} \\
&= \frac{d}{dt} \left[\sum_{j=0}^{N-1} \left(\frac{1}{\lambda_j - t} \right) \left(e^{\sum_{i=0}^{N-1} (\ln(\lambda_i) - \ln(\lambda_i - t))} \right) \right]_{|t=0} \\
&= \left(\sum_{j=0}^{N-1} -1 \times \frac{-1}{(\lambda_j - t)^2} \right) \left(e^{\sum_{i=0}^{N-1} \ln(\lambda_i) - \sum_{i=0}^{N-1} \ln(\lambda_i - t)} \right) \\
&\quad + \left(\sum_{j=0}^{N-1} \frac{1}{\lambda_j - t} \right) \left(\sum_{j=0}^{N-1} \frac{1}{\lambda_j - t} \right) \left(e^{\sum_{i=0}^{N-1} \ln(\lambda_i) - \sum_{i=0}^{N-1} \ln(\lambda_i - t)} \right)_{|t=0} \\
&= \sum_{j=0}^{N-1} \frac{1}{\lambda_j^2} + \left(\sum_{j=0}^{N-1} \frac{1}{\lambda_j} \right)^2
\end{aligned}$$

Now we will find the variance of T :

$$\text{Var}(T) = \sum_{j=0}^{N-1} \frac{1}{\lambda_j^2}$$

This example illustrates that analytical techniques can be applied to find all moments of the birth process regardless of population size.

1.3.5 Conditional Yule Process

Consider a Yule process with $X_0 = i$. Given that $X_t = i + k$, we intend to find the conditional distribution of the birth times of the k individuals born in $(0, t)$.

Birth parameter: $\lambda_n = n\lambda, n \geq 0$.

Let τ_j be the time taken for the population to go from size j to $j + 1$. Since a Yule process is a pure birth process we can take $\tau_j, j \geq i$, as independent with $\tau_j \sim \exp(\lambda_j)$. We will directly quote Ross (1996) p. 236 result:

$$P(X_t = j) = P_{ij}(t) = \binom{j-1}{i-1} e^{-i\lambda t} (1 - e^{-\lambda t})^{j-1}, \quad j \geq i \geq 1. \quad (1.1)$$

Now we will explain how this can be proved by mathematical induction.

$$\begin{aligned} f_{\tau_i}(t) &= \lambda_i e^{-\lambda_i t} = i\lambda e^{-i\lambda t} \\ f_{\tau_{i+1}}(t) &= \lambda_{i+1} e^{-\lambda_{i+1} t} = (i+1)\lambda e^{-(i+1)\lambda t} \\ &\vdots \\ f_{\tau_{i+k-1}}(t) &= \lambda_{i+k-1} e^{-\lambda_{i+k-1} t} = (i+k-1)\lambda e^{-(i+k-1)\lambda t} \end{aligned}$$

Now:

$$P(\tau_i \leq t) = 1 - e^{-i\lambda t}, \text{ Note that this is equivalent to } P(X_t \geq i + 1).$$

$$\begin{aligned} P(\tau_i + \tau_{i+1} \leq t) &= \int_0^t P(\tau_i + \tau_{i+1} \leq t | \tau_i = x) (i\lambda e^{-i\lambda x}) dx \\ &= \int_0^t P(\tau_{i+1} \leq t - x) i\lambda e^{-i\lambda x} dx \\ &= \int_0^t (1 - e^{-(i+1)\lambda(t-x)}) i\lambda e^{-i\lambda x} dx \\ &= 1 - e^{-i\lambda t} - \int_0^t i\lambda e^{-\lambda[(i+1)(t-x)+ix]} dx \\ &= 1 - e^{-i\lambda t} - ie^{-(i+1)\lambda t} \int_0^t \lambda e^{\lambda x} dx \\ &= 1 - e^{-i\lambda t} - ie^{-(i+1)\lambda t} (e^{\lambda t} - 1) \\ &= 1 - e^{-i\lambda t} - ie^{-i\lambda t} + ie^{-(i+1)\lambda t}, \text{ Which is equivalent to } P(X_t \geq i + 2). \end{aligned}$$

This implies that:

$$\begin{aligned} P(X_t = i + 1) &= P(X_t \geq i + 1) - P(X_t \geq i + 2) \\ &= 1 - e^{-i\lambda t} - 1 + e^{-i\lambda t} + ie^{-i\lambda t} - ie^{-(i+1)\lambda t} \\ &= ie^{-i\lambda t} (1 - e^{-\lambda t}) \end{aligned}$$

So by mathematical induction the result can be confirmed. Now we proceed to the next aspect of the problem. Let us define $S_j := \inf\{t : X_t = j\}$, $j > i$ and $S_i = 0$. So S_j is the time at which the j th birth occurs and this implies $S_j = \tau_i + \tau_{i+1} + \dots + \tau_{j-1}$. We seek to find the conditional joint distribution of S_{i+1}, \dots, S_{i+k} given that $X_t = i + k$. We treat densities as probabilities and assume that $s_j + h_j < s_{j+1}$,

$j > i$, then we obtain for $0 \leq s_{i+1} \leq \dots \leq s_{i+k} \leq t$, and let P be such that:

$$\begin{aligned} P &= P(S_{i+1} \in (s_{i+1}, s_{i+1} + h_{i+1}), S_{i+2} \in (s_{i+2}, s_{i+2} + h_{i+2}), \dots, S_{i+k} \in (s_{i+k}, s_{i+k} + h_{i+k}) | X_t = i + k) \\ &= P(\tau_i \in (s_{i+1}, s_{i+1} + h_{i+1}), \tau_i + \tau_{i+1} \in (s_{i+2}, s_{i+2} + h_{i+2}), \tau_i + \tau_{i+1} + \tau_{i+2} \in (s_{i+3}, s_{i+3} + h_{i+3}), \dots \\ &\quad \dots, \tau_i + \tau_{i+1} + \dots + \tau_{i+k-1} \in (s_{i+k}, s_{i+k} + h_{i+k}) | X_t = i + k) \end{aligned}$$

Since $h_j, i + 1 \leq j \leq i + k$, are small:

$$\approx \frac{P(\tau_i \in (s_{i+1}, s_{i+1} + h_{i+1}), \tau_{i+1} \in (s_{i+2} - s_{i+1}, s_{i+2} - s_{i+1} + h_{i+2}), \tau_{i+2} \in (s_{i+3} - s_{i+2}, s_{i+3} - s_{i+2} + h_{i+3}), \dots, \dots, \tau_{i+k-1} \in (s_{i+k} - s_{i+k-1}, s_{i+k} - s_{i+k-1} + h_{i+k}), \tau_{i+k} > t - s_{i+k})}{P(X_t = i + k)}$$

Then by independence:

$$\begin{aligned} &= \frac{P(\tau_i \in (s_{i+1}, s_{i+1} + h_{i+1})) P(\tau_{i+1} \in (s_{i+2} - s_{i+1}, s_{i+2} - s_{i+1} + h_{i+2})) \dots \\ &\quad \dots P(\tau_{i+k-1} \in (s_{i+k} - s_{i+k-1}, s_{i+k} - s_{i+k-1} + h_{i+k})) P(\tau_{i+k} > t - s_{i+k-1})}{P(X_t = i + k)} \\ &= \frac{f_{\tau_i}(s_{i+1}) h_{i+1} \times f_{\tau_{i+1}}(s_{i+2} - s_{i+1}) h_{i+2} \times \dots \\ &\quad \dots \times f_{\tau_{i+k}}(s_{i+k} - s_{i+k-1}) h_{i+k} \times (1 - F_{\tau_{i+k}}(t - s_{i+k}))}{P(X_t = i + k)} \\ &= \frac{i \lambda e^{-i \lambda (s_{i+1})} h_{i+1} \times (i+1) \lambda e^{-(i+1) \lambda (s_{i+2} - s_{i+1})} h_{i+2} \times \dots \\ &\quad \dots \times (i+k-1) \lambda e^{-(i+k-1) \lambda (s_{i+k} - s_{i+k-1})} h_{i+k} \times e^{-(i+k) \lambda (t - s_{i+k})}}{P(X_t = i + k)} \\ &= C_h \frac{e^{-\lambda(t-s_{i+1})} e^{-\lambda(t-s_{i+2})} \dots e^{-\lambda(t-s_{i+k-1})} e^{-\lambda(t-s_{i+k})} e^{-i \lambda t}}{P(X_t = i + k)} \end{aligned}$$

Where $i(i+1)\dots(i+k-1) = \frac{(i+k-1)!}{(i-1)!}$, and:

$$C_h = h_{i+1} h_{i+2} \dots h_{i+k} \frac{(i+k-1)!}{(i-1)!} \lambda^k$$

Then we denote:

$$C = \frac{C_h}{h_{i+1} h_{i+2} \dots h_{i+k}} = \frac{(i+k-1)!}{(i-1)!} \lambda^k$$

And we see that C is a constant that does not depend on $s_{i+1}, s_{i+2}, \dots, s_{i+k}$.

So to get the joint density function we have:

$$\begin{aligned} \lim_{\substack{h_j \rightarrow 0 \\ j=i+1, i+2, \dots, i+k}} \frac{P}{h_{i+1} h_{i+2} \dots h_{i+k}} &= \frac{C e^{-\lambda(t-s_{i+1})} e^{-\lambda(t-s_{i+2})} \dots e^{-\lambda(t-s_{i+k-1})} e^{-\lambda(t-s_{i+k})} e^{-i \lambda t}}{P(X_t = i + k)} \\ f_{S_{i+1}, S_{i+2}, \dots, S_{i+k} | X_t}(s_{i+1}, s_{i+2}, \dots, s_{i+k} | i + k) &= \frac{C e^{-\lambda(t-s_{i+1})} e^{-\lambda(t-s_{i+2})} \dots e^{-\lambda(t-s_{i+k-1})} e^{-\lambda(t-s_{i+k})} e^{-i \lambda t}}{P(X_t = i + k)} \end{aligned}$$

We know that:

$$\begin{aligned}
P_{i,i+k}(t) &= P(X_t = i + k) \\
&= \binom{i+k-1}{i-1} e^{-i\lambda t} (1 - e^{-\lambda t})^{i+k-i} \\
&= \frac{(i+k-1)!}{(i-1)!k!} e^{-i\lambda t} (1 - e^{-\lambda t})^k
\end{aligned}$$

The probability of equality between $s_i = s_j$ is zero for $i < j$.

