CHANGE POINT PROBLEMS FOR POISSON PROCESSES

by

Clive R. Loader
Stanford University

TECHNICAL REPORT NO. 9
JUNE 1990

Prepared Under Grant MDA 904-89-H-2040
For The National Security Agency

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

CHANGE POINT PROBLEMS SSFOR POISSON PROCESS

by

Clive R. Loader
Stanford University

TECHNICAL REPORT NO. 9
JUNE 1990

PREPARED UNDER GRANT MDA 904-89-H-2040
FOR THE NATIONAL SECURITY AGENCY

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Abstract

Our general topic will be the detection of times at which point processes undergo abrupt changes. The simplest model is to assume events occur according to a Poisson process, whose rate changes at an unknown change point. Questions of interest include testing for the presence of a change point and if we conclude there is a change point, to estimate and find confidence regions for the time and size of the change.

We will use likelihood ratio statistics to address these problems. This in turn leads to non-linear boundary crossing problems for Poisson processes. We will discuss large deviation and local methods to approximate these probabilities, and compare with exact recursive formulae which are much more computationally intensive.

Our boundary crossing methodology can be extended to more complicated problems. We discuss also a model where the rate of the Poisson process is assumed to vary in a log-linear fashion with an unknown change point. The unknown linear term makes inference for this model harder than for the constant parameter model.

Another problem we discuss is detecting changes of hazard rate. For patients in a clinical trial, there may be a period following treatment during which side effects are most likely. The use of a change point model allows us to detect such critical periods. This problem is closely related to the change point problem for a Poisson

process, and we are able to develop tests and approximate confidence regions.

For point processes in two or more dimensions, the natural analogy of a change point is a region of higher intensity. Likelihood ratio tests for a change region lead to questions about boundary crossings by random fields. Assuming the change region to be rectangular, we obtain large deviation approximations to the significance level of tests. The special structure we make use of is that the random fields may be constructed locally as the superposition of independent one dimensional processes. This enables us to apply techniques similar to those used in the one dimensional case.

We will illustrate our methods through numerical examples and reference to several datasets. For our discussion of one dimensional Poisson processes, we use British coal mining accident data and also a second set of industrial accident data. The hazard rate model is fitted to the Stanford heart transplant dataset.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Change point problems are concerned with detecting times at which the mechanism generating a process changes. Historically, most attention has been given to on line procedures for detecting change points, which aim to detect a change as soon as possible after it occurs. Procedures used for this problem include the Cusum test (cf. Siegmund (1985) Section 2.6) and an alternative scheme proposed by Shiryayev (1963).

More recently, attention has been given to fixed sample change point problems. In this context, the process is observed over a fixed time interval, and we are interested in determining whether there were any abrupt changes over this interval. In this thesis, we are interested in fixed sample problems for Poisson processes, for which a change point will represent a time at which the event rate undergoes an abrupt change.

1.1 Statement of the Problems

Given an interval $[0, T]$ and a non-negative rate function $\{\lambda(t); 0 \le t \le T\}$, we define a measure Λ on Borel subsets of $[0, T]$ by

$$\Lambda(A) = \int_A \lambda(t)dt.$$

Suppose a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a random function $X : \mathcal{B}[0, T] \times \Omega \to \mathcal{N}$ satisfying the following properties:

1. For any decreasing sequence $A_1, A_2, \ldots$ of sets in $\mathcal{B}[0, T]$ with $\cap_{n=1}^{\infty} A_n = \emptyset$, $P(X(A_n, .) = 0) = 1 - \Lambda(A_n) + o(\Lambda(A_n))$ and $P(X(A_n, .) = 1) = \Lambda(A_n) + o(\Lambda(A_n))$ as $n \to \infty$.

2. For any disjoint sets $A_1, \ldots, A_n$, the random variables $X(A_1, .), \ldots, X(A_n, .)$ are independent and $X(\cup_{i=1}^{n} A_i, .) = \sum_{i=1}^{n} X(A_i, .)$.

Then X is a **Poisson Process** with rate function $\lambda(t)$. If $\lambda(t) = \lambda$ is constant, then $X(t)$ is a **homogeneous** Poisson process. We will write $X(t)$ to mean $X([0, t], .)$. We assume the reader is familiar with basic properties of the Poisson process: $X(A, .) \sim Po(\Lambda(A))$ and conditional on $X(T) = n$, the event times $T_1, \ldots, T_n$ have the same distribution as the order statistics of a sample of size n from a density $\frac{\lambda(t)}{\Lambda([0,T])}$. See, for example, Feller (1968).

A **change point** is a point at which $\lambda(t)$ is discontinuous. The simplest model we consider is the **constant parameter** model, given by

$$\lambda(t) = \begin{cases} \lambda_0 & 0 \le t \le \tau \\ \lambda_1 & \tau < t \le T \end{cases}. \tag{1.1}$$

Here, λ_0 and λ_1 are unknown parameters, and τ is the unknown change point. We are interested in making inference about τ. For example, we may want to test the null hypothesis of no change. If we conclude there is a change, we would like to find point estimates and confidence regions for τ and other parameters, such as the size of the change which we will measure by $\delta = \log(\lambda_1/\lambda_0)$.

More generally, we may allow $\lambda(t)$ to both vary smoothly over time and have a change point. As a particular example, we will consider a **log-linear** model,

$$\lambda(t) = \begin{cases} e^{a+bt} & 0 \le t \le \tau \\ e^{a+\delta+bt} & \tau < t \le T \end{cases}. \tag{1.2}$$

For sequences of independent Gaussian observations, regression models with change points have been studied by Kim and Siegmund (1989) and our results for analyzing (1.2) will have some similarity. A related model is to have a regression function which is continuous but the slope changes at the change point. Analysis of this model requires somewhat different techniques; see Knowles and Siegmund (1988). We will not study this model here.

The benefit of fitting a log-linear model is that it attempts to distinguish between a smoothly changing rate function and an abruptly changing rate function. If we fit the constant parameter model (1.1) when in fact the rate is varying smoothly, we may well end up detecting a change when there is none. However, if we fit the log-linear model when the constant parameter model is correct, we would expect to lose some power. In fact, we will show in Chapter 3 that the power loss is fairly substantial.

1.2 Likelihood for a Poisson Process

Suppose we observe $X(T) = n$ and event times $T_1, \ldots, T_n$. The likelihood of a rate function λ is given by

$$L(\lambda) = \exp(-\int_0^T \lambda(t)dt) \prod_{i=1}^{n} \lambda(T_i). \tag{1.3}$$

See Cox and Lewis (1966) for a derivation of (1.3). For the constant parameter model, this becomes

$$L(\tau, \lambda_0, \lambda_1) = \lambda_0^{X(\tau)} \lambda_1^{n-X(\tau)} \exp(-\lambda_0 \tau - \lambda_1 (T - \tau)). \tag{1.4}$$

The maximum likelihood estimates of λ_0, λ_1 and τ are chosen to maximize (1.4). For fixed τ, we get

$$\hat{\lambda}_0 = \frac{X(\tau)}{\tau} \tag{1.5}$$

and

$$\hat{\lambda}_1 = \frac{n - X(\tau)}{T - \tau}. \tag{1.6}$$

We find $\hat{\tau}$ by substituting these into (1.4) and maximizing over τ. As a slight generalization, we will restrict the estimate $\hat{\tau}$ to an interval $[\tau_0, \tau_1]$, where $0 \leq \tau_0 < \tau_1 \leq 1$. Formally, we define

$$\hat{\tau} = \lim_{\epsilon \to 0} \inf_{\tau_0 \leq \tau \leq \tau_1} \left\{ \tau : L(\tau, \hat{\lambda}_0, \hat{\lambda}_1) \geq \sup_{\tau_0 \leq t \leq \tau_1} L(t, \hat{\lambda}_0, \hat{\lambda}_1) - \epsilon \right\}. \tag{1.7}$$

This somewhat complicated definition is necessary because $L(\tau, \hat{\lambda}_0, \hat{\lambda}_1)$ is a discontinuous function of τ and so the supremum may not actually be achieved. The inf in (1.7) allows for the (probability 0) event that $L(\tau, \hat{\lambda}_0, \hat{\lambda}_1)$ achieves the same supremum at two different times. It is easily shown that between successive events $L(\tau, \hat{\lambda}_0, \hat{\lambda}_1)$ is a convex function of τ and therefore $\hat{\tau}$ will be one of the event times or one of the endpoints τ_0 and τ_1. We will write $L(\hat{\tau}, \hat{\lambda}_0, \hat{\lambda}_1)$ to mean $\sup_{\tau_0 \leq t \leq \tau_1} L(t, \hat{\lambda}_0, \hat{\lambda}_1)$.

The choice of truncation points τ_0 and τ_1 is fairly arbitrary. When $L(t, \hat{\lambda}_0, \hat{\lambda}_1)$ is large, this is indicating a large difference between $\hat{\lambda}_0$ and $\hat{\lambda}_1$. However, if t is close to 0 then the estimate $\hat{\lambda}_0$ has a large variance and so the difference between $\hat{\lambda}_0$ and $\hat{\lambda}_1$ may not be a reliable indicator of a change point. By choosing $\tau_0 > 0$ and $\tau_1 < t$ we reduce the possibility of spuriously detecting a change near an end point.

We will not discuss in detail the distribution of the estimators. Chernoff and Rubin (1955) show for a wide class of problems involving estimation of a discontinuity in a density that the maximum likelihood estimate will have error of $O_p(\frac{1}{n})$. It follows that the estimates $\hat{\lambda}_0$ and $\hat{\lambda}_1$, evaluated at $\hat{\tau}$, will have the same asymptotic distributions as if τ were known, and are asymptotically independent of $\hat{\tau}$. The limiting distribution of $n(\hat{\tau} - \tau)$ is related to the location of the maximum of a random walk which we will not derive explicitly, but the alert reader will guess the result from our discussion of confidence regions in Chapter 4.

The log-likelihood ratio statistic for testing $\mathcal{H}_0 : \lambda_0 = \lambda_1$ against the alternative $\mathcal{H}_1 : \lambda_0 \neq \lambda_1$ if τ were known is

$$\begin{aligned} l(\tau) &= \log(L(\tau, \hat{\lambda}_0, \hat{\lambda}_1) - L(0, \hat{\lambda})) \\ &= X(\tau) \log\left(\frac{TX(\tau)}{\tau n}\right) + (n - X(\tau)) \log\left(\frac{T(n - X(\tau))}{(T-\tau)n}\right). \end{aligned} \tag{1.8}$$

For unknown τ we reject $\mathcal{H}_0$ if the maximum of $l(t)$ over the interval $[\tau_0, \tau_1]$ exceeds $\frac{1}{2}c^2$ for some positive constant c. The significance level of the test is then

$$\alpha = P_0 \left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{c^2}{2} \,\middle|\, X(T) = n \right\} \tag{1.9}$$

where P_0 is a measure under which $X(t)$ is a homogeneous Poisson process. In (1.9), we condition on $X(T) = n$. Without this conditioning, the distribution of $l(t)$ depends on the unknown rate λ_0. Since under $\mathcal{H}_0$, $X(T)$ is a sufficient statistic for λ_0, conditioning on $X(T)$ will make (1.9) independent of λ_0.

In the testing context, the choice of τ_0 and τ_1 can be considered as trading off power against various alternatives. If we take $\tau_0 = 0$ and $\tau_1 = T$, then randomness near the endpoints will mean we require a large value of c to attain a fixed significance level, and hence the test will have low power for detecting changes near the middle of $[0, T]$. Alternatively, if we choose $\tau_0 > 0$ and $\tau_1 < T$, we will be able to choose a smaller value of c thereby increasing the power to detect changes near the middle of $[0, T]$, but decreasing the power to detect changes near an endpoint. The use of likelihood ratio tests for change point problems is sometimes criticized because of the need to truncate and alternative methods are proposed. However, we view the truncation as an acknowledgement that changes near an endpoint will be difficult to detect, regardless of what test is used.

1.3 Likelihood Testing for the Log-linear Model

For the log-linear model, (1.3) becomes

$$L(\tau, a, \delta, b) = \exp\left(an + \delta(n - X(\tau)) + b\sum_{i=1}^{n} T_i - \frac{e^a}{b}(e^{b\tau} - 1 + e^{\delta}(e^{bT} - e^{b\tau}))\right)$$

where $n = X(T)$ and $T_1, \ldots, T_n$ are the event times. For fixed τ, we can find maximum likelihood estimates of a, δ and b by solving

$$\begin{aligned}
\sum_{i=1}^{n} T_i &= X(\tau)\frac{\tau e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} + (n - X(\tau))\frac{Te^{\hat{b}T} - \tau e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}} - \frac{n}{\hat{b}} && (1.10) \\
\hat{a} &= \log\left(\frac{\hat{b}X(\tau)}{e^{\hat{b}\tau} - 1}\right) \\
\hat{\delta} &= \log\left(\frac{\hat{b}(n - X(\tau))}{e^{\hat{b}T} - e^{\hat{b}\tau}}\right) - \hat{a} \\
&= \log\left(\frac{(n - X(\tau))(e^{\hat{b}\tau} - 1)}{X(\tau)(e^{\hat{b}T} - e^{\hat{b}\tau})}\right) && (1.11)
\end{aligned}$$

There is no closed form solution for $\hat{b}$, so (1.10) must be solved numerically. To establish the existence of $\hat{b}$, we note if $X(\tau) = m$, then $(n - m)\tau \leq \sum_{i=1}^{n} T_i \leq$

$m\tau + (n-m)T$, and the right hand side of (1.10) converges to these limits as $\hat{b} \to -\infty$ and $\hat{b} \to \infty$. The derivative of (1.10) with respect to $\hat{b}$ is positive and hence $\hat{b}$ is unique. Although (1.10) (and many other equations in this thesis) takes the form $0/0$ or $\infty - \infty$ when $\hat{b} = 0$, we can define (1.10) at $\hat{b} = 0$ by continuity.

We find the maximum likelihood estimate $\hat{\tau}$ of τ by maximizing $L(\tau, \hat{a}, \hat{\delta}, \hat{b})$ over an interval $[\tau_0, \tau_1]$. A formal definition of $\hat{\tau}$ is similar to (1.7).

The log-likelihood ratio process for the log-linear model is

$$\begin{aligned} l(t) &= X(t)\log\left(\frac{\hat{b}X(t)(e^{\hat{b}_0 T}-1)}{\hat{b}_0 n(e^{\hat{b}t}-1)}\right) + (n - X(t))\log\left(\frac{\hat{b}(n-X(t))(e^{\hat{b}_0 T}-1)}{\hat{b}_0 n(e^{\hat{b}T}-e^{\hat{b}t})}\right) \\ &\quad +(\hat{b}-\hat{b}_0)\sum_{i=1}^{n} T_i. \end{aligned} \tag{1.12}$$

Note that in (1.12), $\hat{b}$ depends on t, and $\hat{b}_0$ is the maximum likelihood estimate of b under $\mathcal{H}_0$ and satisfies

$$\sum_{i=1}^{n} T_i = n\left(\frac{Te^{\hat{b}_0 T}}{e^{\hat{b}_0 T}-1} - \frac{1}{\hat{b}_0}\right). \tag{1.13}$$

Again, the likelihood ratio test rejects $\mathcal{H}_0$ if the maximum of the likelihood ratio statistic is too large. The significance level is defined by

$$\alpha = P_0\left\{\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{c^2}{2} \,\middle|\, X(T) = n, \sum_{i=1}^{n} T_i = ny\right\}. \tag{1.14}$$

Under $\mathcal{H}_0$ there are two nuisance parameters a and b. Since $X(T)$ and $\sum_{i=1}^{n} T_i$ are sufficient statistics for these parameters, conditioning on these in (1.14) removes the dependency on a and b.

1.4 Summary

Equations (1.9) and (1.14) are both written as the probability that a stochastic process $l(t)$ exceeds a level $\frac{1}{2}c^2$. To relate α and c we need to be able to evaluate or approximate these boundary crossing probabilities. In Chapter 2, we present a collection of general methods that are useful for boundary crossing calculations. After reviewing known results for Gaussian processes, we derive an approximate method for evaluating boundary crossing probabilities for a wide class of locally Poisson processes.

The methods are applied to obtain significance level and power approximations for the likelihood ratio tests in Chapter 3. In Chapter 4, we derive confidence regions for the change point and joint confidence regions for the change point and other parameters. These give rise to similar boundary crossing problems, which are also approximated by the methods of Chapter 2.

The remaining Chapters discuss a variety of related problems and techniques. In Chapter 5, we discuss a model from survival analysis, where the hazard rate is assumed to change at an unknown time. This model is closely related to the constant parameter model for Poisson processes.

The Kolmogorov-Smirnov test is probably the most widely used statistical procedure which involves boundary crossings. When the null distribution is completely specified, several exact and approximate methods to evaluate the significance level are available. In practice, the null distribution will usually involve unknown parameters, and very few results for this case have been published. In Chapter 6 we apply our boundary crossing methods to this problem, which provides a vast improvement over existing techniques when there are nuisance parameters.

In Chapter 7 we consider some generalizations of change point problems. In particular, we consider circular problems, for which there are two change points, and problems on planar regions, where we have a change region; that is, a region with a higher event rate. The results have heuristic extensions to higher dimensions which will be deferred until Appendix A.

Finally, in Chapter 8 we contrast our methods with other approaches to change point problems. We also mention some extensions and other statistical problems where our methods may be useful. These methods are aimed at reducing strong parametric assumptions. We discuss mixing local fitting with change point estimation, recursive partitioning of multidimensional processes and some permutation methods that may be appropriate for non-Poisson renewal processes.

Many technical details are deferred to appendices. In Appendix A, we will give a rigorous treatment of our boundary crossing approximations, and derive second order corrections. Appendix B studies large deaviation approximations; in particular, we are interested in approximating conditional distributions arising from sums of independent random vectors. We will study computational methods for implementing various procedures in Appendix C.

