Lecture Notes on Stochastic Processes in Biostatistics: Applications to Infectious Diseases

Ira M. Longini, Jr.
Department of Biostatistics
Rollins School of Public Health
Emory University
Atlanta, GA

Michael G. Hudgens
Fred Hutchinson Cancer Research Center
Seattle, WA

January 1, 2003

CONTENTS

To see a World in a Grain of Sand
And a Heaven in a Wild Flower,
Hold Infinity in the palm of your hand
And Eternity in a hour. .....

W. Blake, Auguries of Innocence, 1803

0.1 Preface

These notes have grown out of a one-semester course in applied stochastic processes that I have taught over the last twenty-four years. I taught the course for the first time at the Universidad del Valle in Cali, Colombia in 1977. At that time, I also was with the International Center Medical Research and Training in Cali were I worked on projects on tropical diseases. My formal education in stochastic processes was not adequate to deal with the statistics generated by these infectious disease problems. The application of the ideas of stochastic processes to infectious disease problems mostly involved modeling exercises rather than statistical inference from data. At that time, the only book on the mathematics of epidemics that had a good stochastic processes basis was *The Mathematical Theory of Infectious Diseases* by Norman Bailey [1]. For me, this book provided a foundation for the analytic study of infectious disease problems. My interpretation of Bailey's basic approach to solving the problems associated with the analysis of infectious disease statistics is i) frame the problem mathematically, ii) carry out a qualitative analysis of the deterministic equations, iii) carry out the qualitative analysis of the stochastic equations, iv) carry out inference with the stochastic equations from field data, if available. Usually, the deterministic formulation of the process, as either differential or difference equations, is more tractable than the stochastic formulation. Thus, an analysis of the deterministic equations can lead to basic insights about the dynamics of the mean of the process. For linear processes, the solution to the deterministic process is exactly the mean of the analogous stochastic process. For nonlinear processes, the solution to the deterministic process may serve as an approximation. In some cases, the investigator can start with step iii. Generally, the stochastic equations are needed in order formulate likelihood functions or other estimating functions used in inference. These notes are partially based on the sound approach of applying steps i - iv above.

More recent books, *Analysis of Infectious Diseases* [3] and *Stochastic Epidemic Models and Their Statistical Analysis*[**?**], make use of stochastic processes to analyze infectious disease data. However, neither book can serve as a general reference for stochastic processes. These lecture notes are intended to fill this gap. Infectious disease problems offer an excellent paradigm on which to teach applied stochastic process in biostatistics. A susceptible individual makes the potential transition to infected through his or her interaction with infected individuals. If infection occurs, this individual then makes further transitions to other infected states with the possibility

of eventual recovery. These transitions, for both susceptible and infected individuals, evolve over time and space. Such a complex probabilistic process is best described through the use of stochastic processes. Through such illustrations, we present classical material on a host of stochastic processes including branching processes, Markov processes, birth and death processes and martingales. We develop material for a course in applied stochastic process in a uniform, logical manner moving from discrete to continuous processes, but we anchor the material with illustrations mostly from infectious disease problems.

The flow and material for the classical stochastic processes backbone of these lecture notes is similar to that in Chiang[4] and Karlin and Taylor [**?**]. Other important stochastic processes background texts are Bailey[**?**], Bhat[**?**] and Ross[**?**]. There are few books about inference on stochastic processes [**?**][**?**], and Basawa and Rao[**?**] is a good reference in this area.

Students taking this course should have a basic grounding in probability theory and mathematical statistics. In addition, some basic knowledge in real analysis and differential equations is helpful, but not necessary. A willingness to develop intuition about the nature of dynamic systems is important and curiosity about the natural world is essential.

Ira M. Longini, Jr.
Atlanta
January 1, 2003

Chapter 1
INTRODUCTION

We give the formal definition of a stochastic process as follows:

Definition 1.0.1 *We define a stochastic process as a collection of random variables* $\{X(t), t \in T\}$*, defined on a state space (set)* S*, where each random variable is indexed by a parameter (index parameter)* t *which varies in an index set* T*.*

Thus, the random variable $X(t)$ has a range that we will refer to as the state space given by

Definition 1.0.2 *We define the state space as the set of possible values* $X(t)$ *can take on.*

Some examples of state spaces are as follows:

Example 1.0.3 $S = \{0, 1, 2, \ldots\}$ *integer valued on a discrete state space.*

Example 1.0.4 $S = \{x : -\infty < x < \infty\}$ *continuous state space.*

Example 1.0.5 $S = \{\mathbf{x} : \mathbf{x} \in \mathcal{R}^k\}$ k *- vector process*

The domain, t, of random variable $X(t)$ is the index parameter.

Definition 1.0.6 *We define the index parameter* t *and the index set* T*.*

Example 1.0.7 $T = \{0, \pm 1, \pm 2, \ldots\}$ *discrete parameter process.*

Example 1.0.8 $T = \{t : -\infty < t < \infty\}$ *continuous parameter process.*

Based on the state space and the index set, the stochastic process $\{X(t), t \in T\}$ can be classified into four possible categories as shown in table 1.

Table 1: Categories of stochastic processes

		Index Set T	
		Discrete	Continuous
State Space	Discrete	I	II
S	Continuous	III	IV

Example 1.0.9 *An example of a category I process is the Galton-Watson branching process. Let Z_n be the random variable for the population size in the n^{th} generation. Then $S = \{0, 1, 2, \ldots\}$ and $T = \{0, 1, 2, \ldots\}$.*

Example 1.0.10 *An example of a category II process is a linear birth and death process. Let $X(t)$ be the random variable for the population at time t. Then $S = \{0, 1, 2, \ldots\}$ and $T = \{t : 0 \leq t < \infty\}$.*

Example 1.0.11 *An example of a category IV process is diffusion process (Brownian motion). Imagine a particle moving randomly in one dimensional. Let $X(t)$ be the random variable for the position of the particle on the real line at time t. Then $S = \{x : -\infty < x < \infty\}$ and $T = \{t : 0 \leq t < \infty\}$.*

Examples of category III processes are relatively rare in biostatistical applications, but we will briefly encounter one in chapter 9 on Gibbs sampling.

Chapter 2
PRELIMINARIES

2.1 Probability Generating Function

The probability generating function (pgf) is one of the major analytical tools we will use to work with stochastic processes on discrete state spaces.

Definition 2.1.1 *Let X be a nonnegative integer-valued random variable such that $P[X = k] = p_k$, $k = 0, 1, 2, \ldots$ is a probability mass function (pmf). Then the probability generating function is given by*

$$g_X(s) = E\left(s^X\right) = \sum_{k=0}^{\infty} s^k p_k. \tag{2.1}$$

We note the following properties of the pgf

$$g_X(1) = \sum_{k=0}^{\infty} p_k = 1$$

$$g_X^{'}(s) = \sum_{k=1}^{\infty} k s^{k-1} p_k, \; g_X^{'}(0) = p_1$$

$$g_X^{''}(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} p_k, \; g_X^{''}(0) = 2p_2$$

In general, $g_X^r(s) = \sum_{k=r}^{\infty} k(k-1)(k-2)\cdots(k-r+1) s^{k-r} p_k$, $g_X^r(0) = r! p_r$, so that $p_r = \frac{g_X^r(0)}{r!}$ and we see how the pgf literally "generates" probabilities. Therefore the pmf and pgf have one-to-one correspondence. Thus, if we are able to find the pgf of a stochastic process, then, at least theoretically, we can recover the pmf of the process. This is important since in some cases, it is difficult or impossible to derive the pmf of a process directly, but it is possible to derive its pgf.

The moments are also easily attainable from the pgf

$$g_X^{'}(1) = \sum_{k=1}^{\infty} k p_k = E(X)$$

$$g_X''(1) = \sum_{k=2}^{\infty} k(k-1)p_k = E(X(X-1))$$

implying that $Var(X) = g_X''(1) + g_X'(1) - \left[g_X'(1)\right]^2$. Similarly, for the r^{th} factorial moment about the origin we have

$$E[X(X-1)\cdots(X-r+1)] = g_X^r(1).$$

Thus, we can recover all the moments about the origin from the pgf.

Example 2.1.2 *Let X be a Bernoulli random variable such that the pmf of X is given by $p_1 = p$ and $p_0 = 1-p$. Then the pgf of X is given by*

$$g_X(s) = \sum_{k=0}^{\infty} s^k p_k = 1 - p + ps$$

Example 2.1.3 *Let X have a Poisson distribution such that*

$$P[X = k] = \frac{e^{-\lambda}\lambda^k}{k!},\ k = 0, 1, 2, \ldots$$

Then the pgf of X is given by

$$g_X(s) = \sum_{k=0}^{\infty} s^k p_k = e^{-\lambda(1-s)}$$

2.2 Convolution

Convolutions are distributions of sums of random variables. Such distributions can be difficult to find directly, but the pgf of such a sum is just the product of the pgf's involved. This is generally a relatively simple operation.

Let X, Y be independent random variables such that if $P[X = i] = p_i$ and $P[Y = j] = q_j$ then $P[X = i, Y = j] = p_{ij} = p_i q_j$. Let $Z = X + Y$. Then $P[Z = k] = r_k = \sum_{i=0}^{k} p_i q_{k-i} = \sum_{j=0}^{k} p_{k-j} q_j$. We write a convolution as $\{r_k\} = \{p_k\} * \{q_k\}$. Note that

$$\begin{aligned} g_Z(s) &= \textstyle\sum_{k=0}^{\infty} s^k r_k = \sum_{k=0}^{\infty} s^k \sum_{i=0}^{k} p_i q_{k-i} \\ &= \textstyle\sum_{i=0}^{\infty} \sum_{k=i}^{\infty} s^k p_i q_{k-i} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s^{i+j} p_i q_j \\ &= \textstyle\sum_{i=0}^{\infty} s^i p_i \sum_{j=0}^{\infty} s^j q_j \end{aligned}$$

Therefore,

$$g_Z(s) = g_X(s) \cdot g_Y(s).$$

In general, suppose $X_1, \ldots, X_n$ are mutually independent random variables and $Z_n = X_1 + \cdots + X_n$. Let $p_{k_j} = P[X_j = k]$ and $r_k = P[Z_n = k]$. Then Z_n has probability distribution given by $\{r_k\} = \{p_{k_1}\} * \cdots * \{p_{k_n}\}$and pgf $g_{z_n}(s) = g_{x_1}(s) \cdots g_{x_n}(s)$. Furthermore, if $X_1, \ldots, X_n$ are i.i.d. with pgf $g(s)$, then $g_{Z_n}(s) = [g(s)]^n$.

Example 2.2.1 *Let Z_n be a binomial random variable which is the sum of n Bernoulli random variables. Then the pgf of Z_n is given by*

$$g_{Z_n}(s) = [g(s)]^n = [1 - p + ps]^n$$

2.2.1 Compound distributions

Now suppose $Z_N = X_1 + \cdots + X_N$ where $X_1, ..., X_N$ are i.i.d. with pgf $g(s)$ and N is a random variable with pgf $h(s) = E\left(s^N\right) = \sum_{k=0}^{\infty} s^k P[N = k]$. Thus, Z_N is a random sum of i.i.d. random variables. Then the pgf of Z_N is given by

$$\begin{aligned} G(s) &= E\left(s^{Z_N}\right) = \textstyle\sum_{k=0}^{\infty} s^k P[Z_N = k] \\ &= \textstyle\sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} P[Z_N = k | N = j] \cdot P[N = j] \\ &= \textstyle\sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} P[X_1 + \cdots + X_j = k] \cdot P[N = j] \\ &= \textstyle\sum_{j=0}^{\infty} P[N = j] \sum_{k=0}^{\infty} P[X_1 + \cdots + X_j = k] \cdot s^k \\ &= \textstyle\sum_{j=0}^{\infty} P[N = j] \cdot (g(s))^j = h(g(s)) \end{aligned}$$

2.3 Exercises

Exercise 2.3.1 *Use the equation $G(s) = h(g(s))$ to show that $E[Z_N] = E[N] \cdot E[X]$ and $Var(Z_N) = E(N) \cdot Var(X) + Var(N) \cdot E[X]^2$.*

Exercise 2.3.2 *Consider the following experiment: A cell is given a single dose of pathogenic organisms, where the number of organisms in a dose is a random variable N which follows a Poisson distribution with parameter λ. Given that a dose has been administered, each organism survives with probability p. In addition, the probability that any particular organism survives is independent of the fate of the other $N - 1$ organisms in the dose. Let Y_N be a random variable for the number of surviving organisms in a dose.*

a. Find the pmf of Y_N.

b. If the cell dies when more than r organisms survive from a single dose, what is the probability that the cell dies when infected by a single dose?

c. Suppose n cells are given a single dose each. What is the probability that k of the cells survive (where $k \leq n$)?

Chapter 3
GALTON-WATSON (GW) BRANCHING PROCESS

The theory of branching processes goes back to the middle of the nineteenth century. This theory has been applied to problems in genetics, evolution, physics and epidemic theory just to name a few areas. In 1869, Francis Galton posed "the problem of the extinction of families"[**?**]. He presented this challenge [**?**] to his acquaintance, the Reverend H.W. Watson, who attacked the problem with generating functions and functional iteration[**?**]. Unfortunately, the good reverend got the wrong answer. Their joint efforts lead to more difficulties[**?**], but they helped lay the groundwork for the correct solutions as we see in the next section. A good history and rigorous modern treatment of the branching process can be found in the book by Guttorp[**?**]. An important earlier book on the topic is by Harris[**?**].

3.1 Probability distribution, Expectations, Extinction

Let Z_n be the population size in generation n and X be the number of offspring for an individual with pmf $P[X=k]=p_k$, $k=0,1,2,\ldots$, $E[X]=\theta$, and $Var[X]=\sigma^2<\infty$. Let X_j be the random variable for the number of offspring for the j^{th} individual such that $Z_{n+1}=\sum_{j=1}^{Z_n}X_j$. Let the pgf of X be given by $g(s)=\sum_{k=0}^{\infty}s^k p_k$ and let $g_n(s)=\sum_{k=0}^{\infty}P[Z_n=k]s^k$ be the pgf of Z_n for $n=1,2,3\ldots$ It follows that

$$g_{n+1}(s)=g_n(g(s))=g(g_n(s))$$

the latter equality holding only when $Z_0=1$. (Note that by $Z_0=1$ we mean $P[Z_0=1]=1$). We see that $g_0(s)=s$ when $Z_0=1$, $g_0(s)=s^2$ when $Z_0=2$, and in general, $g_0(s)=s^{i_0}$ when $Z_0=i_0$. The expected value of the Z_n is given by

$$E[Z_n]=g_n^{'}(1)=\left[g^{'}(1)\right]^n=\theta^n$$

since $g_n^{'}(1)=g^{'}(1)\cdot g_{n-1}^{'}(g(1))=g^{'}(1)\cdot g_{n-1}^{'}(1)$ implies that $g_n^{'}(1)=\cdots=\left[g^{'}(1)\right]^n$. We could arrive at the same solution via conditional expectation by noting that

$E[Z_{n+1} \mid Z_n] = E[X_1 + \cdots + X_{Z_n} \mid Z_n] = \theta \cdot Z_n$. It follows that

$$\begin{aligned} E[Z_n] &= E[E[Z_n \mid Z_{n-1}]] \\ &= E[\theta \cdot Z_{n-1}] = \theta \cdot E[Z_{n-1}] \\ &= \theta \cdot E[E[Z_{n-1} \mid Z_{n-2}]] = \theta^2 \cdot E[Z_{n-2}] = \cdots = \theta^n. \end{aligned}$$

We note the underlying deterministic system given by $z_{n+1} = \theta \cdot z_n$, $z_0 = 1$ has the solution $z_n = \theta^n$. However as n tends to ∞, the deterministic and stochastic processes have very different behavior. The limiting value of $E[Z_n]$ is given by:

$$\lim_{n\to\infty} E[Z_n] = \begin{cases} 0 & \text{if } \theta < 1 \text{ (subcritical)} \\ 1 & \text{if } \theta = 1 \text{ (critical)} \\ \infty & \text{if } \theta > 1 \text{ (supercritical)} \end{cases}.$$

It can also be shown that $Var(Z_{n+1}|Z_n) = \sigma^2 Z_n$, from which it follows that

$$Var[Z_n] = \begin{cases} \sigma^2\theta^{n-1}\left(\frac{1-\theta^n}{1-\theta}\right) & \text{if } \theta \neq 1 \\ n\sigma^2 & \text{if } \theta = 1 \end{cases}.$$

Thus when $\theta = 1$, the mean remains constant while the variance tends to infinity, assuring extinction.

For $0 < p_0 < 1$, let $q_n = P[Z_n = 0] = g_n(0)$ be the probability of extinction by the n^{th} generation. Note $g_{n+1}(0) = g(g_n(0))$ implies that $q_{n+1} = g(q_n)$. Since $g(s)$ is a strictly monotone increasing function for $s \in (0,1)$, it follows that

$$q_1 = p_0 < g(q_1) = q_2 < g(q_2) = q_3 < \cdots < q_n < q_{n+1} < \cdots$$

Thus $\{q_n\}$ is a monotonically increasing sequence bounded above by 1, implying the existence of

$$\pi = \lim_{n\to\infty} q_n$$

for $0 < \pi \leq 1$. If $p_0 + p_1 < 1$, then $g''(s) = \sum_{k=0}^{\infty} k(k-1)s^{k-2}p_k > 0$. Thus the equation $s = g(s)$ has one or two solutions for $s \in (0,1]$. (Insert graphs here.) We leave it as an exercise to show that the probability of extinction, π, is the smallest positive root of $s = g(s)$. Here we refer to graphs to show that $\pi = 1$ when $g'(1) = \theta \leq 1$ and $\pi < 1$ when $g'(1) = \theta > 1$. To summarize the behavior of this process: if $\theta \leq 1$, then eventual extinction is certain, but if $\theta > 1$ then population goes extinct with probability with probability π which is the smallest positive root of $s = g(s)$. The population does not go extinct with probability $1 - \pi$. In this case, the population size will go to infinity. This very fundamental result can be thought of as a type of folk theorem for population processes. The counter intuitive message is that finite populations may go extinct even if $\theta > 1$.

It is instructive to compare this behavior with the corresponding deterministic system. As we mentioned earlier, the path of the corresponding deterministic system is the same as the mean of the stochastic system, *i.e.*, $z_n = \theta^n$. The behavior of the deterministic system is as follows: if $\theta < 1$, then extinction occurs, but if $\theta \geq 1$ then the population does not go extinct. This simpler behavior is not realistic for finite populations. Thus, the deterministic model is not realistic unless the population size becomes very large. At this point, it provides a good approximation to reality. This is a very common characteristic of deterministic formulations.

Thus far we have assumed that $Z_0 = 1$. Suppose now that instead $Z = i_0$ where $i_0 > 1$. It follows that $g_0(s) = s^{i_0}$ and that $E[Z_n] = i_0\theta^n$. In this case, if $\theta \leq 1$ then extinction occurs with probability 1, while if $\theta > 1$ then the probability of extinction is π^{i_0}.

Example 3.1.1 ***Survival and Family Names (male lines)****: Let p_k be the probability that a newborn boy becomes the progenitor of k other boys. What is the probability that the family name goes extinct? Lotka uses the following estimated offspring distribution to solve the problem: $p_0 = 0.4825$, $p_k = (0.2126)(0.5893)^{k-1}, k \geq 1$.*

Example 3.1.2 ***Genes and Mutation****: Let the number of descendants of a mutant gene K have a Poisson distribution with parameter λ. Then $g(s) = e^{\lambda(s-1)}$ and we can easily calculate the extinction probability. For example, if $\lambda = 2$, then $\pi = 0.203$.*

Example 3.1.3 $S \to I \to R$ ***Epidemic (Susceptible, Infected, Removed):*** *Let Z_n be the number of people infected in generation n, and θ be the average number of people that an infected person infects during his or her infectious period in a fully susceptible population. (θ is called the basic reproductive number). Let X_n be the number of susceptible people in generation n, so that $X_{n+1} = X_n - Z_n$. Then one version of the celebrated threshold theorem of epidemics is as follows: if $\theta \leq 1$, there is no epidemic with probability one. If $\theta > 1$, then there is no epidemic with probability π^{i_0} and an epidemic with probability $1 - \pi^{i_0}$, where i_0 is the number of initial infected people.*

3.2 Inference on the GW branching process

3.2.1 The explosion set

In this section we describe methods for estimating the offspring distribution and its moments. Generally such estimation will be done from an observed realization $\{Z_0, Z_1, \cdots, Z_n\}$. Technically, such estimators are consistent only if the observed realization is part of the explosion set. The explosion set is that set of sequences for $\theta > 1$ that do not go extinct. In the following sections, we will assume that we are doing inference on the explosion set.

3.2.2 Time Series Methods

Our main goal is to do inference on the mean and variance of the offspring distribution θ and σ^2. Based on the conditional expectation of the GW branching process, $E[Z_{n+1} \mid Z_n] = \theta \cdot Z_n$, we consider the first order autoregressive process

$$Z_n = \theta Z_{n-1} + Y_n$$

where $\{Y_n\}_{n=1}^{\infty}$ is a sequence of uncorrelated random variables such that $E(Y_i) = 0$ and $Var(Y_i) = \sigma^2$. We let

$$SSE = \sum_{k=1}^{n} (Z_k - \theta Z_{k-1})^2 = \sum_{k=1}^{n} Y_k^2$$

so that

$$\frac{\partial SSE}{\partial \theta} = -\sum_{k=1}^{n} 2(Z_k - \theta Z_{k-1}) Z_{k-1}$$

Setting to 0, we get the following estimate for θ :

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} (Z_k \cdot Z_{k-1})}{\sum_{k=1}^{n} (Z_{k-1})^2}$$

which is serial the correlation of lag one.