Hence we see that the conditional density of $S_{i+1}, S_{i+2}, \dots, S_{i+k}$ given that $X_t = i + k$ is given by:

$$f_{S_{i+1}, S_{i+2}, \dots, S_{i+k} | X_t}(s_{i+1}, s_{i+2}, \dots, s_{i+k} | i + k) = k! \prod_{j=i+1}^{i+k} f_{S_j}(s_j), \quad 0 < s_{i+1} < \dots < s_{i+k} < t.$$

Where:

$$f_{S_j}(s_j) = \begin{cases} \frac{\lambda e^{-\lambda(t-s_j)}}{1 - e^{-\lambda t}}, & 0 \leq s_j \leq t \\ 0, & \text{otherwise.} \end{cases}$$

We have illustrated that analytically, we can find all relevant transition probabilities in a pure birth process, even when we condition upon the number of lives at time t .

1.4 Pure Death Processes

In this section we apply *Lemmas 1,2,3* to two important pure death process examples. The first one is a *m to 0 pure death process* and we consider the total time it takes for the process to reach the absorption state 0. The second example is a *linear continuous-time Markovian pure death process*, which will provide us with a contrast to the non-linear continuous-time Markovian pure death process.

1.4.1 m to 0 Pure Death Process

Consider a pure death process with death rate λ_n when there are n individuals, $n = 1, 2, \dots$ and $\lambda_0 = 0$. Start with m individuals, we seek to find the expected time that all of them die out.

Death parameter: $\lambda_n, n = 1, 2, \dots$

We can represent the situation with the following Generator Matrix:
(see Bhat & Miller [?] (2002), p. 212.)

$$\begin{array}{cccccccc}
 & 0 & 1 & 2 & \dots & m-2 & m-1 & m \\
 \begin{array}{l} 0 \\ 1 \\ 2 \\ \vdots \\ m-2 \\ m-1 \\ m \end{array} & \left(\begin{array}{cccccccc}
 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
 \lambda_1 & -\lambda_1 & 0 & \dots & 0 & 0 & 0 & 0 \\
 0 & \lambda_2 & -\lambda_2 & \dots & 0 & 0 & 0 & 0 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
 0 & 0 & 0 & \dots & -\lambda_{m-2} & 0 & 0 & 0 \\
 0 & 0 & 0 & \dots & \lambda_{m-1} & -\lambda_{m-1} & 0 & 0 \\
 0 & 0 & 0 & \dots & 0 & \lambda_m & -\lambda_m & 0
 \end{array} \right)
 \end{array}$$

We define $T_A := \inf\{t : X_t \in A\}$ and we treat A as the absorption event so $\{0\} = A$.

We require $E(T_A | X_0 = m)$, and we denote τ_i as the time it takes for the population to go from size i to $i - 1$. So we have time intervals $\tau_m, \tau_{m-1}, \dots, \tau_1$ with conditional intensities $\lambda_m, \lambda_{m-1}, \dots, \lambda_1$ respectively. Since we are dealing with a pure death process, the process will only enter state 0 if it has passed through all preceding states $j = m, m - 1, \dots, 1$. So $\inf\{t : X_t \in A\}$ can also be written as the sum of the time it takes for the process to pass through and exit states $j = m, m - 1, \dots, 1$. So we have:

$$T_A = \sum_{i=0}^{m-1} \tau_{m-i}$$

So T_A represents the total time it takes for the population to reach size 0.

We have defined $\tau_i := V_i - U_i$ as in the *Pure Birth Process* section. We see that a pure death process can be viewed as a reversed pure birth process. When the pure birth process is reversed, τ_j is still independent of $\tau_i, i < j$, by *Lemma 1*; τ_j is still memoryless by *Lemma 2*; and $\tau_j \sim \exp(\lambda_j)$ for all j by *Lemma 3*.

Now we proceed to find $E(T_A|X_0 = m)$:

$$\begin{aligned} E(T_A|X_0 = m) &= E(\tau_m + \tau_{m-1} + \dots + \tau_1|X_0 = m) \\ &= E(\tau_m|X_0 = m) + E(\tau_{m-1}|X_0 = m) + \dots + E(\tau_1|X_0 = m) \end{aligned}$$

We know that $\tau_m \sim \exp(\lambda_m)$ is based on the assumption that $X_0 = m$. So we can write $\tau_m = (\tau_m|X_0 = m) \sim \exp(\lambda_m)$. This implies that:

$$\begin{aligned} (\tau_{m-1}|X_0 = m) &\sim \exp(\lambda_{m-1}) \\ (\tau_{m-2}|X_0 = m) &\sim \exp(\lambda_{m-2}) \\ &\vdots \\ (\tau_1|X_0 = m) &\sim \exp(\lambda_1) \end{aligned}$$

So we have:

$$E(T_A|X_0 = m) = \frac{1}{\lambda_m} + \frac{1}{\lambda_{m-1}} + \dots + \frac{1}{\lambda_1} = \sum_{i=1}^m \frac{1}{\lambda_i}$$

Thus we have shown that the expected total survival time, for any population size, can be obtained solely from conditional death intensities.

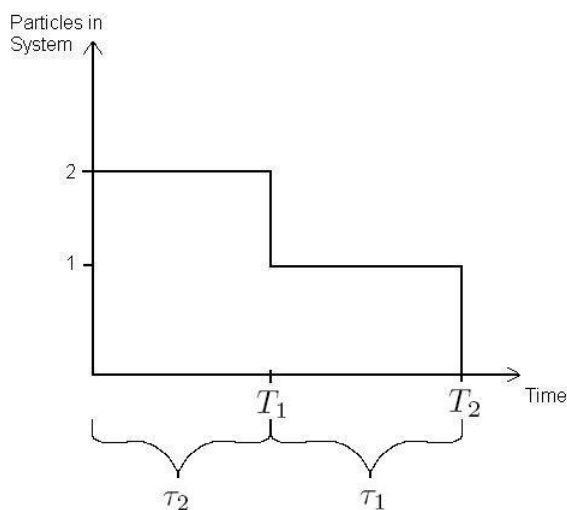
1.4.2 Linear Continuous-Time Markovian Pure Death Process

A pure death process starting with n particles each with equally likely chance to die at any time T_i , $i = 1, 2, \dots, n$. The death time for particle j is S_j , $j = 1, 2, \dots, n$. When there are k particles, the death rate is $k\mu$; We seek to find the joint density of (T_1, T_2, \dots, T_n) , and the joint density of (S_1, S_2, \dots, S_n) .

Let us denote μ_k as the death intensity for each particle i , $i = 1, 2, \dots, k$, when there are k particles in the system.

We begin by considering the case when $n = 2$. When there are two particles in the system the average particle death rate is $\frac{2\mu}{2} = \mu$ for the two particles $i = 1, 2$. When there is only one particle left in the system the average particle death rate is $\frac{\mu}{1} = \mu$ for the last of the particles $i = 1, 2$. So the average death rates form a constant sequence.

Graphical Representation:



So we have:

$$\begin{aligned} \min(S_1, S_2) &= T_1 = \inf\{t : N_t = 1\} \\ \max(S_1, S_2) &= T_2 = \inf\{t : N_t = 0\} \end{aligned}$$

From earlier definitions we see that $T_1 = \tau_2$ and $T_2 - T_1 = \tau_1$. By *Lemma 1*, $T_2 - T_1$ is independent of T_1 . $T_2 - T_1$ and T_1 are memoryless by *Lemma 2*. This implies that by *Lemma 3*, $T_2 - T_1 \sim \exp(\mu)$ and $T_1 \sim \exp(2\mu)$.

Now we consider the joint distribution of T_1 and $T_2 - T_1$:

$$\begin{aligned}
& P(T_1 \in (t_1, t_1 + h_1), T_2 \in (t_2, t_2 + h_2)) \\
&= \int_0^\infty P(T_1 \in (t_1, t_1 + h_1), T_2 \in (t_2, t_2 + h_2) | T_1 = s) f_{T_1}(s) ds \\
&= \int_{t_1}^{t_1+h_1} P(T_2 - T_1 \in (t_2 - s, t_2 - s + h_2) | T_1 = s) f_{T_1}(s) ds, \text{ Since } T_2 - T_1 \text{ is independent of } T_1. \\
&= \int_{t_1}^{t_1+h_1} \underbrace{P(T_2 - T_1 \in (t_2 - s, t_2 - s + h_2))}_{\int_{t_2-s}^{t_2-s+h_2} f_{T_2-T_1}(t-s) dt} f_{T_1}(s) ds
\end{aligned}$$

Now we have:

$$\begin{aligned}
f_{T_1, T_2-T_1}(t_1, t_2 - t_1) &= \lim_{\substack{h_i \rightarrow 0 \\ i=1,2}} \frac{1}{h_1 h_2} P(T_1 \in (t_1, t_1 + h_1), T_2 \in (t_2, t_2 + h_2)) \\
&= \lim_{\substack{h_i \rightarrow 0 \\ i=1,2}} \frac{1}{h_1 h_2} \int_{t_1}^{t_1+h_1} \int_{t_2-s}^{t_2-s+h_2} f_{T_2-T_1}(t-s) dt f_{T_1}(s) ds \\
&= f_{T_2-T_1}(t_2 - t_1) f_{T_1}(t_1) \\
&= \mu e^{-\mu(t_2-t_1)} \times 2\mu e^{-2\mu t_1} \\
&= 2\mu^2 e^{-\mu(t_1+t_2)}, \text{ Which we will denote as } \textit{Result 2.A.1.}
\end{aligned}$$

So we can say that:

$$f_{T_1, T_2}(t_1, t_2) = 2\mu^2 e^{-\mu(t_1+t_2)}, \text{ for all } t_1 \text{ \& } t_2.$$

We note that $P(T_1 = T_2) = 0$, which implies that $P(S_1 = S_2) = 0$ since we are dealing with a continuous distribution. This is depicted in *Figure A* ahead, where the volume of the shape generated by $S_1 = S_2$ is a zero-volume shape.