We use two datasets to illustrate the tests and confidence intervals developed. Lucas (1985) gives the dates of 178 industrial accidents at a manufacturing plant between 1970 and 1980. Lucas analyses this dataset using a cusum procedure, and concludes there is a change point. However, this dataset seems more naturally suited to fixed sample rather than sequential procedures. The cumulative accident count is shown in Figure 1.1. The rate does not appear to be constant, with many more events occurring in the first five years than in the last 5 years.

The second dataset has been studied by several authors. Maguire, Pearson and Wynn (1952) give the dates of accidents in British coal mines resulting in 10 or more deaths. The original dataset was extended and corrected by Jarrett (1979). This version lists 191 accidents between 15 March, 1851 and 22 March, 1962. The cumulative accident count is shown in Figure 1.2. We see there is clear evidence of inhomogeneity. Cox and Lewis (1966) fit parametric models (without change points) to the original data and conclude that both linear and quadratic terms are

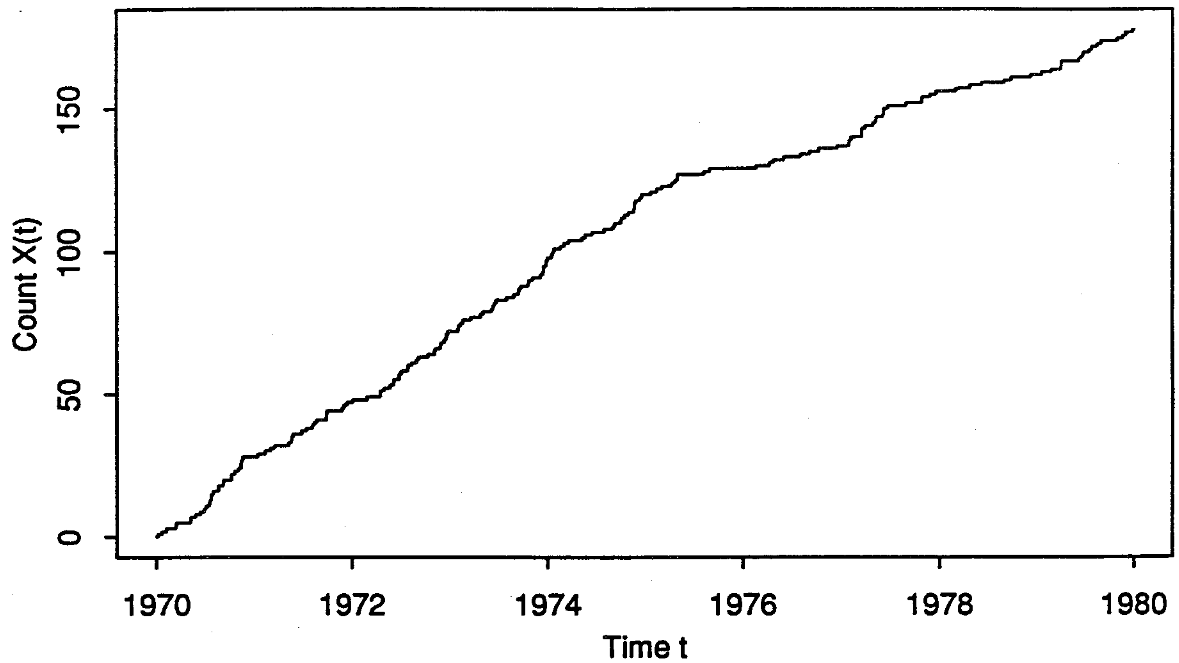


Figure 1.1: Lucas' Industrial Accident Data

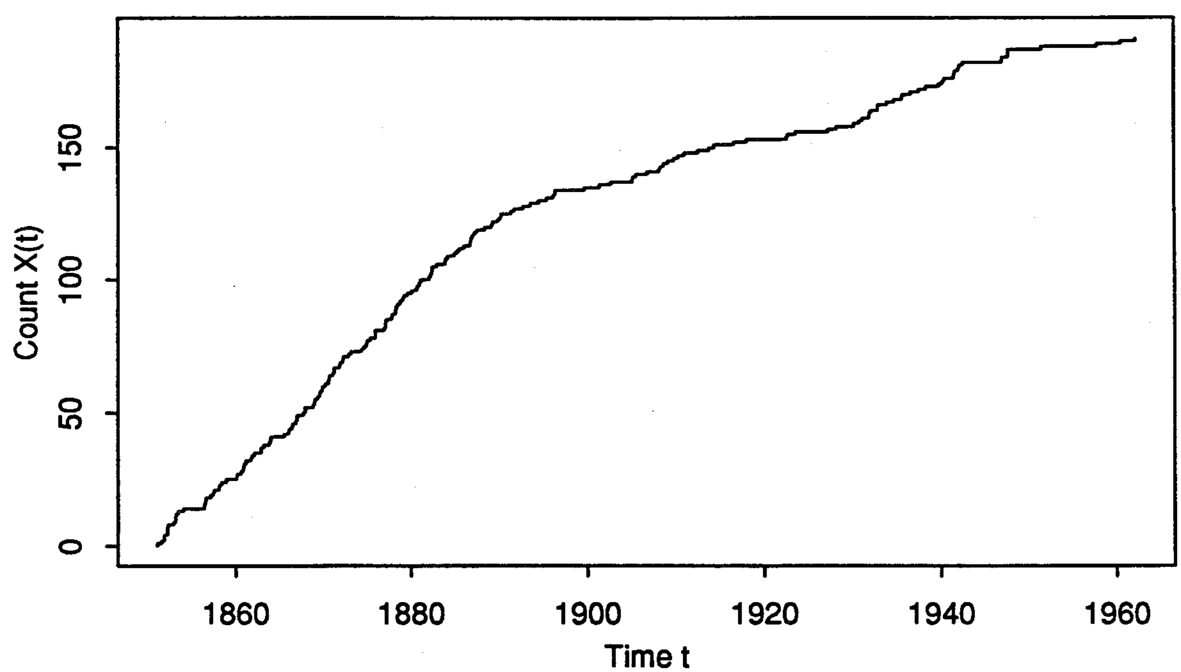


Figure 1.2: British Coal Mining Accident Data

significant in the model $\log(\lambda(t)) = a + b_1 t + b_2 t^2$. Akman and Raftery (1986), using a Bayesian approach, fit the constant parameter model (1.1) and conclude both that the change point is significant and the model fits the data better than the parametric models considered by Cox and Lewis.

In Chapter 5, we will use Stanford heart transplant data to illustrate the hazard rate change point model. This dataset has not been analyzed with a change point model previously.

Chapter 2

Boundary Crossing Probabilities

In this chapter we discuss methods for approximating boundary crossing probabilities. We begin with a discussion of results for a class of Gaussian processes that behave locally like a Brownian motion. Since the results here are not new, we refer to other work for proofs. In later sections, we present methods that are directly applicable to our discussion of Poisson processes, first for linear boundaries and then using large deviation methods to derive results for nonlinear boundaries.

2.1 Gaussian Processes

Let $\{X(t), 0 \leq t \leq T\}$ be a Gaussian process with mean 0 and covariance function $\sigma(s,t)$. If $X(t)$ behaves locally like a Brownian motion, Durbin (1985) gives the following result for the density $p(t)$ of the first passage time $\mathcal{T} = \inf\{t : X(t) \geq a(t)\}$ for a smooth boundary $a(t)$:

$$p(t) = \lim_{s \to t^-} \frac{1}{t-s} E((a(s) - X(s))I(s,X)|X(t) = a(t))f(t) \tag{2.1}$$

where

$$I(s, X) = \begin{cases} 1 & X(u) < a(u), 0 < u \leq s \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

and $f(t)$ is the density of $X(t)$ on the boundary $a(t)$ at time t. The expectation in (2.1) is usually difficult to evaluate. Hence Durbin develops a series of approximations. The first (and simplest) approximation is obtained by setting the indicator function $I(s, X)$ equal to 1. This gives the approximation

$$\begin{aligned} p_1(t) &= \lim_{s \to t^-} \frac{a(s) - E(X(s)|X(t) = a(t))}{t - s} f(t) \\ &= \left(\frac{a(t)}{\sigma(t,t)} \left. \frac{\partial \sigma(s,t)}{\partial s} \right|_{s \to t^-} - a'(t) \right) f(t). \end{aligned} \tag{2.3}$$

Definition 2.1 *Suppose we have a family of boundaries $a_\theta(t), \theta > 0$. Usually we will be concerned with families that recede to ∞ as $\theta \to \infty$. Let $p_\theta(t)$ be the first passage density of $X(t)$ to $a_\theta(t)$ and $\tilde{p}_\theta(t)$ be approximations to $p_\theta(t)$. Then $\tilde{p}_\theta(t)$ will be said to be a* **first order asymptotic** *approximation if*

$$\tilde{p}_\theta(t) = p_\theta(t)(1 + o(1)) \quad \text{as } \theta \to \infty. \tag{2.4}$$

Moreover, if (2.4) holds uniformly in t, then the approximation is said to be **uniformly first order asymptotic**. *In this case, we can obtain first order asymptotic approximations to the distribution of T by integrating $\tilde{p}_\theta(t)$.*

Under fairly general conditions, Jennen and Lerche (1981) show for Brownian motion that (2.3) provides a uniformly first order asymptotic approximation to $p(t)$. Since (2.3) is based on local approximations, similar results will presumably hold for other Gaussian processes for which (2.1) holds.

Our main interest will be in boundaries defined only for $t \geq t_0 > 0$. In this case, Jennen and Lerche (1981) show that (2.3) still provides a uniform first order asymptotic approximation provided t is bounded away from t_0. In many cases, integrating

(2.3) still provides a first order asymptotic approximation to the distribution of $\mathcal{T}$, and correction for the truncation appears in second order terms. See Jennen (1985) for more discussion of this point. However, it seems reasonable to incorporate the marginal probability of exceeding the boundary at time t_0, giving the approximation to the distribution of $\mathcal{T}$:

$$P(\mathcal{T} < t) = P(X(t_0) > a(t_0)) + \int_{t_0}^{t} p_1(u)du. \tag{2.5}$$

In the special case $a(t) = c\sqrt{\sigma(t,t)}$, we get

$$a'(t) = \frac{c}{2\sqrt{\sigma(t,t)}} \left(\frac{\partial \sigma(s,t)}{\partial s} + \frac{\partial \sigma(s,t)}{\partial t} \right)_{s \to t-}$$

and therefore

$$p_1(t) = \frac{c\phi(c)}{2\sigma(t,t)} \left(\frac{\partial \sigma(s,t)}{\partial s} - \frac{\partial \sigma(s,t)}{\partial t} \right)_{s \to t-}. \tag{2.6}$$

The difference of the partial derivatives of $\sigma(s,t)$ in (2.6) is a measure of the quadratic variability of $X(t)$. For processes which behave locally like Brownian motion, this will be strictly positive.

For two boundary problems, approximations to the hitting probability are obtained by evaluating (2.5) for each of the boundaries and adding the resulting approximations. This neglects the probability of crossing both boundaries; however, for problems of statistical interest this is usually negligible.

Numerical methods for evaluating boundary crossing probabilities for Gaussian processes have been discussed by several authors, usually based on solutions of integral equations. A particularly useful method is that of Park and Schuurmann (1976). They discuss only Brownian motion (both one and two boundary problems), but modification of their method to general Gaussian Markov processes and truncated stopping rules is straightforward. We will discuss the method in detail in Appendix C.

2.2 Poisson Processes and Linear Boundaries

In this section we evaluate

$$P\left\{\sup_{t>0}(aX(t)+\gamma t) > c\right\} \tag{2.7}$$

where $X(t)$ is a Poisson process with constant rate λ and a, γ and c are constants with $c \geq 0$ and $a\lambda + \gamma < 0$. With these constraints, we know (cf. Chung (1974) Theorem 8.3.4) that (2.7) is less than 1, at least for c large enough. There are two distinct cases which we consider separately. The simpler case is $a < 0$ and $\gamma > 0$, for which the jumps of $aX(t) + \gamma t$ are downward and hence the process cannot overshoot c. This case is treated in Theorem 2.1. The case $a > 0$ and $\gamma < 0$ are more complicated. An exact expression is given in Theorem 2.2 below. However, this is not always suitable for numerical computation, and an alternative approximation is derived in Theorem 2.3.

These results are not new; similar results may be found in Dvoretzky, Kiefer and Wolfowitz (1953) who consider the two boundary case and Pyke (1959). The approximation given in Theorem 2.3 is used in Siegmund (1988a). However, our approach appears to be simpler than that taken by other authors.

Theorem 2.1 *Suppose $a < 0$ and $\gamma > 0$. Then (2.7) is equal to*

$$\exp\left(\frac{c}{\gamma}(\lambda' - \lambda)\right) \tag{2.8}$$

where

$$\frac{a\lambda}{\gamma} + \log(\lambda) = \frac{a\lambda'}{\gamma} + \log(\lambda') \tag{2.9}$$

and $\lambda' < \frac{\gamma}{|a|}$.

Proof: Let $\mathcal{M} = \sup_{t>0} aX(t) + \gamma t$ and $\mathcal{T}_c = \inf\{t : aX(t) + \gamma t > c\}$. Since $a < 0$ implies all jumps of $aX(t) + \gamma t$ are downward, we apply the strong Markov property to the process restarted at time $\mathcal{T}_c$ to get

$$P(\mathcal{M} > c + d | \mathcal{M} > c) = P(\mathcal{M} > d).$$

But this lack of memory property characterizes the exponential distribution. Therefore,

$$P(\mathcal{M} > c) = e^{-\beta c} \tag{2.10}$$

for some $\beta > 0$. For small δ ($c > \gamma\delta$), we get

$$\begin{aligned}
e^{-\beta c} &= P(\mathcal{M} > c) \\
&= P(\mathcal{M} > c | X(\delta) = 0)e^{-\lambda\delta} + P(\mathcal{M} > c | X(\delta) = 1)\lambda\delta e^{-\lambda\delta} + o(\delta) \\
&= e^{-\beta(c-\gamma\delta)}e^{-\lambda\delta} + e^{-\beta(c-a-\gamma\delta)}\lambda\delta e^{-\lambda\delta} + o(\delta) \\
1 &= e^{(\beta\gamma-\lambda)\delta} + \lambda\delta e^{a\beta} + o(\delta).
\end{aligned}$$

Letting $\delta \to 0$, we get

$$\lambda e^{a\beta} = \lambda - \gamma\beta$$

which reduces to (2.9) if we let $\beta = \frac{1}{\gamma}(\lambda - \lambda')$, and (2.10) gives (2.8). Since $f(\lambda) = \frac{a\lambda}{\gamma} + \log(\lambda)$ is concave and maximized at $\gamma\lambda = |a|$, (2.9) uniquely determines λ'.

□

Lemma 2.1 *Suppose $a > 0$ and $\gamma < 0$. Let $\mathcal{T}_+ = \inf\{t : aX(t) + \gamma t > 0\}$. Then*

$$P(\mathcal{T}_+ < \infty) = \frac{\lambda a}{|\gamma|} \tag{2.11}$$

and conditional on $\mathcal{T}_+ < \infty$, $aX(\mathcal{T}_+) + \gamma\mathcal{T}_+$ is distributed uniformly on $[0, a]$.

Proof: Note $\mathcal{T}_+ < \infty$ is equivalent to $\sup_{j>0} aj + \gamma T_j > 0$. Also, if $S_j = aj + \gamma T_j$, then $\{S_j, j \geq 0\}$ is a random walk, and S_1 has density

$$\frac{\lambda}{|\gamma|} e^{\frac{\lambda}{|\gamma|}(x-a)}$$

for $x < a$. Let $\tau_+ = \inf\{n : S_n > 0\}$, $\tau_- = \inf\{n \geq 1 : S_n \leq 0\}$ and

$$G_\pm(\theta) = E(e^{i\theta S_{\tau_\pm}}; \tau_\pm < \infty).$$

We have

$$E(e^{i\theta S_1}) = e^{i\theta a} E(e^{i\theta\gamma T_1}) = \frac{\lambda e^{i\theta a}}{\lambda - i\theta\gamma}$$

and since S_1 has an exponential lower tail, so does S_{τ_-}. Therefore,

$$G_-(\theta) = \frac{\lambda}{\lambda - i\theta\gamma}. \tag{2.12}$$

Applying the Wiener-Hopf factorization (cf. Siegmund (1985) Theorem 8.41) we get

$$(1 - G_+(\theta))\left(1 - \frac{\lambda}{\lambda - i\theta\gamma}\right) = 1 - \frac{\lambda}{\lambda - i\theta\gamma} e^{i\theta a}$$

and therefore

$$G_+(\theta) = \frac{\lambda a}{|\gamma|} \frac{e^{i\theta a} - 1}{i\theta a}.$$

Letting $\theta \to 0$ gives (2.11). Dividing by $\frac{\lambda a}{|\gamma|}$ shows the conditional distribution of $aX(\mathcal{T}_+) + \gamma\mathcal{T}_+$ has characteristic function $\frac{1}{i\theta a}(e^{i\theta a} - 1)$ and therefore is distributed uniformly on $[0, a]$.

□

Theorem 2.2 *Suppose $a > 0$ and $\gamma < 0$. Then (2.7) equals*

$$h\left(\frac{\lambda a}{|\gamma|}, \frac{\mu a}{|\gamma|}\right) = 1 - \left(1 - \frac{\lambda a}{|\gamma|}\right) \sum_{j=0}^{\infty} \left(\frac{\lambda a}{|\gamma|}\right)^j P(aU_j < c) \tag{2.13}$$

where U_j is the convolution of j independent $\mathcal{U}[0,1]$ random variables.

Proof: Let J be the number of finite ascending ladder times. Then J has a geometric distribution, and by Lemma 2.1,

$$P(J=j) = \left(1-\frac{\lambda a}{|\gamma|}\right)\left(\frac{\lambda a}{|\gamma|}\right)^j, \quad j=0,1,\ldots.$$

Also by Lemma 2.1, $P(\mathcal{M}<c|J=j) = P(aU_j<c)$, from which the result follows.