Recall that $E(Z_n \mid Z_{n-1}) = \theta Z_{n-1}$ and $Var(Z_n \mid Z_{n-1}) = \sigma^2 Z_{n-1}$, suggesting that we might consider the autoregressive type model:

$$Z_n = \theta Z_{n-1} + U_n \sqrt{Z_{n-1}}.$$

It follows that

$$U_n = \left(\frac{Z_n - \theta Z_{n-1}}{\sigma \sqrt{Z_{n-1}}} \right) \sigma = \left(\frac{Z_n - E(Z_n | Z_{n-1})}{\sqrt{Var(Z_n | Z_{n-1})}} \right) \sigma \quad .$$

Again we can minimize the error sum of squares to get an estimate of θ :

$$SSE = \sum_{k=1}^{n} U_k^2 = \sum_{k=1}^{n} \frac{(Z_k - \theta Z_{k-1})^2}{Z_{k-1}} \tag{3.1}$$

$$\frac{\partial SSE}{\partial \theta} = -\sum_{k=1}^{n} 2(Z_k - \theta Z_{k-1})$$

Setting to 0, we get:

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}} \tag{3.2}$$

However we need maximum likelihood to get $Var\left(\hat{\theta}_n\right)$. Based on (**??**), an estimator of the variance of the offspring distribution is

$$\widehat{\sigma}_n^2 = \frac{SSE}{n} = \frac{1}{n} \sum_{k=1}^{n} \frac{(Z_k - \theta Z_{k-1})^2}{Z_{k-1}}$$

3.2.3 Likelihood-based methods

A clever idea by Harris[**?**] is as follows. Assume $Z_0 = 1$. Observe $Z_1, Z_2, \ldots$ and let $Z_{nr} = \#$ of individuals in the n^{th} generation who produce r offspring. Then $Z_n = \sum_{r=0}^{\infty} Z_{nr}$ and $Z_{n+1} = \sum_{r=0}^{\infty} r Z_{nr}$. The joint conditional density is an infinite multinomial:

$$P\left[Z_{n0}, Z_{n1}, \ldots \mid Z_n\right] = \left(\frac{Z_n!}{\prod_{r=0}^{\infty} Z_{nr}!}\right) \prod_{r=0}^{\infty} p_r^{Z_{nr}}$$

which gives rise to the following likelihood function for n generations:

$$L(p_0, p_1, \ldots) = \prod_{k=0}^{n-1} \left[\left(\frac{Z_k!}{\prod_{r=0}^{\infty} Z_{kr}!}\right) \prod_{r=0}^{\infty} p_r^{Z_{kr}} \right] = c \prod_{k=0}^{n-1} \prod_{r=0}^{\infty} p_r^{Z_{kr}}$$

where c is some constant. The MLE is given by

$$\hat{p}_r = \frac{\sum_{k=0}^{n-1} Z_{kr}}{\sum_{k=0}^{n-1} Z_k}.$$

And since

$$\theta = \sum_{r=0}^{\infty} r p_r,$$

it follows that the MLE of θ is

$$\begin{aligned} \hat{\theta}_n &= \textstyle\sum_{r=0}^{\infty} r \hat{p}_r = \frac{\sum_{r=0}^{\infty} r \sum_{k=0}^{n-1} Z_{kr}}{\sum_{k=0}^{n-1} Z_k} \\ &= \frac{\sum_{k=0}^{n-1} \sum_{r=0}^{\infty} r Z_{kr}}{\sum_{k=0}^{n-1} Z_k} \\ &= \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}} \end{aligned}$$

which is the same answer as equation(3.2). However we still can not determine the variance and distribution of this estimator, so we make a weak assumption: namely that the offspring distribution follows a generalized power series distribution.

Definition 3.2.1 *A discrete random variable X has a generalized power series distribution (GPSD) if:*

$$p_x = P\left[X = x\right] = \frac{a_x \lambda^x}{f\left(\lambda\right)} \text{ for } x \in T$$

where

$$f\left(\lambda\right) = \sum_{x \in T} a_x \lambda^x, \ \lambda > 0, \ a_x \geq 0.$$

Example 3.2.2 *Let* $a_x = 1/x!$, $T = \{0, 1, 2, \ldots\}$, $f(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$. *Then* $p_x = \frac{\lambda^x}{x!} e^{-\lambda}$, *which we recognize as the Poisson probability mass function.*

Example 3.2.3 *Let* $T = \{0, 1, 2, \ldots, n\}$, $a_x = \binom{n}{x}$, *which gives binomial where* $\lambda = \frac{p}{1-p}$.

Example 3.2.4 *For some positive integer* c, *let* $T = \{c, c+1, 2, \ldots, n\}$, $a_x = \binom{n}{x}$, *which gives truncated binomial.*

If X has a GPSD, then:

$$\theta = E[X] = \frac{\lambda f'(\lambda)}{f(\lambda)}$$

since

$$f'(\lambda) = \frac{1}{\lambda} \sum_{x \in T} a_x x \lambda^x$$

It is also easy to show that $Var[X] = \sigma^2 = \lambda \frac{d\theta}{d\lambda}$.

Assuming X has a GPSD, the likelihood for the branching process now becomes:

$$\begin{aligned} L_n(p_0, p_1, \ldots) &= c \prod_{k=0}^{n-1} \prod_{r=0}^{\infty} p_r^{Z_{kr}} \\ &= c \prod_{k=0}^{n-1} \prod_{r=0}^{\infty} \left[\frac{a_r \lambda^r}{f(\lambda)}\right]^{Z_{kr}} \\ &= c_1 \prod_{k=0}^{n-1} \prod_{r=0}^{\infty} \frac{\lambda^{rZ_{kr}}}{f(\lambda)^{Z_{kr}}} \\ &= c_1 \frac{\lambda^{\sum_{k=0}^{n-1} \sum_{r=0}^{\infty} rZ_{kr}}}{f(\lambda)^{\sum_{k=0}^{n-1} \sum_{r=0}^{\infty} Z_{kr}}} \\ &= c_1 \frac{\lambda^{\sum_{k=1}^{n} Z_k}}{f(\lambda)^{\sum_{k=1}^{n} Z_{k-1}}}. \end{aligned}$$

Taking the natural logarithm:

$$l_n = c_2 + \sum_{k=1}^{n} [Z_k \ln \lambda - Z_{k-1} \ln f(\lambda)] \tag{3.3}$$

thus

$$\begin{aligned} \frac{\partial l_n}{\partial \theta} &= \frac{\partial l_n}{\partial \lambda} \frac{\partial \lambda}{\partial \theta} \\ &= \frac{1}{\sigma^2} \sum_{k=1}^{n} \left[Z_k - \frac{Z_{k-1} \lambda f'(\lambda)}{f(\lambda)}\right] \\ &= \frac{1}{\sigma^2} \sum_{k=1}^{n} [Z_k - Z_{k-1}\theta] \end{aligned}$$

Setting to zero,

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}},$$

which is the same as equation (3.2). The observed information is given by:

$$I_n(\theta) = -\frac{\partial^2 \ln L_n}{\partial \theta^2} = \frac{\sum_{k=1}^{n} Z_{k-1}}{\sigma^2}$$

implying that asymptotically:

$$Var\left(\hat{\theta}_n\right) = \frac{\hat{\sigma}^2}{\sum_{k=1}^{n} Z_{k-1}}$$

where $\hat{\sigma}^2$ is obtained from time series i.e.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=0}^{n}\left(Z_k - \hat{\theta} Z_{k-1}\right)^2.$$

The expected or Fisher information is given by:

$$\mathcal{I}_n(\theta) = \frac{E\left[\sum_{k=1}^{n} Z_{k-1}\right]}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=0}^{n-1}\theta^k = \frac{1}{\sigma^2}\left(\frac{1-\theta^n}{1-\theta}\right)$$

implying

$$Var\left(\hat{\theta}_n\right) \approx \hat{\sigma}^2\left(\frac{1-\hat{\theta}}{1-\hat{\theta}^n}\right).$$

If we further assume that the offspring distribution $\sim$ Poisson, then

$$Var\left(\hat{\theta}_n\right) \approx \hat{\theta}\left(\frac{1-\hat{\theta}}{1-\hat{\theta}^n}\right).$$

3.3 Epidemics as GW branching process

This example is taken from Becker[**?**]. Let Z_n be the number of infected people in generation n, with $Z_0 = i_0$. Let θ be the basic reproductive number. That is, θ is the average number of individuals that an infective could infect in a fully susceptible population. The threshold theorem (see *example 3.1.3*) is as follows: For $\theta \leq 1$ there is no epidemic, and for $\theta > 1$ there is an epidemic with probability $1 - \pi^{i_0}$ where π is the smallest root of $g(s) = s$. Suppose we want to estimate θ from the sequence $Z_1, Z_2, Z_3 \ldots$. Since in finite populations, the number of susceptibles tends to decrease as the epidemic evolves, we must modify the Galton-Watson process so that the offspring distribution changes with time accordingly.

Let N be the original number of susceptibles and $Y_n = \sum_{i=1}^{n} Z_i$ be the cumulative number of infections through the n^{th} generation. Further let $\mathbf{Z}_n = (Z_1, Z_2, \ldots, Z_n)$ be the history of the number of infections in each generation up

to the n^{th} generation. We will assume that an infected person is infected for only one generation and is subsequently removed from the pool of susceptibles. Then in the first generation we let mean of the offspring distribution be

$$E_1[X] = \theta,$$

and for the n^{th} generation, $n > 1$,

$$E_n[X] = \theta \cdot g_n(\mathbf{Z}_{n-1}) = \mu_n$$

where $g_n(\cdot)$ can be any monotonically decreasing function in n and μ_n is the reproductive number for the n^{th} generation. That is, μ_n is the average number of people a person in generation n infects over his or her infectious period. An intuitive choice for $g_n(\mathbf{Z}_{n-1})$ is the fraction of the population not infected at time $n-1$:

$$g_n(\mathbf{Z}_{n-1}) = \max\left[1 - \frac{Y_{n-1}}{N}, 0\right].$$

Similarly let $Var(Z_n) = \sigma_n^2 = \sigma^2 \cdot h_n(\mathbf{Z}_{n-1})$ where $\sigma^2 = Var(X)$. One choice for h_n is $h_n \equiv g_n$, a Poisson like assumption. Assume X_i has a GPSD. Then we have

$$p_n(X_i) = \frac{a_n(x) \cdot [c_n(\lambda)]^x}{f_n[c_n(\lambda)]} = \frac{a_n c_n^x}{f_n}$$

the latter equality being notational convention. Now

$$f_n = \sum_x a_n c_n^x, \text{ and } f_n' = \frac{df_n}{d\lambda} = \frac{c_n'}{c_n}\sum_x a_n x c_n^x$$

so that

$$\mu_n = \frac{c_n}{c_n'} \cdot \frac{f_n'}{f_n} = \theta \cdot g_n(\mathbf{Z}_{n-1})$$
$$\sigma_n^2 = \frac{c_n}{c_n'} \cdot \mu_n' = \sigma^2 h_n(\mathbf{Z}_{n-1})$$

Making the approximation

$$P[\mathbf{Z}_n|\mathbf{Z}_{n-1}] \approx \frac{(c_n)^{Z_n}}{(f_n)^{Z_{n-1}}}$$

we approximate the log likelihood by

$$\ln L_n \approx \sum_{k=1}^{n} \{Z_k \ln(c_k) - Z_{k-1} \ln(f_k)\}.$$

Taking the derivative with respect to θ and setting equal to zero, one can show that

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} (Z_k g_k / h_k)}{\sum_{k=1}^{n} (Z_{k-1} g_k^2 / h_k)}.$$

Further it can be shown that the observed information is given by:

$$I_n(\theta) = \frac{1}{\hat{\theta}_n} \sum_{k=1}^{n} \frac{Z_k g_k}{\sigma_k^2}$$

If we make the Poisson assumption that $h_k \equiv g_k$, then

$$\hat{\theta}_n = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1} g_k} = \frac{Y_n}{\sum_{k=1}^{n} Z_{k-1} g_k} \tag{3.4}$$

and

$$I_n(\theta) = \frac{Y_n}{\hat{\theta}_n \sigma^2}$$

Note that 3.4 reduces to usual estimator for θ 3.2 when $g_k = 1$ for all k. Thus, the g_k are weights that adjust for the change in susceptibles as the epidemic progresses.

Example 3.3.1 *An outbreak of smallpox in a closed, unvaccinated community in Abakalbiki, Nigeria, 1967. The incubation period for smallpox is 9-15 days (average of 12). The data is clustered into 12 day intervals as follows:.*

generation k	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
# cases (Z_k)	*1*	*1*	*7*	*6*	*3*	*8*	*4*	*0*

The overall attack rate is $\frac{29}{119} = 0.24$. *There were* $N = 119$ *people in the community and by the last generation* $Y_7 = 29$. *It can be shown that* $\hat{\theta}_Z = 1.14$ *with* $Var\left(\hat{\theta}_Z\right) = \frac{(1.14)^2}{29} = 0.045$. *The corresponding 95% confidence interval is given by:* $\hat{\theta} = 1.14 \pm 0.41$. *The estimate of* π *is* $\hat{\pi} = 0.76$. *Thus, smallpox is not very infectious with a basic reproductive number barely above one.*

More examples of branching process-like estimation of θ can be found in Saunders[26], Longini[14] and Becker[3].

3.4 Exercises

Exercise 3.4.1 *In a Galton-Watson branching process, the number of offspring per individual has a binomial distribution with parameters 2, p. Starting with a single individual, calculate:*

a. the extinction probability

b. the probability that the population becomes extinct for the first time in the third generation

Exercise 3.4.2 *For the Galton-Watson branching process, prove that* $Var[Z_{n+1}] = \sigma^2\theta^n \sum_{j=0}^{n} \theta^j$ *using identity (5.34) in Theorem 7 on page 19 of Chiang's book. (Hint: use induction).*

Exercise 3.4.3 *A stochastic process* $\{W_n\}_{n=0}^{\infty}$ *is said to be a Martingale process (with respect to itself) if* $E[|W_i|] < \infty$ *for all* n *and if* $E[W_{n+1}|W_0, W_1, \ldots, W_n] = W_n$.

a. Let $X_1, X_2, \ldots$ *be independent random variables with 0 mean and let* $W_n = \sum_{i=1}^{n} X_i$. *Assume* $E[|X_i|] < \infty$. *Then prove* $\{W_n\}_{n=1}^{\infty}$ *is a Martingale.*

b. Let $X_1, X_2, \ldots$ *be independent random variables with* $E[X_i] = 1$ *and let* $W_n = \Pi_{i=1}^{n} X_i$. *Then prove* $\{W_n\}_{n=1}^{\infty}$ *is a Martingale.*

c. Let $\{Z_n\}_{n=0}^{\infty}$ *be generated by the Galton-Watson branching process and* $W_n = \frac{Z_n}{\theta^n}$, $n = 0, 1, 2, \ldots$ *Then prove* $\{W_n\}_{n=1}^{\infty}$ *is a Martingale.*

Exercise 3.4.4 *For the Galton-Watson branching process, prove that the probability of extinction,* π, *is the smallest positive root of* $s = g(s)$.

Exercise 3.4.5 *Let* X *have a generalized power series distribution.*

a. Find the probability generating function of X *in terms of* $f(\cdot)$.

b. If $T = \{0, 1, 2, \ldots\}$, $a_x = 1$ *for all* $x \in T$, *and* $0 < \lambda < 1$, *then what is the pmf of X?*

c. Given the random sample $X_1, X_2, \ldots, X_n$, *what is the MLE of* λ*?*

d. Give the approximate variance of $\hat{\lambda}$ *from part c.*

Exercise 3.4.6 *Consider a subcritical Galton-Watson branching process that we wish to study when* n *is large. Since we know that the population will be extinct for large* n, *one approach is to study the process conditioned on non-extinction. Let* $G_n(s)$ *be the pgf of* Z_n *conditioned on the event that the population is not extinct at generation* n. *Find the expression for* $G_n(s)$ *in terms of* $g_n(s)$ *and* $g_n(0)$, *where* $g_n(s)$ *is the unconditional pgf of* Z_n.

Chapter 4
RANDOM WALKS

4.1 Simple Random Walks

Let $T = (0, 1, 2, \ldots)$, $S = (\ldots -2, -1, 0, 1, 2, \ldots)$ and Z_n = position of particle after n jumps. Define $Z_n = X_1 + \cdots + X_n$ where

$$P[X_i = k] = \begin{cases} p \text{ if } k = 1 \\ q = 1 - p \text{ if } k = -1 \\ 0 \text{ otherwise} \end{cases} .$$

Assume $Z_0 = 0$. Then $E[X_i] = p - q$, $Var[X_i] = 4pq$,and $g_i(s) = \sum_{k=-\infty}^{\infty} s^k p_k = ps + \frac{q}{s}$. (Note that in our original definition of the pgf, k was restricted to the non-negative integers whereas here we use the more general form which allows k to be any integer. Some of the properties derived earlier for the pgf do not necessarily continue to hold.) It follows that

$$\begin{aligned} G_{Z_n}(s) = (ps + qs^{-1})^n &= \sum_{i=0}^{n} \binom{n}{i} (ps)^i (qs^{-1})^{n-i} \\ &= \sum_{i=0}^{n} \binom{n}{i} p^i q^{n-i} s^{2i-n} \\ &= \sum_{k} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} s^k, \end{aligned}$$

where the last sum is taken over $k = -n, -n+2, -n+4, \ldots, n-2, n$. Therefore,

$$P[Z_n = k] = \begin{cases} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}} & \text{for } k = -n, -n+2, -n+4, \ldots, n-2, n \\ 0 & \text{otherwise} \end{cases} .$$

Note that for any simple random walk $P[Z_n = 0] = \binom{n}{\frac{n}{2}} (pq)^{\frac{n}{2}}$ when n is even and 0 when n is odd. Thus we can only return to the origin on an even numbered step.

Example 4.1.1 *A symmetric random walk is a special case of a simple random walk and is given by $p = q = \frac{1}{2}$. In which case, $P[Z_n = k] = \binom{n}{\frac{n+k}{2}} \left(\frac{1}{2}\right)^n$.*

We can also define a general random walk by letting:

$$P[X_i = k] = \begin{cases} p_i \text{ if } k = 1 \\ r_i \text{ if } k = 0 \\ q_i = 1 - p_i - r_i \text{ if } k = -1 \\ 0 \text{ otherwise} \end{cases} . \tag{4.1}$$

A simple random walk is then just a special case of a general random walk where $r_i = 0$ and $p_i = p$ for all i.

4.2 Difference Equations

First we review solving difference equations. In general, a difference equation is given by:

$$x_{n+k} + a_{k-1}x_{n+k-1} + \cdots + a_0 x_n = g_n$$

When $g_n = 0$ we have a set of homogeneous difference equations. Otherwise the equations are non-homogeneous. Define the characteristic polynomial as

$$c(\lambda) = \lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_0$$

To solve, we let $c(\lambda) = 0$. The k solutions (or roots), $\lambda_1, \lambda_2, \ldots, \lambda_k$, give rise to the solution

$$x_n = c_1\lambda_1^n + \cdots + c_k\lambda_k^n$$

where the c_i's are found from the initial conditions.

Example 4.2.1 *Example: Fibonnacci sequence 1, 1, 2, 3, 5, 8, 13, 21, ... is given by*

$$x_{n+2} - x_{n+1} - x_n = 0,\ x_1 = x_2 = 1,\ n = 1, 2, 3, \ldots$$

Let $x_n = \lambda^n$ *and solve the equation*

$$\lambda^2 - \lambda - 1 = 0$$

which yields two roots: $\lambda_1 = \frac{1+\sqrt{5}}{2} = 1.618$, $\lambda_2 = \frac{1-\sqrt{5}}{2} = -0.618$. *Next we solve the equations:*

$$\begin{aligned} c_1\lambda_1 + c_2\lambda_2 &= 1 \\ c_1\lambda_1^2 + c_2\lambda_2^2 &= 1 \end{aligned}$$

and arrive at the solution:

$$x_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

Example 4.2.2 *Let* $x_{n+2} + x_n = 0$, $x_0 = 1$, $x_1 = 0$, *i.e., -1, 0, 1, 0, -1, 0, 1, 0,... Then* $x_{n+2} = -x_n$ *gives* $\lambda^2 = -1$ *implying the characteristic roots are* $\lambda = \pm i$. *Using initial conditions, the closed form solution is*

$$x_n = \frac{1}{2}\left[i^n + (-i)^n\right].$$

4.3 Gambling Systems

4.3.1 Gambler's Ruin

We now define the scenario for gambler's ruin. Consider to gamblers, A and B, with respective initial fortunes a and b and probabilities of winning p and q such that $p+q=1$. We wish to find R_a, the probability of gambler A's ruin if gambler A has initial wealth a and the game is played until either A or B is broke. Let

$$R_x = p \cdot R_{x+1} + q \cdot R_{x-1},\ x = 1, 2, \ldots, a+b-1$$

where $R_0 = 1$, $R_{a+b} = 0$. The above gives rise to the following homogeneous difference equation:

$$R_{x+2} - \frac{1}{p} R_{x+1} + \frac{q}{p} R_x = 0$$

Thus we solve the characteristic polynomial $c(\lambda) = \lambda^2 - \frac{1}{p}\lambda + \frac{q}{p} = (\lambda - 1)\left(\lambda - \frac{q}{p}\right)$ which has two roots $\lambda_1 = 1$, $\lambda_2 = \frac{q}{p}$. If $\lambda_1 = \lambda_2$, the system is said to be indeterminate, so we assume $\lambda_1 \neq \lambda_2$ (i.e. $p \neq q$). In which case, $R_x = c_1 + c_2 \left(\frac{q}{p}\right)^x$ where c_1 and c_2 are determined by the boundary conditions: $c_1 + c_2 = 1$ and $c_1 + c_2 \left(\frac{q}{p}\right)^{a+b} = 0$. It can be shown that the probability of the gambler's ruin given starting at a, is

$$R_a = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^{a+b}}{1 - \left(\frac{q}{p}\right)^{a+b}} \text{ when } q \neq p$$

Via L'Hopital's rule, it can be shown that $R_a = \frac{b}{b+a}$ when $q = p = \frac{1}{2}$.

What if the opponent is infinitely wealthy?

$$\lim_{b\to\infty} R_a = \begin{cases} 1 & \text{if } p \leq q \\ \left(\frac{q}{p}\right)^a & \text{if } p > q \end{cases}$$

We let $W_a = 1 - R_a$ be the probability of winning given start at a.

Example 4.3.1 *Roulette – Vegas style wheel has 18 red, 18 black and 1 green for a total of 37 slots. The gambler bets on red or black, thus* $p = \frac{18}{37}$ *and* $q = \frac{19}{37}$. *Furthermore, suppose gambler has an initial wealth of \$100 (i.e.* $a = 100$) *and the house has \$1000 (i.e.* $b = 1000$) *(or, more realistically, the gambler will play until she is ruined, or wins \$1000) and on each turn the gambler wages \$1. Then* $1 - R_{100} = W_{100} = 3.29 \cdot 10^{-23}$ *is the probability of the gambler winning. An interesting caveat of this example is that a slight change in the rules (Monaco style), namely green delays a win or loss till the next spin, does not change the result above.*

4.3.2 Expected Duration

Let the expected duration of the game if gambler A has wealth x be given by

$$D_x = p \cdot D_{x+1} + q \cdot D_{x-1} + 1,\ D_0 = D_{a+b} = 0.$$

This gives rise to the following non-homogeneous difference equation

$$p \cdot D_{x+2} - D_{x+1} + q \cdot D_x = -1$$

or

$$D_{x+2} - \left(\frac{1}{p}\right) D_{x+1} + \left(\frac{q}{p}\right) D_x = -\frac{1}{p}.$$

The solution is of the form D_x = general(homogeneous part) + particular = $c_1 + c_2 \left(\frac{q}{p}\right)^x + f_x$. In general, f_x is difficult to ascertain. However, when the non-homogeneous part is a constant, in this case $-\frac{1}{p}$, $f_x = \alpha x$. From the following

$$p \cdot \alpha\,(x+2) - \alpha\,(x+1) + q \cdot \alpha\,(x) = -1$$

we arrive at $\alpha = \frac{1}{q-p}$ when $q \neq p$. Thus $f_x = \frac{x}{q-p}$, so that

$$D_x = c_1 + c_2 \left(\frac{q}{p}\right)^x + \frac{x}{q-p},\ q \neq p$$

where c_1 and c_2 are attained from the boundary conditions. Similarly we resolve when $p = q$.

$$D_a = \begin{cases} \frac{a}{q-p} - \frac{a+b}{q-p}\left[\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}}\right] & \text{when } q \neq p \\ \left(\frac{a+b}{2}\right)^2 & \text{when } q = p \end{cases}$$

Example 4.3.2 *Revisiting the roulette example, where $a + b = 1100$, we find that $D_{100} = 3700$.*

4.3.3 Discrete-time martingales

We motivate martingales by considering different betting strategies. Assume that there is some positive probability $p > 0$ of winning and let the outcome of each game be

$$X_i = \begin{cases} 1 & \text{if win i}^{th} \text{ game} \\ \text{-1} & \text{if lose i}^{th} \text{ game} \end{cases}.$$

Let the amount wagered on the i^{th} game be $W_i = W_i\left(\mathbf{X}_{i-1}\right)$ where $\mathbf{X}_{i-1} = (X_1, X_2, \ldots, X_{i-1})$ is the history of wins and losses. Also let Y_i be the fortune after the i^{th} game, *i.e.*,

$$Y_n = Y_0 + \sum_{k=1}^{n} W_k X_k.$$

It follows that

$$E[Y_n] = Y_0 + E\left(\sum_{k=1}^{n} W_k X_k\right).$$

Now we consider two cases based on the gambler's initial wealth.