Now let us consider the joint density for S_1, S_2 :

Figure A:

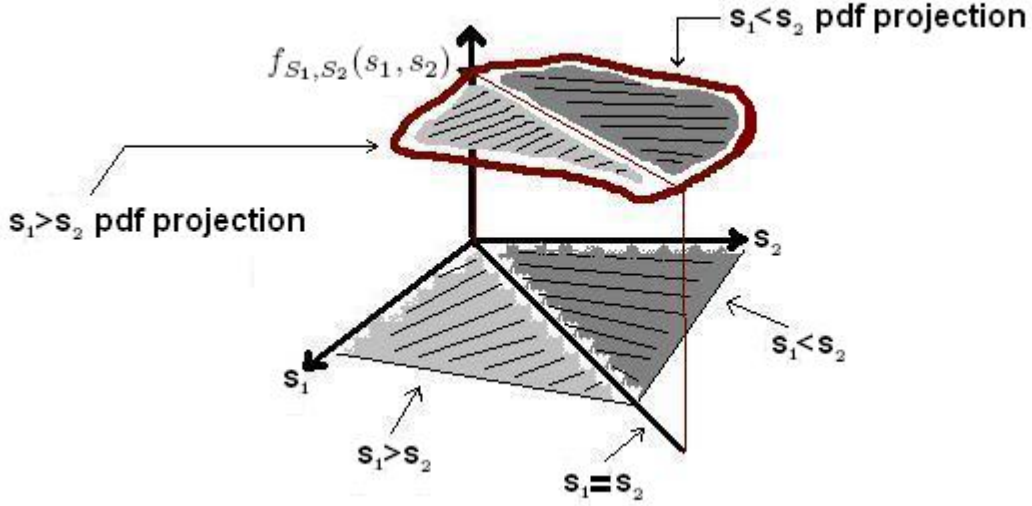


Figure A is a graphical representation of the behaviour of the joint density of S_1 and S_2 . The joint density generated by $S_1 < S_2$ is mirrored in the $S_1 > S_2$ case, across the line $S_1 = S_2$. This implies that $S_1 < S_2$ and $S_1 > S_2$, have equal weights in the distribution space. Thus we can say:

$$f_{S_1, S_2}(s_1, s_2) = f_{S_1, S_2}(s_2, s_1), \text{ By Symmetry.} \quad (1.2)$$

So to find the joint density function we consider:

$$\begin{aligned} & P(\min(S_1, S_2) \in (t_1, t_1 + dt_1), \max(S_1, S_2) \in (t_2, t_2 + dt_2)) \\ &= P(T_1 \in (t_1, t_1 + dt_1), T_2 \in (t_2, t_2 + dt_2)) \\ &= P(T_1 \in (t_1, t_1 + dt_1), T_2 - T_1 \in (t_2 - t_1, t_2 - t_1 + dt_2)) \\ &= 2\mu^2 e^{-\mu(t_1+t_2)} dt_1 dt_2, \text{ Recall Result 2.A.1.} \end{aligned}$$

We also see that: $P(\min(S_1, S_2) \in (t_1, t_1 + dt_1), \max(S_1, S_2) \in (t_2, t_2 + dt_2))$

$$\begin{aligned} &= P(S_1 \in (t_1, t_1 + dt_1), S_2 - S_1 \in (t_2 - t_1, t_2 - t_1 + dt_2)) + P(S_2 \in (t_1, t_1 + dt_1), S_1 - S_2 \in (t_2 - t_1, t_2 - t_1 + dt_2)) \\ &= P(S_1 \in (t_1, t_1 + dt_1), S_2 \in (t_2, t_2 + dt_2)) + P(S_2 \in (t_1, t_1 + dt_1), S_2 \in (t_2, t_2 + dt_2)) \\ &= 2f_{S_1, S_2}(t_1, t_2) dt_1 dt_2, \text{ By Symmetry.} \end{aligned}$$

This implies that:

$$\begin{aligned} f_{S_1, S_2}(t_1, t_2) &= \frac{1}{2} \frac{1}{dt_1 dt_2} 2\mu^2 e^{-\mu(t_1+t_2)} dt_1 dt_2 \\ &= \mu^2 e^{-\mu(t_1+t_2)} = f_{S_1, S_2}(t_2, t_1) \end{aligned}$$

This yields the result:

$$f_{S_1, S_2}(s_1, s_2) = \mu^2 e^{-\mu(s_1+s_2)}, \text{ for all } s_1 \text{ and } s_2. \quad (1.3)$$

Now let us consider the relationship between S_1 and S_2 :

$$\begin{aligned} P(S_2 > t, S_1 > s) &= \int_{s_1 \in (s, \infty)} \int_{s_2 \in (t, \infty)} f_{S_1, S_2}(s_1, s_2) ds_1 ds_2 \\ &= \int_s^\infty \int_t^\infty \mu^2 e^{-\mu(s_1+s_2)} ds_1 ds_2 \\ &= \mu^2 \int_s^\infty e^{-\mu s_1} \left(\int_t^\infty e^{-\mu s_2} ds_2 \right) ds_1 \\ &= \mu^2 \int_s^\infty e^{-\mu s_1} \left[-\frac{e^{-\mu s_2}}{\mu} \right]_t^\infty ds_1 \\ &= \mu^2 \left[0 + \frac{e^{-\mu t}}{\mu} \right] \left[0 + \frac{e^{-\mu s}}{\mu} \right] \\ &= e^{-\mu(s+t)}, \text{ Denote this as } \textit{Result 2.A.2}. \end{aligned}$$

Now we consider:

$$P(S_2 > t)P(S_1 > s) = \int_t^\infty f_{S_2}(s_2) ds_2 \int_s^\infty f_{S_1}(s_1) ds_1 \quad (1.4)$$

We know from elementary probability theory that $f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x, y) dy$.

So we have:

$$\begin{aligned} f_{S_2}(s_2) &= \int_0^\infty f_{S_1, S_2}(s_1, s_2) ds_1, \text{ Since } 0 \leq S_1 < \infty. \\ &= \int_0^\infty \mu^2 e^{-\mu(s_1+s_2)} ds_1 \\ &= \mu^2 e^{-\mu s_2} \int_0^\infty e^{-\mu s_1} ds_1 \\ &= \mu^2 e^{-\mu s_2} \left[-\frac{e^{-\mu s_1}}{\mu} \right]_0^\infty \\ &= \mu e^{-\mu s_2} \end{aligned}$$

Thus:

$$\int_t^\infty \mu e^{-\mu s_2} ds_2 = e^{-\mu t}$$

Analogously we can then say:

$$P(S_1 > s) = e^{-\mu s}$$

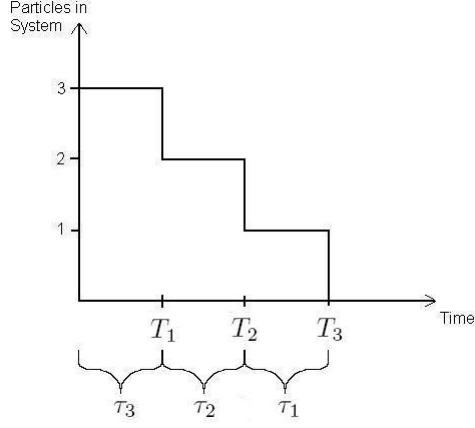
So $P(S_2 > t)P(S_1 > s) = e^{-\mu(s+t)}$. So when we compare this to *Result 2.A.2* we see that:

$$P(S_2 > t, S_1 > s) = P(S_2 > t)P(S_1 > s) \quad (1.5)$$

So we see that S_2 is independent of S_1 .

Now we consider the $n = 3$ case.

Graphical Representation:



$\mu_1, \frac{\mu_2}{2}, \frac{\mu_3}{3}$ forms a constant sequence. We shall now denote the order statistics:
 $S_{(1)} = T_1 = \inf\{t : N_t = 2\}$: Smallest of S_1, S_2, S_3 .
 $S_{(2)} = T_2 = \inf\{t : N_t = 1\}$: Second smallest of S_1, S_2, S_3 .
 $S_{(3)} = T_3 = \inf\{t : N_t = 0\}$: Third smallest (largest) of S_1, S_2, S_3 .

Recalling previous definitions we can see that $T_1 = \tau_3$, $T_2 - T_1 = \tau_2$, $T_3 - T_2 = \tau_1$. So we see that T_1 , $T_2 - T_1$, $T_3 - T_2$ are independent of each other, memoryless and have exponential distributions such that $T_1 \sim \exp(3\mu)$, $T_2 - T_1 \sim \exp(2\mu)$ and $T_3 - T_2 \sim \exp(\mu)$, by *Lemmas 1,2,3* respectively.

For the joint density of T_1 , $T_2 - T_1$, $T_3 - T_2$, we have observed values t_1 , $t_2 - t_1$, $t_3 - t_2$ respectively such that $t_3 > t_2 > t_1$. So we consider:

$$\begin{aligned} & P(T_1 \in (t_1, t_1 + h_1), T_2 \in (t_2, t_2 + h_2), T_3 \in (t_3, t_3 + h_3)) \\ &= \int_0^\infty \int_0^\infty P(T_3 - T_2 \in (t_3 - s, t_3 - s + h_3), T_2 - T_1 \in (t_2 - r, t_2 - r + h_2), T_1 \in (t_1, t_1 + h_1) | T_1 = r, T_2 = s) \\ & \quad \times f_{T_1, T_2 - T_1}(r, s - r) ds dr \\ &= \int_{t_1}^{t_1 + h_1} \int_{t_2 - r}^{t_2 - r + h_2} P(T_3 - T_2 \in (t_3 - s, t_3 - s + h_3) | T_1 = r, T_2 = s) f_{T_1, T_2 - T_1}(r, s - r) ds dr, \end{aligned}$$

Due to Independence of Increments.

$$= \int_{t_1}^{t_1 + h_1} \int_{t_2 - r}^{t_2 - r + h_2} \underbrace{P(T_3 - T_2 \in (t_3 - s, t_3 - s + h_3))}_{\int_{t_3 - s}^{t_3 - s + h_3} f_{T_3 - T_2}(t - s) dt} f_{T_1, T_2 - T_1}(r, s - r) ds dr$$

We found earlier that $f_{T_1, T_2 - T_1}(r, s - r) = f_{T_1}(r)f_{T_2 - T_1}(s - r)$.