□

We can obtain an alternative form for (2.13) using the distribution of U_j given by (B.19). Writing $\frac{c}{a}=m+y$ where m is an integer and $0\le y<1$, we get

$$\begin{aligned} P(aU_j<c) &= \frac{1}{(j-1)!}\sum_{k=0}^{m\wedge j}\binom{j}{k}(-1)^k\int_k^{m+y}(x-k)^{j-1}dx \\ &= \frac{1}{j!}\sum_{k=0}^{m\wedge j}\binom{j}{k}(-1)^k(m+y-k)^j \end{aligned}$$

and hence

$$\begin{aligned} \sum_{j=0}^{\infty}\left(\frac{\lambda a}{|\gamma|}\right)^j P(aU_j<c) &= \sum_{k=0}^{[c/a]}\sum_{j=k}^{\infty}\frac{1}{j!}\binom{j}{k}(-1)^k\left(\frac{\lambda a}{|\gamma|}(\frac{c}{a}-k)\right)^j \\ &= \sum_{k=0}^{[c/a]}\frac{(-1)^k}{k!}\left(\frac{\lambda}{|\gamma|}(c-ak)\right)^k e^{\frac{\lambda}{|\gamma|}(c-ak)}. \end{aligned} \tag{2.14}$$

Substituting (2.14) into (2.13) expresses the result of Theorem 2.3 as a finite oscillating sum. If $\frac{\lambda c}{|\gamma|}$ is large, the summation (2.14) may be unstable and hence should be used with care. This expression is given by Pyke (1959).

Theorem 2.3 *Suppose $a>0$ and $\gamma<0$. Then*

$$P\left\{\sup_{t>0}(aX(t)+\gamma t)>c\right\} \sim \frac{1-\frac{\lambda a}{|\gamma|}}{\frac{\lambda' a}{|\gamma|}-1}\exp\left(\frac{c}{\gamma}(\lambda'-\lambda)\right) \quad \textit{as } c\to\infty$$

where λ' is defined by (2.9) and $\lambda' > \frac{|\gamma|}{a}$.

Proof: Let $\mathcal{T}_c = \inf\{j : S_j > c\}$ where S_j is as defined in the proof of Lemma 2.1. Then by Wald's likelihood ratio identity (cf. Siegmund (1985) p13),

$$
\begin{aligned}
P_\lambda(\mathcal{T}_c < \infty) &= \sum_{n=1}^{\infty} \int_{\{\mathcal{T}_c = n\}} \frac{dP_\lambda}{dP_{\lambda'}} dP_{\lambda'} \\
&= \sum_{n=1}^{\infty} \int_{\{\mathcal{T}_c = n\}} \left(\frac{\lambda}{\lambda'}\right)^n e^{-\frac{1}{\gamma}(\lambda-\lambda')(S_n - an)} dP_{\lambda'} \\
&= e^{-\frac{c}{\gamma}(\lambda-\lambda')} E_{\lambda'}(\exp(-\frac{1}{\gamma}(\lambda-\lambda')(S_{\mathcal{T}_c} - c))) \qquad (2.15) \\
&\sim e^{-\frac{c}{\gamma}(\lambda-\lambda')} \frac{1 - E_{\lambda'}(\exp(-\frac{1}{\gamma}(\lambda-\lambda')S_{\tau_+}))}{\frac{1}{\gamma}(\lambda-\lambda')E_{\lambda'}(S_{\tau_+})} \quad \text{as } c \to \infty. \qquad (2.16)
\end{aligned}
$$

We have used (8.46) of Siegmund (1985) to obtain (2.16) from (2.15). We can evaluate (2.16) by likelihood ratio arguments. Using Wald's equation, we have

$$
E_{\lambda'}(S_{\tau_+}) = E_{\lambda'}(\tau_+)E_{\lambda'}(S_1) = \frac{a + \frac{\gamma}{\lambda'}}{P_{\lambda'}(\tau_- = \infty)}
$$

and

$$
\begin{aligned}
P_{\lambda'}(\tau_- = \infty) &= 1 - E_\lambda\left\{\exp(\frac{1}{\gamma}(\lambda-\lambda')S_{\tau_-})\right\} \\
&= 1 - \frac{\lambda}{\lambda'}
\end{aligned}
$$

by (2.12), so

$$
E_{\lambda'}(S_{\tau_+}) = \frac{a\lambda' + \gamma}{\lambda' - \lambda}. \qquad (2.17)
$$

Similarly,

$$
\begin{aligned}
E_{\lambda'}\left\{\exp(-\frac{1}{\gamma}(\lambda-\lambda')S_{\tau_+})\right\} &= P_\lambda(\tau_+ < \infty) \\
&= \frac{\lambda a}{|\gamma|} \qquad (2.18)
\end{aligned}
$$

by Lemma 2.1. Substituting (2.17) and (2.18) into (2.16) completes the proof.

□

Theorem 2.3 provides an approximation to probabilities expressed exactly by Theorem 2.2 and (2.14). When $[c/a]$ is large the series in (2.13) converges slowly and the summation (2.14) is unstable. In this situation the approximate form performs well. Numerical studies suggest the approximate form should be sufficiently accurate for most purposes whenever $c > 2a$. The approximate form is also useful for some of our power and confidence region calculations in later chapters.

Suppose now that $\lambda \to \infty$ and $\gamma \to \infty$ such that $\frac{\lambda}{\gamma}$ is fixed. Then (2.8) and (2.13) are invariant. However, the drift of $aX(t) + \gamma t$ is tending to $-\infty$, so if the process does exceed c, we would expect the crossing to be for small t. This suggests that under suitable regularity conditions, the results of this section will hold asymptotically when c is a function of t and $X(t)$ is only locally Poisson at $t = 0$. We will study this further in Appendix A.

2.3 Non-Linear Boundary Crossing Probabilities for Poisson Processes

Suppose now $\{X(t), 0 \leq t \leq T\}$ is a homogeneous Poisson process, and the continuous differentiable boundary $a(t)$ satisfies $a(t) < t/T$ and $ta'(t) > a(t)$ for all t. For events A in the σ-field generated by $\{X(t), 0 \leq t \leq T\}$ we define

$$P^{(n)}(A) = P(A|X(T) = n).$$

The first passage time is $\mathcal{T} = \inf\{t : X(t) \leq na(t)\}$. For constants $\tau_0 \leq \tau_1$ we are interested in approximating the probability

$$P^{(n)}(\tau_0 \leq \mathcal{T} \leq \tau_1). \tag{2.19}$$

Our approach taken to approximate (2.19) will be based on techniques developed by Woodroofe (1976). We first note that $\mathcal{T}$ has a discrete distribution and hence (2.19) can be written as a finite sum. The individual terms are then approximated using a local linearization and random walk results.

Let $m = [na(T)]$ and for $0 \leq j < m$ we define t_j to be the unique solution of $j = na(t)$. Since the boundary $a(t)$ is increasing and $X(t)$ is increasing and integer valued, $\mathcal{T}$ can only occur at t_j for some j. Choose integers $m_0 = \lceil na(\tau_0) \rceil$ and $m_1 = \lfloor na(\tau_1) \rfloor$ such that $\tau_0 \leq t_j \leq \tau_1$ for $m_0 \leq j \leq m_1$. We can write

$$\{\mathcal{T} < T\} = \cup_{j=m_0}^{m_1} \{\mathcal{T} = t_j\}. \tag{2.20}$$

Note the union in (2.20) is a disjoint union. We therefore have

$$\begin{aligned} P^{(n)}(\tau_0 \leq \mathcal{T} < \tau_1) &= \sum_{j=m_0}^{m_1} P^{(n)}(\mathcal{T} = t_j) \\ &= \sum_{j=m_0}^{m_1} P^{(n)}(\mathcal{T} = t_j, X(t_j) = j) \\ &= \sum_{j=m_0}^{m_1} P^{(n)}(\mathcal{T} = t_j | X(t_j) = j) P^{(n)}(X(t_j) = j). \end{aligned} \tag{2.21}$$

The conditional distribution of $X(t_j)$ given $X(T)$ is binomial, so we can evaluate $P^{(n)}(X(t_j) = j)$. Suppose we approximate $a(t)$ by its tangent at t_j, $a^*(t_j - t) = a(t_j) - ta'(t_j)$, and approximate $\{X(t_j) - X(t_j - t), t \geq 0\}$ by a Poisson process with rate $\lambda = \frac{j}{t_j}$. Then Lemma 2.1 gives the approximation

$$P^{(n)}(\mathcal{T} = t_j | X(t_j) = j) \approx 1 - \frac{a(t_j)}{t_j a'(t_j)}. \tag{2.22}$$

A more formal justification for (2.22) will be given in Appendix A, where we show (2.22) holds as a limiting relation if $n \to \infty$ with t_j fixed. Combining (2.22) with (2.21) gives

$$P^{(n)}(\tau_0 \leq \mathcal{T} < \tau_1) \approx \sum_{j=m_0}^{m_1} \left(1 - \frac{a(t_j)}{t_j a'(t_j)}\right) \binom{n}{j} \left(\frac{t_j}{T}\right)^{na(t_j)} \left(\frac{T - t_j}{T}\right)^{n(1-a(t_j))} \tag{2.23}$$

$$\approx \int_{\tau_0}^{\tau_1} \left(a'(t) - \frac{a(t)}{t}\right) \frac{n\Gamma(n+1)}{\Gamma(na(t)+1)\Gamma(n(1-a(t))+1)}$$
$$\left(\frac{t}{T}\right)^{na(t)} \left(\frac{T-t}{T}\right)^{n(1-a(t))} dt$$
$$\approx \int_{\tau_0}^{\tau_1} \left(a'(t) - \frac{a(t)}{t}\right) \sqrt{\frac{n}{2\pi a(t)(1-a(t))}}$$
$$\left(\frac{t}{Ta(t)}\right)^{na(t)} \left(\frac{T-t}{T(1-a(t))}\right)^{n(1-a(t))} dt. \qquad (2.24)$$

In Appendix A, we will establish rigorously that under suitable regularity conditions, (2.24) provides a first order asymptotic approximation to the probability of crossing $a(t)$. Similar results hold for truncated stopping times, when we define $\mathcal{T}(\tau_0) = \inf\{t \geq \tau_0 : X(t) \leq na(t)\}$. In this case we need to study the endpoint correction $P^{(n)}(X(\tau_0) \leq na(\tau_0))$. Since $X(\tau_0)$ has a binomial distribution, this can be evaluated either by summing the appropriate binomial probabilities or using the tail probability approximation given by Lemma B.2. In many cases, (2.24) still provides first order asymptotic approximations for truncated stopping times, and the endpoint correction appears as a higher order term. A further discussion of second order corrections is deferred to Appendix A.

We note this method has some similarity between the method developed here for Poisson processes and Durbin's (1985) result for Gaussian processes. Letting $I(s, X)$ be defined by (2.2), we have

$$P^{(n)}(\mathcal{T} = t_j | X(t_j) = j)$$
$$= \lim_{s \to t_j^-} E(I(s,X) | X(t_j) = j)$$
$$= \frac{1}{na'(t_j)} \lim_{s \to t_j^-} \frac{n}{t_j - s} E\left((a(t_j) - a(s)) I(s,X) | X(t_j) = j\right)$$
$$= \frac{1}{na'(t_j)} \lim_{s \to t_j^-} \frac{1}{t_j - s} E\left((X(s) - na(s)) I(s,X) | X(t_j) = j\right) \qquad (2.25)$$

since if $X(t_j) = j$, $I(s, X) = 1$ and $s > t_{j-1}$ then $X(s) = j = na(t_j)$. Therefore, (2.21) is a version of (2.1) for Poisson processes. The extra factor of $na'(t_j)$ in (2.25) disappears when we approximate the sum over j by an integral over t. Our tangent approximation (2.22) can be obtained by deleting the indicator function $I(s, X)$ from (2.25) and therefore we interpret (2.24) as a Poisson process version of (2.6).

The local tangent approximation (2.22) can easily be extended to locally Poisson processes. Informally, by a locally Poisson process, we mean that for any t and small values of s, the process $\{X(t) - X(t-s); s > 0\}$ behaves like a Poisson process. We will make this definition more rigorous in Appendix A. The local event rate is given by

$$n\mu(t, a(t)) = \lim_{s \to t^-} \frac{1}{t-s} E(X(t) - X(s) | X(t) = na(t))$$

and we have $f(t) = P(X(t) = na(t))$ and a continuous approximation $f_c(t)$ to $f(t)$. Then (2.22) becomes

$$P(\mathcal{T} = t_j | X(t_j) = na(t_j)) \approx 1 - \frac{\mu(t, a(t))}{a'(t)}$$

and (2.24) becomes

$$P^{(n)}(\mathcal{T} < T) \approx \int_{t_0}^{T} n(a'(t) - \mu(t, a(t))) f_c(t) dt. \tag{2.26}$$

Particular examples of where such an extension is useful is in analysis of the likelihood ratio test for the log-linear model (see Section 3.2) and in our discussion of Kolmogorov-Smirnov testing with nuisance parameters in Chapter 6. We will discuss methods for deriving suitable approximations for $\mu(t, a(t))$ and $f_c(t)$ in Appendix B.

2.4 Recursive Formulae

Suppose we observe a Poisson process $\{X(t), 0 < t < T\}$, conditional on $X(T) = n$. We are interested in evaluating probabilities of the form

$$P(a_i \leq T_i \leq b_{i-1}, i = 1, \ldots, n). \tag{2.27}$$

We assume $0 \leq a_1 \leq \ldots \leq a_n \leq T$, $0 \leq b_0 \leq \ldots \leq b_{n-1} \leq T$ and $a_i \leq b_{i-1}, 1 \leq i \leq n$. Our interest in (2.27) arises because the significance level (1.9) and various other probabilities related to the constant parameter model can be expressed in this form. We can without loss of generality assume $X(t)$ is homogeneous; if a non-homogeneous rate function $\lambda(t)$ is specified up to a multiplicative constant, then the probability integral transform can be used to transform to the homogeneous case.

There are several algorithms for evaluating (2.27); see for example Section 9.3 of Shorack and Wellner (1986) and Noé (1972). The method we derive here is essentially a two boundary version of the algorithm of Durbin (1973a), Section 2.5.

We define $\mathcal{T}_+ = \inf\{j : X(a_j) = j\}$ and $\mathcal{T}_- = \inf\{j : X(b_j) = j\}$. Let $q_j = P(\mathcal{T}_+ = j, \mathcal{T}_- > j)$ and $p_j = P(\mathcal{T}_- = j, \mathcal{T}_+ > j)$. Then, using the strong Markov property, we get

$$\begin{aligned}
q_j &= P(X(a_j) = j) && -P(X(a_j) = j, \mathcal{T}_+ < j, \mathcal{T}_- \geq \mathcal{T}_+) \\
& && -P(X(a_j) = j, \mathcal{T}_- < j, \mathcal{T}_+ > \mathcal{T}_-) \\
&= P(X(a_j) = j) && -\sum_{i=1}^{j-1} P(X(a_j) = j | X(a_i) = i) q_i \\
& && -\sum_{i=0}^{j-1} P(X(a_j) = j | X(b_i) = i) p_i
\end{aligned} \tag{2.28}$$

and similarly

$$p_j = P(X(b_j) = j) \quad -\sum_{i=1}^{j} P(X(b_j) = j | X(a_i) = i) q_i$$

$$-\sum_{i=0}^{j-1} P(X(b_j) = j | X(b_i) = i) p_i. \tag{2.29}$$

The kernels of (2.28) and (2.29) are binomial probabilities. For example,

$$P(X(b_j) = j | X(a_i) = i) = \binom{n-i}{j-i} \left(\frac{b_j - a_i}{T - a_i}\right)^{j-i} \left(\frac{T - b_j}{T - a_i}\right)^{n-j}.$$

Equations (2.28) and (2.29) enable us to compute in order $q_0, p_1, q_1, p_2, \ldots, q_{n-1}, p_n$. The probability (2.27) is given by

$$1 - \sum_{j=1}^{n} q_j - \sum_{j=0}^{n-1} p_j.$$

The computation of p_j and q_j involve sums of $O(j)$ terms. Summing over j, we see this is an $O(n^2)$ algorithm, with the amount of computation essentially independent of the boundaries. By contrast, the amount of computation needed to evaluate Noé's (1972) algorithm depends on the configuration. If a_i is the solution of $na(t) = i$ and b_j is the solution of $nb(t) = j$ for some smooth functiona a and b, then Noé's algorithm is $O(n^3)$. However, if $a(t) - b(t) \to 0$ as $n \to \infty$, then the computation is reduced.

2.5 Random Walk Results

We conclude this chapter with a discussion of some random walk results that will be useful for some of the problems presented in Chapter 7. We assume $X_1, X_2, \ldots$ are a sequence of *i.i.d.* random variables such that $P(X_1 = 1) = p$, $P(X_1 = -1) = q$ and $P(X_1 = 0) = r$ with $p + q + r = 1$ and $p < q$. We let $\{S_n; n \geq 0\}$ be the associated random walk. All the techniques are reasonably standard so we only sketch proofs. See Feller (1966) or Chung (1974) for more details.

Lemma 2.2 *Let $p_j = P(S_n = j$ for some $n \geq 0)$. Then*

$$p_j = \begin{cases} \left(\frac{p}{q}\right)^j & j \geq 0 \\ 1 & j < 0 \end{cases}.$$

Proof: For $j \geq 1$, we have

$$\begin{aligned} p_j &= E(P(S_n = j \text{ for some } n \geq 0 | X_1)) \\ &= pp_{j-1} + rp_j + qp_{j+1} \\ \Rightarrow 0 &= pp_{j-1} - (p+q)p_j + qp_{j+1}. \end{aligned} \tag{2.30}$$

The difference equation (2.30) has a general solution

$$p_j = a + b\left(\frac{p}{q}\right)^j,$$

and the boundary conditions $p_j \to 0$ as $j \to \infty$ and $p_0 = 1$ imply $a = 0$ and $b = 1$. For $j < 0$, the strong law of large numbers shows $p_j = 1$.