Case 1. Suppose the gambler has infinite wealth initially, *i.e.,* $Y_0 = \infty$. Let $W = W_1$ be the initial wager, and henceforth wager the following:

$$W_k = \begin{cases} W \cdot 2^{k-1} & \text{if } X_i = -1 \text{ for } i = 1, 2, \ldots, k-1 \\ 0 & \text{otherwise} \end{cases}$$

and suppose that and that gambler wins for the first time on the n^{th} game. Then the gambler wins $W \cdot 2^{n-1}$ and accrues a loss (from the first $n-1$ games) of $W + W \cdot 2 + \cdots + W \cdot 2^{n-2} = W \sum_{i=0}^{n-2} 2^i = W\left(\frac{1-2^{n-1}}{1-2}\right) = W(2^{n-1} - 1)$ for a net winning of W.

Case 2. Now suppose $Y_0 = W(2^m - 1)$ and that the game is fair, *i.e.*, $p = q = \frac{1}{2}$. Now let the betting strategy be

$$W_k = \begin{cases} W \cdot 2^{k-1} & \text{if } X_i = -1 \text{ for } i = 1, 2, \ldots, k-1 \text{ and } k \leq m \\ 0 & \text{otherwise} \end{cases}.$$

In which case,

$$\begin{aligned} E[\text{winnings}] &= W \cdot P[\text{win at least one game}] - W(2^m - 1) \cdot P[\text{lose first } m \text{ games}] \\ &= W(1 - 2^{-m}) - W(2^m - 1)\, 2^{-m} = 0 \end{aligned}.$$

For either case, it is easy to see that $Y_{n+1} = Y_n + W_{n+1} X_{n+1}$. It follows that

$$\begin{aligned} E(Y_{n+1} \mid \mathbf{X}_n) &= E(Y_n \mid \mathbf{X}_n) + E(W_{n+1} X_{n+1} \mid \mathbf{X}_n) \\ &= Y_n + W_{n+1} E(X_{n+1} \mid \mathbf{X}_n) \\ &= Y_n + W_{n+1} E(X_{n+1}) \\ &= Y_n \end{aligned} \tag{4.2}$$

where the last equality holds when $p = q = \frac{1}{2}$. This is the principle driving force behind the martingale. In this case, it follows that

$$E(Y_{n+1}) = E(E(Y_{n+1} \mid \mathbf{X}_n)) = E(Y_n) = \cdots = Y_0.$$

Definition 4.3.3 *A stochastic process* $\{Y_n\}$ *is said to be a martingale with respect to* $\mathbf{X}_n = \{X_1, X_2, \ldots, X_n\}$ *if* $E(Y_{n+1} \mid \mathbf{X}_n) = Y_n$ *and* $E(|Y_n|) < \infty$ *for all* n. *It follows that* $E(Y_{n+k} \mid \mathbf{X}_n) = Y_n$ *and via reasoning similar to above that* $E(Y_n) = Y_0$.

Martingale derives it's name from a strap used to restrict the movement of the horse's head. We can see by the definition that the history $\mathbf{X}_n$ restricts the conditional expectation of Y_n. Note that in the betting strategies above, case 2 is a martingale and case 1 is not.

Theorem 4.3.4 *Martingale Convergence Theorem: If* $\{Y_n\}$ *is a martingale such that for some* $M < \infty$ $E\{|Y_n|\} \leq M$, *for all* n, *then with probability 1,* $lim_{n\to\infty} Y_n$ *exists, and is finite, i.e.,* $Y_n \to^{a.s.} Y$ *or* Y_n *converges almost surely to* Y. *(see Section 6.4 in Ross[?] for the proof).*

Definition 4.3.5 *Zero Mean Martingales (ZMM):* $\{Y_n\}$ *is a martingale such that* $E[Y_n] = 0$.

Theorem 4.3.6 *If* $\{Y_n\}$ *is a ZMM, then* Y_n *converges in distribution to a time-transformed Brownian motion as* $n \to \infty$.

Example 4.3.7 *Returning to the GW branching process we let the score function be*

$$Y_n(\theta) = \frac{1}{\sigma^2}\sum_{k=1}^{n}[Z_k - Z_{k-1}\theta] = \frac{1}{\sigma^2}\sum_{k=1}^{n}[Z_k - E(Z_k \mid Z_{k-1})]$$

and let $X_k = Z_k - \theta Z_{k-1}$ *such that* $E[X_k] = 0$. *Then it follows that* $E[Y_n(\theta)] = 0$, *and hence is a ZMM. Thus,* $Y_n(\theta)$ *converges to a time-transformed Brownian motion with mean 0. This makes* $Y_n(\theta)$ *a good estimating equation. This implies that* $\widehat{\theta_n} \to^{as} \theta$, *and is, therefore, a strongly consistent estimator. We will apply this theory in detail in Chapter 7.*

For many sequential stochastic processes, the score functions can be written in the form

$$Y_n(\theta) = \frac{\partial \ln L_n}{\partial \theta} = \sum_{k=1}^{n}[Z_k - E(Z_k \mid Z_1, ..., Z_{k-1})],$$

and under mild regularity conditions $Y_n(\theta)$ will be a *ZMM.* This property can be used to find strongly consistent estimators of θ.

4.4 Exercises

Exercise 4.4.1 *Suppose we have a Galton-Watson branching process. We know that an estimator for* θ *is* $\hat{\theta}_n = \frac{\sum_{k=1}^{n} Z_k}{\sum_{k=1}^{n} Z_{k-1}}$.

a. Show that $\hat{\theta}_n$ *is a sufficient statistic for* θ *(Assume that a generalized power series distribution for the offspring distribution).*

b. Show that $\hat{\theta}_n \to^{a.s.} \theta$ *(Hint: Use the Martingale Convergence Theorem and the Toeplitz Lemma.)*

Chapter 5
DISCRETE-TIME MARKOV CHAINS

5.1 Transition Probabilities, Classifications, Asymptotics

In this chapter we derive the properties of discrete-time Markov chains. Throughout we will assume the stochastic process moves through discrete state space S and has support on the nonnegative integers, i.e., $T = \{0, 1, 2, \ldots\}$. First we start with the fundamental definitions of Markov processes.

Definition 5.1.1 *A stochastic process $\{X_n\}$ is said to be a Markov chain if it has the following property: $P[X_t \mid X_0, \ldots, X_{t-1}] = P[X_t \mid X_{t-1}]$.*

Definition 5.1.2 *The one step transition probability: $P[X_{\alpha+1} = i_{\alpha+1} \mid X_\alpha = i_\alpha] = p_{i_\alpha, i_{\alpha+1}}$. For simplicity, in the time homogeneous case we write $P[X_{\alpha+1} = j \mid X_\alpha = i] = p_{ij}$. Note that $\sum_{i_\beta \in S} p_{i_\alpha, i_\beta} = 1$.*

Definition 5.1.3 *The absolute probability is $P[X_\alpha = i_\alpha] = a_{i_\alpha}$.*

For a Markov chain $\{X_n\}$ and $\alpha_0 < \alpha_1 < \cdots < \alpha_n < \beta$,

$$\begin{aligned} P[X_{\alpha_0}, X_{\alpha_1}, \ldots, X_{\alpha_n}, X_\beta] &= P[X_\beta \mid X_{\alpha_0}, \ldots, X_{\alpha_n}] \cdots P[X_{\alpha_1} \mid X_{\alpha_0}] P[X_{\alpha_0}] \\ &= P[X_\beta \mid X_{\alpha_n}] \cdots P[X_{\alpha_1} \mid X_{\alpha_0}] P[X_{\alpha_0}] \\ &= a_{i_{\alpha_0}} p_{i_{\alpha_0}, i_{\alpha_1}} p_{i_{\alpha_1}, i_{\alpha_2}} \cdots p_{i_{\alpha n}, i_\beta} \end{aligned} \tag{5.1}$$

Definition 5.1.4 *For $S = \{1, 2, 3, \ldots\}$ the one step transition probability matrix is given by:*

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{12} & p_{22} & \\ \vdots & & \ddots \\ & & \end{bmatrix}$$

Note that each row of $\mathbf{P}$ sums to unity, i.e., $\sum_{j \in S} p_{ij} = 1$.

Example 5.1.5 *Recall in the Gambler's Ruin that the state space is* $S = \{0, 1, 2, \ldots, a+b\}$. *The* $(a+b+1) \times (a+b+1)$ *one step transition probability matrix is given by:*

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & & 0 \\ q & 0 & p & 0 & & & \\ 0 & q & 0 & p & & & \\ \vdots & & & & \ddots & & \\ & & & & q & 0 & p \\ 0 & & & & 0 & 0 & 1 \end{bmatrix}$$

Example 5.1.6 *Suppose we have a branching process with a Poisson offspring distribution. The state space is* $S = \{0, 1, 2, \ldots, \}$ *and the corresponding one step transition probability matrix is given by:*

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ p_{10} & p_{11} & p_{12} & p_{13} & & \\ p_{20} & p_{21} & p_{22} & p_{23} & & \\ \vdots & & & & \ddots & \\ & & & & & \\ & & & & & \end{bmatrix}$$

Definition 5.1.7 *The* n *step transition probability is* $P[X_{\alpha+n} = j \mid X_\alpha = i] = p_{ij}(n)$ *and* $\mathbf{P}(n)$ *is the* n *step transition probability matrix with* ij^{th} *entry* $p_{ij}(n)$. *We employ the convention that* $\mathbf{P}(1) = \mathbf{P}$.

We now derive the Chapman-Kolmogorov (CK) equation. For any $i, k \in S$,

$$\begin{aligned} p_{ik}(m+n) &= P[X_{\alpha+m+n} = k \mid X_\alpha = i] \\ &= \textstyle\sum_{j \in S} P[X_{\alpha+m+n} = k, X_{\alpha+m} = j \mid X_\alpha = i] \\ &= \textstyle\sum_{j \in S} P[X_{\alpha+m+n} = k \mid X_{\alpha+m} = j, X_\alpha = i] P[X_{\alpha+m} = j \mid X_\alpha = i] \\ &= \textstyle\sum_{j \in S} P[X_{\alpha+m+n} = k \mid X_{\alpha+m} = j] P[X_{\alpha+m} = j \mid X_\alpha = i] \\ &= \textstyle\sum_{j \in S} p_{ij}(m) p_{jk}(n). \end{aligned} \tag{5.2}$$

Therefore,

$$\mathbf{P}(m+n) = \mathbf{P}(m) \cdot \mathbf{P}(n).$$

We use the convention $p_{ii}(0) = 1$. Note that $\mathbf{P}(n) = \mathbf{P} \cdot \mathbf{P}(n-1) = \mathbf{P}^n$, which is not numerically stable in general.

Example 5.1.8 *Consider the 2 step transition probability matrix for* $S = \{1, 2\}$,

$$\mathbf{P}(2) = \begin{bmatrix} p_{11}(2) & p_{12}(2) \\ p_{21}(2) & p_{22}(2) \end{bmatrix} = \begin{bmatrix} p_{11}^2 + p_{11}p_{12} & p_{11}p_{12} + p_{12}p_{22} \\ p_{11}p_{12} + p_{12}p_{22} & p_{11}p_{12} + p_{22}^2 \end{bmatrix} = \mathbf{P}^2.$$

Example 5.1.9 *Consider a Markov chain model for the transmission of binary code with* $S = \{0, 1\}$ *such that*

$$\mathbf{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix},$$

where $0 < p < 1$. *It can be shown that*

$$\mathbf{P}^n = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(p-q)^n & \frac{1}{2} - \frac{1}{2}(p-q)^n \\ \frac{1}{2} - \frac{1}{2}(p-q)^n & \frac{1}{2} + \frac{1}{2}(p-q)^n \end{bmatrix}.$$

Therefore

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Definition 5.1.10 *State* j *is accessible from* i *there exists some* $n > 0$ *such that* $p_{ij}(n) > 0$. *If state* j *is accessible from* i, *we will use the notation:* $i \to j$.

Definition 5.1.11 *If* $i \to j$ *and* $j \to i$, *then* i *and* j *are said to communicate* $(i \longleftrightarrow j)$.

Theorem 5.1.12 *If* $i \to j$ *and* $j \to k$, *then* $i \to k$.

Proof. We know there exists m such that $p_{ij}(m) > 0$ and there exist n such that $p_{jk}(n) > 0$. Thus by the CK equation, $p_{ik}(m+n) = \sum_{j' \in S} p_{ij'}(m)\, p_{j'k}(n) \geq p_{ij}(m)\, p_{jk}(n) > 0$. ■

Definition 5.1.13 *A communicating class* $C(i)$ *is defined to be the set of all states* j *which communicate with* i. *That is,* $j \in C(i)$ *iff* $i \longleftrightarrow j$.

Theorem 5.1.14 *A communicating class follows an equivalence relation*

1. *Reflexivity:* $i \longleftrightarrow i$
2. *Symmetry: If* $i \longleftrightarrow j$ *then* $j \longleftrightarrow i$.
3. *Transitivity: If* $i \longleftrightarrow j$, $j \longleftrightarrow k$ *then* $i \longleftrightarrow k$

Definition 5.1.15 *The first passage time probability is given by*

$$f_{ij}(n) = P[X_n = j, X_m \neq j; m = 1, \ldots, n-1 \mid X_0 = i].$$

One step transition and first passage time probabilities are related as follows: $f_{ij}(1) = p_{ij}(1) = p_{ij}$. We also use the convention that $f_{ij}(0) = 0$ for $i \neq j$. In general,

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l)\, p_{jj}(n-l).$$

Notice that

$$f_{ij}(n) = p_{ij}(n) - \sum_{l=1}^{n-1} f_{ij}(l)\, p_{jj}(n-l),$$

and that the probability we ever reach j from i is given by $f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n)$, $0 \leq f_{ij} \leq 1$. Often we are interested in when $f_{ij} = 1$ when $i \neq j$, in which case $\{f_{ij}(n)\}_{n=0}^{\infty}$ can be thought of as a proper distribution, and the mean first passage time can be calculated:

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}(n).$$

If $f_{ij} < 1$, then the mean first passage time does not exist.

Definition 5.1.16 *State i is a transient state iff $f_{ii} < 1$.*

Definition 5.1.17 *State i is an absorbing state iff $f_{ii} = 1$.*

Example 5.1.18 *In the Gambler's ruin problem, states $1, 2, \ldots, a+b-1$ are transient states while states 0 and $a+b$ are absorbing states.*

Definition 5.1.19 *State i is a recurrent state iff $f_{ii} = 1$. Note that all absorbing states are recurrent.*

Example 5.1.20 *Consider a random walk. If $p = q$, then $f_{ii} = 1$ so that all states are recurrent. If $p \neq q$, then $f_{ii} < 1$ so that all states are transient.*

Definition 5.1.21 *A recurrent state i is nonnull iff $\mu_{ii} < \infty$.*

Example 5.1.22 *Recall the general random walk defined given by 4.1. A state i is a reflecting barrier on the left (on the right) if $q_i = 0 (q_i > 0)$ and $p_i > 0 (p_i = 0)$. Suppose we have a general random walk on $S = \{a, a+1, \ldots, b\}$ where a and b are left and right reflecting barriers respectively. Then all states in S are nonull recurrent.*

Definition 5.1.23 *A recurrent state i is null iff $\mu_{ii} \to \infty$.*

Example 5.1.24 *All states in a symmetric random walk without reflecting barriers are null recurrent.*

Definition 5.1.25 *Let* $t = \gcd\{n : p_{ii}(n) > 0\}$ *where* gcd *is the greatest common divisor. Then state* i *is periodic with period* t *if* $t > 1$.

Example 5.1.26 *Both states given by the following one step transition matrix*

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

are periodic with period 2.

Example 5.1.27 *A SRW with* $p = q = \frac{1}{2}$ *is also of period 2. Recall that* $p_{00}(n) = \binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n$ *when* n *is even and* 0 *when* n *is odd.*

Definition 5.1.28 *A recurrent state* i *is ergodic if* $t = \gcd\{n : p_{ii}(n) > 0\} = 1$.

Example 5.1.29 *For* $0 < p < 1$, *both states given by the following one step transition probability matrix is ergodic:*

$$\mathbf{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix}.$$

Example 5.1.30 *A general random walk with at least one* $r_i > 0$.

Thus states may be classified as follows:

Transient	Recurrent		
	Recurrent null	Recurrent Nonnull	
		Periodic	Ergodic
Ex. SRW $p \neq q$	Ex. SRW $p = q = \frac{1}{2}$	Ex. $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	Ex. $\mathbf{P} = \begin{bmatrix} \mathrm{p} & \mathrm{q} \\ \mathrm{q} & \mathrm{p} \end{bmatrix}$

Theorem 5.1.31

State i *is transient iff* $\sum_{n=0}^{\infty} p_{ii}(n) < \infty$.
State i *is recurrent iff* $\sum_{n=0}^{\infty} p_{ii}(n) \to \infty$.

Theorem 5.1.32

If state i *is transient or recurrent null, then* $\lim_{n\to\infty} p_{ii}(n) = 0$.
If state i *is recurrent nonnull with period* t, *then* $\lim_{n\to\infty} p_{ii}(nt) = \frac{t}{\mu_{ii}}$.
If state i *is ergodic, the* $\lim_{n\to\infty} p_{ii}(n) = \frac{1}{\mu_{ii}}$.

Example 5.1.33 *Recall the period 2 transition probability matrix given in example 5.1.26. Then by theorem 5.1.32,* $\lim_{n\to\infty} p_{ii}(2t) = \frac{2}{\mu_{ii}}$ *implies that* $\mu_{ii} = 2$.

Theorem 5.1.34

If state j is transient or recurrent null, then for all i $\lim_{n\to\infty} p_{ij}(n) = 0$.
If state j is ergodic, then for all i $\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\mu_{jj}}$.
(Note in the latter case, j must be reachable from i).

Definition 5.1.35 *A set C is closed iff $\sum_{j\in C} p_{ij} = 1$ for all $i \in C$.*

Definition 5.1.36 *A closed set of communicating states is a class or irreducible Markov chain.*

Example 5.1.37 *The following one step transition probability matrices gives rise to an irreducible Markov chain:*

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \mathbf{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix} \text{ for } 0 < p < 1.$$

Example 5.1.38 *The gambler's ruin is not an irreducible Markov chain since states 0 and $a + b$ are absorbing states and do not communicate with the other states.*

Definition 5.1.39 *A probability distribution $\{\pi_i\}$ of a Markov chain C is stationary iff*

$$\pi_j = \sum_{i\in C} \pi_i p_{ij} \text{ for } j \in C \text{ and } \sum_{i\in C} \pi_i = 1.$$

Theorem 5.1.40 *If C is an ergodic, irreducible Markov chain, then*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j > 0$$

exists and are independent of state i. Furthermore, the limiting distribution $\{\pi_j\}$ is stationary. Conversely, if a stationary distribution of an irreducible Markov chain exists, then each state in C is ergodic and the stationary distribution is the limiting distribution of the chain.

In matrix notations, for $\mathbf{\Pi} = [\pi_1, \pi_2, \ldots]$, $\mathbf{\Pi} = \mathbf{\Pi P}$ and $\mathbf{\Pi C} = 1$ where $\mathbf{C} = [1, 1, \ldots]^T$.

Example 5.1.41 *For a 2 state process on $S = \{1, 2\}$ the stationary distribution can be found in general. That is,*

$$[\pi_1, \pi_2] \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = [\pi_1, \pi_2] \text{ and } \pi_1 + \pi_2 = 1$$

implies

$$\pi_1 = \frac{p_{21}}{p_{21} + p_{12}} \text{ and } \pi_2 = \frac{p_{12}}{p_{21} + p_{12}}.$$

Example 5.1.42 *Societal classes: 1 upper, 2 middle, 3 lower.*

$$\mathbf{P} = \begin{bmatrix} 0.448 & 0.484 & 0.068 \\ 0.054 & 0.699 & 0.247 \\ 0.011 & 0.503 & 0.486 \end{bmatrix}$$

yields $\mathbf{\Pi} = [0.067, 0.624, 0.309]$ *[**?**].*

5.1.1 Absorbing Chains.

Let N be a random variable for the number of steps to absorption. In this section we discuss inference on time until absorption which relies on partitioning the one step transition probability matrix. First we consider the simpler, scalar example which can be thought of as having one transient state and one absorbing state.

Example 5.1.43 *Suppose N has a geometric mass function i.e. $p_N(n) = q^{n-1}p$, $n = 1, 2, \ldots$ for $0 < p < 1$, so that the pgf is given by $g_N(s) = ps\sum_{j=1}^{\infty} q^{j-1}s^{j-1} = \frac{ps}{1-qs} = (1-qs)^{-1} ps$. It follows that $E[N] = \frac{1}{p} = (1-q)^{-1}$ and that $Var[N] = \frac{q}{p^2} = q(1-q)^{-2}$.*

Now consider the more general case where we have a reducible Markov chain with r states. Let $s < r$ transient and $r - s$ absorbing. We let $C_1 = \{\text{transient states}\}$ and $C_2 = \{\text{absorbing states}\}$ and partition the one step transition probability matrix according to C_1 and C_2. We illustrate this partitioning in the following example:

Example 5.1.44 *Consider the gambler's ruin scenario with $a + b = 4$ such that the transient states are $C_1 = \{1, 2, 3\}$, the absorbing states are $C_2 = \{0, 4\}$, $r = 5$, $s = 3$, and*

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We then partition P as follows:

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} Trans. & Trans. \to Abs. \\ \underline{0} & Abs. \end{bmatrix} \qquad (5.3) \\ &= \left[\begin{array}{ccc|cc} 0 & p & 0 & q & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & 0 & p \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \\ &= \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \end{aligned}$$

so that

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{Q}^n & (\mathbf{I} + \mathbf{Q} + ... + \mathbf{Q}^{n-1})\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Note that $\mathbf{Q}$ is a substochastic matrix (i.e., the sum of at least one row < 1), so that $\lim_{n\to\infty} \mathbf{Q}^n = \mathbf{0}$. Thus

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \mathbf{0} & (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

We now calculate mean time to absorption (note: this could be accomplished using difference equations as discussed earlier). Let m_i be the mean time to absorption given that $X_0 = i, i \in C_1$,then

$$\begin{aligned} m_i &= E\left[N \mid X_0 = i\right] \\ &= \textstyle\sum_{j\in C} p_{ij} E\left(N \mid X_1 = j, X_0 = i\right) \\ &= \textstyle\sum_{j\in C_1} p_{ij} E\left(N \mid X_1 = j, X_0 = i\right) + \sum_{j\in C_2} p_{ij} E\left(N \mid X_1 = j, X_0 = i\right) \\ &= \textstyle\sum_{j\in C_1} p_{ij}(1 + m_j) + \sum_{j\in C_2} p_{ij} \cdot 1 \\ &= 1 + \textstyle\sum_{j\in C_1} p_{ij} m_j. \end{aligned}$$

In matrix form, we let $\mathbf{M} = [m_1, m_2, \ldots, m_s]^T$ and $\mathbf{C} = [1, 1, \ldots, 1]^T$. It follows that

$$\mathbf{I}_s\mathbf{M} = \mathbf{C} + \mathbf{Q}\mathbf{M},$$

where $\mathbf{Q}$ is the $s \times s$ submatrix for the transition states and $\mathbf{I}_s$ is the $s \times s$ identity matrix. This is equivalent to

$$\mathbf{M} = \left(\mathbf{I}_s - \mathbf{Q}\right)^{-1} \mathbf{C}. \tag{5.4}$$

Example 5.1.45 *Revisiting gambler's ruin from example 5.1.44*

$$\mathbf{I} - \mathbf{Q} = \begin{bmatrix} 1 & -p & 0 \\ -q & 1 & -p \\ 0 & -q & 1 \end{bmatrix},$$

so that for $p = \frac{2}{3}$, *we get* $\mathbf{M} = [3.4, 3.6, 2.2]^T$.