So now we have:

$$\begin{aligned}
f_{T_1, T_2 - T_1, T_3 - T_2}(t_1, t_2 - t_1, t_3 - t_2) &= \lim_{\substack{h_i \rightarrow 0 \\ i=1,2,3}} \frac{1}{h_1 h_2 h_3} P(T_1 \in (t_1, t_1 + h_1), T_2 \in (t_2, t_2 + h_2), T_3 \in (t_3, t_3 + h_3)) \\
&= \lim_{\substack{h_i \rightarrow 0 \\ i=1,2,3}} \frac{1}{h_1 h_2 h_3} \int_{t_1}^{t_1+h_1} \int_{t_2-r}^{t_2-r+h_2} \int_{t_3-s}^{t_3-s+h_3} f_{T_3 - T_2}(t - s) f_{T_2 - T_1}(s - r) f_{T_1}(r) dr ds dt \\
&= f_{T_3 - T_2}(t_3 - t_2) f_{T_2 - T_1}(t_2 - t_1) f_{T_1}(t_1) \\
&= \mu e^{-\mu(t_3 - t_2)} \times 2\mu e^{-2\mu(t_2 - t_1)} \times 3\mu e^{-3\mu t_1} \\
&= 6\mu^3 e^{-\mu(t_1 + t_2 + t_3)}, \text{ Which we will denote as } \textit{Result 2.B.1}.
\end{aligned}$$

Now let us consider the joint distribution for S_1, S_2, S_3 :

Figure B:

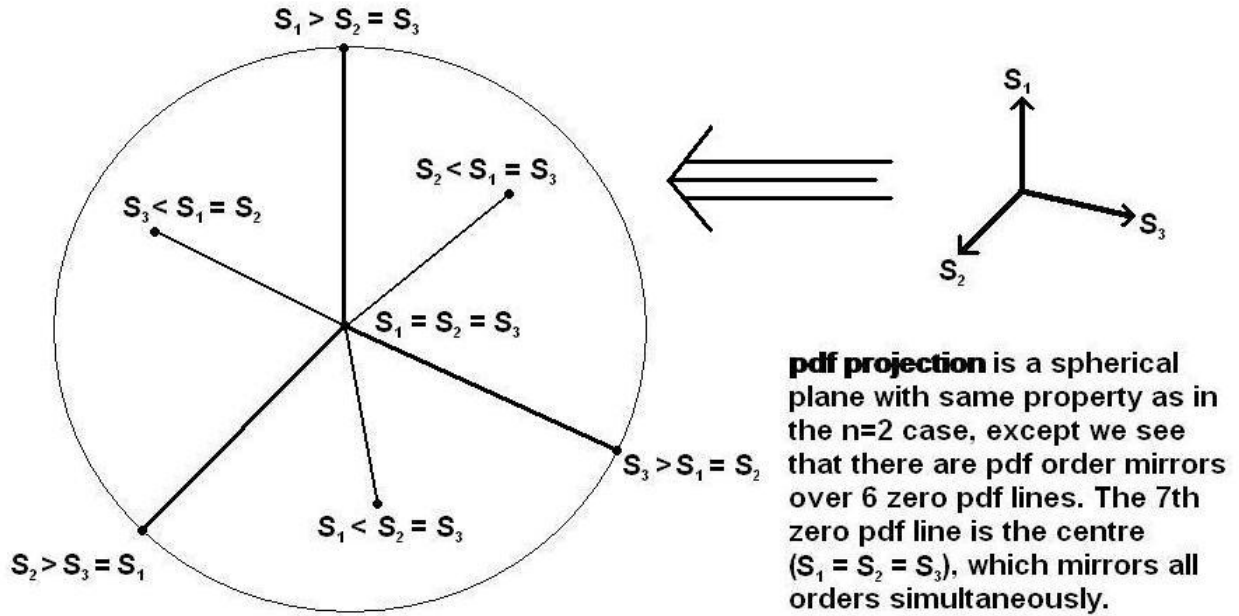


Figure B depicts a three-dimensional sphere with 6 lines that produce "zero volume shapes" when projected from the S_1, S_2, S_3 plane. If we study the diagram further we see that the joint density function generated by a given order (say) $S_3 < S_2 < S_1$ is mirrored in the case $S_3 < S_1 < S_2$, across the line $S_3 < S_1 = S_2$. In addition we observe that the joint density function generated by $S_3 < S_2 < S_1$ is also mirrored in the other five cases ($S_3 < S_1 < S_2, S_2 < S_3 < S_1, S_2 < S_1 < S_3, S_1 < S_3 < S_2, S_1 < S_2 < S_3$), across the line (point) $S_1 = S_2 = S_3$. So we have that all 6 orders (where none of S_1, S_2, S_3 are equal), have equal weights in the distribution space. Thus we can say (By Symmetry):

$$\begin{aligned}
 f_{S_1, S_2, S_3}(s_1, s_2, s_3) &= f_{S_1, S_2, S_3}(s_1, s_3, s_2) \\
 &= f_{S_1, S_2, S_3}(s_2, s_1, s_3) \\
 &= f_{S_1, S_2, S_3}(s_2, s_3, s_1) \\
 &= f_{S_1, S_2, S_3}(s_3, s_1, s_2) \\
 &= f_{S_1, S_2, S_3}(s_3, s_2, s_1)
 \end{aligned}$$

So to find the joint density function we consider:

$$\begin{aligned}
P(S_{(1)} \in (t_1, t_1 + dt_1), S_{(2)} \in (t_2, t_2 + dt_2), S_{(3)} \in (t_3, t_3 + dt_3)) \\
&= P(T_1 \in (t_1, t_1 + dt_1), T_2 \in (t_2, t_2 + dt_2), T_3 \in (t_3, t_3 + dt_3)) \\
&= P(T_1 \in (t_1, t_1 + dt_1), T_2 - T_1 \in (t_2 - t_1, t_2 - t_1 + dt_2), T_3 - T_2 \in (t_3 - t_2, t_3 - t_2 + dt_3)) \\
&= 6\mu^3 e^{-\mu(t_1+t_2+t_3)} dt_1 dt_2 dt_3, \text{ Recall } \textit{Result 2.B.1}.
\end{aligned}$$

Also we have that:

$$\begin{aligned}
P(S_{(1)} \in (t_1, t_1 + dt_1), S_{(2)} \in (t_2, t_2 + dt_2), S_{(3)} \in (t_3, t_3 + dt_3)) \\
&= 6f_{S_1, S_2, S_3}(t_1, t_2, t_3) dt_1 dt_2 dt_3, \text{ By Symmetry.}
\end{aligned}$$

This implies that:

$$\begin{aligned}
f_{S_1, S_2, S_3}(t_1, t_2, t_3) &= \frac{1}{6} \frac{1}{dt_1 dt_2 dt_3} 6\mu^3 e^{-\mu(t_1+t_2+t_3)} dt_1 dt_2 dt_3 \\
&= \mu^3 e^{-\mu(t_1+t_2)} \\
&= f_{S_1, S_2, S_3}(t_1, t_3, t_2) = \dots = f_{S_1, S_2, S_3}(t_3, t_2, t_1)
\end{aligned}$$

This yields the result:

$$f_{S_1, S_2, S_3}(s_1, s_2, s_3) = \mu^3 e^{-\mu(s_1+s_2+s_3)}, \text{ for all } s_1, s_2, s_3. \quad (1.6)$$

Now lets consider the relationship between S_1, S_2, S_3 .

$$\begin{aligned}
P(S_1 > t, S_2 > s, S_3 > r) &= \int_{S_1 \in (t, \infty)} \int_{S_2 \in (s, \infty)} \int_{S_3 \in (r, \infty)} f_{S_1, S_2, S_3}(s_1, s_2, s_3) ds_1 ds_2 ds_3 \\
&= \int_t^\infty \int_s^\infty \int_r^\infty \mu^3 e^{-\mu(s_1+s_2+s_3)} ds_1 ds_2 ds_3 \\
&= \mu^3 \int_t^\infty \int_s^\infty e^{-\mu(s_1+s_2)} ds_1 ds_2 \left[-\frac{e^{-\mu s_3}}{\mu} \right]_r^\infty \\
&= \mu^3 \left[0 + \frac{e^{-\mu t}}{\mu} \right] \left[0 + \frac{e^{-\mu s}}{\mu} \right] \left[0 + \frac{e^{-\mu r}}{\mu} \right] \\
&= e^{-\mu(t+s+r)}, \text{ Denote this as } \textit{Result 2.B.2}.
\end{aligned}$$

Now we consider:

$$P(S_1 > t)P(S_2 > s)P(S_3 > r) = \int_t^\infty f_{S_1}(s_1) ds_1 \int_s^\infty f_{S_2}(s_2) ds_2 \int_r^\infty f_{S_3}(s_3) ds_3 \quad (1.7)$$

Then since $0 \leq (s_2, s_3) < \infty$, we have:

$$\begin{aligned}
f_{S_3}(s_3) &= \int_0^\infty \int_0^\infty f_{S_1, S_2, S_3}(s_1, s_2, s_3) ds_1 ds_2 \\
&= \int_0^\infty \int_0^\infty \mu^3 e^{-\mu(s_1+s_2+s_3)} ds_1 ds_2 \\
&= \mu^3 e^{-\mu s_3} \int_0^\infty e^{-\mu s_1} ds_1 \int_0^\infty e^{-\mu s_2} ds_2 \\
&= \mu^3 e^{-\mu s_3} \left[-\frac{e^{-\mu s_1}}{\mu} \right]_0^\infty \left[-\frac{e^{-\mu s_2}}{\mu} \right]_0^\infty \\
&= \mu e^{-\mu s_3}
\end{aligned}$$

Then we have:

$$P(S_3 > r) = e^{-\mu r}$$

Analogously we also have:

$$\begin{aligned} P(S_1 > t) &= e^{-\mu t} \\ P(S_2 > s) &= e^{-\mu s} \end{aligned}$$

So $P(S_1 > t)P(S_2 > s)P(S_3 > r) = e^{-\mu(t+s+r)}$ and when we compare this to *Result 2.B.2* we see that:

$$P(S_1 > t, S_2 > s, S_3 > r) = P(S_1 > t)P(S_2 > s)P(S_3 > r) \quad (1.8)$$

So we see that S_1, S_2, S_3 are all independent of each other.