□

Now let $\tau_+ = \inf\{n \geq 1 : S_n \geq 0\}$ and $\tau_- = \inf\{n : S_n < 0\}$. Then

$$\begin{aligned} P(S_n = 0 \text{ for some } n \geq 1) &= P(\tau_+ < \infty) \\ &= 1 - \frac{1}{E(\tau_-)} \\ &= 1 - \frac{p-q}{-1} \\ &= 1 - |p-q|. \end{aligned} \tag{2.31}$$

By a geometric distribution argument, the expected number of n for which $S_n = 0$ (including $n = 0$) is $\frac{1}{|p-q|}$ and by a strong Markov argument, the expected number of n for which $S_n = j$ is $\frac{p_j}{|p-q|}$. The expected number of n for which $S_n \geq 0$ is therefore

$$\begin{aligned} \sum_{n=0}^{\infty} P(S_n \geq 0) &= \sum_{j=0}^{\infty} \frac{p_j}{|p-q|} \\ &= \frac{q}{(p-q)^2}. \end{aligned} \tag{2.32}$$

These random walk results can also be applied to Poisson processes. Suppose $X_1(t)$ and $X_2(t)$ are independent Poisson processes with rates λ and μ, $\lambda < \mu$, and let $Z(t) = X_1(t) - X_2(t)$. Then we can regard $Z(t)$ as a Poisson process with rate $\lambda + \mu$ and random jump sizes $Z_i = Z(T_i) - Z(T_i^-)$ with $P(Z_i = 1) = 1 - P(Z_i = -1) = \frac{\lambda}{\lambda+\mu}$. From Lemma 2.2 we get

$$\begin{aligned} P(Z(t) \leq 0 \text{ for all } t > 0) &= 1 - P(\exists t : Z(t) = 1) \\ &= 1 - \frac{\lambda}{\mu}. \end{aligned} \tag{2.33}$$

Suppose now we observe $Z(t)$ only at times $i\delta, i = 0, 1, 2, \ldots$ for some $\delta > 0$. Since $\{Z(t) \leq 0 \; \forall \; t > 0\} = \lim_{\delta \to 0} \{Z(i\delta) \leq 0, i = 0, 1, \ldots\}$, we have

$$\lim_{\delta \to 0} P(Z(i\delta) \leq 0, i = 0, 1, \ldots) = 1 - \frac{\lambda}{\mu}. \tag{2.34}$$

We also have

$$\begin{aligned} & P(Z(i\delta) < 0, i = 1, 2, \ldots) \\ \leq \; & P(Z(\delta) = -1) P(Z(i\delta) \leq -1, i = 1, 2, \ldots | Z(\delta) = -1) + P(Z(\delta) \leq -2) \\ = \; & \mu\delta(1 - \frac{\lambda}{\mu}) + o(\delta) \\ = \; & \delta(\mu - \lambda) + o(\delta). \end{aligned}$$

A similar lower bound establishes

$$P(Z(i\delta) < 0, i = 1, 2, \ldots) = \delta(\mu - \lambda)(1 + o(1)) \tag{2.35}$$

as $\delta \to 0$.

For our discussion of the Poisson clumping heuristic in Appendix A, we will need

$$\begin{aligned} E \int_0^\infty I(Z(t) = 0) dt &= \frac{1}{\lambda + \mu} \frac{1}{\frac{\mu}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu}} \\ &= \frac{1}{\mu - \lambda} \end{aligned} \tag{2.36}$$

and

$$
\begin{aligned}
E\int_0^\infty I(Z(t) \geq 0)dt &= \frac{1}{\lambda+\mu}\frac{\mu/(\lambda+\mu)}{((\lambda-\mu)/(\lambda+\mu))^2} \\
&= \frac{\mu}{(\lambda-\mu)^2}.
\end{aligned} \tag{2.37}
$$

which we obtain using (2.32) together with the mean delay between jumps of $\frac{1}{\lambda+\mu}$.

Chapter 3

Significance and Power Calculations

The methods of the previous chapter are used to derive approximations to the significance level and power of the likelihood ratio tests for a change point. A variety of methods are used which vary according to computational simplicity and accuracy, and we compare the results.

3.1 Significance Levels for the Constant Parameter Model

Expanding the likelihood ratio statistic (1.8) in a Taylor series around $TX(t) = nt$ shows that under the null hypothesis of no change, for large n,

$$l(t) \approx \frac{u(t)^2}{2n\sigma(t,t)}$$

where

$$u(t) = X(t) - \frac{nt}{T} \tag{3.1}$$

and

$$\sigma(s,t) = \frac{s}{T}\left(1 - \frac{t}{T}\right), \quad 0 \le s \le t \le T. \tag{3.2}$$

Applying central limit type theorems shows as $n \to \infty$, $\frac{1}{\sqrt{n}}u(t)$ converges weakly to a Gaussian process $W(t)$ with mean 0 and covariance function $\sigma(s,t)$. We will not attempt to rigorously establish the weak convergence of $\frac{1}{\sqrt{n}}u(t)$ here, but refer the reader to Billingsley (1968) for details.

The covariance function (3.2) is that of a Brownian Bridge on $[0,T]$. We use (2.5) and (2.6) to obtain a central limit approximation to the significance level of the likelihood ratio test,

$$\begin{aligned} \alpha &\approx P_0^{(n)}\left\{\sup_{\tau_0 \le t \le \tau_1} (|W(t)| - c\sqrt{t(T-t)}) \ge 0\right\} \\ &\approx c\phi(c)\log\left(\frac{\tau_1(T-\tau_0)}{\tau_0(T-\tau_1)}\right) + 2(1-\Phi(c)) \end{aligned} \tag{3.3}$$

where $\phi(x)$ and $\Phi(x)$ are the standard normal density and distribution functions respectively. Here, P_0 denotes a measure under which $X(t)$ is a homogeneous Poisson process on $[0,T]$ and

$$P_0^{(n)}(A) = P_0(A|X(T) = n)$$

for events A in the sigma field $\mathcal{F}_T$ generated by $\{X(t); 0 \le t \le T\}$.

The approximation (3.3) is simple and easy to evaluate. Since it is based only on the covariance function of $X(t)$ this approximation will hold for a wide class of change point problems; however this generality is also a weakness in that the approximation will not always be a good one. Akman and Raftery (1986) have used the Gaussian approximation (3.3) to study the asymptotic behaviour of (3.1).

An alternative approximation can be derived from (2.24). For $0 < t < T$, let

$$f(t,p) = p \log\left(\frac{pT}{t}\right) + (1-p)\log\left(\frac{(1-p)T}{T-t}\right). \tag{3.4}$$

We can write the log-likelihood ratio statistic (1.8) as

$$l(t) = nf\left(t, \frac{X(t)}{n}\right).$$

Differentiating (3.4) we get

$$\frac{\partial f(t,p)}{\partial p} = \log\left(\frac{pT}{t}\right) - \log\left(\frac{(1-p)T}{T-t}\right)$$

and

$$\frac{\partial^2 f(t,p)}{\partial p^2} = \frac{1}{p(1-p)} > 0.$$

Hence for fixed t, $f(t,p)$ is a convex function of p. Moreover,

$$f\left(t, \frac{t}{T}\right) = \left.\frac{\partial f(t,p)}{\partial p}\right|_{p=\frac{t}{T}} = 0.$$

We fix $\eta > 0$ and define p_t and q_t to satisfy

$$f(t,q_t) = f(t,p_t) = \frac{\eta^2}{2} \tag{3.5}$$

subject to $p_t < \frac{t}{T} < q_t$. The convexity of $f(t,p)$ in p implies that if p_t and q_t exist, then they are unique. It is easy to see p_t will exist for $t/T \geq 1 - e^{-\frac{1}{2}\eta^2}$ and q_t will exist for $t/T \leq e^{-\frac{1}{2}\eta^2}$. When they do not exist, we set $p_t = 0$ and $q_t = 1$.

If we let $c = \eta\sqrt{n}$, then $l(t) \leq \frac{1}{2}c^2$ if and only if $np_t \leq X(t) \leq nq_t$. The significance level of the likelihood ratio test can therefore be written

$$\begin{aligned} \alpha &= P_0^{(n)}\left(\inf_{\tau_0 \leq t \leq \tau_1}(X(t) - np_t) < 0 \cup \sup_{\tau_0 \leq t \leq \tau_1}(X(t) - nq_t) > 0\right) \\ &\approx P_0^{(n)}\left(\inf_{\tau_0 \leq t \leq \tau_1}(X(t) - np_t) < 0\right) + P_0^{(n)}\left(\sup_{\tau_0 \leq t \leq \tau_1}(X(t) - nq_t) > 0\right) \end{aligned}$$

where we have neglected the probability of crossing both boundaries. We can use (2.24) directly to approximate the probability of crossing p_t. We get

$$\begin{aligned}
& P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1}(X(t) - np_t) < 0\right) \\
\approx\ & P_0^{(n)}(X(\tau_0) \le np_{\tau_0}) + \int_{\tau_0}^{\tau_1}(p_t' - \frac{p_t}{t})\sqrt{\frac{n}{2\pi p_t(1-p_t)}}e^{-\frac{1}{2}c^2}dt \\
=\ & P_0^{(n)}(X(\tau_0) \le np_{\tau_0}) + c\phi(c)\int_{\tau_0}^{\tau_1}\frac{p_t' - \frac{p_t}{t}}{\eta\sqrt{p_t(1-p_t)}}dt. \qquad (3.6)
\end{aligned}$$

Here, we assume $p_{\tau_0} > 0$; otherwise, we replace τ_0 by $T(1 - \exp(-\frac{1}{2}\eta^2))$, which is the solution of $p_t = 0$. The derivative p_t' can be found by implicit differentiation of (3.4). We get

$$\frac{\partial f(t,p)}{\partial p}\frac{dp}{dt} + \frac{\partial f(t,p)}{\partial t} = 0$$

which gives

$$\frac{dp}{dt} = -\frac{\frac{1-p}{T-t} - \frac{p}{t}}{\log(\frac{pT}{t}) - \log(\frac{(1-p)T}{T-t})}.$$

Since $X(\tau_0) \sim \mathcal{B}(n, \tau_0)$ under $P_0^{(n)}$, we can evaluate the end point correction either by adding the exact probabilities or using the tail approximation given by (B.18).

The probability of crossing q_t can be approximated by time reversal. This gives

$$\begin{aligned}
& P_0^{(n)}\left(\sup_{\tau_0 \le t \le \tau_1}(X(t) - nq_t) > 0\right) \\
\approx\ & P_0^{(n)}(X(\tau_1) \ge nq_{\tau_1}) + c\phi(c)\int_{\tau_0}^{\tau_1}\frac{q_t' - \frac{1-q_t}{T-t}}{\eta\sqrt{q_t(1-q_t)}}dt. \qquad (3.7)
\end{aligned}$$

Adding (3.6) and (3.7) gives an approximation to (1.9), the significance level of the likelihood ratio test. We note that in the special case $\tau_1 = T - \tau_0$, (3.6) and (3.7) are equal.

Second order approximations can be obtained by considering the curvature of p_t to adjust the tangent approximation (2.22). We also need to study the other sources of approximation; in particular, the endpoint corrections. The details will be deferred to Appendix A.

We note the Gaussian approximation (3.3) can be derived from (3.6) by letting $\eta \to 0$. We get

$$\begin{aligned} p_t &= t - \eta\sqrt{t(1-t)} + o(\eta) \\ p_t' &= 1 - \eta\frac{1-2t}{2\sqrt{t(1-t)}} + o(\eta) \end{aligned}$$

which combined with the ordinary central limit theorem gives

$$P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1}(X(t) - np_t) < 0\right) \approx 1 - \Phi(c) + c\phi(c)\int_{\tau_0}^{\tau_1}\frac{1}{2t(1-t)}dt.$$

The significance level (1.9) can also be expressed in the form (2.27). If there is no truncation, we simply take a_j to be the solution (in t) of $nq_t = j$ and b_j to be the solution of $np_t = j$. When there is truncation, we must replace a_j by 0 if $a_j < t_0$, by t_1 if $a_j > t_1$ and replace b_j by t_0 if $b_j < t_0$ and by T if $b_j > t_1$. This technique has been applied to change point problems by Worsley (1986).

In Tables 3.1 and 3.2, we give approximations to the significance level, for $n = 50$ and $n = 100$, $T = 1$ and with 10% and 20% truncation. The two Gaussian approximations represent the approximation (3.3) and the 'exact' Gaussian probability, computed numerically from an integral equation; see Appendix C.

The Gaussian approximations are performing well, even for large values of c. In some cases the Gaussian approximations are better than the large deviation approximations. This is very different to the situation in other statistical problems. For example, Kim (1988) has found the Gaussian approximations are typically about twice the true probability for some discrete time change point problems. We note the difference between the first order Gaussian approximation (3.3) and the exact Gaussian results are usually much less than the overall error, indicating the vastly

	Method	c				
		2.0	2.5	3.0	3.5	4.0
	Gaussian	0.5200	0.2050	0.0611	0.01389	0.00242
	Exact Gau.	0.4646	0.1896	0.0578	0.01337	0.00236
$n = 50$	Exact	0.4399	0.1771	0.0563	0.01430	0.00230
	Large Dev'n	0.5160	0.1939	0.0679	0.01793	0.00285
	2nd Order	0.4313	0.1782	0.0569	0.01460	0.00234
$n = 100$	Exact	0.4431	0.1782	0.0539	0.01258	0.00225
	Large Dev'n	0.4976	0.1976	0.0591	0.01343	0.00233
	2nd Order	0.4333	0.1770	0.0539	0.01261	0.00227

Table 3.1: Constant Parameter Significance Levels: 10% Truncation

	Method	c				
		2.0	2.5	3.0	3.5	4.0
	Gaussian	0.3449	0.1339	0.0396	0.00893	0.00155
	Exact Gau.	0.3479	0.1330	0.0391	0.00885	0.00154
$n = 50$	Exact	0.3245	0.1243	0.0361	0.00835	0.00150
	Large Dev'n	0.3178	0.1249	0.0371	0.00842	0.00147
	2nd Order	0.3046	0.1207	0.0357	0.00836	0.00153
$n = 100$	Exact	0.3305	0.1255	0.0369	0.00816	0.00143
	Large Dev'n	0.3345	0.1257	0.0392	0.00847	0.00143
	2nd Order	0.3079	0.1216	0.0365	0.00810	0.00145

Table 3.2: Constant Parameter Significance Levels: 20% Truncation

increased computation necessary for the exact method is probably not worth the effort.

The large deviation methods perform well in most cases and are very accurate in the tails, which is the important region for testing purposes. The second order approximation is usually better than the first order method.

All approximations (except the exact formulae) neglect the probability of crossing both boundaries and obtain the approximation for the two boundary case by adding the one boundary approximations. The effect of this is usually minor. For $c = 2$ and 10% truncation, the probability of crossing both boundaries is about 0.012 for both sample sizes, and much less in all other cases.

The score statistic $u(t)$ and the likelihood ratio statistic $l(t)$ are plotted for Lucas' accident data in Figure 3.1. The dotted lines on the score statistic plot represent ± 2 standard deviations. There is very strong evidence of a change point around 1975. The likelihood ratio process achieves a maximum of 12.07 at $t = 1975.334$. This value is highly significant ($p < 0.0001$) regardless of which approximation is used. At this time the point estimates for the rates are $\hat{\lambda}_0 = 23.81$ and $\hat{\lambda}_1 = 10.93$, giving $\hat{\delta} = -0.779$. Moreover, the score process $u(t)$ has roughly a triangular shape, which is the expected behaviour under the constant parameter model. This suggests that a change point model is a reasonable choice.

Similar plots for the coal mining data are shown in Figure 3.2. Since end-points for the observation interval are not precisely defined, we take the first and last observations to be the end-points and do not count these events. Again, there is strong evidence of a change point. The score process $u(t)$ again has a triangular trend, and the likelihood ratio process achieves its maximum of 36.24 at $t = 1890.19$. The point estimates are $\hat{\lambda}_0 = 3.181$ and $\hat{\lambda}_1 = 0.902$, giving $\hat{\delta} = -1.260$.