It can also be shown that $Var\left[\mathbf{M}\right] = \mathbf{M}^{(2)} - \mathbf{M}^2$ where $\mathbf{M}^{(2)} = \left(\mathbf{I} - \mathbf{Q}\right)^{-1}\left(2\mathbf{M} - \mathbf{C}\right)$.

We now present an alternative derivation using the properties of the pgf for a simpler case. Suppose we have r states such that state 0 is an absorbing state and states $\{1, 2, \ldots, r-1\}$ are communicating transient states. That is, $C_1 = \{1, 2, \ldots, r-1\}$ and $C_2 = \{0\}$. Let $p_i\left(n\right) = P\left[N = n \mid X_0 = i\right] = \sum_{j\in C_1} p_{ij}\left(n-1\right) p_{j0}$. Then

$$\mathbf{P}\left(n\right) = \begin{bmatrix} p_1\left(n\right) \\ p_2\left(n\right) \\ \vdots \\ \\ p_{r-1}\left(n\right) \end{bmatrix} = \mathbf{Q}^{n-1}\mathbf{R}\ ,$$

so that the pgf for N is given by

$$\begin{aligned}\mathbf{f}(s) &= [f_1(s), f_2(s), \ldots, f_{r-1}(s)]^T \\ &= \sum_{j=1}^{\infty} \mathbf{Q}^{j-1} s^j \mathbf{R} \\ &= \sum_{j=1}^{\infty} (\mathbf{Q}s)^{j-1} s\mathbf{R} \\ &= (\mathbf{I} - \mathbf{Q}s)^{-1} s\mathbf{R}.\end{aligned} \quad . \tag{5.5}$$

We leave it as an exercise to derive the mean time to absorption from equation 5.5.

5.2 Algebraic treatment

As mentioned earlier, direct evaluation of $\mathbf{P}^n$ is numerically unstable. Here we use tools from linear algebra to derive the closed form (i.e. numerically stable) of the n-step transition probability matrix.

Let $\mathbf{W}$ be a $s \times s$ matrix

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1s} \\ w_{21} & w_{22} & & \\ \vdots & & \ddots & \\ w_{s1} & w_{s2} & \cdots & w_{ss} \end{bmatrix}.$$

Definition 5.2.1 *The cofactor W_{ij} of the $(i,j)th$ element of $\mathbf{W}$ is given by $(-1)^{i+j}$ times the determinant of the submatrix obtained by deleting the i^{th} row and j^{th} column of $\mathbf{W}$.*

Definition 5.2.2 *The adjoint of $\mathbf{W}$ is given by*

$$\mathbf{W}^+ = \begin{bmatrix} W_{11} & W_{21} & \cdots & W_{s1} \\ W_{12} & W_{22} & & W_{s2} \\ \vdots & & \ddots & \\ W_{1s} & W_{2s} & \cdots & W_{ss} \end{bmatrix}$$

Definition 5.2.3 *If there exist a non-zero vector $\mathbf{t}$ and a scalar λ such that $\mathbf{W}\mathbf{t} = \lambda\mathbf{t}$, then λ is called an eigenvalue of $\mathbf{W}$ and $\mathbf{t}$ the eigenvector. An equivalent form is given by $(\lambda\mathbf{I} - \mathbf{W})\mathbf{t} = 0$. Letting $\mathbf{A} = \lambda\mathbf{I} - \mathbf{W}$, then $\mathbf{A}\mathbf{t} = 0$ has a solution if and only if $\mathbf{A}$ is singular. That is, if and only if $|\mathbf{A}| = 0$ where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$.*

Definition 5.2.4 *$|\mathbf{A}| = 0$ is the characteristic equation of $\mathbf{W}$.*

Thus eigenvalues are the roots of the characteristic equation. We will rely heavily on the following theorem.

Theorem 5.2.5 *If $\lambda_1, \lambda_2, \ldots, \lambda_s$ are distinct eigenvalues of $\mathbf{W}$, then the matrix of corresponding eigenvectors $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_s]$ is invertible.*

From the following equations

$$\mathbf{WT} = [\mathbf{Wt}_1, \mathbf{Wt}_2, \ldots, \mathbf{Wt}_s] = [\lambda_1\mathbf{t}_1, \lambda_2\mathbf{t}_2, \ldots, \lambda_s\mathbf{t}_s]$$

we arrive at

$$\mathbf{T}^{-1}\mathbf{WT} = [\lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \ldots, \lambda_s\mathbf{e}_s] = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_s \end{bmatrix},$$

where $\mathbf{e}_i$ is a vector of 0's with a 1 in the i^{th} position (e.g. $\mathbf{e}_1 = [1, 0, 0, \ldots, 0]$). Therefore,

$$\mathbf{W}^n = \mathbf{T} \begin{bmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_s^n \end{bmatrix} \mathbf{T}^{-1}. \tag{5.6}$$

Theorem 5.2.6 *Let $\mathbf{A}(\lambda_j) = \lambda_j\mathbf{I} - \mathbf{W}$ be the characteristic matrix of λ_j. Then any non-zero column of $\mathbf{A}^+(\lambda_j)$ is an eigenvector for λ_j.*

Example 5.2.7 *Suppose*

$$\mathbf{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix} \text{ for } 0 < p < 1,$$

such that

$$\mathbf{A} = \begin{bmatrix} \lambda - p & -q \\ -q & \lambda - p \end{bmatrix}.$$

Setting the determinant to zero, i.e., $|\mathbf{A}| = 0$, gives rise to two roots: $\lambda_1 = 1$ and $\lambda_2 = p - q$. It follows that

$$\mathbf{A}^+(1) = \begin{bmatrix} q & q \\ q & q \end{bmatrix} \text{ and } \mathbf{A}^+(p-q) = \begin{bmatrix} -q & q \\ q & -q \end{bmatrix}.$$

From theorem 5.2.6, a matrix of eigenvectors is given by

$$\mathbf{T} = \begin{bmatrix} q & q \\ q & -q \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \mathbf{P}(n) &= \begin{bmatrix} q & q \\ q & -q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (p-q)^n \end{bmatrix} \tfrac{1}{2q} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \tfrac{1}{2} \begin{bmatrix} 1 + (p-q)^n & 1 - (p-q)^n \\ 1 - (p-q)^n & 1 + (p-q)^n \end{bmatrix}. \end{aligned}$$

Chiang derived the following explicit form for the ij^{th} entry of $\mathbf{P}(n)$:

$$P_{ij}(n) = \sum_{l=1}^{s} A_{ji}(\lambda_l)\,\lambda_l^n \frac{1}{\prod_{\substack{m=1 \\ m\neq l}}^{s}(\lambda_l - \lambda_m)} \tag{5.7}$$

Example 5.2.8 *Revisiting example 5.2.7:*

$$\begin{aligned} P_{11}(n) &= \textstyle\sum_{l=1}^{2} A_{11}(\lambda_l) \frac{\lambda_l^n}{\prod_{\substack{m=1 \\ m\neq l}}^{2}(\lambda_l-\lambda_m)} \\ &= A_{11}(\lambda_1)\,\lambda_1^n\left(\tfrac{1}{\lambda_1-\lambda_2}\right) + A_{11}(\lambda_2)\,\lambda_2^n\left(\tfrac{1}{\lambda_2-\lambda_1}\right) \\ &= q\left(\tfrac{1}{1-p+q}\right) + (-q)(p-q)^n\left(\tfrac{1}{p-q-1}\right) \\ &= \tfrac{1}{2} + \tfrac{1}{2}(p-q)^n. \end{aligned}$$

Furthermore, Chiang shows that

$$\lim_{n\to\infty} P_{ij}(n) = \frac{A_{jj}(1)}{\sum_{k=1}^{s} A_{kk}(1)}$$

Finally, we note that for matrices with non-distinct eigenvalues, diagonalization as in equation 5.6 is not possible, and one must resort to the Jordan from.

5.3 Inference

In this section, we will introduce likelihood based inference on discrete state space and index set, homogeneous Markov chains. Generally one can perform inference on a single observed sequence over time or on many observed sequences over time.

5.3.1 Inference on a single sequence

We assume that we are estimating a set of parameters, $\boldsymbol{\theta}$, from a single observed sequence $\{x_1, x_2, \ldots, x_n\}$, from a single population. Then the likelihood function is

$$L_n(\boldsymbol{\theta}) = \prod_{k=0}^{n} P(X_k \mid X_{k-1}).$$

Then the score functions are

$$S_n(\boldsymbol{\theta}) = \frac{\partial \ln L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{k=0}^{n} P(X_k \mid X_{k-1})^{-1} \frac{\partial P(X_k \mid X_{k-1})}{\partial \boldsymbol{\theta}}.$$

Example 5.3.1 *Galton-Watson branching process. Then* $P(Z_k \mid Z_{k-1}) \Finv \frac{\lambda^{Z_k}}{f(\lambda)^{Z_{k-1}}}$, *and the log likelihood is given by (3.3) in section 3.2.3.*

Example 5.3.2 *The Reed-Frost model. See (5.26) in section 5.5.3.*

5.3.2 Inference on multiple observed sequences

We let n_{ij} be the number of observed transitions from state i to state $j, (i,j) \in S$. Based on these data, we can estimate the elements of one step transition matrix, $\mathbf{P}$. Suppose the Markov chain is finite with s states. Then the problem reduces to a $s \times s$ contingency table. The probability of the outcome $\{n_{i1}, n_{i2}, \ldots, n_{is}\}$ follows a multinomial distribution with probability

$$\frac{n_{i\cdot}!}{n_{i1}!n_{i2}!\ldots n_{is}!} p_{i1}^{n_{i1}} p_{i2}^{n_{i2}} \ldots p_{is}^{n_{is}},$$

where $n_{i\cdot} = \sum_{j=1}^{s} n_{ij}$. The likelihood function for $\mathbf{P}$, conditional on $\{n_{i1}, n_{i2}, \ldots, n_{is}\}$ is

$$L(\mathbf{P}) = c \prod_{i=1}^{s} \prod_{ji=1}^{s} p_{ij}^{n_{ij}},$$

where c is a constant. This is the likelihood for multinomial data. It follows directly that the maximum likelihood estimates are $\widehat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}$,with estimated variances $\widehat{var}(\widehat{p}_{ij}) \simeq \frac{\widehat{p}_{ij}(1-\widehat{p}_{ij})}{n_{i\cdot}}$, $(i,j) \in S$.

5.4 The chain binomial model

Let time be discrete and indexed $t = 0, 1, ...$ Let S_t be the number of individuals at risk for the event of interest (*e.g.*, infection, death) at the beginning of time interval t, and I_t be the number that experienced the event of interest at the beginning of time interval t. The event has a duration of at least one time interval. We let $p_t = 1 - q_t = f(t, \boldsymbol{\theta}, I_t)$ be the probability that an at-risk individual has a new event at the beginning of time interval $t+1$, with parameters $\boldsymbol{\theta}$. As shown, this probability can be a function, $f(\cdot)$, of t and I_t. We usually start with a closed population of $n = S_0 + I_0$ individuals. Then I_{t+1} is a binomial random variable that follows the conditional probability mass function

$$\Pr(I_{t+1} = i_{t+1} \mid S_t = s_t, p_t) = \binom{s_t}{i_{t+1}} p_t^{i_{t+1}} q_t^{s_t - i_{t+1}}, s_t \geq i_{t+1} \ . \tag{5.8}$$

In many cases, S_t is updated via the relationship

$$S_{t+1} = S_t - I_{t+1}, \tag{5.9}$$

although other relationships are possible (see below). The conditional expectation and variance of I_{t+1}, respectively, are

$$E(I_{t+1} \mid s_t, p_t) = s_t p_t, \tag{5.10}$$

$$var(I_{t+1} \mid s_t, p_t) = s_t p_t q_t. \tag{5.11}$$

Equations (5.8,5.9) form the classical chain-binomial model. Formal mathematical treatment of the model involves formulation of the discrete, two-dimensional Markov chain $\{S_t, I_t\}_{t=0,1,...}$. I_t is the (binomial) random variable of interest, and S_t is updated using (5.9). The probability of a particular chain, $\{i_0, i_1, i_2, ..., i_r\}$, is given by the product of conditional binomial probabilities from (5.8) as

$$\begin{aligned}\Pr(I_1 &= i_1 \mid S_0 = s_0, p_0)\Pr(I_2 = i_2 \mid S_1 = s_1, p_1) \cdots \Pr(I_r = i_r \mid S_{r-1} = s_{r-1}, p_{r-1}) \\ &= \prod_{t=0}^{r-1} \binom{s_t}{i_{t+1}} p_t^{i_{t+1}} q_t^{s_t - i_{t+1}}.\end{aligned}$$

The conditional expected value of I_{t+1} (5.10) suggests the deterministic system of first-order difference equations

$$\begin{aligned} i_{t+1} &= s_t p_t, \\ s_{t+1} &= s_t - i_{t+1}, \end{aligned} \tag{5.12}$$

which can be analyzed as an approximation to the mean of the sample paths of the stochastic process $\{S_t, I_t\}_{t=0,1,...}$. This system reduces to

$$s_t = s_{t-1} q_{t-1} = s_0 \prod_{\ell=0}^{t-1} q_\ell, \tag{5.13}$$

which is analyzed using methods from discrete mathematics (*e.g.*, see Frauenthal[9] and Longini[14]).

5.5 The Reed-Frost Model

This section is taken from Longini[16].

5.5.1 History

The probabilistic form of the Reed-Frost epidemic model was introduced by the biostatistician Lowell J. Reed and the epidemiologist Wade Hampton Frost around 1930 as a teaching tool at Johns Hopkins University. It was developed as a mechanical model consisting of colored balls and wooden shoots. Although Reed and Frost never published their results, the work is described in articles and books by others (see Chapters 14 and 18 in Bailey[1] and Chapters 2 and 3 in Becker [3]). An excellent description of the early Reed-Frost model is given by Fine [8]. The deterministic version of the Reed-Frost model has been traced back to the Russian epidemiologist P.D. En'ko who used the model to analyze epidemic data in the 1880's (see Dietz [7]). The Reed-Frost version of the chain binomial and its extensions is used to study the dynamics of epidemics in small populations, such as families or day care centers, and to estimate transmission probabilities from epidemic data.

5.5.2 Formulation

In this case, S_t is the number of susceptible persons at the beginning of time interval t and I_t is the number of persons who where newly infected at the beginning of time interval t. An infected person is infectious for exactly one time interval and then is removed, *i.e.*, becomes immune. Thus, a person infected at the beginning of time interval t, will be infectious to others until the beginning of time interval $t+1$. We let R_t be the number of removed persons at the beginning of time interval t, and then, by definition

$$R_{t+1} = R_t + I_t = \sum_{r=0}^{t} I_r. \tag{5.14}$$

Since the population is closed, we have $S_t + I_t + R_t = n$ for all t. We let $p = 1 - q$ be the probability that any two specified people make sufficient contact in order to transmit the infection, if one is susceptible and the other infected, during one time interval. We note that p is a form of the secondary attack rate. We assume random mixing. Then, if during time interval t there are I_t infectives, then the probability that a susceptible will escape being infected over the time interval is q^{I_t}, and the probability that they will become a new case at the beginning of time interval $t+1$ is $1 - q^{I_t}$. Thus, $q_t = q^{I_t}$, and substituting into (5.8) yields

$$\Pr(I_{t+1} = i_{t+1} \mid S_t = s_t, I_t = i_t) = \binom{s_t}{i_{t+1}} \left(1 - q^{i_t}\right)^{i_{t+1}} q^{i_t(s_t - i_{t+1})}, s_t \geq i_{t+1} \ . \tag{5.15}$$

The epidemic process starts with $I_0 > 0$, and terminates at stopping time T, where

$$T = \inf_{t \geq 0} \{t : S_t I_t = 0\} . \tag{5.16}$$

Table 1 shows the possible chains for a population of size 4 with one initial infective, *i.e.*, $S_0 = 3,\ I_0 = 1$.

Table 1: *Possible individual chains when* $S_0 = 3,\ I_0 = 1$

Chain	Probability	Final Size
$\{i_0, i_1, i_2, ..., i_T\}$		R_T
$\{1\}$	q^3	1
$\{1,1\}$	$3pq^4$	2
$\{1,1,1\}$	$6p^2q^4$	3
$\{1,2\}$	$3p^2q^3$	3
$\{1,1,1,1\}$	$6p^3q^3$	4
$\{1,1,2\}$	$3p^3q^2$	4
$\{1,2,1\}$	$3p^3q\,(1+q)$	4
$\{1,3\}$	p^3	4

The probability of no epidemic is defined as the probability that there will be no further cases beyond the initial cases. This probability is

$$\Pr(I_1 = 0 \mid S_0 = s_0, p_0) = q^{i_0 s_0}. \tag{5.17}$$

For example if $S_0 = 10,\ I_0 = 1$, and $p = 0.05$, then the probability of no further cases beyond the initial case is 0.599. From (5.10), the conditional expected number of new cases in time interval t is $\mathrm{E}(I_{t+1} \mid s_t, p_t) = s_t\,(1 - q^{i_t})$. On the average, the epidemic process will not progress very far if the expected number of cases in the first generation is less than or equal to one, *i.e.*, $\mathrm{E}(I_1 \mid s_0, p_0) = s_0\,(1 - q^{i_0}) \leq 1$. In many cases, $i_0 = 1$, so that there will be few secondary cases if $s_0 p \leq 1$. Then, for example, if $S_0 = 10,\ I_0 = 1$, there will be few secondary cases if $p \leq 0.1$.

From (5.13), the deterministic counterpart of the Reed-Frost model is

$$s_t = s_0 q^{\sum_{\ell=0}^{t-1} i_\ell}, \tag{5.18}$$

which has been thoroughly analyzed (*e.g.*, see Frauenthal[9] and Longini[14]).

In some cases, the distribution of the total number of cases, R_T, is the random variable of interest. We let J be the random variable for the total number of cases in addition to the initial cases, so that $R_T = J + I_0$. If we let $S_0 = k$ and $I_0 = i$, then the probability of interest is

$$\Pr\left(J = j \mid S_0 = k, I_0 = i\right) = m_{ijk,} \tag{5.19}$$

where $\sum_{j=0}^{k} m_{ijk} = 1$. Then, based on probability arguments (*e.g.*, see Bailey [1]), we have the recursive expression

$$m_{ijk} = \binom{k}{j} m_{ijj} q^{(i+j)(k-j)},\ j < k \tag{5.20}$$

and

$$m_{ikk} = 1 - \sum_{j=0}^{k-1} m_{ijk}. \tag{5.21}$$

The Reed-Frost model has several extensions and special cases. If it is hypothesized that the probability that a susceptible becomes infected does not depend on the number of infectives that he or she is exposed to, then

$$p_t = \begin{cases} p, & \text{if} \quad I_t > 0, \\ 0, & \text{if} \quad I_t = 0. \end{cases} \tag{5.22}$$

This model is know as the Greenwood model [10].

Longini and Koopman [23] modified the Reed-Frost model for the common case where there is a constant source of infection from outside the population that does not depend on the number of infected persons in the population. We let $a_t = 1 - b_t$ be

the probability that a susceptible person is infected during interval t due to contacts with infected persons outside the population, where

$$\begin{aligned} a_t &> 0 \quad \text{if} \quad t \leq T, \\ a_t &= 0 \quad \text{if} \quad t > T, \end{aligned}$$

and T is a stopping time. Then $p_t = 1 - b_t q^{I_t}$. If we let $B = \prod_{t=0}^{T} b_t$, then B is the probability that a person escapes infection from sources outside of the population over the entire period $[0, T]$. We then define CPI $= 1-$ B as the community probability of infection. Longini and Koopman derive probability mass function

$$m_{ijk} = \binom{k}{j} m_{ijj} B^{(k-j)} q^{(i+j)(k-j)}, \ j < k. \tag{5.23}$$

Usually, $i = 0$ for this model. This model reduces to (5.20) when $B = 1$.

Another extension of the Reed-Frost model is for infectious diseases that do not confer immunity following infection. In this case, there is no removed state so that $S_t + I_t = n$. Then, since $S_{t+1} = n - I_{t+1}$, the model is a discrete, one-dimensional Markov chain $\{I_t\}_{t=0,1,\ldots}$. The transition probabilities for this process are

$$\Pr(I_{t+1} = i_{t+1} \mid I_t = i_t) = \binom{n - i_t}{i_{t+1}} \left(1 - q^{i_t}\right)^{i_{t+1}} q^{i_t(n - i_t - i_{t+1})}, i_t - i_{t+1} \ \leq n. \tag{5.24}$$

In this case, the disease in question can become "endemic." An interesting analytical question involves the study of the mean stopping time for the endemic process. From (5.12), the deterministic counterpart of this model is

$$i_{t+1} = (n - i_t)\left(1 - q^{i_t}\right), \tag{5.25}$$

which is a form of the discrete logistic function. The stochastic behavior of (5.24) has been analyzed by Longini[13], and the dynamics of (5.25) have been analyzed by Cooke, *et al.*[6].

There are many other extensions of the Reed-Frost model depending on the particular infectious disease being analyzed, but a further key extension is to allow the infectious period to extend over several time intervals. In this case $p_t = f(t, \boldsymbol{\theta}, I_0, I_1, ..., I_t)$, and $\{S_t, I_t\}_{t=0,1,\ldots}$ is not a Markov chain. Special methods are used to analyze this model (Saunders[26]).

5.5.3 Inference

Data are usually in the form of observed chains, $\{i_0, i_1, ..., i_r\}$, for one or more populations, or final sizes, R_T, for more than one population. With respect to the former data form, suppose that we have N populations and let $\{i_{k0}, i_{k1}, ..., i_{kr}\}$ be the observed chain for the k^{th} population. Then, from (5.1), the likelihood function for estimating $p = 1 - q$ is

$$\mathrm{L}(p) = \prod_{k=1}^{N} \prod_{t=0}^{r-1} \binom{s_{kt}}{i_{kt+1}} \left(1 - q^{i_{kt}}\right)^{i_{kt+1}} q^{i_{kt}(s_{kt} - i_{kt+1})}, \tag{5.26}$$

For final value data, let a_{ijk} be the observed frequencies of the m_{ijk}, from (5.23), $i = 1, ..., I$; $k = 1, ..., K$; and $j = 1, ..., k$. Then the likelihood function for estimating p and B is

$$\mathrm{L}(p, B) = \prod_{i=1}^{I} \prod_{k=1}^{K} \prod_{j=0}^{k} m_{ijk}^{a_{ijk}}. \tag{5.27}$$

The logarithms of (5.26,5.27) are maximized using standard scoring routines *(e.g.*, Bailey[1], Becker[3], Longini, *et al.*[23], [24]), or the corresponding generalized linear model (see Becker [3], Haber, *et al.*[11]*).* Extensions involve making both p and the CPI functions of covariates, such as age, level of susceptibility or vaccination status. Bailey[1] *(Sec.* 14.3) gives an example where (5.26) is used to estimate $\widehat{p} = 0.789 \pm 0.015$ (estimate ± 1 standard error) for the household spread of measles among children. In the case of the household spread of influenza, Longini, *et al.*[24] use (5.27) to estimate $\widehat{p} = 0.260 \pm 0.030$ for persons with no prior immunity and $\widehat{p} = 0.021 \pm 0.026$ for persons with some prior immunity. In addition, they estimate $\widehat{\mathrm{CPI}} = 0.164 \pm 0.015$ and $\widehat{\mathrm{CPI}} = 0.092 \pm 0.013$ for persons with no and some prior immunity, respectively.