Now we consider the n -element case:

We know that $(T_1, T_2 - T_1, \dots, T_n - T_{n-1}) = (\tau_n, \tau_{n-1}, \dots, \tau_1)$ respectively, and here we have that $T_1 \sim \exp(n\mu)$, $T_2 - T_1 \sim \exp((n-1)\mu), \dots, T_n - T_{n-1} \sim \exp(\mu)$. Thus:

$$\begin{aligned} f_{T_1, T_2, \dots, T_n}(t_1, t_2, \dots, t_n) &= f_{T_1, T_2 - T_1, \dots, T_n - T_{n-1}}(t_1, t_2 - t_1, \dots, t_n - t_{n-1}) \\ &= f_{T_1}(t_1) f_{T_2 - T_1}(t_2 - t_1) \dots f_{T_n - T_{n-1}}(t_n - t_{n-1}), \text{ Due to independence of exponentials.} \\ &= n\mu e^{-n\mu t_1} (n-1)\mu e^{-(n-1)\mu(t_2 - t_1)} \dots \mu e^{-\mu(t_n - t_{n-1})} \\ &= n! \mu^n e^{-\mu(t_1 + t_2 + \dots + t_n)} \end{aligned}$$

Now S_1, S_2, \dots, S_n :

$$\begin{aligned} f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) &= \frac{f_{S_{(1)}, S_{(2)}, \dots, S_{(n)}}(t_1, t_2, \dots, t_n)}{n!} \\ &= \frac{\sum_{\text{all } \pi} f_{s_{\pi(1)}, s_{\pi(2)}, \dots, s_{\pi(n)}}(s_1, s_2, \dots, s_n)}{n!}, \text{ } \pi \text{ represents permutations of } 1, 2, \dots, n. \\ &= \mu^n e^{-\mu(s_1 + s_2 + \dots + s_n)} \end{aligned}$$

Thus we have:

$$f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) = \mu^n e^{-\mu(s_1 + s_2 + \dots + s_n)}, \text{ for all } s_1, s_2, \dots, s_n. \quad (1.9)$$

Now lets consider the relationship between S_1, S_2, \dots, S_n :

$$\begin{aligned} P(S_1 > z_1, S_2 > z_2, \dots, S_n > z_n) &= \int_{z_1}^{\infty} \int_{z_2}^{\infty} \dots \int_{z_n}^{\infty} f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_n \\ &= \int_{z_1}^{\infty} \int_{z_2}^{\infty} \dots \int_{z_n}^{\infty} \mu^n e^{-\mu(s_1 + s_2 + \dots + s_n)} ds_1 ds_2 \dots ds_n \\ &= \mu^n \left[0 + \frac{e^{-\mu z_1}}{\mu} \right] \left[0 + \frac{e^{-\mu z_2}}{\mu} \right] \dots \left[0 + \frac{e^{-\mu z_n}}{\mu} \right] \\ &= e^{-\mu(z_1 + z_2 + \dots + z_n)} \end{aligned}$$

Now we consider:

$$P(S_1 > z_1)P(S_2 > z_2) \dots P(S_n > z_n) = \int_{z_1}^{\infty} f_{S_1}(s_1) ds_1 \int_{z_2}^{\infty} f_{S_2}(s_2) ds_2 \dots \int_{z_n}^{\infty} f_{S_n}(s_n) ds_n \quad (1.10)$$

Then since $0 \leq s_j < \infty$ for all $1 \leq j \leq n$ we have:

$$\begin{aligned}
f_{S_n}(s_n) &= \int_0^\infty \int_0^\infty \dots \int_0^\infty f_{S_1, S_2, \dots, S_n}(s_1, s_2, \dots, s_n) ds_1 ds_2 \dots ds_{n-1} \\
&= \int_0^\infty \int_0^\infty \dots \int_0^\infty \mu^n e^{-\mu(s_1 + s_2 + \dots + s_n)} ds_1 ds_2 \dots ds_{n-1} \\
&= \mu^n e^{-\mu s_n} \left[-\frac{e^{-\mu s_1}}{\mu} \right]_0^\infty \left[-\frac{e^{-\mu s_2}}{\mu} \right]_0^\infty \dots \left[-\frac{e^{-\mu s_{n-1}}}{\mu} \right]_0^\infty \\
&= \mu e^{-\mu s_n}
\end{aligned}$$

Thus we have:

$$P(S_n > z_n) = e^{-\mu z_n}$$

Analogously we also have:

$$\begin{aligned}
P(S_1 > z_1) &= e^{-\mu z_1} \\
P(S_2 > z_2) &= e^{-\mu z_2} \\
&\vdots \\
P(S_{n-1} > z_{n-1}) &= e^{-\mu z_{n-1}}
\end{aligned}$$

This implies that $P(S_1 > z_1, S_2 > z_2, \dots, S_n > z_n) = P(S_1 > z_1)P(S_2 > z_2)\dots P(S_n > z_n)$, and thus we see that S_1, S_2, \dots, S_n are all independent of each other.

Chapter 2

Intensity Relationships and Associated Dependence Properties in a Pure Death Process

This chapter will attempt to summarise the works of Faddy (1985), Ball and Donnelly (1987), Brown and Donnelly (1993).

Faddy (1985) conjectured that, in a pure death process, variability of death processes increase (decreases), relative to the linear case, if the death rates $\mu_0, \mu_1, \dots, \mu_N$ form a concave (convex) sequence. Faddy (1985) uses a variability function $V_t = \frac{Var(X_t)}{E(X_t)(1 - \frac{E(X_t)}{N})}$ to determine the process variability relative to the linear case. It is asserted that in the linear case $V_t = 1$ and $V_t \geq (\leq) 1$ if $\mu_0, \mu_1, \dots, \mu_N$ form a concave (convex) sequence, for all t . If $V_t > (<) 1$ then the process is more (less) variable than the linear case.

Ball and Donnelly (1987) announce that they settle the Faddy (1985) conjecture and clarify the implication of concave (convex) death rates. They [Ball and Donnelly 1987 p. 759] introduce a definition which states that a sequence $\mu_1, \mu_2, \dots, \mu_N$ is said to be superlinear (sublinear) if $\frac{\mu_n}{n} \leq (\geq) \frac{\mu_{n+1}}{n+1}$, for $n = 1, 2, \dots, N - 1$, with strict inequality (in both cases) for at least one n . A theorem [see Theorem 1. Ball and Donnelly 1987] follows this definition and states that for $n = 2, 3, \dots, N$ and any $t_1, t_2, \dots, t_n > 0$, if $\mu_1, \mu_2, \dots, \mu_N$ form a sublinear (superlinear) sequence then:

$$E \left(\prod_{k=1}^n I_k(t_k) \right) > (<) \prod_{k=1}^n E(I_k(t_k))$$

Where $I_k(t_k)$ is an indicator function which is equal to 1 if particle k still exists in the population, at time t_k . Using this theorem they claim to show that a decreasing (increasing) average death rate $\left(\frac{\mu_n}{n}\right)$, $n = 1, 2, \dots, N$, implies that the fates of individuals are positively (negatively) correlated.

Brown and Donnelly (1993) claim that the proof in Ball and Donnelly (1987) has an error. They claim that the second equation in the proof of *Theorem 1. Ball and Donnelly (1987)* depicts a situation where the conditional probabilities cannot be written in terms of conditional hazard rates for $n = 2$. Nevertheless, Brown and Donnelly (1993) insist that the result (*Theorem 1.*) is true. Lefèvre and Milhaud (1990) take

advantage of the conditional independence structure and provide a much more general result. For a pure death process with optional death times T_1, T_2, \dots, T_n , and observations $0 \leq t_1, t_2, \dots, t_n$, we have:

$$P(T_1 > t_1 \cap T_2 > t_2 \cap \dots \cap T_n > t_n) \geq (<) P(T_1 > t_1) P(T_2 > t_2) \dots P(T_n > t_n)$$

if the average death rates $\mu_1, \frac{\mu_2}{2}, \dots, \frac{\mu_n}{n}$ form a decreasing (increasing) sequence.

(This is denoted as Theorem 3. in Brown and Donnelly (1993) and is proven using *multivariate total positivity of order 2*)

For our analysis we will have that if individual death rates $\frac{\mu_n}{n}, \frac{\mu_{n-1}}{n-1}, \dots, \mu_1$ form an increasing (decreasing) sequence, then interparticle correlation is positive (negative).

Chapter 3

Analysing Life Data

Life tables are used to depict the death experience of the population at each age. Life table data is collected at discrete-time intervals and the exact death times of the subjects are not explicitly stated. So an assumption regarding the death times is made so that we can apply the theory in discrete time. Assuming uniform distribution of deaths within each age interval, we seek to use the increasing/decreasing individual death rates, obtained from the data, to identify the effect on other lives within the population.