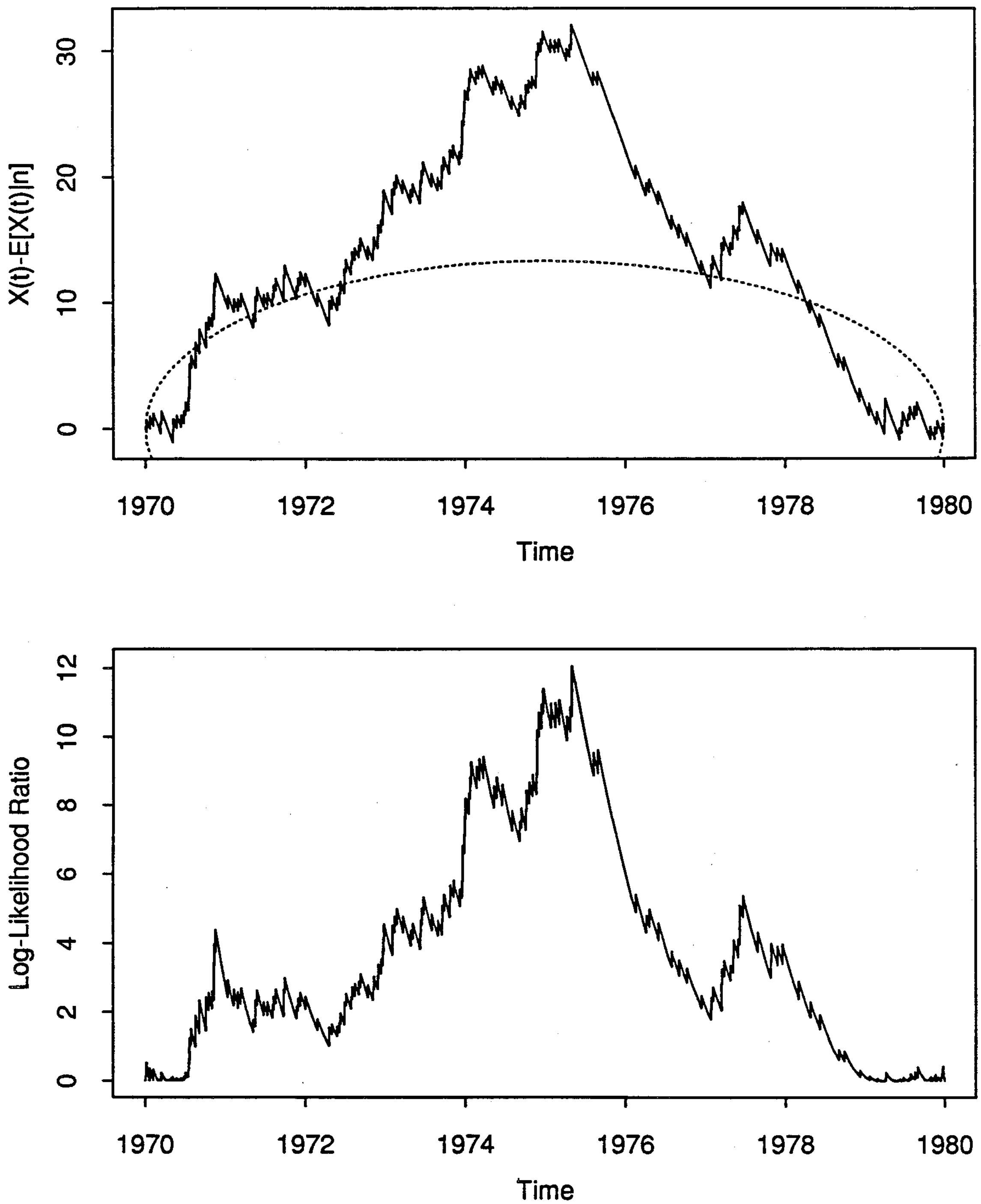


Figure 3.1: Lucas' Accident Data: $u(t)$ and $l(t)$ for Constant Parameter Model

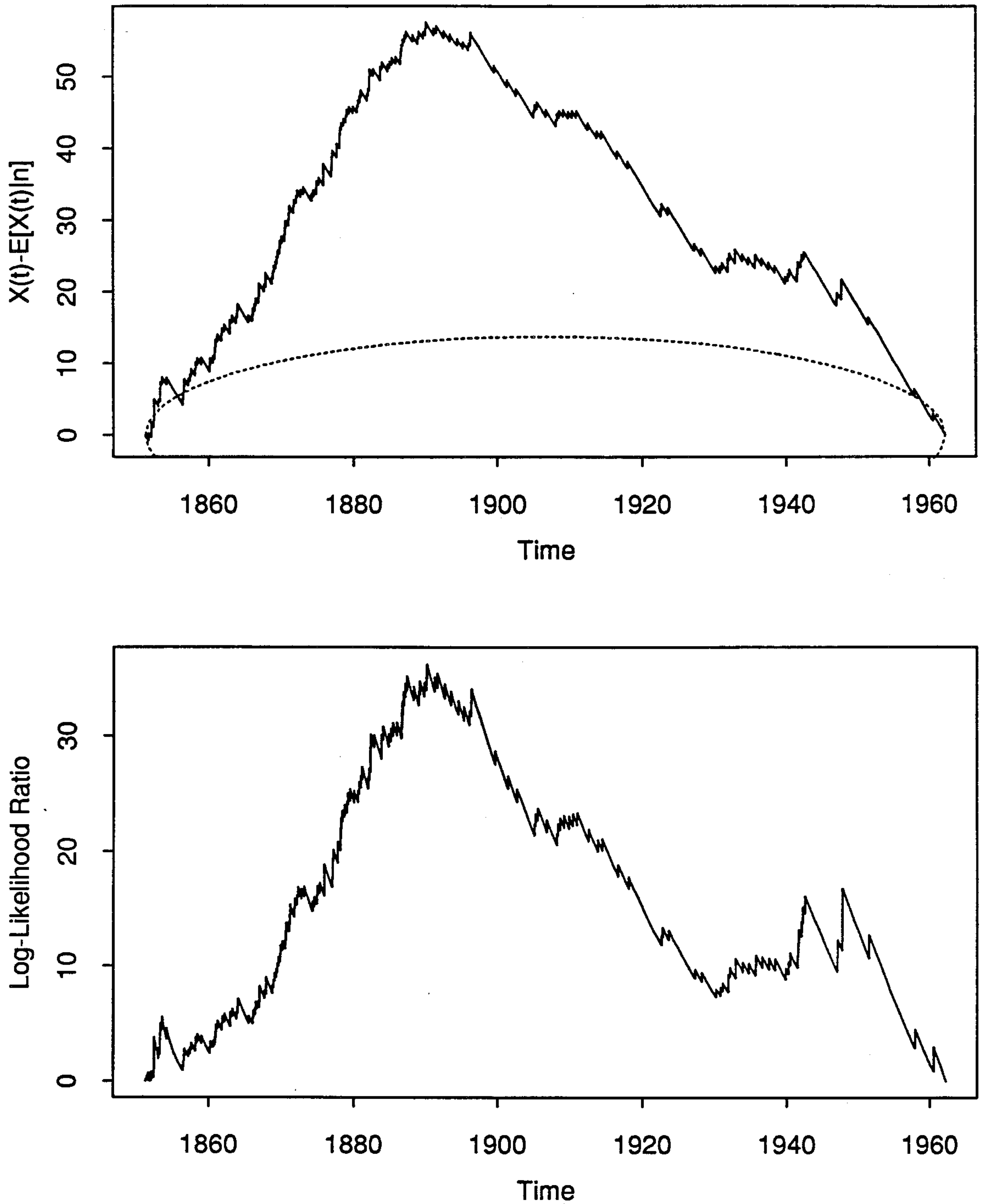


Figure 3.2: Coal Mining Data: $u(t)$ and $l(t)$ for Constant Parameter Model

3.2 Significance Levels for the Log-Linear Model

In this section we show how the methods of the previous section can be applied to approximate the significance level (1.14) of the likelihood ratio test for the log-linear model. We begin by developing a a Gaussian approximation then compare with the more complicated large deviation approximation.

The likelihood ratio process $l(t)$ is given by (1.12). If

$$X(t) = n\frac{e^{\hat{b}_0 t} - 1}{e^{\hat{b}_0 T} - 1},$$

then $\hat{b} = \hat{b}_0$ and $l(t) = 0$. Expanding (1.12) in a Taylor series and using the consistency of $\hat{b}_0$ as an estimate of b, shows that under $\mathcal{H}_0$,

$$l(t) \approx \frac{u(t)^2}{2n\sigma(t,t)}$$

where the score process $u(t)$ is given by

$$u(t) = X(t) - n\frac{e^{\hat{b}_0 t} - 1}{e^{\hat{b}_0 T} - 1}$$

and

$$\sigma(s,t) = \frac{(e^{bs} - 1)(e^{bT} - e^{bt})}{(e^{bT} - 1)^2} - \frac{g(s,b)g(t,b)}{\frac{1}{b^2} - \frac{T^2}{(e^{bT}-1)^2}} \tag{3.8}$$

for $0 \leq s \leq t \leq T$, where

$$g(t,b) = \frac{Te^{bT} - te^{bt} - (T-t)e^{b(T+t)}}{(e^{bT} - 1)^2}.$$

Moreover, it can be shown (c.f. Durbin, 1973b) that under $\mathcal{H}_0$, $\frac{1}{\sqrt{n}}u(t)$ converges weakly to a Gaussian process with mean 0 and covariance function $\sigma(s,t)$. The limiting covariance can be derived informally by noting

$$\frac{e^{\hat{b}_0 t} - 1}{e^{\hat{b}_0 T} - 1} = \frac{e^{bt} - 1}{e^{bT} - 1} + (\hat{b}_0 - b)g(t,b) + o_p(\frac{1}{\sqrt{n}})$$

and using the central limit theorem to obtain asymptotic distributions of $X(t)$ and $\hat{b}_0 - b$. The weak convergence of $\frac{1}{\sqrt{n}}u(t)$ is a special case of the general results given by Durbin (1973b). Also, $\frac{1}{\sqrt{n}}u(t)$ is asymptotically independent of $\sum_{i=1}^{n} T_i$ and hence the limiting distribution holds conditionally on $S_n = ny$ and we replace b by $\hat{b}_0$ in (3.8). We have

$$\begin{aligned} \frac{\partial\sigma(s,t)}{\partial s} - \frac{\partial\sigma(s,t)}{\partial t}\bigg|_{s\to t-} &= \frac{\hat{b}_0 e^{\hat{b}_0 s}(e^{\hat{b}_0 T} - e^{\hat{b}_0 t})}{(e^{\hat{b}_0 T} - 1)^2} + \frac{\hat{b}_0 e^{\hat{b}_0 t}(e^{\hat{b}_0 s} - 1)}{(e^{\hat{b}_0 T} - 1)^2}\bigg|_{s\to t-} \\ &= \frac{\hat{b}_0 e^{\hat{b}_0 t}}{e^{\hat{b}_0 T} - 1} \end{aligned}$$

and using (2.6) we get the Gaussian approximation to the significance level (1.14),

$$\alpha \approx 2(1 - \Phi(c)) + c\phi(c)\int_{\tau_0}^{\tau_1} \frac{1}{\sigma(t,t)} \frac{\hat{b}_0 e^{\hat{b}_0 t}}{e^{\hat{b}_0 T} - 1} dt.$$

The tangent approximation (2.26) for locally Poisson processes can be used to approximate the significance level (1.14) for the log-linear model, although the details are somewhat complicated. We need to find the boundary at time t and also a large deviation approximation to the distribution of $X(t)$, given $X(T) = n$ and $\sum_{i=1}^{n} T_i = ny$.

Lemma 3.1 *Let b_0 be defined by*

$$y = \frac{Te^{b_0 T}}{e^{b_0 T} - 1} - \frac{1}{b_0}$$

and $\eta = \frac{c}{\sqrt{n}}$. Define p_t and q_t to be the values of p obtained by simultaneously solving

$$y = p\frac{te^{bt}}{e^{bt} - 1} + (1-p)\frac{Te^{bT} - te^{bt}}{e^{bT} - e^{bt}} - \frac{1}{b} \tag{3.9}$$

$$\frac{\eta^2}{2} = p\log\left(\frac{pb(e^{b_0 T} - 1)}{b_0(e^{bt} - 1)}\right) + (1-p)\log\left(\frac{(1-p)b(e^{b_0 T} - 1)}{b_0(e^{bT} - e^{bt})}\right) + (b - b_0)y \tag{3.10}$$

for p and b, subject to $p_t < \frac{e^{b_0 t}-1}{e^{b_0 T}-1} < q_t$. Then

$$l(t) \le \frac{c^2}{2} \iff np_t \le X(t) \le nq_t. \tag{3.11}$$

If p_t and q_t exist, then they are unique. If p_t does not exist, we set $p_t = 0$ and if q_t does not exist, we set $q_t = 1$, and (3.11) still holds.

Proof: Suppose $\frac{1}{n} l(t) = \frac{1}{2}\eta^2$. Let $p = \frac{X(t)}{n}$ and $\hat{b}$ be the maximum likelihood estimate of b, given there is a change point at t. We know from (1.10) that (3.9) must hold, and from (1.12) that (3.10) must hold. Let $f(t,p,b)$ denote the RHS of (3.10) and for fixed p, choose b_p to satisfy (3.9). Then $f(t,p,b_p) = \sup_b f(t,p,b)$ and is a convex function of p. Also, for fixed t, $f(t,p,b_p)$ is a convex function of p which guarantees the uniqueness of p_t and q_t. If p_t does not exist, then $f(t,p,b_p) < \frac{1}{2}\eta^2$ for all $p < \frac{e^{b_0 t}-1}{e^{b_0 T}-1}$ and so we can take $p_t = 0$. The case where q_t does not exist is similar.

□

Using Lemma 3.1, we can write

$$\begin{aligned}
& P_0^{(n)}\left(\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{1}{2}c^2 \mid S_n = ny\right) \\
= \; & P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1} (X(t) - np_t) \le 0 \cup \sup_{\tau_0 \le t \le \tau_1} (X(t) - nq_t) \ge 0 | S_n = ny\right) \\
\approx \; & P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1} (X(t) - np_t) \le 0 \mid S_n = ny\right) \\
& + P_0^{(n)}\left(\sup_{\tau_0 \le t \le \tau_1} (X(t) - nq_t) \ge 0 \mid S_n = ny\right)
\end{aligned} \tag{3.12}$$

where $S_n = \sum_{i=1}^n T_i$. We want to approximate the probabilities (3.12) using the tangent approximation (2.26) for locally Poisson processes. To attain this, we need to evaluate three components: the derivative p'_t, the local event rate $\mu(t, p_t)$ and a continuous approximation $f_c(t)$ to the conditional distribution of $X(t)$.

Finding the derivative p_t' is straightforward. Letting $g(t,p,b)$ be the right side of (3.9), we differentiate (3.10) with respect to t to obtain

$$\begin{aligned}
0 &= \frac{\partial f(t,p_t,\hat{b})}{\partial p_t}\frac{dp_t}{dt} + \frac{\partial f(t,p_t,\hat{b})}{\partial \hat{b}}\frac{\partial b}{\partial t} + \frac{\partial f(t,p_t,\hat{b})}{\partial t} \\
&= \left(\log\left(\frac{p_t\hat{b}(e^{b_0T}-1)}{b_0(e^{\hat{b}t}-1)}\right) - \log\left(\frac{(1-p_t)\hat{b}(e^{b_0T}-1)}{b_0(e^{\hat{b}T}-e^{\hat{b}t})}\right)\right)\frac{dp_t}{dt} \\
&\quad +(y-g(t,p_t,\hat{b}))\frac{d\hat{b}}{dt} - \frac{p_t\hat{b}e^{\hat{b}t}}{e^{\hat{b}t}-1} + \frac{(1-p_t)\hat{b}e^{\hat{b}t}}{e^{\hat{b}T}-e^{\hat{b}t}}.
\end{aligned}$$

By (3.9), we have $y - g(t,p_t,\hat{b}) = 0$ and hence

$$\frac{dp_t}{dt} = \left.\frac{be^{bt}(\frac{p}{e^{bt}-1} - \frac{1-p}{e^{bT}-e^{bt}})}{\log(\frac{pb}{e^{bt}-1}) - \log(\frac{(1-p)b}{e^{bT}-e^{bt}})}\right|_{p=p_t,b=\hat{b}}.$$

We need to approximate the local event rate given $X(t) = np_t$ and $S_n = ny$,

$$n\mu(t,p_t) = \lim_{s\to t^-}\frac{1}{t-s}E^{(n)}(X(t)-X(s)|X(t)=np_t, S_n=ny). \tag{3.13}$$

By sufficiency, (3.13) does not depend on b or δ, and therefore will be approximately the unconditional event rate at time t when $\delta = \hat{\delta}$ and $b = \hat{b}$, which suggests

$$\begin{aligned}
\mu(t,p_t) &\approx \frac{be^{\hat{b}t}}{e^{\hat{b}t}-1+e^{\hat{\delta}}(e^{\hat{b}T}-e^{\hat{b}t})} \\
&= p_t\frac{\hat{b}e^{\hat{b}t}}{e^{\hat{b}t}-1}.
\end{aligned} \tag{3.14}$$

A more formal derivation of (3.14) will be given in Appendix B.

It remains to approximate the conditional distribution under $\mathcal{H}_0$, of $X(t)$ given $X(T) = n$ and $\sum_{i=1}^n T_i = ny$. This can be obtained by regarding $T_1,\ldots,T_n$ as order statistics of a sample of size n from a density

$$f_{b,\delta}(x) = \exp\left(bx + \delta I(x>t) - \psi(\delta,b)\right)\frac{1}{T}I_{[0,T]}(x)$$

with $\delta = 0$, where

$$\psi(\delta,b) = \log\left(\frac{1}{bT}(e^{bt}-1+e^{\delta}(e^{bT}-e^{bt}))\right).$$

Applying large deviation methods discussed in Appendix B gives the approximation

$$\begin{aligned} & P_0^{(n)}\left(X(t) = np_t | \sum_{i=1}^{n} T_i = ny\right) \\ = & \frac{1}{\sqrt{2\pi n}}\sqrt{-\frac{\psi_{22}(0,\hat{b}_0)}{\psi_{22}(\hat{\delta},\hat{b})}\frac{\partial\hat{\delta}}{\partial p}} \exp(-l(t))(1+o(1)) \end{aligned} \tag{3.15}$$

as $n \to \infty$, where $l(t)$, $\hat{\delta}$ and $\hat{b}$ are all calculated under the assumption $\tau = t$ and $X(t) = np_t$. Although (3.15) is a density with respect to counting measure on $[0,\ldots,n]$, for large n we can for the purposes of integration treat it as a density with respect to Lebesgue measure. Explicit forms for $\psi_{22}(0,\hat{b}_0)$, $\psi_{22}(\hat{\delta},\hat{b})$ and $\frac{\partial\hat{\delta}}{\partial p}$ are given by (B.25), (B.26) and (B.27) respectively.

Substituting (3.14) and (3.15) into (2.26), we get the tangent approximation

$$\begin{aligned} & P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1}(X(t) - np_t) < 0 | S_n = ny\right) \\ \approx \; & P_0^{(n)}(X(\tau_0) < np_{\tau_0} | S_n = ny) \\ & + c\phi(c)\int_{t_0}^{t_1}\frac{1}{\eta}\left(p_t' - \frac{p_t b e^{bt}}{e^{bt}-1}\right)\sqrt{-\frac{\psi_{22}(0,\hat{b}_0)}{\psi_{22}(\hat{\delta},\hat{b})}\frac{\partial\hat{\delta}}{\partial p}}dt. \end{aligned} \tag{3.16}$$

The boundary probability is calculated using Lemma B.2. The probability of crossing q_t can be approximated by time reversal. Unfortunately there is no simplification for the case $\tau_1 = T - \tau_0$ now.