5.6 Life Tables

The chain binomial model forms the statistical underpinnings of the life table (see Chapter 10 in Chiang [5]). In this case, p_t simply depends on the time interval. Then, S_t is the random variable of interest, which is formulated in terms of the interval survival probabilities $q_t = 1 - p_t$. Many important life table indices are functions of q_t. For example, the probability that an individual who starts in the cohort at time 0, is still alive at the end of time interval r, denoted q_{0r}, is $q_{0r} = \prod_{t=0}^{r} q_t$. The expected number alive at the beginning of time interval $r + 1$ is $E(S_{r+1}) = s_0 q_{0r}$. This model is a discrete, one-dimensional Markov chain $\{S_t\}_{t=0,1,...}$. From (5.8), we see that the chain binomial model for S_t is simply

$$\Pr(S_{t+1} = s_{t+1} \mid S_t = s_t) = \binom{s_t}{s_{t+1}} q_t^{s_{t+1}} p_t^{s_t - s_{t+1}}, s_t \geq s_{t+1}\ . \tag{5.28}$$

From (5.1), the probability of a particular chain $\{s_0, s_1, s_2, \ldots, s_r\}$ is

$$\begin{aligned} \Pr(S_1 &= s_1 \mid S_0 = s_0)\Pr(S_2 = s_2 \mid S_1 = s_1) \cdots \Pr(S_r = s_r \mid S_{r-1} = s_{r-1}) \qquad (5.29) \\ &= \prod_{t=0}^{r-1} \binom{s_t}{s_{t+1}} q_t^{s_{t+1}} p_t^{s_t - s_{t+1}}. \end{aligned}$$

For an observed chain $\{s_0, s_1, s_2, \ldots, s_r\}$, (5.29) is the likelihood function for estimating $\{q_0, q_1, \ldots, q_r\}$. The maximum likelihood estimators are

$$\widehat{q}_t = \frac{s_{t+1}}{s_t}, \tag{5.30}$$

while the approximate variances, for large S_0, are

$$\text{var}(\widehat{q}_t) \approx \frac{p_t q_t}{\text{E}(S_t)}. \tag{5.31}$$

In addition, the $\widehat{q}_t$ are unique, unbiased estimates of the q_t, and $\text{cov}(\widehat{q}_t, \widehat{q}_\ell) = 0,\ t \neq \ell$. Estimators of most of the life table functions are based on the estimators $\widehat{q}_t$.

5.7 HIV-progression model

Longini[15] defined the stages of HIV infection based on T4 cell counts as shown in the table below:

Stages of HIV infection

Stage	T4 cell count range
1	> 899
2	700 - 899
3	500 - 699
4	200 - 499
5	0 - 200
6	Deceased

We assume that the T4 cell decline is monotonically decreasing so that the one-step transition matrix has the following form:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ 0 & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ 0 & 0 & p_{33} & p_{34} & p_{35} & p_{36} \\ 0 & 0 & 0 & p_{44} & p_{45} & p_{46} \\ 0 & 0 & 0 & 0 & p_{55} & p_{56} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

An estimate of this matrix from U.S. Army data is

$$\mathbf{P} = \begin{bmatrix} 0.48 & 0.18 & 0.20 & 0.14 & 0 & 0 \\ 0 & 0.49 & 0.23 & 0.26 & 0.02 & 0 \\ 0 & 0 & 0.53 & 0.45 & 0.02 & 0 \\ 0 & 0 & 0 & 0.79 & 0.19 & 0.02 \\ 0 & 0 & 0 & 0 & 0.70 & 0.30 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The eigenvalues of this matrix are $\boldsymbol{\rho} = [\ 0.48, 0.49, 0.53, 0.79, 0.7, 1]$, and the

matrix of eigenvectors is

$$\mathbf{T} = \begin{bmatrix} 1 & 18.0 & 24.7 & 2.8419 & -13.28 & 1 \\ 0 & 1.0 & 5.75 & 2.1936 & -8.5101 & 1 \\ 0 & 0 & 1.0 & 1.7308 & -5.4706 & 1 \\ 0 & 0 & 0 & 1.0 & -2.1111 & 1 \\ 0 & 0 & 0 & 0 & 1.0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

5.8 Endemic Reed-Frost Model

Consider the endemic model described in (5.24). This model is a one-dimensional Markov chain $\{I_t\}_{t=0,1,...}$, on the state space $S = \{0, 1, ..., n\}$, with transition probabilities

$$p_{ij} = \begin{cases} \binom{n-i}{j} (1-q^i)^j \, q^{i(n-i-j)}, & \text{if, } i+j \ \leq n \\ 0, & \text{if, } i+j \ > n \end{cases}. \tag{5.32}$$

State 0 is an absorbing state and state n is an isolated state. States $1, ..., n-1$ are transient states. Thus, given that the process starts in one of the transient states, it will eventually end up in state 0. The one-step transition matrix is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} & \cdots & p_{1n-1} & p_{10} \\ p_{21} & p_{22} & \cdots & p_{2r} & \cdots & 0 & p_{20} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} & \vdots & \vdots & p_{r0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n-1,1} & 0 & 0 & 0 & 0 & 0 & p_{n-1,0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ where } r \leq \tfrac{n}{2}.$$

We are interested in the mean time to absorption. To study the absorbing chain, we have reordered the states as $S = \{1, 2, ..., n-1, 0\}$.We can now partition matrix $\mathbf{P}$ into the absorbing chain form (5.3), where

$$\mathbf{Q} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} & \cdots & p_{1n-1} \\ p_{21} & p_{22} & \cdots & p_{2r} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{n-1,1} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} p_{10} \\ p_{20} \\ \vdots \\ p_{r0} \\ \vdots \\ p_{n-1,0} \end{bmatrix}, \text{ and } \mathbf{I} = 1.$$

Then, (5.4) can be used to compute the mean time to absorption, *i.e.*, length of the endemic process. For example, suppose $p = 0.05$ and $n = 5$, then given one initial infective, i.e., $I_0 = 1$,the mean time to absorption is $m_1 = 1.2$ time units; given two initial infectives, $m_2 = 1.3$; and for three initial infectives, $m_3 = 1.3$. The following table shows m_1 for different values of n.

Expected Length of the Endemics Process
With One Initial Infective

Population size n	Expected length m_1
5	1.2
10	1.6
15	2.3
20	3.8
25	8.7
30	45.2
35	839.6
40	52,385

Once the population size approaches around 40, the mean time to absorption jumps to a very long time. Thus, the infectious disease becomes endemic for population sizes of about 40 or larger. We can see this by examining the components of vector $\mathbf{R}$, $\{p_{i0}\}$, where $p_{i0} = q^{i(n-i)}$. When $n \to \infty$, then $p_{i0} \to 0$, for fixed i, and $\mathbf{R} \to \mathbf{0}$.Thus, the chain becomes recurrent, *i.e.*, the disease becomes endemic, as n gets large.

5.9 Exercises

Exercise 5.9.1 *The Chapman-Komogorov equations 5.2 may be extended as follows:*

$$p_{il}(n_1 + n_2 + n_3) = \sum_j \sum_k = p_{ij}(n_1)\, p_{jk}(n_2)\, p_{kl}(n_3)$$

Verify this equation.

Exercise 5.9.2 *Let $\{X_n\}_{n=0}^{\infty}$ be a homogeneous Markov chain.*

a. Show that

$$P[X_m = j \mid X_{m+1} = i, X_{m+2} = i_2, \ldots, X_{m+n} = i_n] = P[X_m = j \mid X_{m+1} = i]$$

(Note that $P[X_m = j \mid X_{m+1} = i]$ are the one-step transition probabilities for the "reverse" Markov chain).

b. Let $b_{ij} = P[X_m = j \mid X_{m+1} = i]$ be the elements of the reverse one-step transition probability matrix $\mathbf{B}$. Find an expression for b_{ij} in terms of p_{ij} and $a_j(n)$ from the forward Markov chain.

Exercise 5.9.3 *Consider a simple random walk. Prove that if $p \neq q$, then state 0 is transient (i.e. $f_{00} < 1$) and that if $p = \frac{1}{2}$, then state 0 is recurrent (i.e. $f_{00} = 1$). Use theorem 5.1.31 and Stirling's formula to approximate the factorial term.*

Exercise 5.9.4 *Suppose we have a homogeneous, reducible Markov chain with r states, one of which is absorbing state and the others being communicating transient states. Let N be the random variable for the number of steps to absorption. We showed in class that the pgf for N is given by $\mathbf{f}(s) = (\mathbf{I} - s\mathbf{Q})^{-1} s\mathbf{R}$. Using $\mathbf{f}(s)$ show that*

a. $\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{C}$

b. $\mathbf{M}^{(2)} = (\mathbf{I} - \mathbf{Q})^{-1} (2\mathbf{M} - \mathbf{C})$

Exercise 5.9.5 *Suppose we have an endemic process for a group of size $n = 3$, $s = (0, 1, 2, 3)$.*

a. Find the eigenvalues of the one-step transition matrix.

b. Write out $p_{11}(n)$, $p_{12}(n)$, and $p_{10}(n)$.

Exercise 5.9.6 *Consider a staged disease process where an individual passes through stages $1, \ldots, k-1$ before reaching the absorbing state $k > 1$. The one-step transition probabilities are*

$$p_{ij} = \begin{cases} p_i & \text{if } j = i + 1 \\ q_i & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

$p_i + q_i = 1$, $0 < p_i < 1$, $i = 1, 2, \ldots, k-1$, and $p_{kk} = 1$, $k > 1$.

a. Classify the states and the Markov chain.

b. Find the eigenvalues of the one-step transition probability matrix and interpret them.

c. Suppose $k = 3$, find $p_{11}(n)$, $p_{12}(n)$, and $p_{13}(n)$.

Exercise 5.9.7 *The natural history of carcinoma of the cervix is as follows:*

$$\textit{Dysplasia} \rightarrow \textit{ Carcinoma in sites} \rightarrow \textit{ Invasive cancer}$$

Suppose a cohort of women with dysplasia are followed on a yearly basis for the rest of their lives. The following information is known about the yearly transitions between states:

i. 20% of those with dysplasia will convert to cancer in sites, while 80% remain dysplasia.

ii. 10% of those with cancer in sites will develop invasive cancer; 10% of those with cancer in sites will revert back to dysplasia; while the rest, i.e. 80%, maintain cancer in sites.

iii. All those with invasive cancer will remain in that state.

The progression of the women with dysplasia to other forms of the disease can be modeled as a homogeneous Markov process with three discrete states and in discrete time (one year intervals).

a. Given the one-step transition probability matrix $\mathbf{P}$ for this process.

b. Find the eigenvalues of this system.

c. Give the mean number of years for a woman to develop invasive cancer from the dysplastic state.

Now assume that 80% of those women with invasive cancer will be successfully treated, and will return to the cancer in sites state in one year.

d. Classify the states of this Markov chain, and the chain itself.

e. On the average, what proportion of the cohort will be in each of the three states after a long period of time?

Exercise 5.9.8 *Consider the "endemic" process discussed in section 5.5. That is, let I_t be the number of infectives at time t, n be the population size, $p = (1 - q)$ be the transmission probability. Then the deterministic system is*

$$I_{t+1} = (n - I_t)\left(1 - q^{I_t}\right), \ t = 1, 2, \ldots$$

and $0 < I_0 < n$.

a. Describe the state space of the system.

b. Find the fixed points of the system.

c. Describe the behavior of the system as $t \to \infty$. Specifically, under what conditions does

$$\lim_{t\to\infty} I_t = 0 \text{ and } \lim_{t\to\infty} I_t = I^* > 0.$$

Hint: Transform the system by $X_t = \frac{I_t}{n}$ and $\beta = -n \ln q$.

Chapter 6
CONTINUOUS-TIME MARKOV CHAINS

Suppose now that the stochastic process still assumes values on a discrete state space, e.g. $S = (0, 1, 2, \ldots)$, but that time is now continuous, i.e. $T = (t : 0 \leq t < \infty)$.

Continuous-time processes of this type can be used to model birth and death type processes as well as infectious diseases. We first give the definition for the continuous time analog of the discrete time n step transition probability.

Definition 6.0.1 *Let the conditional probability that the process is in state k at time t given it is in state i at time 0 be given by $p_{ik}(0,t) = P[x(t) = k \mid x(0) = i]$. For simplicity, we will use the notation $p_k(t) = p_{ik}(0,t)$.*

6.1 Poisson process

Let $X(t)$ be the number of individuals alive at time t with initial condition $X(0) = 0$. Then define

$$P[X(t+\Delta) = k \mid X(t) = j] = \begin{cases} \lambda\Delta + o(\Delta) & \text{if } k = j+1 \\ 1 - \lambda\Delta + o(\Delta) & \text{if } k = j \\ o(\Delta) & \text{otherwise} \end{cases},$$

where $o(\Delta)$ is defined such that $\lim_{\Delta\to 0} \frac{o(\Delta)}{\Delta} = 0$. It follows that

$$p_k(t+\Delta) = [1 - \lambda\Delta] p_k(t) + \lambda\Delta p_{k-1}(t) + o(\Delta) \text{ for } k \geq 1, \tag{6.1}$$

and

$$p_0(t+\Delta) = [1 - \lambda\Delta] p_0(t) + o(\Delta). \tag{6.2}$$

From equation(6.2) it follows that

$$\lim_{\Delta\to 0} \frac{p_0(t+\Delta) - p_0(t)}{\Delta} = -\lambda p_0(t),$$

implying

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t),$$

and in turn

$$p_0(t) = e^{-\lambda t}.$$

Likewise, from equation (6.1) we would like to solve the differential equations:

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t), \ k = 1, 2, 3, \ldots$$

Note that

$$\frac{dp_1(t)}{dt} = -\lambda p_1(t) + \lambda p_0(t) = -\lambda p_1(t) + \lambda e^{-\lambda t},$$

from which one can show

$$p_1(t) = e^{-\lambda t}\lambda t.$$

Via induction, it follows that

$$p_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \ k = 0, 1, 2, \ldots$$

We now solve for the pgf of $X(t)$,

$$G_X(s,t) = \sum_{k=0}^{\infty} p_k(t) s^k.$$

Taking derivative with respect to time t,

$$\begin{aligned}
\tfrac{\partial G_x(s,t)}{\partial t} &= \textstyle\sum_{k=0}^{\infty} \frac{dp_k(t)}{dt} s^k \\
&= \textstyle\sum_{k=0}^{\infty} -\lambda p_k(t) s^k + \sum_{k=1}^{\infty} \lambda p_{k-1}(t) s^k \\
&= -\lambda \textstyle\sum_{k=0}^{\infty} p_k(t) s^k + \lambda s \sum_{k=1}^{\infty} p_{k-1}(t) s^{k-1} \\
&= -\lambda G_X(s,t) + \lambda s G_X(s,t) \\
&= -\lambda(1-s) G_X(s,t),
\end{aligned}$$

with initial condition $G_X(s,0) = 1$. Integrating

$$\int \frac{\partial G_X(s,t)}{G_X(s,t)} = \int -\lambda(1-s)\partial t,$$

we arrive at

$$G_X(s,t) = e^{-\lambda(1-s)t},$$

which we recognize as the generating function of the Poisson distribution.

We may generalize the Poisson process by replacing λ by $\lambda(t)$, such that we do not have a time homogeneous Markov chain. In which case,

$$G_X(s,t) = \exp\left[-(1-s)\int_0^t \lambda(w)\,dw\right].$$

Often $\Lambda(t) = \int_0^t \lambda(w)\,dw$ is called the cumulative hazard function. In this case the following is easily shown:

$$p_k(t) = \frac{e^{-\Lambda(t)}\left[\Lambda(t)\right]^k}{k!},$$

and corresponding expectation

$$E\left[X(t)\right] = \Lambda(t).$$

6.2 Birth and death processes

A general birth and death process is defined as follows:

$$P\left[X(t+\Delta) = k \mid X(t) = j\right] = \begin{cases} \lambda_j\Delta + o(\Delta) & \text{if } k = j+1 \\ \mu_j\Delta + o(\Delta) & \text{if } k = j-1 \\ 1 - \lambda_j\Delta - \mu_j\Delta + o(\Delta) & \text{if } k = j \\ o(\Delta) & \text{otherwise} \end{cases}. \quad (6.3)$$

Example 6.2.1 *If we let $\lambda_j = \lambda$, $\mu_j = 0$ and $i = 0$, then we see that a Poisson process is a special case of a general birth and death process.*

Equation 6.3 gives rise to the following differential equations:

$$\frac{dp_j(t)}{dt} = -\left(\lambda_j + \mu_j\right)p_j(t) + \lambda_{j-1}p_{j-1}(t) + \mu_{j+1}p_{j+1}(t) \text{ for } j > 0, \quad (6.4)$$

$$\frac{dp_0(t)}{dt} = -\left(\lambda_0 + \mu_0\right)p_0(t) + \mu_1 p_1(t),$$

with initial conditions $p_i(0) = 1$.

To find the stationary distribution of a birth-death process, we take the limits of both sides of (6.4) to get

$$0 = -\left(\lambda_j + \mu_j\right)\pi_j + \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1},$$

with constraint

$$\sum_{j=0}^{\infty}\pi_j = 1,$$

which has a solution if and only if the process is ergodic.

6.2.1 Linear Birth Process

A pure linear birth process is a special case of (6.3) with $\mu_j = 0$ and $\lambda_j = j\lambda$ for $j = 0, 1, 2, \ldots$ which gives rise to

$$\frac{dp_j(t)}{dt} = -j\lambda p_j(t) + (j-1)\lambda p_{j-1}(t) \text{ for } j \geq 0.$$

We will attack this problem using pgf's. Let

$$G_X(s,t) = \sum_{j=0}^{\infty} p_j(t)\, s^j,$$

such that

$$\begin{aligned}
\tfrac{\partial G_X(s,t)}{\partial t} &= \textstyle\sum_{j=0}^{\infty} \tfrac{dp_j(t)}{dt} s^j \\
&= -\lambda \textstyle\sum_{j=0}^{\infty} j p_j(t)\, s^j + \lambda \sum_{j=1}^{\infty} (j-1)\, p_{j-1}(t)\, s^j \\
&= -\lambda s \textstyle\sum_{j=0}^{\infty} j p_j(t)\, s^{j-1} + \lambda s^2 \sum_{j=1}^{\infty} (j-1)\, p_{j-1}(t)\, s^{j-2} \\
&= -\lambda s \tfrac{\partial G_X(s,t)}{\partial s} + \lambda s^2 \tfrac{\partial G_X(s,t)}{\partial s}.
\end{aligned}$$

With the initial condition $G_X(s,0) = s^i$, it can be shown that

$$G_X(s,t) = s^i \left[\frac{e^{-\lambda t}}{1 - s\left(1 - e^{-\lambda t}\right)}\right]^i,$$

which suggest that $X(t)$ is the sum of a constant i and a negative binomial with parameters i and $e^{-\lambda t}$. That is,

$$p_j(t) = \binom{j-1}{j-i} e^{-\lambda t i} \left(1 - e^{-\lambda t}\right)^{j-i},\ j \geq i$$

$$E[X(t)] = ie^{\lambda t},$$
$$Var[X(t)] = ie^{\lambda t}\left(e^{\lambda t} - 1\right).$$

6.2.2 Linear Death Process

A pure linear death process is a special case of (6.3) with $\mu_j = j\mu$ and $\lambda_j = 0$ for $j = 0, 1, 2, \ldots$ That is,

$$P[X(t+\Delta) = k \mid X(t) = j] = \begin{cases} j\mu\Delta + o(\Delta) & k = j-1 \\ 1 - j\mu\Delta + o(\Delta) & k = j \\ o(\Delta) & \text{otherwise} \end{cases}.$$

It follows that

$$\frac{dp_j(t)}{dt} = -j\mu p_j(t) + (j+1)\mu p_{j+1}(t) \text{ for } j \geq 0$$

with initial condition $p_i(0) = 1$. It can be shown that

$$\frac{\partial G_X(s,t)}{\partial t} = \mu(1-s)\frac{\partial G_X(s,t)}{\partial s}.$$

Using the initial condition $G_X(s,0) = s^i$, it follows that

$$G_X(s,t) = \left[1 - e^{-\mu t} + e^{-\mu t}s\right]^i,$$

which we recognized as the pgf of a binomial. Thus,

$$p_{ij}(t) = \binom{i}{j}\left(e^{-\mu t}\right)^j\left(1-e^{-\mu t}\right)^{i-j},\ j = 0,1,\ldots,i,$$

$$E[X(t)] = ie^{-\mu t},$$

$$Var[X(t)] = ie^{-\mu t}\left(1-e^{-\mu t}\right).$$

If we replace μ by $\mu(t)$, then the cumulative hazard function becomes

$$M(t) = \int_0^t \mu(w)\,dw,$$

and

$$p_j(t) = \binom{i}{j}e^{-M(t)j}\left(1-e^{-M(t)}\right)^{i-j},\ j = 0,1,\ldots,i.$$

When $i = 1$, the probability of surviving at time t is $p_1(t)$, which is commonly known as the survival function

$$S(t) = e^{-M(t)}.$$

6.2.3 Linear Birth-Death Process

A linear birth-death process is also a special case of (6.3) with $\lambda_j = j\lambda$ and $\mu_j = j\mu$ such that

$$\frac{dp_j(t)}{dt} = -j(\mu+\lambda)p_j(t) + (j-1)\lambda p_{j-1}(t) + (j+1)\mu p_{j+1}(t) \text{ for } j \geq 1,$$

$$\frac{dp_0(t)}{dt} = \mu p_1(t),$$

which give rise to

$$E[X(t)] = ie^{(\lambda-\mu)t}. \tag{6.5}$$

Note the underlying deterministic process $\frac{dX(t)}{X(t)} = (\lambda - \mu)\, dt$ yields $X(t) = ie^{(\lambda-\mu)t}$. Taking the limit as $t \to \infty$ of equation 6.5, we have

$$\lim_{t\to\infty} E[X(t)] = \begin{cases} \infty & \lambda > \mu \\ i & \lambda = \mu \\ 0 & \lambda < \mu \end{cases}.$$

Further it can be shown that

$$Var(X(t)) = \begin{cases} i\left(\frac{\lambda+\mu}{\lambda-\mu}\right) e^{(\lambda-\mu)t}\left[e^{(\lambda-\mu)t} - 1\right] & \text{if } \lambda \neq \mu \\ 2i\lambda t & \text{if } \lambda = \mu \end{cases}. \tag{6.6}$$

Taking the limit as $t \to \infty$ of equation 6.6, we have

$$\lim_{t\to\infty} Var(X(t)) = \begin{cases} \infty & \lambda \geq \mu \\ 0 & \lambda < \mu \end{cases}.$$

Suppose we let $\theta = \frac{\lambda}{\mu}$ be the average number of offspring of an individual per lifetime. Then if $\theta \leq 1$, the extinction probability is one, whereas if $\theta > 1$, then the probability of extinction is $\left(\frac{1}{\theta}\right)^i$.