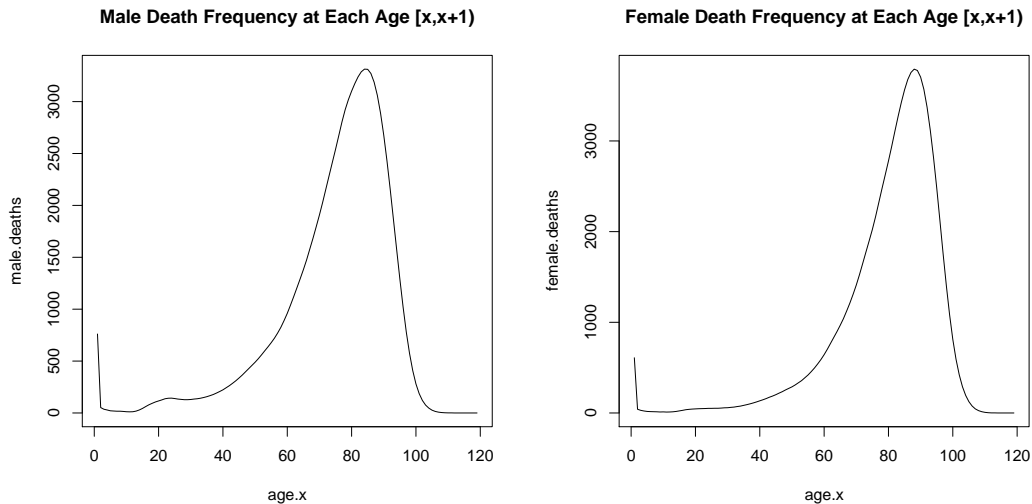
3.1 The Data

The following life data was obtained from the United States Social Security Service [?], accessed 12/9/2007. This data describes the death experience of males and females who live in the United States. Data such as this is usually used by the life insurance industry to determine life expectancies of its customers. Life tables have been used as evidence that young males aged 0-60 have, on average, a higher death rate than women within the same age interval. However, at later ages, on average, females have a slightly higher death rate than males.

Life insurance companies use the information they obtain from life tables as a factor in determining the appropriate, "actuarially fair", premium for the policyholder. An actuarially fair premium calculation takes into account the risk of claim of the client (death risk in this case), the profit target of the insurance company and the reserving requirements which the company must maintain in order to avoid defaulting on claims. So for the life insurance company, life expectancy is very important. (see Evstigneev and Haneveld [?] (1999) for further discussion)

The general insurance industry also uses life data to set premiums, especially in motor vehicle insurance. Life tables have been used as evidence that male drivers from 18-25 have increasing death rates. Thus motor vehicle insurance companies would naturally set premiums higher for a male within this age interval compared to a female within this age interval.

This paper seeks to use life data to find whether an increasing/decreasing individual death rate, for a group of lives, affects other lives within the same population. Below we have a graphical representation for the number of male deaths (left) and number of female deaths (right).



3.2 Life Data Analysis

We seek to apply the theory in the *Preliminaries* chapter, and so we seek to determine the sequence produced by the average death rates. We know that $T_1 \sim \exp(\mu_N)$, $T_2 - T_1 \sim \exp(\mu_{N-1})$, ..., $T_N - T_{N-1} \sim \exp(\mu_1)$, where $N = 100000$ is the total population (number of newborns aged exactly 0). This implies the likelihood function:

$$L(\tilde{\mu}; \tilde{t}) = \prod_{i=2}^N \mu_{N-i+1} e^{-\mu_{N-i+1}(t_i - t_{i-1})} \times \mu_N e^{-\mu_N t_1}$$

Where we have vectors $\tilde{\mu} = (\mu_N, \mu_{N-1}, \dots, \mu_1)$ and $\tilde{t} = (t_1, t_2, \dots, t_N)$. The log-likelihood which follows:

$$l(\tilde{\mu}; \tilde{t}) = \sum_{i=2}^N [\ln(\mu_{N-i+1}) - \mu_{N-i+1}(t_i - t_{i-1})] + \ln(\mu_N) - \mu_N t_1$$

When we take the partial derivative with respect to μ_N we obtain:

$$\frac{\partial l}{\partial \mu_N} = \frac{1}{\mu_N} - t_1$$

When we set this partial derivative equal to zero to find the maximum likelihood estimate (MLE) we obtain $\hat{\mu}_N = \frac{1}{t_1}$. When we take the partial derivative with respect to μ_{N-1} :

$$\frac{\partial l}{\partial \mu_{N-1}} = \frac{1}{\mu_{N-1}} - (t_2 - t_1)$$

We set this equal to zero and obtain the MLE $\hat{\mu}_{N-1} = \frac{1}{t_2 - t_1}$. This implies that for $2 \leq i \leq N$:

$$\hat{\mu}_{N-i+1} = \frac{1}{t_i - t_{i-1}}$$

Each age interval is exposed to a specific death frequency. Let us denote k_i , for $i = 0, 1, \dots, \omega$, to be the number of deaths which occur during age $[i, i + 1)$, where ω represents the maximum (limiting) age of the population. This implies that t_{k_i} is the final death time within the age interval. Let us define $K_i := \sum_{j=0}^i k_j$ and we denote ν_i to represent the transition intensity (towards death) of an individual during age $[i, i + 1)$. The data does not allow us to distinguish between death times within an age interval (since deaths recorded at discrete ages) so we assume that death times, within the age interval, are chosen randomly and uniformly conditional upon the fact that a certain number of deaths occur within the age interval. More specifically, we are assuming that if $(S_1, S_2, \dots, S_{k_i}) \in [i, i + 1)$, $i = 0, 1, \dots, \omega$, then, strictly within the age interval, S_j 's, $j = 1, 2, \dots, k_i$, are independent identically distributed continuous random variables such that $S_j \sim U(i, i + 1)$. This suggests we cannot distinguish different death intensities within an age interval so we have that ν_i is equivalent to $\frac{\mu_{N-K_i-1}}{N-K_i-1} = \frac{\mu_{N-K_{i-1}-1}}{N-K_{i-1}-1} = \dots = \frac{\mu_{N-K_i+1}}{N-K_i+1}$.

Let us first consider age $[0, 1)$. The death intensity of the k_0 individuals within this age interval, has the following attribute:

$$\begin{aligned} \mu_N &= N\nu_0 \\ \mu_{N-1} &= (N-1)\nu_0 \\ &\vdots \\ \mu_{N-k_0+1} &= (N-k_0+1)\nu_0 \end{aligned}$$

Recall:

$$\hat{\mu}_N = \frac{1}{t_1} = \hat{\mu}_{N-1} = \frac{1}{t_2 - t_1}, \dots, \hat{\mu}_{N-k_0+1} = \frac{1}{t_{k_0} - t_{k_0-1}}$$

From here we rewrite and substitute the attribute defined above and obtain:

$$t_1 = \frac{1}{\hat{\mu}_N} = \frac{1}{N\nu_0}, t_2 - t_1 = \frac{1}{\hat{\mu}_{N-1}} = \frac{1}{(N-1)\nu_0}, \dots, t_{k_0} - t_{k_0-1} = \frac{1}{\hat{\mu}_{N-k_0+1}} = \frac{1}{(N-k_0+1)\nu_0}$$

This implies that:

$$\begin{aligned} t_{k_0} &= t_1 + (t_2 - t_1) + \dots + (t_{k_0} - t_{k_0-1}) \\ &= \frac{1}{\nu_0} \left(\frac{1}{N} + \frac{1}{N-1} + \dots + \frac{1}{N-k_0+1} \right) \\ \nu_0 &= \frac{\sum_{i=1}^{k_0} \frac{1}{N-i+1}}{t_{k_0}} \end{aligned}$$

If we are able to find t_{k_0} , we would be able to obtain ν_0 and thus obtain $\mu_N, \mu_{N-1}, \dots, \mu_{N-k_0+1}$.

Now let us consider age interval $[i, i+1)$ for $i = 2, 3, \dots, \omega$. We know that ν_i is equivalent to all individual death intensities within the age interval. It follows that $\mu_{N-j+1} = (N-j+1)\nu_i$ for $j = K_{i-1} + 1, K_{i-1} + 2, \dots, K_i$. This implies, by the maximum likelihood estimate $t_j - t_{j-1} = \frac{1}{\hat{\mu}_{N-j+1}} = \frac{1}{(N-j+1)\nu_i}$, that:

$$t_{K_i} = \frac{1}{\nu_i} \sum_{j=K_{i-1}+1}^{K_i} \frac{1}{N-j+1} \quad (3.1)$$

$$\nu_i = \frac{\sum_{j=K_{i-1}+1}^{K_i} \frac{1}{N-j+1}}{t_{K_i}} \quad (3.2)$$

If we can find t_{K_i} , we can obtain ν_i and all corresponding μ_{N-j+1} 's.

3.3 Uniform Distribution of Final Death Times

Lemma 4

Let X_1, X_2, \dots, X_m be independent and each have distribution $U(i, i + 1)$, then:
 $E[\max(X_1, X_2, \dots, X_m)] = \frac{(i+1)m+i}{m+1}$.

Proof:

$$\begin{aligned} P(\max(X_1, X_2, \dots, X_m) \leq t) &= P(X_1 \leq t, X_2 \leq t, \dots, X_m \leq t) \\ &= [P(X_1 \leq t)]^m, \text{ Since } X_1, \dots, X_m \text{ are independent.} \\ &= \left[\frac{t-i}{i+1-i} \right]^m = (t-i)^m, \quad i < t < i+1. \end{aligned}$$

Thus:

$$f_{\max(X_1, X_2, \dots, X_m)}(t) = m(t-i)^{m-1}$$

Now we have:

$$\begin{aligned} E[\max(X_1, X_2, \dots, X_m)] &= \int_i^{i+1} tm(t-i)^{m-1} dt \\ &\quad \text{Let } u = t - i \\ &= m \int_0^1 (u+i)u^{m-1} du \\ &= m \left[\frac{1}{m+1} u^{m+1} + \frac{i}{m} u^m \right]_0^1 \\ &= \frac{m}{m+1} + i \\ &= \frac{(i+1)m+i}{m+1} \quad \text{- End Proof.} \end{aligned}$$

In this section we assume that S_1, S_2, \dots, S_{k_i} are independent with identical distribution

$(S_1, S_2, \dots, S_{k_i}) \sim U(i, i + 1)$. This implies, by *Lemma 4*, $E(\max(S_1, S_2, \dots, S_{k_i})) = E(T_{K_i}) = \frac{(i+1)k_i+i}{k_i+1}$.
 However, if we find that cells don't have a significant amount of deaths, we create the following lemma.

Lemma 5

Let X_1, X_2, \dots, X_m be independent and each have distribution $U(i, b)$, where i is the beginning of the merged age intervals and b is the last age in the merged age interval, then:
 $E[\max(X_1, X_2, \dots, X_m)] = \frac{bm+i}{m+1}$.