Table 3.3 gives approximations to the significance level, for $T = 1$, $n = 100$, $t_0 = 0.1$, $t_1 = 0.9$ and $y = 0.5$. Since no exact method is available, we compare with the results of a simulation. See Appendix C for more detail of the simulations. When c is small the approximations do not perform particularly well. However, for larger values of c which are of interest for significance testing, the approximations are acceptable and the large deviation approximation is performing better than the Gaussian approximation.

The poor performance for large c is in part due to the probability of crossing both boundaries. For $y = 0.5$, the simulation gave probabilities of 0.3413, 0.0855,

y	Method	c				
		2.0	2.5	3.0	3.5	4.0
0.5	Simulation	0.6728	0.3225	0.1054	0.0271	0.0040
	Gaussian	1.2276	0.4921	0.1483	0.0339	0.0059
	Large Dev'n	1.1126	0.4353	0.1290	0.0292	0.0051
0.6	Simulation	0.6808	0.3264	0.1105	0.0253	0.0046
	Gaussian	1.2328	0.4942	0.1489	0.0340	0.0059
	Large Dev'n	1.1477	0.4377	0.1325	0.0298	0.0051
0.7	Simulation	0.6903	0.3350	0.1107	0.0254	0.0044
	Gaussian	1.2521	0.5021	0.1513	0.0346	0.0060
	Large Dev'n	1.1355	0.4485	0.1307	0.0286	0.0048
Simulation s.e.		0.005	0.005	0.003	0.002	0.0007

Table 3.3: Significance Level Approximations for Log-Linear Model

0.0174, 0.0015 and 0.0002 of crossing both boundaries. In most cases, this counts for over half of the error of the large deviation approximation.

The score process and likelihood ratio process for Lucas' accident data, calculated for the log-linear model, are shown in Figure 3.3. The likelihood ratio process attains a maximum of 5.58, at $t = 1975.334$. With 10% truncation, the attained significance level calculated using the tangent approximation is 0.051. The point estimate of δ is $\hat{\delta} = -1.005$.

The score statistic and likelihood ratio process for the coal mining dataset are shown in Figure 3.4. The maximum of the likelihood ratio process is 6.27, at $t = 1890.19$. The attained significance level is 0.027. This indicates the presence of a change point. Moreover, the likelihood ratio process has a second peak which indicates the possibility of a second change point around 1930. The point estimate of δ is $\hat{\delta} = -1.0226$, which is slightly larger (closer to 0) than we obtained with the constant parameter model.

The expected behaviour of the score process under $\mathcal{H}_1$ for the log-linear model is difficult to quantify. Since we are essentially subtracting the conditional mean of $X(t)$ given $\sum_{i=1}^{n} T_i$, the process $u(t)$ is constrained to integrate to 0. This means if $u(\tau)$ is positive, we must have $u(t) < 0$ for some other values of t to compensate. Calculating central limit approximations to $E(X(t))$ under alternatives typically gives curves which are sharply peaked at τ, and therefore only go slightly negative for other values of t. This agrees well with the behaviour of $u(t)$ in Figure 3.3 but not so well with the behaviour in Figure 3.4, where $u(t)$ remains negative for a fairly substantial period between 1920 and 1940, again indicating the possibility of a second change. Since we excced the lower boundary during this period, this indicates an increase in the accident rate at this second change point, compared with the overall decrease suggested by the negative slope $\hat{b}$. This second change is

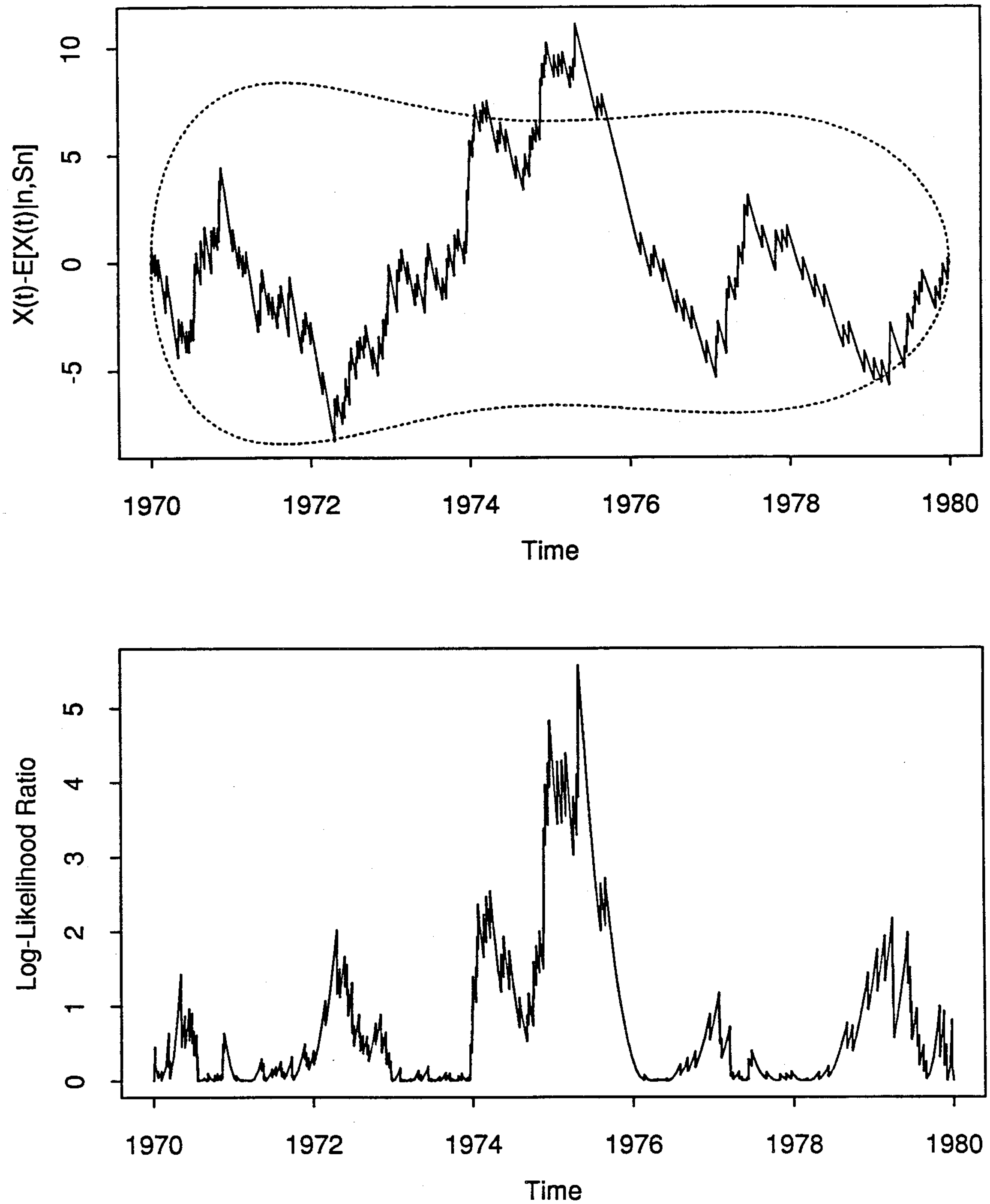
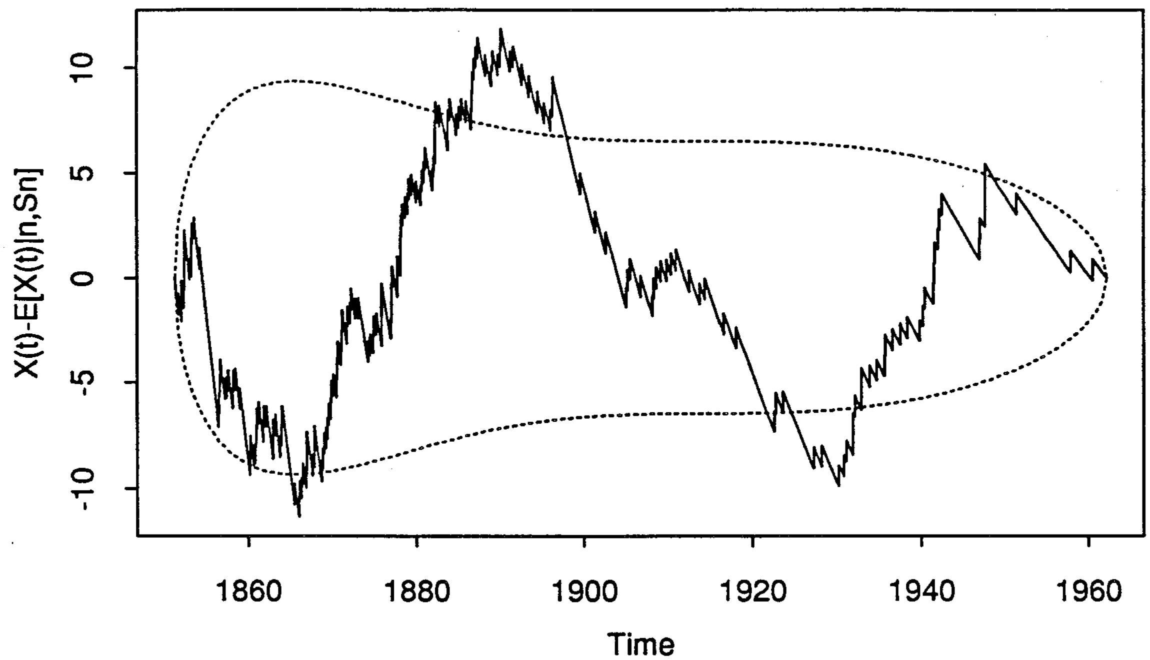


Figure 3.3: Lucas' Accident Data: $u(t)$ and $l(t)$ for Log-linear Model



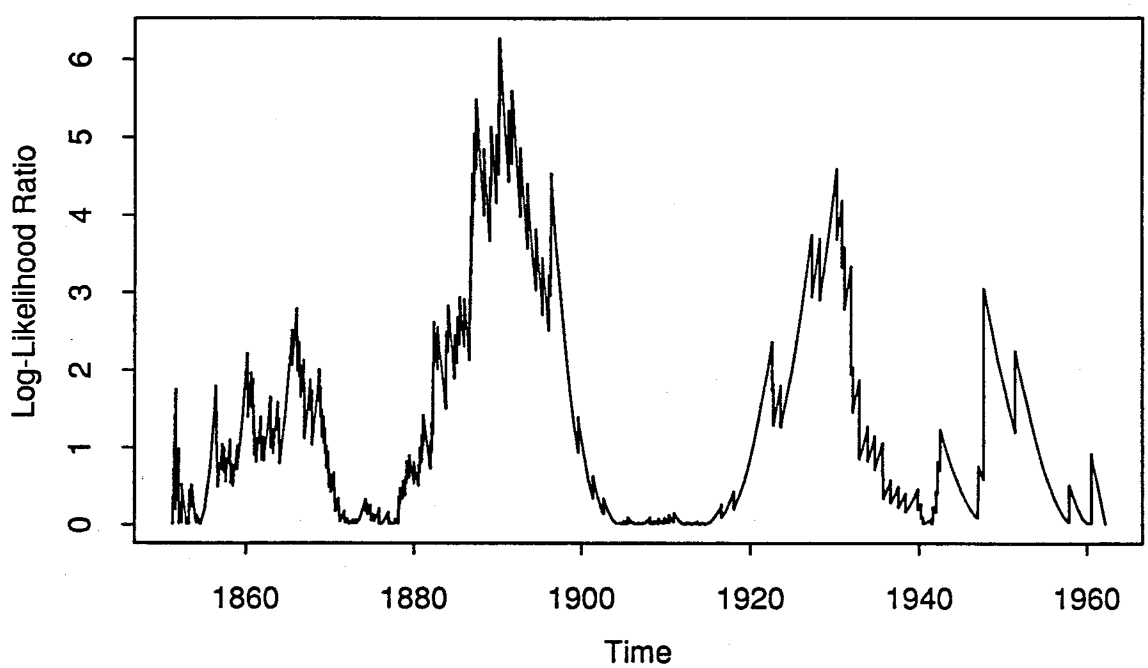


Figure 3.4: Coal Mining Data: $u(t)$ and $l(t)$ for Log-linear Model

completely masked by this negative slope when we fit the constant parameter model.

3.3 Power Calculations for the Constant Parameter Model

Let R be the rejection region for the likelihood ratio test,

$$R = \left\{ \sup_{\tau_0 \leq t \leq \tau_1} l(t) \geq \frac{c^2}{2} \right\}.$$

The power of the likelihood ratio test is

$$\beta(\tau, \delta) = P^{(n)}_{\tau,\delta}(R) \tag{3.17}$$

where $P_{\tau,\delta}$ is a measure under which $X(t)$ has a rate $\lambda(t)$ given by (1.1) with $\delta = \log(\lambda_1/\lambda_0)$ and $P^{(n)}_{\tau,\delta}(A) = P_{\tau,\delta}(A|X(T) = n)$ for $A \in \mathcal{F}_T$. By sufficiency, (3.17) depends on λ_0 and λ_1 only through δ. Conditioning on $X(\tau)$ we get

$$\begin{aligned} \beta(\tau, \delta) &= \sum_{j=0}^{n} P^{(n)}_{\tau}(R|X(\tau) = j) P^{(n)}_{\tau,\delta}(X(\tau) = j) \\ &= P^{(n)}_{\tau,\delta}\left(l(\tau) \geq \frac{1}{2}c^2\right) + \sum_{j=j_0}^{j_1} P^{(n)}_{\tau,\delta}(R|X(\tau) = j) P^{(n)}_{\tau,\delta}(X(\tau) = j) \end{aligned} \tag{3.18}$$

for suitable j_0 and j_1. We concentrate on evaluating the terms

$$P^{(n)}_{\tau,\delta}(R|X(\tau) = j). \tag{3.19}$$

by two methods. The first is a large deviation method. We assume $n \to \infty$ and $j \to \infty$ such that $\frac{j}{n}$ is fixed, and use tangent approximations similar to those used for significance calculations. The second method uses a local linearization around the change point, and is similar to the method used by James, James and Siegmund

(1987) for normally distributed random variables. In both cases, we assume $c \to \infty$ with $\eta = c/\sqrt{n}$ fixed. Since by sufficiency (3.19) does not depend on δ, we will suppress the δ in future equations.

Both these methods have strengths and weaknesses. The later method is based on the observation that asymptotically only values of j for which $l(\tau)$ is close to the boundary will contribute to the sum and the major contribution to these probabilities will come from paths for which $l(t)$ crosses the boundary close to τ. However, this approximation is less accurate than the first method for boundaries and sample sizes of interest. There are two disadvantages for the large deviation approximation. Firstly, the approximation is not uniform in j for values of j such that $l(t)$ is close to the boundary, so a precise asymptotic justification for this method may not be possible. Secondly, this method involves much more computation.

We define a measure $P_\tau^{(j,n)}$ by

$$P_\tau^{(j,n)}(A) = P_\tau(A|X(\tau) = j, X(T) = n).$$

for events $A \in \mathcal{F}_T$. Since conditional on $X(\tau)$, $\{X(t), 0 \le t \le \tau\}$ and $\{X(t), \tau \le t \le T\}$ are independent, we can write

$$P_\tau^{(n)}(R|X(\tau) = j) = 1 - (1 - P_\tau^{(j,n)}(R_1))(1 - P_\tau^{(j,n)}(R_2)) \tag{3.20}$$

where

$$\begin{aligned} R_1 &= \left\{ \sup_{\tau_0 \le t \le \tau} l(t) \le \frac{1}{2}c^2 \right\} \\ R_2 &= \left\{ \sup_{\tau \le t \le \tau_1} l(t) \le \frac{1}{2}c^2 \right\} \end{aligned}$$

and under $P_\tau^{(j,n)}$, R_1 and R_2 are independent.

We begin by discussing the large deviation method. Suppose $j, n \to \infty$ with $\frac{j}{n}$ fixed. Then

$$P_\tau^{(j,n)}\left(\sup_{\tau_0 \le t \le \tau} l(t) \ge \frac{1}{2}c^2 \right)$$

$$\approx \; P_\tau^{(j,n)}\left(\inf_{\tau_0\le t\le\tau}(X(t)-np_t)<0\right)+P_\tau^{(j,n)}\left(\inf_{\tau_0\le t\le\tau}(X(t)-nq_t)>0\right)$$

where p_t and q_t are defined by (3.5). Using the large deviation tangent approximation (2.24) gives

$$\begin{aligned} & P_\tau^{(j,n)}\left(\inf_{\tau_0\le t\le\tau}(X(t)-np_t)<0\right) \\ \approx \; & \int_{\tau_0}^{\tau}(p_t'-\frac{p_t}{t})\sqrt{\frac{j}{2\pi p_t(\frac{j}{n}-p_t)}}e^{-nl_2(t)}dt \end{aligned} \tag{3.21}$$

and similarly,

$$\begin{aligned} & P_\tau^{(j,n)}\left(\inf_{\tau\le t\le\tau_1}(X(t)-np_t)<0\right) \\ \approx \; & \int_{t'}^{\tau_1}(p_t'-\frac{p_t-\frac{j}{n}}{t-\tau})\sqrt{\frac{j}{2\pi(1-p_t)(p_t-\frac{j}{n})}}e^{-nl_2(t)}dt \end{aligned} \tag{3.22}$$

where $t'>\tau$ is the solution of $np_t=j$. Here, the terms involving $nl_2(t)$ arise from the conditional (binomial) distribution of $X(t)$ given $X(\tau)=j=nx$ and $X(T)=n$,

$$l_2(t)=p_t\log\left(\frac{\tau p_t}{tx}\right)+(x-p_t)\log\left(\frac{\tau(x-p_t)}{(\tau-t)x}\right)$$

if $t<\tau$ and

$$l_2(t)=(p_t-x)\log\left(\frac{(T-\tau)(p_t-x)}{(t-\tau)(1-x)}\right)+(1-p_t)\log\left(\frac{(T-\tau)(1-p_t)}{(1-t)(1-x)}\right)$$

if $t>\tau$. An alternative interpretation useful for comparison with the log-linear model is that $nl_2(t)$ is the log- likelihood ratio statistic for testing the null hypothesis of a single change point at τ against the alternative of two change points at t and τ, with $X(\tau)=j$ and $X(t)=np_t$. We can approximate the probability of crossing q_t by time reversal. Since conditional on $X(T)$, $X(\tau)$ has a binomial distribution, we can now approximate the power using (3.21), (3.22) and (3.18).