Example 6.2.2 *Consider an S-I-R epidemic where $X(t)$, $Y(t)$, and $Z(t)$ are the number of susceptible, infected and removed individuals in the population at time t, respectively. Let $X(t) + Y(t) + Z(t) = n + i$, $X(0) = n$, $Y(0) = i$, $Z(0) = 0$. Since population size is fixed, this is a 2-dimensional stochastic process.*

$$P[X(t+\Delta) = x - 1,\ Y(t+\Delta) = y + 1 \mid X(t) = x, Y(t) = y] = \lambda X(t) Y(t) \Delta + o(\Delta)$$

$$P[\ Y(t+\Delta) = y - 1, Z(t+\Delta) = z + 1 \mid Y(t) = y,\ Z(t) = z] = \gamma Y(t) \Delta + o(\Delta)$$

where $\frac{1}{\gamma}$ is the average length of the infectious period. This system is intractable, but at the onset of the epidemic, when $X(t)$ is large and $Y(t)$ is small, we approximate the system by the linear birth and death theory developed earlier. Namely, let $\theta = \frac{\lambda}{\gamma}$, and again, if $\theta \leq 1$, the extinction probability is one, whereas if $\theta > 1$, then the probability of extinction is $\left(\frac{1}{\theta}\right)^i$.

6.3 Kolmogorov differential equations

For any continuous time Markov process, the Kolmogorov differential equations are given by

$$\frac{dp_k(t)}{dt} = \sum_{j\in S} p_j(t)\, v_{jk}, \tag{6.7}$$

with the restriction that $\sum_{k\in S} v_{jk} = 0$ and $v_{jk} > 0$ whenever $j \neq k$. Thus it follows immediately that $v_{jj} = -\sum_{\substack{k\in S \\ k\neq j}} v_{jk}$. Therefore, we can write

$$P\left[X\left(t+\Delta\right)=k \mid X\left(t\right)=j\right] = \begin{cases} v_{jk}\Delta + o\left(\Delta\right) & \text{if } k \neq j \\ 1 - \sum_{j\neq k} v_{jk}\Delta + o\left(\Delta\right) & \text{if } k = j \\ o\left(\Delta\right) & \text{otherwise} \end{cases}$$

Example 6.3.1 *In the linear birth-death process,*

$$v_{jk} = \begin{cases} j\lambda & k = j+1 \\ j\mu & k = j-1 \\ -j\left(\lambda+\mu\right) & k = j \\ 0 & \textit{otherwise} \end{cases}$$

Definition 6.3.2 *The infinitesimal generator* $\mathbf{V}$ *is given by*

$$\mathbf{V} = \begin{bmatrix} v_{00} & v_{01} & \cdots \\ v_{10} & v_{11} & \\ \vdots & & \ddots \end{bmatrix}$$

Example 6.3.3 *The infinitesimal generator of the linear birth-death process is*

$$\mathbf{V} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ \mu & -\left(\lambda+\mu\right) & \lambda & 0 & \\ 0 & 2\mu & -2\left(\lambda+\mu\right) & 2\lambda & \\ 0 & 0 & & \ddots & \\ \vdots & & & & \end{bmatrix}.$$

Example 6.3.4 *A time homogeneous Poisson process has infinitesimal generator*

$$\mathbf{V} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \\ 0 & 0 & -\lambda & \lambda & \\ \vdots & & & \ddots & \end{bmatrix}.$$

6.4 Algebraic Treatment

Now suppose the state space is $S=\{1,2,\ldots,s\}$ and let $\mathbf{P}(t)$ be a $s \times s$ matrix with ij^{th} element $p_{ij}(t)$. Similarly, let $\frac{d\mathbf{P}(t)}{dt}$ be a $s \times s$ matrix with ij^{th} element $\frac{dp_{ij}(t)}{dt}$. Note that $\mathbf{P}(0) = \mathbf{I}$, the identity matrix. The task at hand becomes solving

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t)\mathbf{V},$$

which is the matrix equivalent to (6.7). We motivate the solution by first considering the one dimensional analog:

$$\frac{dp(t)}{dt} = vp(t),\ p(0) = 1,$$

which has solution

$$p(t) = e^{vt} = 1 + vt + \frac{(vt)^2}{2!} + \cdots$$

This suggests that we let $\mathbf{P}(t)$ have the following expansion

$$\begin{aligned} \mathbf{P}(t) &= e^{\mathbf{V}t} \\ &= \mathbf{I} + \mathbf{V}t + \tfrac{\mathbf{V}^2t^2}{2!} + \cdots + \tfrac{\mathbf{V}^kt^k}{k!} + \cdots \\ &= \textstyle\sum_{n=0}^{\infty} \frac{\mathbf{V}^nt^n}{n!} \end{aligned} \tag{6.8}$$

which can be shown to converge uniformly in t provided that $|v_{ij}| \leq M$ for all i, j. Taking the derivative with respect to t of $\mathbf{P}(t)$, we have

$$\frac{d\mathbf{P}(t)}{dt} = \sum_{n=1}^{\infty} \frac{\mathbf{V}^{n-1}t^{n-1}}{(n-1)!}\mathbf{V} = \mathbf{P}(t)\mathbf{V}.$$

We now take advantage of the expansion 6.8 to get a closed form $\mathbf{P}(t)$. First, consider the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\mathbf{V}$ such that

$$\mathbf{V}^n = \mathbf{T}\mathbf{\Lambda}^n\mathbf{T}^{-1} \text{ where } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_s \end{bmatrix},$$

where $\mathbf{T}$ is a matrix of eigenvalues. Note that we assume that all the eigenvalues are

distinct so that the inverse of $\mathbf{T}$ is known to exist. It follows that

$$\begin{aligned}\mathbf{P}(t) = e^{\mathbf{V}t} &= \textstyle\sum_{n=0}^{\infty} \frac{\mathbf{V}^n t^n}{n!} \\ &= \mathbf{T}\left(\textstyle\sum_{n=0}^{\infty} \frac{\mathbf{\Lambda}^n t^n}{n!}\right)\mathbf{T}^{-1} \\ &= \mathbf{T}\begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_s t} \end{bmatrix}\mathbf{T}^{-1} \\ &= \mathbf{T}e^{\mathbf{\Lambda}(t)}\mathbf{T}^{-1},\end{aligned}$$

where

$$e^{\mathbf{\Lambda}(t)} = \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_s t} \end{bmatrix}.$$

It can be shown that the form $\mathbf{P}(t) = \mathbf{T}e^{\mathbf{\Lambda}(t)}\mathbf{T}^{-1}$ gives rise to the following explicit solution

$$P_{ij}(t) = \sum_{l=1}^{s} \frac{A_{ji}(\lambda_l)\, e^{\lambda_l t}}{\prod_{\substack{m=1 \\ m\neq l}} (\lambda_l - \lambda_m)}, \tag{6.9}$$

the continuous-time analog of equation 5.7. Further, it can be shown that the limiting transition probabilities are given by

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j = \frac{v_{jj}}{\sum_l v_{ll}},$$

so that $\mathbf{0} = \mathbf{\Pi V}$ where $\mathbf{\Pi} = [\pi]$.

Example 6.4.1 *Consider the two stage continuous time Markov process with infinitesimal generator*

$$\mathbf{V} = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

Then the characteristic matrix

$$\mathbf{A}(\rho) = \begin{bmatrix} \rho + \lambda & -\lambda \\ -\mu & \rho + \mu \end{bmatrix},$$

gives rise to eigenvalues $\rho_1 = 0$ *and* $\rho_2 = -(\lambda + \mu)$. *Then*

$$\mathbf{A}(0) = \begin{bmatrix} \lambda & -\lambda \\ -\mu & \mu \end{bmatrix},\ \mathbf{A}^{+}(0) = \begin{bmatrix} \mu & \mu \\ \lambda & \lambda \end{bmatrix},$$

implying $[\mu, \lambda]^T$ *is an eigenvector of the eigenvalue* $\rho_1 = 0$. *Similarly,*

$$\mathbf{A}\left(-\left(\lambda+\mu\right)\right) = \begin{bmatrix} -\mu & -\lambda \\ -\mu & -\lambda \end{bmatrix}, \ \mathbf{A}^{+}\left(-\left(\lambda+\mu\right)\right) = \begin{bmatrix} -\lambda & \mu \\ \lambda & -\mu \end{bmatrix},$$

yields $[-1, 1]^T$ *as an eigenvector for* $\rho_2 = -\left(\lambda+\mu\right)$. *Thus*

$$\mathbf{T} = \begin{bmatrix} \mu & -1 \\ \lambda & 1 \end{bmatrix} \text{ and } \mathbf{T}^{-1} = \left(\frac{1}{\lambda+\mu}\right) \begin{bmatrix} 1 & 1 \\ -\lambda & \mu \end{bmatrix},$$

and

$$\begin{aligned} \mathbf{P}\left(t\right) &= \mathbf{T} \begin{bmatrix} 1 & 0 \\ 0 & e^{-(\lambda+\mu)} \end{bmatrix} \mathbf{T}^{-1} \\ &= \left(\tfrac{1}{\lambda+\mu}\right) \begin{bmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda\left(1 - e^{-(\lambda+\mu)t}\right) \\ \mu\left(1 - e^{-(\lambda+\mu)t}\right) & \lambda + \mu e^{-(\lambda+\mu)t} \end{bmatrix}, \end{aligned}$$

implying $\lim_{t\to\infty} \mathbf{P}\left(t\right) = \left(\tfrac{1}{\lambda+\mu}\right) \begin{bmatrix} \mu & \lambda \\ \mu & \lambda \end{bmatrix}$.

6.5 Mean time to absorption

We wish to investigate mean time to absorption in continuous-time Markov chains. Again we motivate with the 1-dimensional case. Let T be the random variable for time to absorption with exponential density $f\left(t\right) = \lambda e^{-\lambda t}$ and corresponding moment generating function

$$m\left(s\right) = \lambda \int_0^\infty e^{-(\lambda-s)t} dt = \frac{\lambda}{\lambda - s} = \lambda\left(\lambda - s\right)^{-1}.$$

The r^{th} moment of T is then found by evaluating

$$\left.\frac{d^r m\left(s\right)}{ds^r}\right|_{s=0} = r!\left(\lambda\right)^{-r}.$$

It follows that the mean time of absorption is

$$E\left(T\right) = \lambda^{-1}.$$

Now suppose there are s transient states and 1 super absorbing state such that the infinitesimal generator $\mathbf{V}$ is a $(s+1) \times (s+1)$ matrix which can be partitioned as follows

$$\mathbf{V} = \begin{bmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & 0 \end{bmatrix},$$

where $\mathbf{Q}$ is the $s \times s$ infinitesimal generator among the transient states. Define a $s \times 1$ vector $\mathbf{f}\left(t\right) = \left[f_1\left(t\right), f_2\left(t\right), \ldots, f_s\left(t\right)\right]^T$ where $f_i\left(s\right)$ is the time to absorption density

from state i such that $\mathbf{f}(t) = e^{\mathbf{Q}t}\mathbf{R}$. Then matrix moment generating function is given by

$$\mathbf{M}(s) = \int_0^\infty e^{st}\mathbf{f}(t)\,dt = \int_0^\infty e^{st}e^{\mathbf{Q}t}dt\ \mathbf{R} = -(\mathbf{Q}+s\mathbf{I})^{-1}\mathbf{R},$$

from which we can obtain the r^{th} moment

$$\mathbf{M}_r = (-1)^r\, r!\mathbf{Q}^{-1}\mathbf{C}$$

$$\mathbf{M}_1 = -\mathbf{Q}^{-1}\mathbf{C}$$

Example 6.5.1 *Suppose we have $s = 2$ transient states and 1 absorbing state and the infinitesimal generator is given by*

$$\mathbf{V} = \begin{bmatrix} -(\lambda_1+\mu) & \lambda_1 & \mu \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then

$$\mathbf{Q}^{-1} = \frac{1}{(\lambda_1+\mu)\lambda_2}\begin{bmatrix} -\lambda_2 & -\lambda_1 \\ 0 & -(\lambda_1+\mu) \end{bmatrix}$$

implying that the mean time to absorption is

$$\mathbf{M}_1 = \begin{bmatrix} \left(\frac{1}{\lambda_1+\mu}\right)\left(1+\frac{\lambda_1}{\lambda_2}\right) \\ \frac{1}{\lambda_2} \end{bmatrix}.$$

Next we investigate the probability of absorption given starting at state i. Our derivation involves embedded Markov chains.

Definition 6.5.2 *A stochastic process $\{X(t)\}$ is right continuous at t if for every $\epsilon > 0$ there exists s such that $X(s) = X(t)$ where $t \le s \le t+\epsilon$.*

Next, suppose we are in state i, at some time T_i such that

$$P_{ii}(\Delta) = 1 + v_{ii}\Delta + o(\Delta)$$

$$\begin{aligned} S_i(t) &= P[T_i > t] \\ &= \lim_{\substack{\Delta\to 0 \\ n\to\infty}} [p_{ii}(\Delta)]^n \\ &= \lim_{\substack{\Delta\to 0 \\ n\to\infty}} \left[1 + v_{ii}\left(\tfrac{t}{n}\right) + o(\Delta)\right]^n \\ &= e^{v_{ii}t} \end{aligned}$$

where $\Delta = \frac{t}{n}$. Then the probability that when a jump is made, we go to state j, is given by:

$$\frac{v_{ij}\Delta + o(\Delta)}{-v_{ii}\Delta + o(\Delta)}$$

where $-v_{ii} = \sum_{j \neq i} v_{ij}$. We then let

$$p_{ij} = \lim_{\Delta \to \infty} \frac{v_{ij} + \frac{o(\Delta)}{\Delta}}{-v_{ii} + \frac{o(\Delta)}{\Delta}} = \frac{v_{ij}}{-v_{ii}}$$

Thus, examining a continuous Markov process only at transition times, we observe the discrete, embedded process defined by $\mathbf{P} = \{p_{ij}\}$.

Example 6.5.3 *Let $\{X(t)\}$ be a continuous birth-death process. Then $\{Y_n\}$, the discrete embedded process, is defined by*

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & \\ 0 & q_2 & 0 & p_2 & \\ \vdots & & & \ddots & \\ & & & & \end{bmatrix}$$

where $p_k = p_{k,k+1} = \frac{\lambda_k}{\lambda_k + \mu_k}$ and $q_k = p_{k,k-1} = \frac{\mu_k}{\lambda_k + \mu_k}$. If we let ω_i be the probability of absorption given starting in state i, then

$$\omega_i = \frac{\lambda_i}{\lambda_i + \mu_i}\omega_{i+1} + \frac{\mu_i}{\lambda_i + \mu_i}\omega_{i-1}, \ w_0 = 1$$

Furthermore, if we assume that we have a linear birth-death process i.e. $p_k = \frac{\lambda}{\lambda+\mu}$ and $q_k = \frac{\mu}{\lambda+\mu}$, then the embedded process is very similar to gambler's ruin scenario. We have

$$\omega_i = \frac{\lambda}{\lambda + \mu}\omega_{i+1} + \frac{\mu}{\lambda + \mu}\omega_{i-1}, \ w_0 = 1$$

Letting $\phi = \frac{\mu}{\lambda}$,

$$\omega_{i+2} - (1 + \phi)\omega_{i+1} + \phi\omega_i = 0$$

which gives rise to

$$\omega_i = \begin{cases} \left(\frac{\mu}{\lambda}\right)^i & \text{if } \lambda > \mu \\ \\ 1 & \text{if } \lambda \leq \mu \end{cases}$$

Now if m_i is the mean time to absorption from state i and p_{ij} is the one step transition probability from the embedded Markov process, then

$$m_i = \sum_{j=1}^{S} p_{ij}m_j - v_{ii}^{-1}, \ m_0 = 0$$

(where we assume that the state space is finite i.e. $S = \{1, 2, \ldots, s\}$.

Example 6.5.4 *Revisiting the general birth-death process*

$$m_i = \frac{\lambda_i}{\lambda_i + \mu_i} m_{i+1} + \frac{\mu_i}{\lambda_i + \mu_i} m_{i-1} + \frac{1}{\lambda_i + \mu_i}$$

Now let $\boldsymbol{\theta} = \left[-v_{ii}^{-1}, \ldots, -v_{ss}^{-1}\right]$, $\mathbf{M} = [m_1, m_2, \ldots, m_s]$ and $\mathbf{Q}$ be the embedded process one step transition probability matrix among transient states. Then

$$\mathbf{IM} = \boldsymbol{\Theta} + \mathbf{QM}$$

implying

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} \boldsymbol{\Theta}$$

6.6 Inference

In this section, we will introduce likelihood based inference on discrete state space and continuous index set, homogeneous Markov chains. Generally one can perform inference on a single observed sequence over time or on many observed sequences over time.

6.6.1 Inference on a single sequence

Assume that we plan to estimate parameter, $\boldsymbol{\theta}$, that have dimension less that the number of states. Let $\{X_{(1)}, X_{(2)}, \ldots X_{(k)}\}$ be the observed successive states that the system passes through with $X_{(k)} \neq X_{(k+1)}$, where the () indicates order statistics. We let T_k be the sojourn time in state k. We know have a bivariate discrete index set process $\{(X_{(k)}, T_k), k = 1, 2, \ldots\}$, which is a Markov process on $s \times [0, \infty)$. The transition probabilities for this process are

$$P\left[X_{(k+1)} = j, T_{k+1} > t \mid X_{(k)} = i,\ T_k = u\right] = \frac{v_{ij}}{-v_{ii}} e^{v_{jj}t},$$

with densities

$$\lim_{\Delta \to 0^+} \left[\frac{P\left(X_{(k+1)} = j, t < T_{k+1} \leq t + \Delta \mid X_{(k)} = i,\ T_k = u\right)}{\Delta} \right] = \frac{v_{ij}}{v_{ii}} v_{jj} e^{v_{jj}t},$$

$$\lim_{\Delta \to 0^+} \left[\frac{P\left(X_{(k+1)} = j, t \leq T_{k+1} < t + \Delta\right)}{\Delta} \right] = v_{jj} e^{v_{jj}t},$$

and survival functions

$$P\left[X_{(k)} = j,\ T_k > t\right] = e^{v_{jj}t}.$$

Example 6.6.1 *Suppose observe the following:* $(1, t_1), (3, t_2), (2, t_3), (3, t_4), (2, t_5^*)$. *Then the likelihood is given by*

$$\begin{aligned} L(\Theta) &= c\left[v_{11}\left(\tfrac{v_{13}}{v_{11}}\right)e^{v_{11}t_1}\right]\left[v_{33}\left(\tfrac{v_{32}}{v_{22}}\right)e^{v_{33}t_2}\right]\left[v_{22}\left(\tfrac{v_{23}}{v_{33}}\right)e^{v_{22}t_3}\right]\left[v_{33}\left(\tfrac{v_{32}}{v_{22}}\right)e^{v_{33}t_4}\right]\left[e^{v_{22}t_5}\right] \\ &= c v_{13} v_{32} v_{23} v_{32} e^{v_{11}t_1} e^{v_{22}(t_3+t_5)} e^{v_{33}(t_2+t_4)}. \end{aligned}$$

In general, let n_{ij} be the observed number of transitions from state i to j, and β_i be the total time spend in state i, where $\Sigma_{i\in S}\beta_i = t$. Then the likelihood function for $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = c \prod_{(i,j)\in S,\ i\neq i} \exp\left(\Sigma_{i\in S} v_{ij}\beta_i\right).$$

If the Markov process is ergodic, then

$$\begin{aligned} E(\frac{\beta_i}{t}) &\longrightarrow {}^p\pi_i, \text{ as } t \longrightarrow \infty, \\ E(\frac{n_{ij}}{t}) &\longrightarrow {}^p\pi_i v_{ij}, \text{ as } t \longrightarrow \infty. \end{aligned}$$

6.6.2 Inference on birth and death processes

We define the elements of infinitesimal generator as

$$v_{ij} = \begin{cases} \lambda_i, & j = i+1 \\ \mu_i, & j = i-1 \\ -(\lambda_i + \mu_i), & j = i \\ 0, & \text{otherwise} \end{cases}, i = 0, 1, ...,$$

where $\mu_0 = 1$. The observations take the form of $b_i = n_{i,i+1}$, the total number of births observed from state i; $d_i = n_{i,i-1}$, the total number of deaths observed from state i; and β_i total time spent in state i. Then the likelihood function for $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, ...)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, ...)$ is

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \prod_{i=0}^{\infty} \lambda_i^{b_i} \prod_{i=0}^{\infty} \mu_i^{d_i} \exp\left\{ -\sum_{i=1}^{\infty} (\lambda_i + \mu_i)\beta_i - \lambda_0\beta_0 \right\},$$

The score functions are

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda_i} &= \frac{b_i}{\lambda_i} - \beta_i, \\ \frac{\partial \ln L}{\partial \mu_i} &= \frac{d_i}{\mu_i} - \beta_i, \end{aligned} \quad , \tag{6.10}$$

resulting in the MLE's $\widehat{\lambda}_i = \frac{b_i}{\beta_i}, i = 0, 1, ...$ and $\widehat{\mu}_i = \frac{d_i}{\beta_i}, i = 1, 2, ...$. The second partials are

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \lambda_i^2} &= -\frac{b_i}{\lambda_i^2}, \\ \frac{\partial^2 \ln L}{\partial \mu_i^2} &= -\frac{d_i}{\mu_i^2}, \\ \frac{\partial^2 \ln L}{\partial \lambda_i \partial \mu_i} &= \frac{\partial^2 \ln L}{\partial \mu_i \partial \lambda_i} = 0. \end{aligned} \tag{6.11a}$$

Example 6.6.2 *Simple Immigration-Emigration Process: In this case $\lambda_k = \lambda$ and $\mu_k = k\mu$, $k = 0, 1, ...$. Using (6.10), the MLE's are $\widehat{\lambda} = \frac{b}{t}$ and $\widehat{\mu} = \frac{d}{\Sigma_{i=1}^{\infty} i\beta_i}$. Note that $\Sigma_{i=1}^{\infty} i\beta_i$ is the total person time under observation in all the non-zero states. From the expected values of (6.11a), the variances of the estimators based on Fisher's information are $var(\widehat{\lambda}) = \frac{\lambda}{t}$ and $var(\widehat{\mu}) = \frac{\mu}{\theta t}$, where $\theta = \frac{\lambda}{\mu}$. The form of the $var(\widehat{\lambda})$ comes from the fact that the births part of the process is a pure birth process. The form of the $var(\widehat{\mu})$ comes from the fact that we have an ergodic process, where θ is the average number of people in the system at equilibrium (see exercise 6.8.3).*

6.7 HIV-progression models

Longini, *et al.*[18] modelled the progression of an infected individual through the stages of infection and ultimately death as a time-homogeneous Markov process in which stages 1 to 4 are transient states, and stage 5 is an absorbing state. Thus we have the following infinitesimal generator

$$\mathbf{V} = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & 0 \\ 0 & 0 & 0 & -\lambda_4 & \lambda_4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with eigenvalues $\rho_1 = -\lambda_1,\ \rho_2 = -\lambda_2, \rho_3 = -\lambda_3, \rho_4 = -\lambda_4,\ \rho_5 = 0$. It follows from equation (6.9) that the transition probabilities among the transient states are given by

$$p_{ik}(t) = (-1)^{k-i} \lambda_i \dots \lambda_{k-1} \sum_{j=i}^{k} \frac{e^{-\lambda_j t}}{\Pi_{\substack{l=i \\ l \neq j}}^{k} (\lambda_j - \lambda_l)},\ i = 1, 2, 3, 4;\ i \leq\ k < 5.$$

The transition probabilities from a transient state i to the absorbing state (death) are

$$p_{i5}(t) = (-1)^{4-i} \lambda_i \dots \lambda_4 \sum_{j=i}^{4} \frac{\left(1 - e^{-\lambda_j t}\right)}{\lambda_j \Pi_{\substack{l=i \\ l \neq j}}^{k} (\lambda_j - \lambda_l)},\ i = 1, 2, 3, 4.$$

See Longini, *et al.* [20][19][17][21] for further use of continuous time Markov models for HIV progression and prediction.