Proof:

$$\begin{aligned} P(\max(X_1, X_2, \dots, X_m) \leq t) &= P(X_1 \leq t, X_2 \leq t, \dots, X_m \leq t) \\ &= P(X_1 \leq t)^m, \text{ Since } X_1, X_2, \dots, X_m \text{ are independent.} \\ &= \left[\frac{t-i}{b-i} \right]^m, \quad i < t < b. \end{aligned}$$

Thus:

$$f_{\max(X_1, X_2, \dots, X_m)}(t) = \frac{m}{(b-i)^m} (t-i)^{m-1}$$

Now we have:

$$\begin{aligned} E[\max(X_1, X_2, \dots, X_m)] &= \int_i^b t \frac{m}{(b-i)^m} (t-i)^{m-1} dt \\ &\text{Let } u = t - i \\ &= \frac{m}{(b-i)^m} \int_0^{b-i} (u+i)u^{m-1} du \\ &= \frac{m}{(b-i)^m} \left[\frac{1}{m+1} u^{m+1} + \frac{i}{m} u^m \right]_0^{b-i} \\ &= \frac{m}{(b-i)^m} \left[\frac{1}{m+1} (b-i)^{m+1} + \frac{i}{m} (b-i)^m \right] \\ &= (b-i) \frac{m}{m+1} + i \\ &= \frac{bm+i}{m+1} \quad \text{- End Proof.} \end{aligned}$$

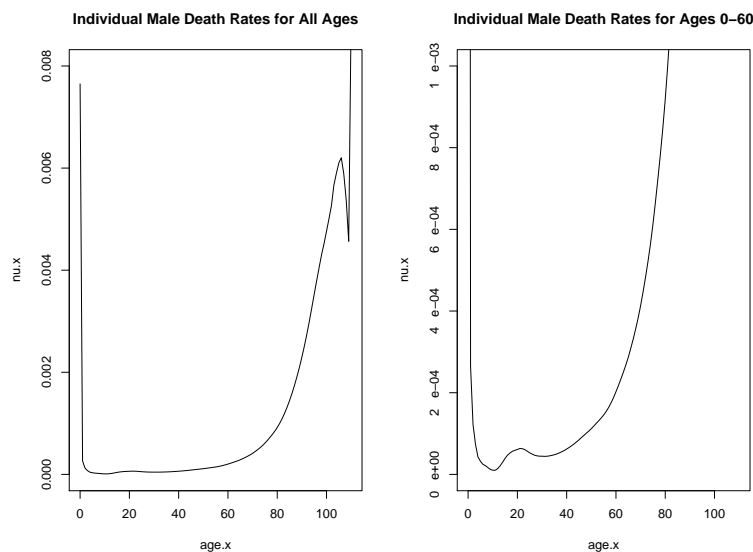
In our analysis we have S_1, S_2, \dots, S_{k_b} are independent with identical distribution $(S_1, S_2, \dots, S_{k_b}) \sim U(i, b)$. This implies, by *Lemma 5*, $E(\max(S_1, S_2, \dots, S_{k_i})) = E(T_{K_b}) = \frac{bm+i}{m+1}$.

Ball and Donnelly [?] (1988) state that if application of an independent particle model results in underrepresentation of variability, then the interparticle correlations are positive and any model modification which addresses this will improve the model. Using *Theorem 3. Brown and Donnelly (1993)* we can find the interparticle correlation from the individual death rate sequence we obtain.

Chapter 4

Results and Discussion of the Analysis

Applying the uniform distribution assumption, and *Lemma 4*, we display the individual death rate for each age.



The death rate diagram on the left depicts the individual death rate (ν_x) for all ages $[x, x + 1)$. The death rate diagram on the right magnifies the trend of the individual death rates in the earlier ages. Both plots strongly support increasing individual death rates as age approaches ω . In order to see exactly where the individual death rate is increasing/decreasing we take first differences of successive ν_x 's and plot a sign graph. The sign graph will plot 1 if the sign difference is positive (increasing individual death rate) and -1 if the sign difference is negative (decreasing individual death rate).



The sign plot shows us that the majority of the age intervals exhibit an increasing individual death rate. However we have not taken into account the magnitude of first differences. To take into account the magnitude of first differences we note that since ν_i depends on the death frequency k_i on $[i, i + 1)$, we will use the expected death frequency to create a new sign plot which will reject insignificant death rate differences.

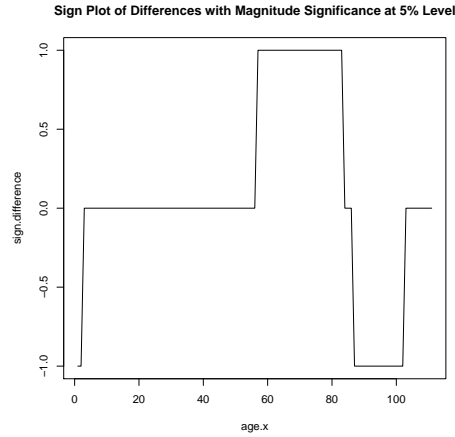
The expected death frequency per age interval is the total number of deaths divided by the number of age intervals. If we denote D_i to be a random variable for the number of deaths in age interval $[i, i + 1)$, then we are saying that $E(D_i) = \frac{N}{\omega}$. For males we have $N=100000$ and $\omega=111$ so $E(D_i)=900.90 \approx 901$. Our null hypothesis is that there is no difference between the death frequencies between successive ages. Our alternative hypothesis is that there is significant difference between the death frequencies between ages. Thus our test statistic is of the form:

$$T = \frac{k_{i+1} - k_i}{E(D_i)} \quad (4.1)$$

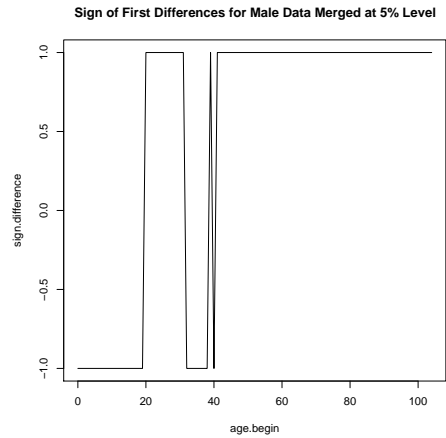
Now if $k_{i+1} - k_i > \alpha \times E(D_i)$ we reject H_0 , otherwise we accept that there is no difference between the death frequencies. We treat α as the random fluctuation which we want to remove from the model. If we set $\alpha = 0.05$ then we have a critical value of $\alpha \times E(D_i) = 0.05 \times 901 = 45.05$. Thus we use the following criteria for the new sign plot:

$$\begin{aligned} k_{i+1} - k_i > 45, & \text{ Sign} = +1. \\ k_{i+1} - k_i \in [-45, 45], & \text{ Sign} = 0. \\ k_{i+1} - k_i < -45, & \text{ Sign} = -1. \end{aligned}$$

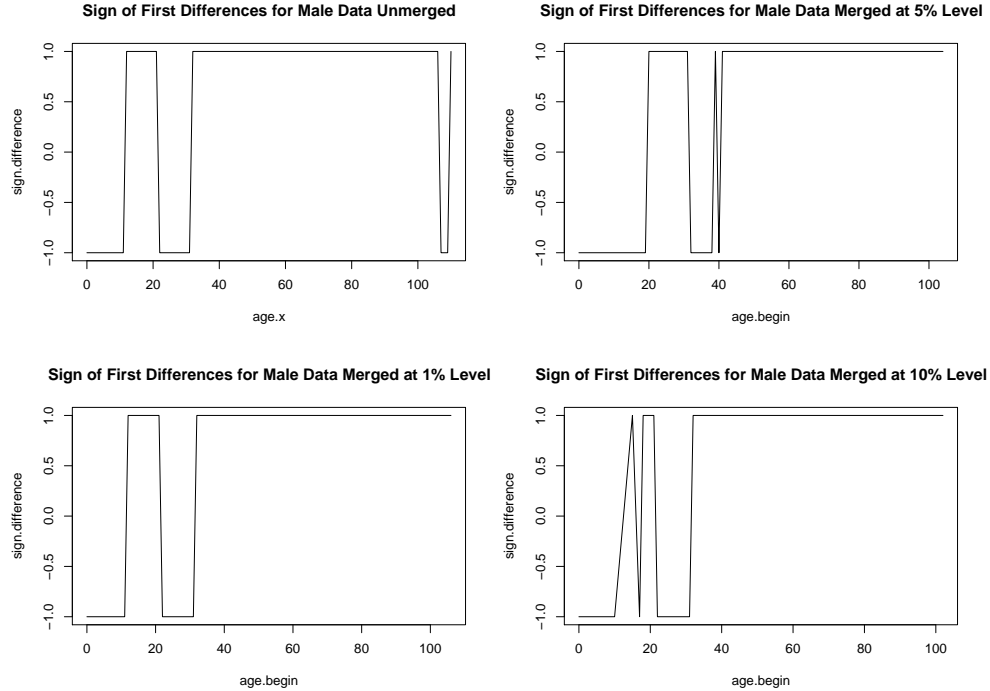
Thus we have a new sign plot:



The plot shows that most of the death frequency differences are insignificant. Specifically, 66 zeros, 27 positives and 18 negatives were obtained. This suggests support for the positive (ie. increasing individual death rates) but we have observed alot of zeros as well. The high number of zeros may suggest that we have cells with relatively small death frequencies. So, subject to this criteria, we merge the cells appropriately and recalculate the relevant individual death rates using *Lemma 5*. Therefore we have the following sign plot:



This plot shows a convincingly positive first difference at ages greater than 40. It was found that a total of 79 out of 96 (82%) groups had a positive first difference. We repeat the cell merging procedure but this time we set $\alpha = 0.01$ and $\alpha = 0.10$ to represent 1% and 10% random fluctuation respectively. We compare the results as follows:

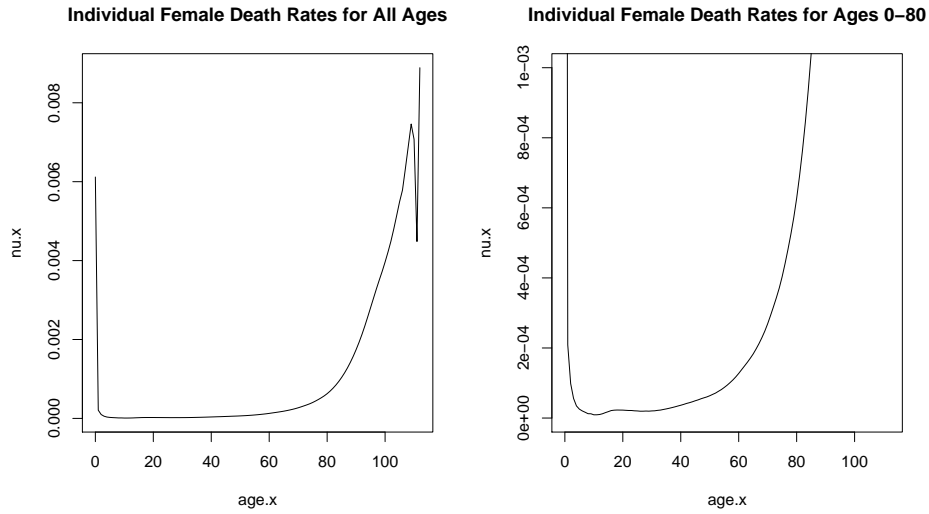


α	Number of Positives	Number of Negatives	Positive Dominance
none	86	25	0.7747748
5%	77	19	0.8020833
1%	85	22	0.7943925
10%	76	15	0.8351648

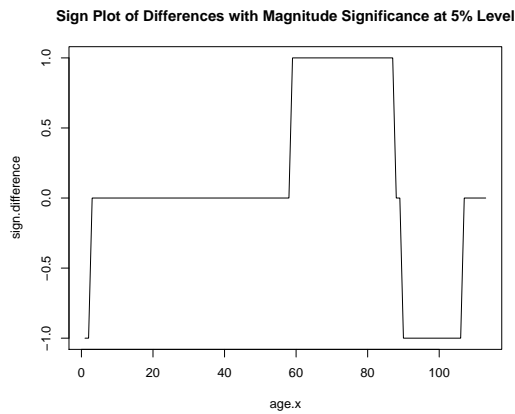
The plots show a clear domination of positive individual death rate sequence. The table shows us that as we increase the random fluctuation parameter (α), the positive dominance percentage increases. The theory suggests that since the individual death rates are increasing we have

$P(S_1 > s_1 \cap S_2 > s_2 \cap \dots \cap S_n > s_n) \geq \prod_{j=1}^n P(S_j > s_j)$. This implies a non-negative correlation between male death times and hence when an individual male dies another male within the population is more likely to die.

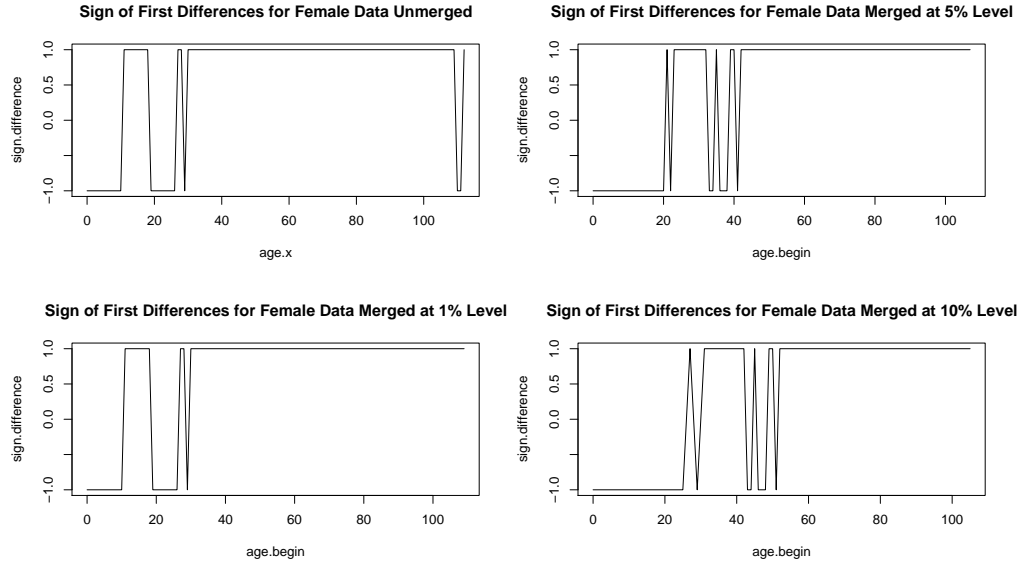
We repeat the analysis for the female population. Initially we have the following individual female death rate for each age, applying *Lemma 4*.



The new significance sign plot is depicted below:



At the 5% significance level we see a similar non-significant first difference dominance as with the male data. Specifically, 65 insignificant counts, 29 positives and 19 negatives. Once again, we proceed to apply *Lemma 5* and merge cells appropriately at levels $\alpha = 5\%, 1\%, 10\%$. For females we have $N = 100000$ and $\omega = 113$. Thus the new critical value is of the form $\alpha \times 884.9558 \approx 885\alpha$. The following results are obtained.



α	Number of Positives	Number of Negatives	Positive Dominance
none	91	22	0.8053097
5%	80	16	0.8333333
1%	90	20	0.8181818
10%	68	16	0.8095238

The plots once again show a clear dominance of positive sign of first differences. However, we see more sign fluctuation for the 5% and the 10% levels, relative to males. This may be due to the fact that female deaths are more evenly distributed across different ages, relative to the male deaths. The table shows us that the highest positive dominance is achieved at the 5% level which is not consistent with the male result of higher dominance with higher α . This may suggest that females and males are exposed to different random fluctuations. Nonetheless, as with the male data, females seem to display an increasing individual death rate sequence. This implies that female death times are positively correlated and thus we have that the death of one female increases the death propensity for the next female in the same population.

Conclusion

The transition probabilities, within birth and death processes, are difficult to obtain analytically. Numerical methods are required to solve the differential equations given by the resulting Kolmogorov equations. It was shown that by considering pure birth and pure death processes, the system can be analytically solved for transition probabilities. We discover that techniques used to solve pure birth processes can be applied to pure death processes and both systems are able to clearly specify transition probabilities for the entire system. We find that in a linear continuous-time Markovian pure death process we are able to define joint distribution functions between optional death times and confirm that interparticle correlation structure exhibits independence.

The conjecture made by Faddy (1985), regarding variability of death processes, was announced to be true by Ball & Donnelly (1987). It was conjectured that if death rates $\mu_0, \mu_1, \dots, \mu_N$ form a concave (convex) sequence then the variability of the model increases (decreases) relative to a linear continuous-time Markovian pure death process. Ball and Donnelly (1987) claim that if $\mu_1, \mu_2, \dots, \mu_N$ form a sublinear (superlinear) sequence, then interparticle correlation is positive (negative), where sublinear (superlinear) implies that $\frac{\mu_n}{n}$ decreases (increases) with n . Brown and Donnelly (1993) found an error in the Ball and Donnelly (1987) proof and show that an alternative approach, using *multivariate total positivity of order 2*, allows a more direct connection between interparticle correlations and death rate sequences. It is found that if individual death rates $\mu_1, \frac{\mu_2}{2}, \dots, \frac{\mu_n}{n}$ form a decreasing (increasing) sequence then the fates of individuals are positively (negatively) correlated.

In our analysis it was found that life data exhibits predominantly increasing individual death rates, but the analysis relied on an independent and identically distributed uniform distribution assumption within age intervals. By using a maximum likelihood estimate it was found that individual death rates are dependent upon the number of deaths within each age interval and the final death time within each age interval. *Lemma 4* and *lemma 5* eliminate the need for performing arbitrary simulation procedures and thus reliable individual death rate estimates are obtained. These estimates are used to model the individual death rate structure governing both males and females in the data set. It was found that males and females have predominantly increasing individual death rates and thus exhibit a positive interparticle correlation structure. This suggests that when a male or female dies within a population, the next male or female, respectively, is more likely to die within a shorter period of time. In other words, the next individual is subject to a higher individual death rate.

Bibliography

- [1] Actuarial Publications Period Life Table, 2003. <http://www.ssa.gov/OACT/STATS/table4c6.html>, accessed 6/9/2007
- [2] F. Ball & P. Donnelly. Interparticle correlation in death processes with application to variability in compartmental models. *Adv. in Appl. Probab.*, 19 (4):755–766, 1987.
- [3] F. Ball & P. Donnelly. A unified approach to variability in compartmental models. (French summary). *Biometrics*, 44 (3):685–694, 1988.
- [4] U. N. Bhat & G. K. Miller. *Elements of Applied Stochastic Processes 3rd Edition*. John Wiley and Sons, Inc. Hoboken, New Jersey, 2002.
- [5] T. C. Brown & P. Donnelly. On Conditional Intensities and on Interparticle Correlation in Nonlinear Death Processes. *Adv. Appl. Prob.*, 25 (1):255–260, 1993.
- [6] I. V. Evstigneev & W. K. Haneveld & L. J. Mirman. Robust insurance mechanisms and the shadow prices of information constraints. (English summary). *J. Appl. Math. Decis. Sci.*, 3 (1):85–128, 1999.
- [7] M. J. Faddy. Nonlinear stochastic compartmental models. *IMA J. Math. Appl. Med. Biol.*, 2 (4):287–297, 1985.
- [8] A. A. Kherani. A direct approach to sojourn times in a busy period of an $M/M/1$ queue., *Queueing System.*, 53 (3):159–169, 2006.
- [9] C. Lefèvre & G. Michaletzky. Interparticle dependence in a linear death process subjected to a random environment. *J. Appl. Probab.*, 27 (3):491–498, 1990.
- [10] C. Lefèvre & X. Milhaud. On the association of the lifelengths of components subjected to a stochastic environment. *Adv. in Appl. Probab.*, 22 (4):961–964, 1990.
- [11] S. M. Ross. *Stochastic processes*. Second edition. John Wiley and Sons, Inc., New York, 1996.