We now turn our attention to using local expansions of the likelihood ratio statistic. Suppose we have a change of size $\delta>0$ at τ, and let $\delta_0=\log(\frac{1-p_\tau}{T-\tau})-$

$\log(\frac{p_T}{\tau}) > 0$ be the maximum likelihood estimate of δ when $X(\tau) = np_\tau$. Now suppose $X(\tau) = np_\tau + x$. Then a Taylor series expansion shows $l(\tau) = \frac{1}{2}n\eta^2 - \delta_0 x + o(1)$ as $n \to \infty$. Moreover, a local linearization (more details will be given in Chapter 4) shows

$$P_\tau^{(np_\tau+x,n)}\left(\sup_{t_0\le t\le t_1} l(t) \ge \frac{n\eta^2}{2}\right) \to 1 - \left(1 - h(\frac{\delta_0}{e^{\delta_0}-1}, \frac{\delta_0 x}{e^{\delta_0}-1})\right)(1-e^{-\delta_0 x}) \quad (3.23)$$

as $n \to \infty$, where $h(y,z)$ is given by (2.13). The exact form of $h(y,z)$ is difficult to work with. However, we show in Theorem 4.1 that

$$e^{\delta_0 x} h\left(\frac{\delta_0}{e^{\delta_0}-1}, \frac{\delta_0 x}{e^{\delta_0}-1}\right) \to \nu(\delta_0) \quad (3.24)$$

as $x \to \infty$, where

$$\nu(\delta) = \frac{1 - \frac{|\delta|}{e^{|\delta|}-1}}{\frac{|\delta|}{1-e^{-|\delta|}} - 1}.$$

Moreover, the limiting approximation (3.24) is usually very good even for small x. Expanding the distribution of $X(\tau)$ gives

$$\begin{aligned}
&P_{\tau,\delta}^{(n)}(X(\tau) = np_\tau + x) \\
&= \binom{n}{np_\tau + x}\left(\frac{\tau}{\tau + e^\delta(T-\tau)}\right)^{np_\tau + x}\left(\frac{e^\delta(T-\tau)}{\tau + e^\delta(T-\tau)}\right)^{n(1-p_\tau)-x} \\
&= \binom{n}{np_\tau}\left(\frac{1-p_\tau}{p_\tau}\right)^x\left(\frac{\tau}{T}\right)^{np_\tau}\left(1-\frac{\tau}{T}\right)^{n(1-p_\tau)} \\
&\quad \times \left(\frac{Te^{\delta(1-p_\tau)}}{\tau + e^\delta(T-\tau)}\right)^n\left(\frac{\tau}{e^\delta(T-\tau)}\right)^x(1+o(1)) \\
&= \frac{\phi(c)}{\sqrt{np_\tau(1-p_\tau)}}\left(\frac{Te^{\delta(1-p_\tau)}}{\tau + e^\delta(T-\tau)}\right)^n e^{(\delta_0-\delta)x}(1+o(1)) \qquad (3.25)
\end{aligned}$$

as $n \to \infty$.

A power approximation is obtained by substituting (3.25) and (3.23) into (3.18) and summing over x. The appropriate values of x are those for which $np_\tau + x$ is an integer. Letting x_i be the solution of $np_\tau + x_i = \lfloor np_\tau \rfloor + i$, we get the approximation

$$\begin{aligned} &P^{(n)}_{\tau,\delta}\left(\sup_{\tau_0<t<\tau_1} l(t) \geq \frac{1}{2}c^2\right) \\ &\sim \; P^{(n)}_{\tau,\delta}(X(\tau) \leq np_\tau) + \frac{\phi(c)}{\sqrt{np_\tau(1-p_\tau)}}\left(\frac{Te^{\delta(1-p_\tau)}}{\tau + e^{\delta}(T-\tau)}\right)^n \\ &\quad \times \sum_{i=1}^{\infty} e^{(\delta_0-\delta)x_i}\left(e^{-\delta_0 x_i} + (1-e^{-\delta_0 x_i})h(\frac{\delta_0}{e^{\delta_0}-1}, \frac{\delta_0 x_i}{e^{\delta_0}-1})\right) \end{aligned} \tag{3.26}$$

as $n \to \infty$. A precise justification of (3.26) involves truncating the sum to $x < X$ for some X before letting $n \to \infty$, then letting $X \to \infty$ and showing the truncated terms of (3.18) are negligible. We omit the details.

Using the approximation (3.24), we can approximate the summation in (3.26) by

$$\begin{aligned} &\sum_{i=1}^{\infty} e^{(\delta_0-\delta)x_i}\left(e^{-\delta_0 x_i} + (1-e^{-\delta_0 x_i})h(\frac{\delta_0}{e^{\delta_0}-1}, \frac{\delta_0 x_i}{e^{\delta_0}-1})\right) \\ &\approx \; (1+\nu(\delta_0))\frac{e^{-\delta x_1}}{1-e^{-\delta}} - \nu(\delta_0)\frac{e^{-(\delta+\delta_0)x_1}}{1-e^{-(\delta+\delta_0)}}. \end{aligned} \tag{3.27}$$

We note the approximation (3.26) assumes $\delta > 0$. For $\delta < 0$, we can derive a similar expression using a local expansion around (τ, q_τ) or use a time reversal argument to get

$$P^{(n)}_{\tau,\delta}\left\{\sup_{\tau_0\leq t\leq\tau_1} l(t) \geq \frac{1}{2}c^2\right\} = P^{(n)}_{1-\tau,-\delta}\left\{\sup_{1-\tau_1\leq t\leq 1-\tau_0} l(t) \geq \frac{1}{2}c^2\right\}.$$

The recursive formulae may also be used for exact power computations. Defining a_i and b_i, $i = 0, \ldots, n$ as in our discussion of significance levels, we can write

$$P^{(n)}_{\tau,\delta}(a_i \leq T_i \leq b_{i-1}, i = 1, \ldots, n) = P^{(n)}_{\tau,\delta}(F_{\tau,\delta}(a_i) \leq F_{\tau,\delta}(T_i) \leq F_{\tau,\delta}(b_i), i = 1, \ldots, n) \tag{3.28}$$

where

$$F_{\tau,\delta}(x) = \begin{cases} \frac{x}{\tau+e^{\delta}(T-\tau)} & x \leq \tau \\ \frac{\tau+e^{\delta}(x-\tau)}{\tau+e^{\delta}(T-\tau)} & x > \tau \end{cases}.$$

Morover, under $P_{\tau,\delta}^{(n)}$, $\{F_{\tau,\delta}(T_i), i = 1, \ldots, n\}$ have the same distribution as the order statistics of n *i.i.d.* $\mathcal{U}[0,1]$ random variables, so we can use (2.28) and (2.29) to evaluate (3.28).

The approximations are compared in Figure 3.5 for $T = 1$, $\tau = 0.5$ and $n = 100$, using $c = 3.07$ and $\tau_0 = 1 - \tau_1 = 0.1$. In the top graph we plot the power approximations and in the bottom graph we have subtracted the marginal probability of $\{l(\tau) \geq \frac{1}{2}c^2\}$. The large deviation approximation is almost indistinguishable from the exact. The local approximation slightly underestimates the power, but since precise power calculations are rarely important this is probably accurate enough for most purposes.

We note that the large deviation method does not have a precise asymptotic justification. In particular, the local expansion shows that asymptotically all contribution to the power comes from values of $X(\tau)$ equal to $np_\tau + O(1)$ and $t = \tau + O(\frac{1}{n})$. In particular, this means that we cannot justify passing from summation to integration. However, Figure 3.5 indicates this is not a serious problem.

3.4 Power Calculations: Log-Linear Model

We can apply similar methods to those in Section 3.3 to approximate the power of the likelihood ratio test for the log-linear model. Let $S_n = \sum_{i=1}^n T_i$ and define

$$\beta(\tau, \delta, y) = P_{\tau,\delta}^{(n)}(R | S_n = ny)$$

where

$$R = \left\{ \sup_{\tau_0 \leq t \leq \tau_1} l(t) \geq \frac{1}{2}c^2 \right\}.$$

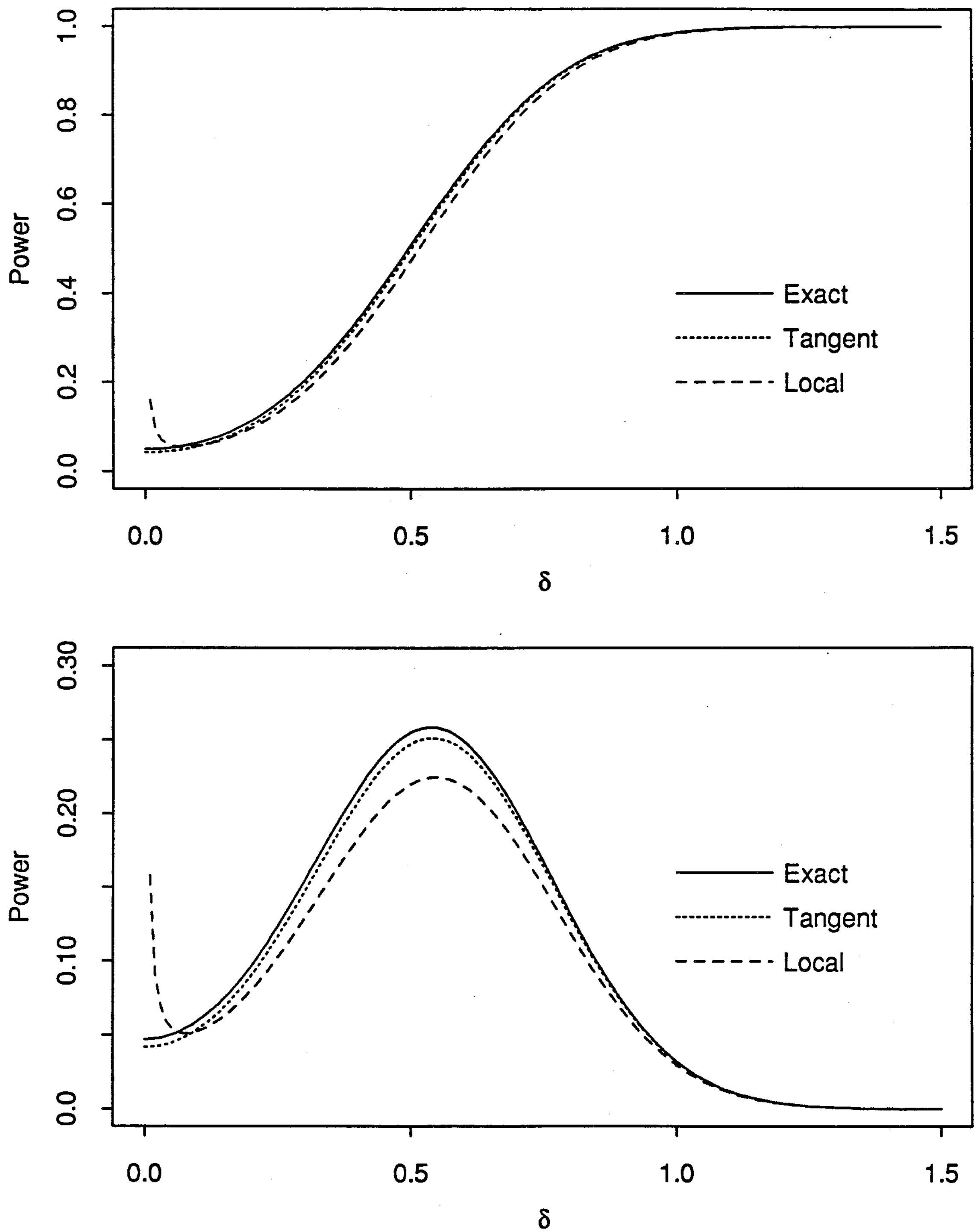


Figure 3.5: Power Approximation Comparisons

As before, we can write

$$\beta(\tau,\delta,y) = P^{(n)}_{\tau,\delta}\left\{l(\tau) \geq \frac{1}{2}c^2 \mid S_n = ny\right\} + \sum_{j=j_0}^{j_1} P^{(j,n)}_{\tau}\left\{R \mid \sum_{i=1}^{n} T_i = ny\right\}. \quad (3.29)$$

We cannot simply split the conditional probabilities in (3.29) as we did in (3.20) since conditional on $\sum_{i=1}^{n} T_i = ny$, the behaviour of $X(t)$ for $t < \tau$ is not independent of the behaviour for $t > \tau$. It is therefore not immediately apparent that we can use the large deviation method in this case, so we begin by discussing the local expansion method.

We can use local independence of the empirical processes on either side of τ to obtain, similarly to (3.23),

$$\begin{aligned} P^{(n)}_{\tau}&\left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) \geq \frac{n\eta^2}{2} | X(\tau) = np_\tau + x, \sum_{i=1}^{n} T_i = ny\right) \\ &\to e^{-\delta_0 x} + (1 - e^{-\delta_0 x}) h\left(\frac{\delta_0}{e^{\delta_0} - 1}, \frac{\delta_0 x}{e^{\delta_0} - 1}\right) \end{aligned}$$

as $n \to \infty$.

We can write

$$\begin{aligned} & P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau + x | S_n = ny) \\ = \quad & \frac{P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau + x | S_n = ny)}{P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau | S_n = ny)} P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau | S_n = ny) \end{aligned}$$

Using large deviation approximations from Appendix B we get

$$\begin{aligned} & \frac{P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau + x | S_n = ny)}{P^{(n)}_{\tau,\delta}(X(\tau) = np_\tau | S_n = ny)} \\ = \quad & \exp\left(-n(l_\delta(\tau, p_\tau + \frac{x}{n}) - l_\delta(\tau, p_\tau))\right)(1 + o(1)) \\ \to \quad & \exp\left(-x \frac{\partial}{\partial p_\tau} l_\delta(\tau, p_\tau)\right) \\ = \quad & e^{(\delta_0 - \delta)x} \end{aligned}$$

as $n \to \infty$, where

$$\begin{aligned} l_\delta(\tau, p) &= p \log\left(\frac{\hat{b}p(e^{\hat{b}_\delta \tau} - 1 + e^{\delta}(e^{\hat{b}_\delta T} - e^{\hat{b}_\delta \tau}))}{\hat{b}_\delta(e^{\hat{b}\tau} - 1)}\right) \\ &\quad +(1-p)\log\left(\frac{\hat{b}(1-p)(e^{\hat{b}_\delta \tau} - 1 + e^{\delta}(e^{\hat{b}_\delta T} - e^{\hat{b}_\delta \tau}))}{\hat{b}_\delta(e^{\hat{b}T} - e^{\hat{b}\tau})}\right) \\ &\quad -\delta(1-p_\tau) + (\hat{b} - \hat{b}_\delta)y \end{aligned}$$

is $\frac{1}{n}$ times the log-likelihood ratio statistic for testing the null hypothesis of a single change of size δ (known) vs the general alternative when $X(\tau) = np$ and $\hat{b}_\delta$ and $\hat{b}$ are the maximum likelihood estiamtes of b under the respective hypothesis. Another application of the large deviation theory shows

$$\begin{aligned} &P_{\tau,\delta}^{(n)}(X(\tau) = np_\tau | S_n = ny) \\ &\approx \left(\frac{e^{\hat{b}_\delta y + \delta(1-p_\tau) - \psi(\hat{b}_\delta, \delta)}}{e^{\hat{b}_0 y - \psi(\hat{b}_0, 0)}}\right)^n P_0^{(n)}(X(\tau) = np_\tau | S_n = ny). \end{aligned}$$

This leads to the power approximation

$$\begin{aligned} &P_{\tau,\delta}^{(n)}\left\{\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{1}{2}c^2 | S_n = ny\right\} \\ &\approx P_{\tau,\delta}^{(n)}\left(X(\tau) \le np_\tau | S_n = ny\right) \\ &\quad + \frac{\phi(c)}{\sqrt{n\sigma(\tau,\tau)}}\left(\frac{e^{\hat{b}_\delta y + \delta(1-p_\tau) - \psi(\hat{b}_\delta, \delta)}}{e^{\hat{b}_0 y - \psi(\hat{b}_0, 0)}}\right)^n \\ &\quad \times \sum_x e^{(\delta_0 - \delta)x}\left(e^{-\delta_0 x} + (1 - e^{-\delta_0 x})h(\frac{\delta_0}{e^{\delta_0} - 1}, \frac{\delta_0 x}{e^{\delta_0} - 1})\right) \end{aligned} \tag{3.30}$$

where

$$\frac{1}{\sigma(\tau,\tau)} = -\frac{\psi_{22}(0, \hat{b}_0)}{\psi_{22}(\hat{\delta}, \hat{b})}\frac{\partial \hat{\delta}}{\partial p}.$$

Exact forms for ψ_{22} and $\frac{\partial \hat{\delta}}{\partial p}$ are given in Appendix B. The summation in (3.30) can be approximated by (3.27).