6.8 Exercises

Exercise 6.8.1 *Pure death with annihilation process. For a time homogeneous pure death process with death rate $\mu_k = k\mu, k = 0, 1, \dots$, suppose that the entire population may be subject independently at time t to sudden annihilation. Let $v\Delta + o(\Delta)$ be the probability that a population, of any size, will be annihilated in the time interval $[t, t + \Delta)$. Also, let $p_i(0) = 1,\ i > 0$.*

a. Construct the state diagram for the process.

b. Classify the states and the Markov process.

c. Write out the system of differential-difference equations for $p_k(t)$, $k = 0, 1, \dots$

d. Solve the system for $p_k(t)$ (Hint: Use induction)

e. Find the survival distribution, $S_T(t) = 1 - F_T(t)$, for the population.

Note: An interesting application of this problem is as follows: A convoy of i ships is carrying supplies to a given destination in time of war. If convoy protection

is ineffective, enemy attacks on the convoy may cause attrition or even total annihilation. Analysis of WWII data has shown that effective convoy protection would result in an attrition rate which is independent of the size of the convoy at a given time.

Exercise 6.8.2 *Consider a time-dependent pure death process where* $\mu_k(t) = k\mu(t)$, $k = 0, 1, \ldots, i$. *Also define* $p_{0l} = e^{-\int_0^{t_l} \mu(\tau)d\tau}$ *and* $p_l = e^{-\int_{t_l}^{t_{l+1}} \mu(\tau)d\tau}$.

a. Give a physical interpretation of p_{0i} *and* p_i *and show that* $p_{0l+1} = p_{0l}p_l$.

b. Consider a sequence of random variables $X(t_1), \ldots, X(t_\omega)$ *defined at the points* $0 \leq t_1 < t_2 < \cdots < t_\omega$. *Derive the joint pmf for* $X(t_1) = k_1, \ldots, X(t_\omega) = k_\omega$. *(Assume* $P[X(t_0) = k_0] = 1$*).*

c. Now suppose that $X(t_1) = k_1, \ldots, X(t_\omega) = k_\omega$ *is a set of observations made at points* $0 \leq t_1 < t_2 < \cdots < t_\omega$ *with* $k_0 \geq k_1 \geq \cdots \geq k_\omega$. *Give the likelihood function.*

d. Find the maximum likelihood estimators $\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_{\omega-1}$ *of the probabilities* $p_0, p_1, \ldots, p_{\omega-1}$. *(Assume* $k_{\omega-1} > 0$*).*

e. Give the asymptotic variances for the mle's $\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_{\omega-1}$; *also give the asymptotic variance-covariance matrix (asymptotic).*

Exercise 6.8.3 *Consider a special case of the general birth-death process where* $\lambda_k(t) = \eta$, $\mu_k(t) = k\mu$, $k = 0, 1, \ldots$ *given in Chiang pages 229-233.*

a. Construct the state diagram for the process.

b. Classify the states and identify the type of Markov chain.

c. Solve the P.D.E.(6.9) and verify that (6.17) is correct.

d. Find the stationary distribution $\{\pi_0, \pi_1, \ldots\}$ *if one exists.*

Chapter 7
COUNTING PROCESSES

Let $N(t)$ be the cumulative number of events at time t. Consider $\{N(t)\}_{t=0}^{\infty}$ where $S = \{0, 1, 2, \ldots\}$ and $N(0) = 0$. Let F_t be the history up to t of the process generated by the $N(t)$ so that F_t is a right continuous σ-field. Note for $\tau < t$, $F_\tau \subset F_t$. Now

$$P[dN(t) = 1|F_t] = A(t)\,dt$$
$$P[dN(t) = 0|F_t] = 1 - A(t)\,dt$$

where $A(t)$ is the intensity of the process.

Example 7.0.1 *If $A(t) = \lambda(t)$ (i.e. does not depend on F_t), then we have an ordinary Poisson process:*

$$P[dN(t) = 1|N(t)] = \lambda(t)\,dt$$
$$P[dN(t) = 0|N(t)] = 1 - \lambda(t)\,dt$$

Example 7.0.2 *S-I-R epidemic. Let $N(t) = S(0) - S(t)$, $F_t = \sigma(S(\tau), I(\tau) : 0 \leq \tau \leq t)$, $A(t) = \beta S(t) I(t)$, so that:*

$$P[dN(t) = 1|N(t), I(t)] = \beta S(t) I(t)\,dt$$
$$P[dN(t) = 0|N(t), I(t)] = 1 - \beta S(t) I(t)\,dt$$

Example 7.0.3 *Linear death process. Let $X(t) = X(0) - N(t)$ and $A(t) = X(t)\lambda(t)$.*

7.1 Continuous Time Martingales

We would like to use Martingale machinery. First note that

$$E[dN(t)|\mathcal{F}_t] = A(t)\,dt$$

and

$$E[N(t)|\mathcal{F}_t] = \int_0^t A(\tau)\,d\tau$$

which is called the compensator. Similarly one can show that

$$Var[dN(t)|\mathcal{F}_t] = A(t)\,dt\,[1 - A(t)\,dt] \cong A(t)\,dt$$

and

$$Var\left[N\left(t\right)|\mathcal{F}_t\right] = \int_0^t A\left(\tau\right) d\tau.$$

Define $M\left(t\right) = N\left(t\right) - \int_0^t A\left(\tau\right) d\tau$, $M\left(0\right) = 0$. Then it follows that

$$E\left[dM\left(t\right)|\mathcal{F}_t\right] = E\left[dN\left(t\right)|\mathcal{F}_t\right] - A\left(t\right) dt = 0$$

and it can also be shown that

$$E\left[M\left(t\right)|\mathcal{F}_t\right] = 0$$

and that

$$Var\left[M\left(t\right)\right] \cong E\left[\int_0^t A\left(\tau\right) d\tau\right]$$

Definition 7.1.1 *$\{M\left(t\right)\}_{t=0}^{\infty}$ is a martingale with respect to $\mathcal{F}_t$ if*

$$i.\ E\left[\left|M\left(t\right)\right|\right] < \infty$$
$$ii.\ E\left[M\left(\tau\right)|\mathcal{F}_t\right] = M\left(t\right) \text{ for all } \tau \geq t$$

Furthermore, if $E\left[M\left(t\right)\right] = M\left(0\right)$, when $M\left(0\right) = 0$, then $M\left(t\right)$ is a Zero Mean Martingale (ZMM).

Theorem 7.1.2 *Martingale Convergence Theorem (MCT): $M\left(t\right) \rightarrow^{a.s.} M$*

Theorem 7.1.3 *Zero Mean Martingale Central Limit Theorem (ZMMCLT): as $t \rightarrow \infty$, $\frac{M(t)}{\sqrt{Var(M(t))}} \rightarrow^{d} N\left(0,1\right)$*

Example 7.1.4 *Time homogeneous Poisson process. We have $M\left(t\right) = N\left(t\right) - \lambda t$ and $Var\left[M\left(t\right)\right] = E\left[N\left(t\right)\right] = \lambda t$. It follows that $\left[N\left(t\right) - \lambda t\right] \rightarrow^{a.s.} 0$, or equivalently*

$$\left[\frac{N\left(t\right)}{t} - \lambda\right] \rightarrow^{a.s.} 0.$$

Therefore

$$\hat{\lambda}\left(t\right) = \frac{N\left(t\right)}{t} \text{ and } \hat{\lambda} \rightarrow^{a.s.} \lambda.$$

From the ZMMCLT,

$$\frac{M\left(t\right)}{\sqrt{\lambda t}} = \frac{N\left(t\right) - \lambda t}{\sqrt{\lambda t}} = \frac{\hat{\lambda}\left(t\right) - \lambda}{\sqrt{\frac{\lambda}{t}}} \sim N\left(0,1\right)$$

and thus

$$Var\left(\hat{\lambda}\right) \cong \frac{\lambda}{t}$$

Definition 7.1.5 *A process* $\{B(t)\}_{t=0}^{\infty}$ *is predictable if determined by* $\mathcal{F}_t$*. Note:* $\{B(t)\}_{t=0}^{\infty}$ *is left continuous.*

Define

$$M^*(t) = \int_0^t B(\tau)\, dM(\tau) = \int_0^t B(\tau)\, dN(\tau) - \int_0^t B(\tau)\, A(\tau)\, d\tau$$

so that $dM^*(t) = B(t)\, dM(t)$. It follows that

$$\begin{aligned} E[dM^*(t)\,|\mathcal{F}_t] &= E[B(t)\, dM(t)\,|\mathcal{F}_{t-}] \\ &= B(t)\, E[dM(t)\,|\mathcal{F}_{t-}] = 0 \end{aligned}$$

so that both $dM^*(t)$ and $M^*(t)$ are ZMM. Finally, note that

$$Var[M^*(t)] = E\left[\int_0^t B(\tau)^2\, dN(\tau)\right] \tag{7.1}$$

Example 7.1.6 *Linear Pure Death Process. Start with a cohort of* n *individuals. Let* $X(t)$ *be the number alive at time* t *and* $N(t)$ *be the cumulative number of deaths by time* t*, so that* $X(t) = n - N(t)$*. Let* $A(t) = X(t)\lambda(t)$ *be the intensity of* $N(t)$*. We want to estimate* $\Lambda(t) = \int_0^t \lambda(\tau)\, d\tau$*, the cumulative hazard function, and in turn,* $s(t) = e^{-\Lambda(t)}$*. Let*

$$M(t) = N(t) - \int_0^t X(\tau)\lambda(\tau)\, d\tau$$

and

$$J(\tau) = \begin{cases} 1 & \text{if } X(\tau) > 0 \\ 0 & \text{if } X(\tau) = 0 \end{cases}.$$

Then $B(\tau) = \frac{J(\tau)}{X(\tau^-)}$ *is a predictable process. Define*

$$\begin{aligned} M^*(t) &= \int_0^t B(\tau)\, dM(\tau) = \int_0^t \frac{1}{X(\tau^-)} dN(\tau) - \int_0^t \lambda(\tau)\, d\tau \text{ if } X(t) > 0 \\ &= \int_0^t \frac{1}{X(\tau^-)} dN(\tau) - \Lambda(t) \end{aligned}$$

Since $M^*(t)$ *is a ZMM, it follows that*

$$\hat{\Lambda}(t) = \int_0^t \frac{1}{X(\tau^-)} dN(\tau)$$

Nelson-Aalen estimator:

$$\begin{aligned} \hat{\Lambda}(t) &= \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{n - N(t) + 1} \\ &\cong \ln\left(\frac{n}{n - N(t)}\right) = \ln\left(\frac{n}{X(t)}\right) \end{aligned}$$

so that

$$-\hat{\Lambda}(t) = \ln\left(\frac{X(t)}{n}\right) = \ln\left(\hat{S}(t)\right),\ \hat{S}(t) = e^{-\hat{\Lambda}(t)}$$

From (7.1), the variance of $M^*(t)$ *is*

$$\begin{aligned} Var[M^*(t)] &= E\left[\frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{[n-N(t)+1]^2}\right] \\ &\cong E\left[\frac{1}{n-N(t)} - \frac{1}{n}\right] \\ &= E\left[\frac{1}{X(t)} - \frac{1}{n}\right] = E\left[\frac{1-\frac{X(t)}{n}}{X(t)}\right] \end{aligned} \tag{7.2}$$

Since $S(t) = 1 - \frac{X(t)}{n}$, *then*

$$Var\left[M^*(t)\right] \cong \frac{1-\hat{S}(t)}{X(t)}$$

and by ZMMCLT

$$\frac{\hat{\Lambda}(t) - \Lambda(t)}{\sqrt{Var\left[M^*(t)\right]}} \tilde{} N(0,1)$$

Example 7.1.7 *Linear Birth-Death Process.*

$$P\left[dX(t) = 1|F_t\right] = \lambda X(t) + o(dt)$$

$$P\left[dX(t) = -1|F_t\right] = \mu X(t) + o(dt)$$

$$P\left[dX(t) = 0|F_t\right] = 1 - (\lambda+\mu)X(t) + o(dt)$$

Then $E\left[dX(t)|F_t\right] = (\lambda-\mu)X(t)\,dt$ *and*

$$E\left[X(t)|F_t\right] = X(0)\,e^{(\lambda-\mu)t}$$

which is similar to the underlying deterministic system:

$$\frac{dX(t)}{dt} = (\lambda-\mu)X(t)$$

$$X(t) = X(0)\,e^{(\lambda-\mu)t}$$

7.2 Inference on continuous-time epidemics

Recall the S-I-R stochastic process $\{S(t), I(t)\}$. Let $n = S(t) + I(t) + R(t)$, $\mathcal{F}_t = \sigma\{S(\tau), I(\tau) : 0 \leq \tau \leq t\}$, $\mathcal{F}_0 = \sigma\{n, 1\}$ and

$$\begin{aligned} P[dS(t) = -1,\ dI(t) = +1|\mathcal{F}_t] &= \lambda I(t) S(t) dt + o(dt) \\ P[dS(t) = 0,\ dI(t) = -1|\mathcal{F}_t] &= \gamma I(t) dt + o(dt) \\ P[dS(t) = 0,\ dI(t) = 0|\mathcal{F}_t] &= 1 - [\lambda S(t) + \gamma] I(t) dt + o(dt) \end{aligned}$$

with $\lambda = \frac{c\beta}{n}$ where c is the number of contacts per unit for an individual and β is the transmission probability.

Counting process. Let $N(t) = n - S(t)$ be the cumulative number infected and $R(t)$ be the cumulative number recovered. Now $\mathcal{F}_t = \sigma\{N(\tau), R(\tau) : 0 \leq \tau \leq t\}$.

$$\begin{aligned} P[dN(t) = 1,\ dR(t) = 0|\mathcal{F}_t] &= \lambda I(t) S(t) dt + o(dt) \\ P[dN(t) = 0,\ dR(t) = 1|\mathcal{F}_t] &= \gamma I(t) dt + o(dt) \end{aligned}$$

Let

$$\begin{aligned} M_1(t) &= N(t) - \lambda \int_0^t I(\tau) S(\tau) d\tau \\ M_2(t) &= R(t) - \gamma \int_0^t I(\tau) d\tau \end{aligned}$$

$$J(t) = \begin{cases} 1 & \text{when } S(t) > 0 \\ 0 & \text{when } S(t) = 0 \end{cases} \tag{7.3}$$

$$B(t) = \frac{J(t^-)}{S(t^-)} \tag{7.4a}$$

Then let

$$M_1^*(t) = \int_0^t B(\tau) dM(\tau) = \int_0^t B(\tau) dN(\tau) - \int_0^t B(\tau) I(\tau) S(\tau) d\tau$$

where we estimate

$$\int_0^t B(\tau) dN(\tau) = \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{S(t^-)} \cong -\ln\left[1 - AR\left(t^-\right)\right]. \tag{7.5}$$

If we let $\theta = \frac{\lambda}{\gamma}$ and define

$$\begin{aligned} M(t) &= M_1^*(t) - \theta M_2(t) \\ &= \int_0^t B(\tau)\, dN(\tau) - \theta R(t) + \lambda \int_0^t I(\tau)\,[1 - J(\tau)]\, d\tau \end{aligned} \tag{7.6}$$

Define stopping time

$$T = \inf\{t \geq 0 : S(t)\,[N(t) - R(t)] = 0\}$$

Then

$$\int_0^T I(\tau)\,[1 - J(\tau)]\, d\tau = 0$$

implying

$$\begin{aligned} M(T) &= -\ln[1 - AR(T)] - \theta R(T) \\ \hat{\theta} &= \frac{-\ln[1 - AR(T)]}{R(T)} = \frac{c\beta}{n\gamma} = \frac{1}{n}\hat{\phi} \\ \hat{\phi} &= \frac{-\ln[1 - AR(T)]}{AR(T)} \end{aligned}$$

Taking the variance of (7.6) we have

$$\begin{aligned} Var[M(t)] &= Var[M_1^*(t)] - \theta^2 Var[M_2(t)] \\ &= E\left[\int_0^T B^2(\tau) dN(\tau)\right] + \theta^2 E[R(t)]. \end{aligned}$$

From (7.2) we have

$$Var[M(t)] \cong E\left[\frac{1 - AR(t)}{R(t)}\right] + \theta^2 E[R(t)].$$

Then the estimated variance for the stopped process is

$$\widehat{Var}[M(t)] \cong \frac{1 - AR(T)}{R(T)} + \widehat{\theta}^2 R(t).$$

By the ZMMCLT

$$\frac{M_1^*(T) - \theta M_2(T)}{\sqrt{Var[M(T)]}} \tilde{} N(0,1).$$

Since

$$\frac{M_1^*(T) - \theta M_2(T)}{\sqrt{\widehat{Var}[M(T)]}} = \frac{\widehat{\theta} - \theta}{\frac{\sqrt{\widehat{Var}[M(T)]}}{M_2(T)}},$$

then

$$Var(\widehat{\theta}) \cong \frac{\left[\frac{1-AR(T)}{R(T)}\right] + \widehat{\theta}^2 R(T)}{R(T)^2}.$$

7.3 Martingale-based approach to estimating vaccine efficacy

This section in based on Longini, *et al.*[21]. We use the martingale approach to estimate vaccine efficacy. Consider an epidemic process in a fixed population of size n. We let n_ν be number of people in vaccination stratum ν, where $\nu = 0$ (unvaccinated), 1 (vaccinated). Then the fraction vaccinated is $f = 1 - \frac{n_0}{n}$. Let β_ν be the per-contact transmission probability for a person in stratum ν. Individuals in the population make an average of c contacts per unit of time. The scaled contact rate is $\lambda = c/n$. The relative risk $\theta = \beta_1/\beta_0$ is the relative susceptibility of people in the vaccination stratum as compared to those in the unvaccinated stratum. We define the vaccine efficacy for susceptibility as VE $= 1-\theta$. Once infected, we assume that people remain ifectious for a average of τ time units. We work with the SIR infectious process so that $S_\nu(t)$, $I_\nu(t)$ and $R_\nu(t)$ are number of susceptible, infected and removed (immune) people, respectively, in stratum ν, at time t,where $S_\nu(t) + I_\nu(t) + R_\nu(t) = n_\nu$. Then the stochastic process of interest is $\{S_0(t), S_1(t), I(t),\, t \geq 0\}$ is on the discrete state space $\mathcal{I}^3$ and has a continuous index set. The transition probabilities are

$$\begin{gathered}
\Pr\{dS_0(t) = -1, dS_1(t) = 0, dI(t) = 1|\mathcal{F}_t\} = \lambda\ \beta_0 I(t)S_0(t)dt + o(dt),\\
\Pr\{dS_0(t) = 0, dS_1(t) = -1, dI(t) = 1|\mathcal{F}_t\} = \lambda\ \beta_1 I(t)S_1(t)dt + o(dt),\\
\Pr\{dS_0(t) = dS_1(t) = 0, dI(t) = -1|\mathcal{F}_t\} = \gamma\ I(t)dt + o(dt),\\
\Pr\{dS_0(t) = dS_1(t) = dI(t) = 0|\mathcal{F}_t\} = 1 - I(t)\{[\textstyle\sum_{\nu=0}^{1}\beta_\nu S_\nu(t)] + \gamma\}dt + o(dt),
\end{gathered} \tag{7.7}$$

where $\mathcal{F}_t$ is the σ- field for the process (7.7) up to and including time t, so that $\mathcal{F}_t = \sigma\{S_0(t), S_1(t), I(\tau) : 0 \leq \tau \leq t\}$.The initial conditions are $\Pr\{I_\nu(0) = 1\} = 1$, for some ν, and the process has stopping times $T_\nu = \inf_{t\geq 0}\{t : S_\nu(t)I(t) = 0\}$, $T = \max\{T_0, T_1\}$. Our goal is to use counting process-based methods to estimate θ of observations on the stochastic process.

The counting process of interest is the cumulative number of infected people in stratum ν by time t,which is $N_\nu(t) = S_\nu(0) - S_\nu(t)$. Then, this counting process is the stochastic process $\{N_0(t), N_1(t),\, t \geq 0\}$ on the discrete space $\mathcal{I}^2$,with continuous index set. The transition probabilities for the counting process are

$$\begin{gathered}
\Pr\{dN_0(t) = 1, dN_1(t) = 0|\mathcal{F}_t\} = \lambda\ \beta_0 I(t)S_0(t)dt + o(dt),\\
\Pr\{dN_0(t) = 0, dN_1(t) = 1|\mathcal{F}_t\} = \lambda\ \beta_1 I(t)S_1(t)dt + o(dt),\\
\Pr\{dN_0(t) = dN_1(t) = 0|\mathcal{F}_t\} = 1 - I(t)[\textstyle\sum_{\nu=0}^{1}\beta_\nu S_\nu(t)]dt + o(dt).
\end{gathered}$$

Then it follow directly that

$$M_\nu(t) = N_\nu(t) - \lambda\ \beta_\nu \int_0^t S_\nu(\tau)I(\tau)d\tau, \nu = 0, 1,$$

is a zero mean martingale (ZMM) with respect to $\mathcal{F}_t$. We construct the stochastic

integrals

$$M_\nu^*(t) = \int_0^t B_\nu(\tau) dM_\nu(\tau) = \int_0^t B_\nu(\tau) dN_\nu(\tau) - \lambda\ \beta_\nu \int_0^t J_\nu(\tau) I(\tau) d\tau, \tag{7.8}$$

which are also ZMM's with respect to $\mathcal{F}_t$, where $J_\nu(\tau)$ and $B_\nu(\tau)$ are defined by (7.3) and (7.4a), respectively. In order to evaluate the integrals in (7.8), we note, as before (7.5), that

$$\int_0^t B_\nu(\tau) dN_\nu(\tau) = \frac{1}{S_\nu(0)} + \frac{1}{S_\nu(0)-1} + ... + \frac{1}{S_\nu(t-)} \cong -\log_e[\frac{S_\nu(t-)}{S_\nu(0)}]. \tag{7.9}$$

Substituting (7.9) into (7.8) yields

$$M_\nu^*(t) \cong -\log_e[\frac{S_\nu(t-)}{S_\nu(0)}] - \lambda\ \beta_\nu \int_0^t J_\nu(\tau) I(\tau) d\tau.$$

Now consider the estimating equation

$$M(t) = M_1^*(t) - \theta\ M_0^*(t),$$

which is a ZMM with respect to $\mathcal{F}_t$,so that $E\{M_1^*(t) - \theta\ M_0^*(t)\} = 0$, which suggests the estimator

$$\hat{\theta} = M_1^*(t)/M_0^*(t).$$

By definition, the stratum-specific final attack rate is $AR_\nu = 1 - [S_\nu(T_\nu)/S_\nu(0)]$. Evaluating this estimator for the stopped process yields

$$\hat{\theta} = \log_e(1 - \hat{A}R_1)/log_e(1 - \hat{A}R_0).$$

From the variation process, the variance of $M_\nu^*(t)$ is

$$Var[M(t)] = Var[M_1^*(t)] + \theta^2 Var[M_0^*(t)], \tag{7.10}$$

since $M_0^*(t)$ and $M_1^*(t)$ are orthogonal, where, as before,

$$Var[M_\nu^*(t)] = E[\int_0^t B_\nu^2(\tau) dN_\nu(\tau)] = \frac{1}{S_\nu(0)^2} + \frac{1}{[S_\nu(0)-1]^2} + ... + \frac{1}{S_\nu(t-)^2} \tag{7.11}$$

$$\cong \frac{1 - \frac{S_\nu(t-)}{S_\nu(0)}}{S_\nu(t-)}, \nu = 0, 1. \tag{7.12}$$

Define the stopped ZMM $M = M(T)$. Then, evaluating (7.10) at (7.11) for the stopped process yields

$$\widehat{Var}[M] = \frac{\widehat{AR_1}}{\widehat{S_1}} + \hat{\theta}^2 \frac{\widehat{AR_0}}{\widehat{S_0}}.$$

By the ZMMCLT (*Theorem 6.1.3*), we have

$$\frac{M_1^* - \theta M_0^*}{\text{Var}[M]^{1/2}} \sim \mathcal{N}(0,1). \tag{7.13}$$

Recognizing that $M_1^* - \theta\ M_0^* = [-\log_e(1-AR_1) + \theta\ \log_e(1-AR_0)]$ and manipulating the right-hand side of (7.13) reveals that

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{\text{Var}[M]}{[\log_e(1-AR_0)]^2}\right)$$

Thus, the large sample variance of $\hat{\theta}$ is $\widehat{var}(\hat{\theta}) = \frac{\text{Var}[M]}{[\log_e(1-AR_0)]^2}$.

Some vaccine efficacy estimates and 95% confidence intervals are given in the table below.

Vaccine coverage (among children with vaccination cards), attack rates, and estimated vaccine efficacy from the measles outbreak in Muyinga, Burundi, July 1988 to January 1989

Age group	Group size		Attack rates		Vaccine efficacy	
months		f	$\widehat{\text{AR}_0}$	$\widehat{\text{AR}_1}$	$\widehat{\text{VE}}$	95% CI
$[9-15]$	199	0.452	0.560	0.178	0.761	$[0.752, 0.770]$
$[16-36]$	533	0.842	0.405	0.134	0.723	$[0.716, 0.731]$
$[37-60]$	432	0.956	0.158	0.080	0.515	$[0.348, 0.683]$
Total	$1,164$			Summary†	0.735	$[0.648, 0.822]$

†Reciprocal variance weighted average. Source: Longini, *et al.* [22] [3][12][21]

Chapter 8
HIDDEN MARKOV CHAINS

How do we write the likelihood when transition times are not observed? For example, suppose we observe $\mathbf{Y} = \{Y_0, Y_1, \ldots, Y_m\}$ at times $\boldsymbol{\tau} = \{\tau_0, \tau_1, \ldots, \tau_m\}$. Let $p_{Y_k Y_{k+1}} = p_{Y_k Y_{k+1}}(\tau_k, \tau_{k+1})$. Then

$$L(\mathbf{Y}) = \prod_{k=0}^{m-1} p_{Y_k Y_{k+1}}.$$

But what if we do not observe $\mathbf{Y}$, but rather something related to $\mathbf{Y}$. That is, say we observe $\mathbf{X}$ where $\mathbf{X} \mid \mathbf{Y} \sim f(x \mid y)$. Then

$$L(\mathbf{X}) = \int_y f(x|y)\, f(y)\, dy.$$

Example 8.0.1 *Suppose $S = \{1, 2\}$, $Y_0 = 1$, we observe $X_0 = 1, X_1 = 1, X_2 = 2$ and*

$$\begin{aligned} f(x = 1|y = 1) &= \alpha \\ f(x = 2|y = 1) &= 1 - \alpha \\ f(x = 1|y = 2) &= 1 - \beta \\ f(x = 2|y = 2) &= \beta \end{aligned}$$

Then

$$\begin{aligned} L(\mathbf{X}) &= p_{11}p_{11}\alpha(1-\alpha) + p_{11}p_{12}\alpha\beta + p_{12}p_{22}(1-\beta)\beta + p_{12}p_{21}(1-\beta)(1-\alpha) \\ &= \textstyle\sum_{Y_1=1}^{2} \sum_{Y_2=1}^{2} p_{1Y_1} p_{Y_1Y_2} f(1|Y_1)\, f(2|Y_2) \\ &= \textstyle\sum_{Y_1=1}^{2} p_{1Y_1} f(1|Y_1) \sum_{Y_2=1}^{2} p_{Y_1Y_2} f(2|Y_2) \end{aligned}$$

Now suppose that Y_0 is not known, so that there are $2^3 = 8$ possible paths (note that previously there were $2^2 = 4$ possible paths). We let p_{Y_0} be the initial distribution. Then

$$L(\mathbf{X}) = \sum_{Y_0=1}^{2} p_{Y_0} f(1|Y_0) \sum_{Y_1=1}^{2} p_{Y_0Y_1} f(1|Y_1) \sum_{Y_2=1}^{2} p_{Y_1Y_2} f(2|Y_2).$$

In general, we have a chain with S states and m observations $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ yielding

$$L(\mathbf{X}) = \sum_{Y_0} p_{Y_0} f(X_0|Y_0) \sum_{Y_1} p_{Y_0 Y_1} f(X_1|Y_1) \ldots \sum_{Y_m} p_{Y_{m-1} Y_m} f(X_m|Y_m)$$

which, due to Baum, *et al.*[2], can be simplified using matrix multiplication. Let

$$\mathbf{f}(X_0) = \begin{bmatrix} p_{Y_0=y_0} f(X_0|y_0) \\ p_{Y_0=y_1} f(X_0|y_1) \\ \vdots \\ p_{Y_0=y_s} f(X_0|y_S) \end{bmatrix}, \ \mathbf{C} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

where $\mathbf{f}(X_0)$ and $\mathbf{C}$ are $s \times 1$ column vectors, and $\mathbf{T}^{(j)}(X_j)$ be the $s \times s$ matrix having (k, ℓ) element $\left\{p_{Y_{j-1}=k, Y_j=\ell} f(X_j|Y_j = \ell)\right\}$. Then

$$L(\mathbf{X}) = \mathbf{f}(X_0)^T \mathbf{T}^{(1)}(X_1) \mathbf{T}^{(2)}(X_2) \cdots \mathbf{T}^{(m)}(X_m) \mathbf{C},$$

Example 8.0.2 *Continuing previous example where* $\mathbf{X} = (1, 1, 2)$,

$$\mathbf{f}(X_0) = \begin{bmatrix} p_1 f(1|1) \\ p_2 f(1|2) \end{bmatrix} = \begin{bmatrix} p_1 \alpha \\ p_2 (1-\beta) \end{bmatrix}$$

$$\mathbf{T}^{(1)}(1) = \begin{bmatrix} p_{11}\alpha & p_{12}(1-\beta) \\ p_{21}\alpha & p_{22}(1-\beta) \end{bmatrix}$$

$$\mathbf{T}^{(2)}(2) = \begin{bmatrix} p_{11}(1-\alpha) & p_{12}\beta \\ p_{21}(1-\alpha) & p_{22}\beta \end{bmatrix}$$

Satten and Longini[25] analyze HIV progression as hidden Markov chains.

Chapter 9
GIBBS SAMPLING

Suppose we are given the joint density $f(x, y_1, \ldots, y_p)$ and would like to find the marginal density

$$f(x) = \int \ldots \int f(x, y_1, \ldots, y_p)\, dy_1 \ldots dy_p. \tag{9.1}$$

The Gibbs sampling technique generates $x_1, x_2, \ldots, x_m \tilde{} f(x)$ without requiring explicit calculation of the integrations required in equation (9.1). Rather the Gibbs sampler only requires knowledge of the conditional distributions.

Example 9.0.3 *Bivariate case.* $(p = 1)$.

$$\begin{aligned} f_X(x) &= \int f_{XY}(x, y)\, dy \\ &= \int f_{X|Y}(x, y)\, f(y)\, dy \\ &= \int f_{X|Y}(x|y) \int f_{Y|X}(y|w)\, f_X(w)\, dw dy \\ &= \int \left[\int f_{X|Y}(x|y)\, f_{Y|X}(y|w)\, dy \right] f_X(w)\, dw \\ &= \int h(x, w)\, f_X(w)\, dw \end{aligned}$$

where

$$h(x, w) = \int f_{X|Y}(x, y)\, f_{Y|X}(y, w)\, dy$$

which is a fixed point integral equation for which $f_X(x)$ is a unique solution.

Gibbs sampling generates a random sample from $f_X(x)$ by sampling from $f_{X|Y}(x, y)$ and $f_{Y|X}(y, x)$. That is, generate sequence $\{Y_0, X_0, Y_1, X_1, \ldots, Y_k, X_k, \ldots\}$ where we first choose X_0 and the sample as follows: $Y_1 = f(y|X_0)$, $X_1 = f(x|Y_1)$, $Y_2 = f(y|X_1), \ldots, X_k = f(x|Y_k)$. It can be shown that as $k \to \infty$, $X_m \to^d f_X(x)$.

Consider the simple case where X and Y are (marginally) Bernoulli random variables, $S = \{0,1\}$, $T = \{0,1,2,\ldots\}$. Since

$$f_{Y|X}(y,x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

we define transition matrix

$$\mathbf{P}_{Y|X} = \begin{bmatrix} \frac{f(0,0)}{f_X(0)} & \frac{f(0,1)}{f_X(0)} \\ \\ \frac{f(1,0)}{f_X(1)} & \frac{f(1,1)}{f_X(1)} \end{bmatrix}.$$

Similarly, we let

$$\mathbf{P}_{X|Y} = \begin{bmatrix} \frac{f(0,0)}{f_Y(0)} & \frac{f(1,0)}{f_Y(0)} \\ \\ \frac{f(0,1)}{f_Y(1)} & \frac{f(1,1)}{f_Y(1)} \end{bmatrix}$$

For $X_0 \to Y_1 \to X_1$, it follows from the CK equations that $P[X_1|X_0] = \sum_{Y_1} P[Y_1|X_0] P[X_1|Y_1]$ and thus

$$\mathbf{P}_{X|X} = \mathbf{P}_{Y|X}\mathbf{P}_{X|Y}.$$

Let

$$\mathbf{f}_X(k) = [f(X_k = 0), f(X_k = 1)]$$

so that

$$\begin{aligned} \mathbf{f}_X(k) &= \mathbf{f}_X(k-1)\mathbf{P}_{X|X} \\ &= \mathbf{f}(0)\mathbf{P}_{X|X}^k \end{aligned}$$

This irreducible, ergodic Markov chain has stationary distribution $\mathbf{f}_X = \mathbf{f}_X\mathbf{P}_{X|X}$. Thus $X_m \to^d f_X(x)$.

Notice that $\mathbf{f}_X = \mathbf{f}_X\mathbf{P}_{X|X}$ is the discrete state space analog of

$$f_X(x) = \int h(x,w) f_X(w)\, dw$$

Theorem 9.0.4 *Ergodic Theorem:*

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} X_i \to^{a.s.} \int x f_X(x)\, dx = EX$$

Example 9.0.5 *The following example comes from Casella and George[?]. Suppose*

$$f(x,y) = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{x+\alpha-1}(1-y)^{n-x+\beta-1},\ x = 0,1,\ldots,n,\ 0 \le y \le 1$$

where $f(x|y)$ *is binomial* (n, y) *and* $f(y|x)$ *is beta* $(x+\alpha, n-x+\beta)$. *Then we*

$$f(x) = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)},\ x = 0, 1, \ldots, n.$$

Multivariate case. $S = \{0, 1, 2, \ldots, n\}$, $T = \{0, 1, 2, \ldots\}$. Suppose $n = 3$. That is, we have X, Y, Z random variables. Start with Y_0, Z_0, X_0. Generate Gibbs $\{Y_0, Z_0, X_0, Y_1, Z_1, X_1, \ldots\}$sequence as follows:

$$\begin{aligned} Y_1 &= f(y|X_0, Z_0) \\ Z_1 &= f(z|X_0, Y_1) \\ X_1 &= f(x|Y_1, Z_1) \\ Y_2 &= f(y|X_1, Z_1) \\ &\vdots \end{aligned}$$

Example: Let N be Poisson with parameter λ.

$$f(x, y, n) \propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1} e^{-\lambda} \frac{\lambda^n}{n!},\ x = 0, 1, \ldots, n,\ 0 \le y \le 1$$

Then $f(x|y, n)$ ˜binomial(n, y), $f(y|x, n)$ ˜beta$(x+\alpha, n-x+\beta)$, and

$$f(n|x, y) \propto e^{-(1-y)\lambda} \frac{[(1-y)\lambda]^{n-x}}{(n-x)!},\ n = x, x+1, \ldots$$

Continuous case. Suppose X and Y are continuous with $S = (-\infty \le x \le \infty,\ -\infty \le y \le \infty)$. Then

$$f_{X_1|X_0}(x_1|x_0) = \int f_{X_1|Y_1}(x_1|y) f_{Y_1|X_0}(y|x_0)\, dy$$

$$f_{X_k|X_0}(x|x_0) = \int f_{X_k|X_{k-1}}(x|t) f_{X_{k-1}|X_0}(t|x_0)\, dt$$

As $k \to \infty$, converges to a stationary point

$$f_{X_k|X_0}(x|x_0) \to f_X(x)$$

$$f_{X_k|X_{k-1}}(x|t) \to h(x, t)$$

$$f_X(x) = \int h(x, t) f_X(x)\, dt$$

fixed point integral equation from before.

Chapter 10
APPENDIX

10.1 Series

$$\text{Newton's Formula:} \sum_{r=0}^{\infty} \binom{n}{r} x^r = (1+x)^n, |x| < 1$$

$$\text{Binomial:} \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r} = (x+y)^n$$

$$\text{Geometric:} \quad \begin{array}{l} \sum_{r=0}^{\infty} x^r = \frac{1}{1-x}, \ |x| < 1 \\ \sum_{r=0}^{n-1} x^r = \frac{1-x^n}{1-x}, \ x \neq 1 \end{array}$$

$$\text{Exponential:} \sum_{r=0}^{\infty} \frac{x^r}{r!} = e^x$$

$$\text{Logarithmic:} \sum_{r=0}^{\infty} \frac{(-1)^r}{r+1} x^{r+1} = \ln(1+x), \ |x| < 1$$

10.2 Inequalities

$$\text{Schwartz (Inequalities):} \ [E(X,Y)]^2 \leq E\left[X^2\right] E\left[Y^2\right]$$

$$\text{Schwartz:} \left|\sum_{r=0}^{n} X_r Y_r\right|^2 \leq \sum_{r=0}^{n} |X_r|^2 \sum_{r=0}^{n} |Y_r|^2$$

$$\text{Absolute Values:} \left|\sum_{r=0}^{n} X_r\right| \leq \sum_{r=0}^{n} |X_r|$$

$$\text{Jensen:} \ E[f(X)] \geq f[E(X)] \ \text{ for } f(\cdot) \text{ a convex function}$$

10.3 Convergence of Sequences and Series

- $\sum_{k=0}^{\infty} \frac{1}{k^p}$ converges if $p > 1$ and diverges $p \leq 1$.
- If $\sum_{k=0}^{\infty} a_k$ converges, then $\lim_{k\to\infty} a_k \to 0$.
- Fatou's Lemma: Let $\{a_k(t)\}$ be a sequence of non-negative numbers where $\lim_{k\to\infty} a_k(t)$ exists for each t in $T = \{0, 1, \ldots\}$, then for all t in T and $k = 0, 1, \ldots,$

$$\sum_{t\in T} \left\{ \lim_{k\to\infty} a_k(t) \right\} \leq \lim_{k\to\infty} \left\{ \sum_{t\in T} a_k(t) \right\}$$

- The Renewal Theorem: Let $\{a_k\}$, $\{b_k\}$ and $\{u_k\}$ be sequences of non-negative numbers with $\sum_{k=0}^{\infty} a_k = 1$, $\sum_{k=0}^{\infty} b_k < \infty$, and with the $\{u_k\}$ sequence bounded. Assume the g.c.d.$\{k : a_k > 0\} = 1$ and assume the renewal equation $u_n - \sum_{k=0}^{n} a_{n-k} u_k = b_n$ holds for all $n = 0, 1, \ldots$ Then the $\lim_{n\to\infty} u_n$ exists. In fact,

$$\begin{array}{cc} \lim_{n\to\infty} u_n = \frac{\sum_{k=0}^{\infty} b_k}{\sum_{k=1}^{\infty} k a_k} & \text{if } \sum_{k=1}^{\infty} k a_k < \infty \\ \lim_{n\to\infty} u_n = 0 & \text{if } \sum_{k=1}^{\infty} k a_k = 0 \end{array}$$

- Toeplitz Lemma: Let $\{a_k\}$ be a sequence of numbers, and let $b_n = \sum_{k=1}^{n} a_k$, such that $\lim_{n\to\infty} b_n \to \infty$. Also, let $\{x_k\}$ be a sequence of numbers such that $\lim_{k\to\infty} x_k \to x < \infty$. Then $\lim_{n\to\infty} \left\{ \frac{1}{b_n} \sum_{k=1}^{n} a_k x_k \right\} \to x$.

10.4 Convergence in distribution

Let $\{F_n\}$ be a sequence of distribution functions (df). If there exists a df F such that, as $n \to \infty$, $F_n(x) \to F(x)$, at every point x at which F is continuous, we say that F_n converges in law (weakly) to F, and we write $F_n \to^L F$. In addition, if $\{X_n\}$ is a sequence of random variables and $\{F_n\}$ is the corresponding sequence of df's, we say that X_n converges in distribution (or law) to X if there exists a random variable X with df F such that $F_n \to^D F$. We write $X_n \to^L X$.

10.5 Convergence in Probability

Let $\{X_n\}$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$. We say that the sequence $\{X_n\}$ converges in probability to the random variable X if, for every $\varepsilon > 0$,

$$\Pr\{|X_n - X| > \varepsilon\} \to 0 \text{ as } n \to \infty.$$

We write $X_n \to^P X$.

10.6 Almost Sure Convergence

Let $\{X_n\}$ be a sequence of random variables. We say that $\{X_n\}$ converges almost surely (a.s.) the random variable X if and only if

$$\Pr\{\omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\} = 1,$$

and we write $X_n \to^{a.s.} X$ or $X_n \to X$ with probability 1. We also have the theorem that $X_n \to^{a.s.} X$ if and only if

$$lim_{n\to\infty} \Pr\{\sup_{m\geq n} |X_n - X| > \varepsilon\} = 0 \text{ for all } \varepsilon > 0.$$

We note that a.s. convergence is stronger than convergence in probability, which in turn is stronger that convergence in law so that

$$X_n \to^{a.s.} X \Longrightarrow X_n \to^{P} X \Longrightarrow X_n \to^{L} X.$$

References

1. N. Bailey. *The mathematical theory of infectious diseases*. Griffin, London, second edition, 1975.
2. L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Statistics*, 41:164–171, 1970.
3. N. Becker. *Analysis of infectious disease data*. Chapman and Hall, New York, 1989.
4. C. Chiang. *An Introduction to Stochastic Processes and Their Applications*. Krieger, Huntington, New York, 1980.
5. C. Chiang. *The Life Table and Its Applications*. Krieger, Malabar, Florida, 1984.
6. K. Cooke, D. Calef, and E. Level. Stability or chaos in discrete epidemic models. In *Nonlinear systems and applications–an international conference*. Academic press, New York, 1977.
7. K. Dietz. The first epidemic model: A historical note on P.D. En'ko. *Austrialian Journal of Statistics*, 30A:56–65, 1988.
8. P. Fine. A commentary on the mechanical analogue to the Reed-Frost epidemic model. *Americal Journal of Epidemiology*, 106:87–100, 1977.
9. J. Frauenthal. *Mathematical models in epidemiology*. Springer-Verlag, Berlin, 1980.
10. M. Greenwood. The statistical measure of infectiosness. *Joural of Hygiene (Cambridge)*, 31:336–351, 1931.
11. M. Haber, I. Longini, and G. Cotsonis. Models for the statistical analysis of infectious disease data. *Biometrics*, 44:163–173, 1988.
12. M. Halloran, M. Haber, and I. Longini. Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology*, 136:328–343, 1992.
13. I. Longini. A chain binomial model of endemicity. *Mathematical Biosciences*, 50:85–93, 1980.
14. I. Longini. The generalized discrete-time epidemic model with immunity: A synthesis. *Mathematical Biosciences*, 82:19–41, 1986.
15. I. Longini. Modeling the decline of CD4+ T-lymphocyte counts in HIV-infected individuals. *Journal of Acquired Immune Deficiency Syndromes*, 3:930–931, 1990.
16. I. Longini. Chain binomial models. In P. Armitage and T. Colton, editors, *The Encyclopedia of Biostatistics*, pages 593 – 597. Wiley, New York, 1998.
17. I. Longini, R. Byers, N. Hessol, and W. Tan. Estimating the stage-specific numbers of HIV infection using a Markov model and back-calculation. *Statistics in Medicine*, 11:831–843, 1992.
18. I. Longini, W. Clark, R. Byers, G. Lemp, J. Ward, W. Darrow, and H. Hethcote. Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, 8:831–843, 1989.
19. I. Longini, W. Clark, L. Gardner, and J. Brudage. The dynamics of CD4+ T-lymphocyte decline in HIV infected individuals: a Markov modeling approach. *Jour-*

nal of Acquired Immune Deficiency Syndromes, 4:1141–1147, 1991.

20. I. Longini, W. Clark, M. Haber, and R. Horsburgh. The stages of HIV infection: Waiting times and infection transmission probabilities. *Lecture Notes in Biomathematics*, 83:111–137, 1989.
21. I. Longini, M. Halloran, and M. Haber. Estimation of vaccine efficacy from epidemics of acute infectious agents under vaccine-related heterogeneity. *Math Biosci*, 117:271–281, 1993.
22. I. Longini, M. Halloran, M. Haber, and R. Chen. Measuring vaccine efficacy from epidemics of acute infectious agents. *Stat Med*, 12:249–263, 1993.
23. I. Longini and J. Koopman. Household and community transmission parameters from final distributions of infections in households. *Biometrics*, 38:115–126, 1982.
24. I. Longini, J. Koopman, M. Haber, and G. Cotsonis. Statistical inference for infectious diseases: Risk-specific household and community transmission parameters. *Am J Epidemiol*, 128:845–859, 1988.
25. G. Satten and I. Longini. Markov chains with measurement error: estimating the "true" course of a marker of HIV disease progression (with discussion). *Applied Statistics*, 45:275–309, 1996.
26. I. Saunders. An approximate maximum likelihood estimator for chain binomial models. *Australian Journal of Statistics*, 22:307–316, 1980.