As before, the local expansion indicates that large deviation methods cannot have a precise asymptotic justification since we cannot pass sums to integrals. However, this also shows that the important contribution to the power comes from values of $t = \tau + O_p(\frac{1}{n})$ and that we can split the two sides similarly to (3.20), This leads to approximations

$$\begin{aligned} &P_\tau^{(j,n)}\left(\inf_{\tau_0 \le t \le \tau_1}(X(t) - np_t) \le 0 | S_n = ny\right) \\ &\approx \int_{\tau_0}^{\tau}\left(p_t' - \frac{p_t b_2 e^{b_2 t}}{e^{b_2 t} - 1}\right)\sqrt{\frac{n|\psi''(b,\delta)|}{2\pi|\psi''_{(2)}(b_2,\delta_{21},\delta_{22})|}}e^{-nl_2(t)}dt \end{aligned} \tag{3.31}$$

and

$$\begin{aligned} &P_\tau^{(j,n)}\left(\inf_{\tau \le t \le \tau_1}(X(t) - np_t) \le 0 | S_n = ny\right) \\ &\approx \int_{\tau}^{\tau_1}\left(p_t' - \frac{(p_t - \frac{j}{n} b_2 e^{b_2 t})}{e^{b_2 t} - e^{b_2 \tau}}\right)\sqrt{\frac{n|\psi''(b,\delta)|}{2\pi|\psi''_{(2)}(b_2,\delta_{21},\delta_{22})|}}e^{-nl_2(t)}dt \end{aligned} \tag{3.32}$$

where $nl_2(t)$ is the log-likelihood ratio statistic for testing the null hypothesis of a single change point at τ against the alternative of two change points at τ and t, b_2, δ_{21} and δ_{22} are the maximum likelihood estimates under this alternative when $X(\tau) = j$ and $X(t) = np_t$ and $\psi_{(2)}$ is the normalizing constant for the appropriate three parameter exponential family,

$$f_{b,\delta_1,\delta_2}(x) = \frac{b\exp(bx + \delta_1 I(x > \tau) + \delta_2 I(x > t))}{(e^{b\tau} - 1) + e^{\delta_1}(e^{bt} - e^{b\tau}) + e^{\delta_1+\delta_2}(e^{bT} - e^{b\tau})} I_{[0,T]}(x).$$

It is easy to show, if $x = \frac{j}{n}$,

$$\begin{aligned} l_2(t) &= p_t \log\left(\frac{b_2 p_t (e^{\hat{b}\tau} - 1)}{\hat{b}x(e^{b_2 t} - 1)}\right) + (x - p_t)\log\left(\frac{b_2(x - p_t)(e^{\hat{b}\tau} - 1)}{\hat{b}x(e^{b_2\tau} - e^{b_2 t})}\right) \\ &\quad + (1 - x)\log\left(\frac{b_2(e^{\hat{b}T} - e^{\hat{b}\tau})}{e^{b_2 T} - e^{b_2\tau}}\right) + (b_2 - \hat{b})y \end{aligned}$$

if $t < \tau$, and a similar expression if $t > \tau$. The equation defining b_2 is

$$y = \frac{j}{n}\frac{\tau e^{b\tau}}{e^{b\tau} - 1} + (p_t - \frac{j}{n})\frac{te^{bt} - \tau e^{b\tau}}{e^{bt} - e^{b\tau}} + (1 - p_t)\frac{Te^{bT} - te^{bt}}{e^{bT} - e^{bt}} - \frac{1}{b}.$$

Approximations to the probability of crossing q_t are found by time reversal.

One question of interest is how much power is lost if we use the log-linear model when the constant parameter model is correct. In Figures 3.6 and 3.7, we compare the power of the two tests, for $\tau = 0.3$ and $\tau = 0.5$. We also include the power of the two sample test that would be appropriate for testing $\mathcal{H}_0 : \lambda_0 = \lambda_1$ if τ were known. The power for the log-linear model is calculated for several values of y. We take $c = 3.0671$ for the constant parameter model and $c = 3.3372, 3.3317$ and 3.3372 for $y = 0.4, 0.5$ and 0.6 respectively for the log-linear model. These values of c are chosen to obtain a significance level of about 0.05, as indicated by the first order large deviation approximations. This nominal significance level is represented by the horizontal line in Figures 3.6 and 3.7.

We see a moderate loss of power, up to about 20%, due to searching for the unknown change point. The test for the log-linear model is substantially less powerful than the test for the constant parameter model, especially for $\tau = 0.5$. When $\tau = 0.3$, there is some dependence on y. Presumably the slightly lower power when $y = 0.6$ is due to the small number of observations that would be expected around τ in this case.

One implication of this power loss is that if we fit the log-linear model when the constant parameter model is correct and there is a moderate sized change, there is a substantial chance of being unable to detect the change. This suggests we should not use the log-linear model unless we are fairly sure there is a linear term and the constant parameter model is inadequate.

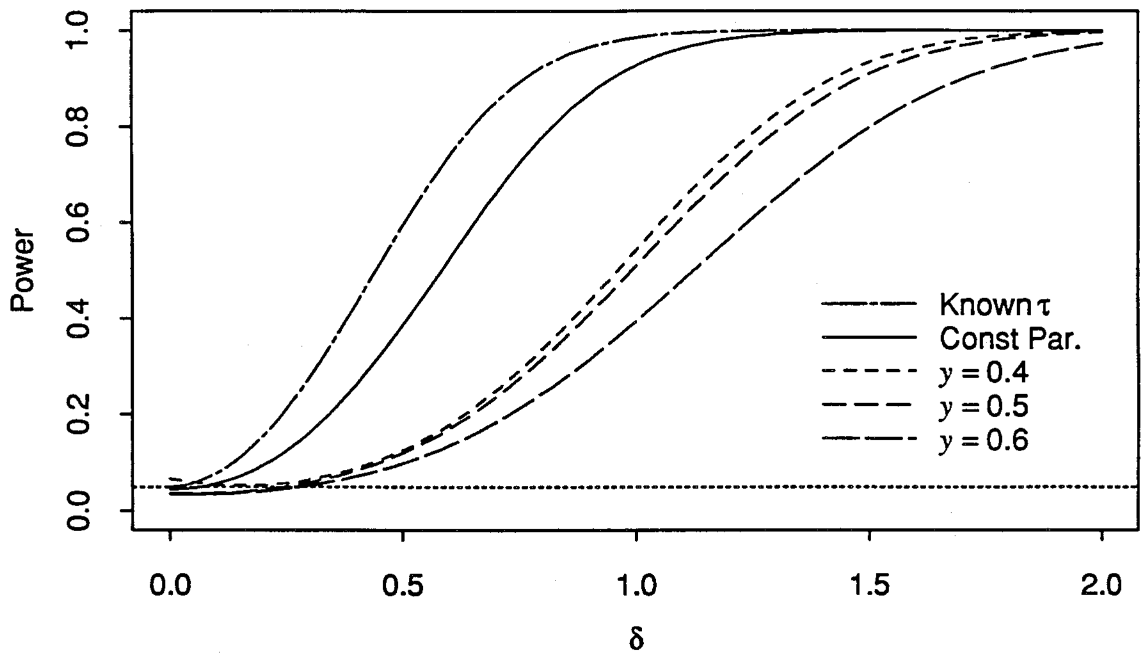


Figure 3.6: Power Comparisons for $n = 100$ and $\tau = 0.3$

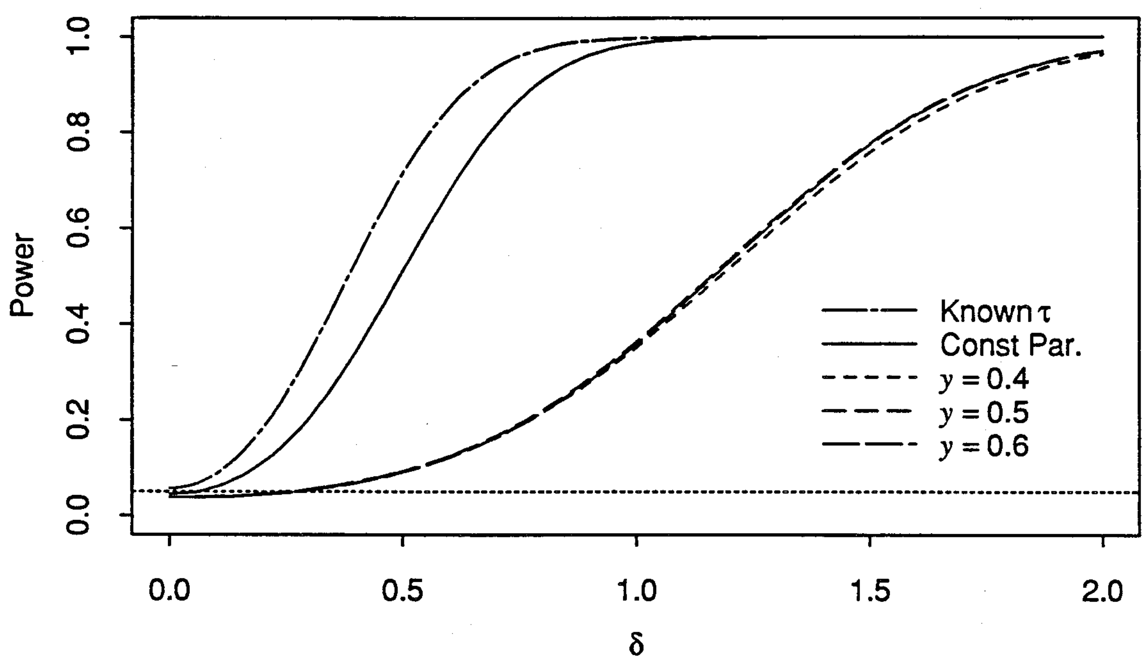


Figure 3.7: Power Comparisons for $n = 100$ and $\tau = 0.5$

Chapter 4

Confidence Regions

After performing the likelihood ratio test, the researcher may conclude there is a change point. Typically, he will want to go further, and answer questions such as 'At what time did the change occur?' and 'How large was the change?'. This chapter discusses confidence regions that will help to answer these questions. The method used is based on the likelihood ratio statistic. This is not the only method that may be used to form confidence regions. See Siegmund (1988a) for comparisons of confidence sets for general change point problems.

4.1 Confidence Regions for the Change Point

We defined the maximum likelihood estimate $\hat{\tau}$ of the change point τ in (1.7). The motivation for this estimator is that we expect $l(\tau)$ to be large, and if t is not close to $\hat{\tau}$, then $l(t)$ should be smaller. We would expect a reasonable confidence region for τ to include those values of t for which $l(t)$ is 'large enough'. This suggests a region of the form

$$I_1 = \{t : l(t) \geq l(\hat{\tau}) - c\} \tag{4.1}$$

for some suitable c. To obtain a $(1-\alpha)100\%$ confidence region, we require $P_\tau^{(m,n)}(\tau \in I_1) = 1 - \alpha$. Here, $P_\tau^{(m,n)}$ denotes a measure under which $X(t)$ is a Poisson process with rate $\lambda(t)$ specified by the constant parameter model (1.1) and change point at τ, conditional on $X(\tau) = m, X(T) = n$. We condition on both $X(\tau)$ and $X(T)$ to remove dependency on the unknown parameters λ_0 and λ_1.

We note $\tau \in I \iff l(\tau) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c$. Therefore,

$$\begin{aligned}
1 - \alpha &= P_\tau^{(m,n)}(\tau \in I_1) \\
&= P_\tau^{(m,n)}\left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau) + c\right) && (4.2) \\
&= P_\tau^{(m,n)}\left(\sup_{\tau_0 \leq t \leq \tau} l(t) \leq l(\tau) + c\right) P_\tau^{(m,n)}\left(\sup_{\tau \leq t \leq \tau_1} l(t) \leq l(\tau) + c\right) && (4.3)
\end{aligned}$$

where (4.3) follows from (4.2) since conditional on $X(\tau) = m$ and $X(T) = n$, $l(\tau)$ is known and $\{l(t), 0 < t < \tau\}$ and $\{l(t), \tau < t < T\}$ are independent. It suffices to evaluate just one of the terms on the right side of (4.3); the second term can then be evaluated by a time reversal argument.

From (4.3) we see that for fixed α, the constant c will depend on τ, m and n. We will consider asymptotic approximations under the condition $m \to \infty$ and $n \to \infty$ with $\rho = \frac{m}{n}$ fixed. We obtain different approximations for c fixed and $c \to \infty$ with $\frac{c}{n}$ fixed. We assume $\tau_0 < \tau < \tau_1$; if τ is outside this interval, then one of the probabilities on the right of (4.3) will be 1 and the range of the supremum in the other will need to be modified.

Theorem 4.1 *Suppose $m \to \infty$, $n \to \infty$ with c fixed. Then if $\rho < \frac{\tau}{T}$ (so $\hat{\delta} > 0$),*

$$\lim_{n \to \infty} P_\tau^{(m,n)}\left(\sup_{\tau \leq t \leq \tau_1} l(t) \geq l(\tau) + c\right) = e^{-c} \tag{4.4}$$

and if $\rho > \frac{\tau}{T}$ *(so* $\hat{\delta} < 0$*),*

$$\lim_{n\to\infty} P_\tau^{(m,n)}\left(\sup_{\tau\le t\le\tau_1} l(t) \ge l(\tau)+c\right) = h\left(\frac{|\hat{\delta}|}{e^{|\hat{\delta}|}-1}, \frac{c}{e^{|\hat{\delta}|}-1}\right) \tag{4.5}$$

$$\approx \nu(\hat{\delta})e^{-c} \tag{4.6}$$

where

$$\begin{aligned} & h\left(\frac{|\delta|}{e^{|\delta|}-1}, \frac{c}{e^{|\delta|}-1}\right) \\ &= \left(1-\frac{|\delta|}{e^{|\delta|}-1}\right)\sum_{j=0}^{\infty}\left(\frac{|\delta|}{e^{|\delta|}-1}\right)^j P(|\delta|U_j > c) \\ &= 1-\left(1-\frac{|\delta|}{e^{|\delta|}-1}\right)\sum_{k=0}^{[\frac{c}{|\delta|}]}\frac{(-1)^k}{k!}\left(\frac{c-k|\delta|}{e^{|\delta|}-1}\right)^k \exp\left(\frac{c-k|\delta|}{e^{|\delta|}-1}\right). \end{aligned} \tag{4.7}$$

and

$$\nu(\delta) = \frac{1-\frac{|\delta|}{e^{|\delta|}-1}}{\frac{|\delta|}{1-e^{-|\delta|}}-1}. \tag{4.8}$$

Proof: (Outline. Details will be given in Appendix A.) Differentiating $l(t)$ with respect to t and $X(t)$ (regarding $X(t)$ as continuous!) shows that for any $k > 0$ and $t < k$,

$$l(\tau+\frac{t}{n}) = l(\tau) + a\left(X(\tau+\frac{t}{n}) - X(\tau)\right) + \gamma t + o_p(1) \tag{4.9}$$

where $o_p(1)$ is uniform in t and

$$\begin{aligned} a = a(\tau,\rho) &= \log\left(\frac{T\rho}{\tau}\right) - \log\left(\frac{T(1-\rho)}{T-\tau}\right) \\ &= -\hat{\delta} \end{aligned} \tag{4.10}$$

$$\begin{aligned} \gamma = \gamma(\tau,\rho) &= \frac{1-\rho}{T-\tau} - \frac{\rho}{\tau} \\ &= \frac{1}{n}(\hat{\lambda}_1 - \hat{\lambda}_0). \end{aligned} \tag{4.11}$$

Conditional on $X(\tau) = m$, $\{X(\tau+\frac{t}{n}) - X(\tau); 0 \le t \le k\}$ converges in law to a Poisson process on $[0,T]$, with rate $\frac{1}{n}\hat{\lambda}_1 = \frac{1-\rho}{T-\tau}$. We have

$$a\frac{\hat{\lambda}_1}{n} + \gamma = \frac{\hat{\lambda}_1}{n}\log\left(\frac{\hat{\lambda}_0}{\hat{\lambda}_1}\right) + \frac{\hat{\lambda}_1 - \hat{\lambda}_0}{n}$$

$$
\begin{aligned}
&< \left(\frac{\hat{\lambda}_0}{\hat{\lambda}_1} - 1\right)\hat{\lambda}_1 + \hat{\lambda}_1 - \hat{\lambda}_0 \\
&= 0 \qquad (4.12)
\end{aligned}
$$

using the inequality $\log(x) < x - 1$ for $x \neq 1$. We can then write

$$
\begin{aligned}
&P\left(\sup_{0<t<k}(aY(t) + \gamma t) > c\right) \\
&\leq \lim_{n\to\infty} P_\tau^{(m,n)}\left(\sup_{\tau<t<\tau_1} l(t) \geq l(\tau) + c\right) \\
&\leq P\left(\sup_{0<t<k}(aY(t) + \gamma t) > c\right) + \limsup_{n\to\infty} P_\tau^{(m,n)}\left(\sup_{\tau+k/n} l(t) \geq l(\tau) + c\right) \qquad (4.13)
\end{aligned}
$$

where $Y(t)$ is a Poisson process with rate $\frac{1-\rho}{T-\tau}$. An application of Lemma A.6 shows the second term on the right of (4.13) converges to 0 as $k \to \infty$. A similar lower bound shows

$$
\lim_{n\to\infty} P_\tau^{(m,n)}\left(\sup_{\tau<t<\tau_1} l(t) \geq l(\tau) + c\right) = P\left(\sup_{0<t<\infty}(aY(t) + \gamma t) > c\right).
$$

Suppose $\rho < \frac{\tau}{T}$. Then we have $a < 0$ and $\gamma > 0$. Moreover, we have

$$
\begin{aligned}
\left(\frac{a\hat{\lambda}_1}{\gamma}\right) + \log(\hat{\lambda}_1) - \log\left(\frac{n}{T}\right) &= \frac{n-m}{T-\tau}\frac{\log(\frac{Tm}{\tau n}) - \log(\frac{T(n-m)}{n(T-\tau)})}{\frac{n-m}{T-\tau} - \frac{m}{\tau}} + \log\left(\frac{T(n-m)}{n(T-\tau)}\right) \\
&= \frac{\frac{n-m}{T-\tau}\log(\frac{Tm}{\tau n}) - \frac{m}{\tau}\log(\frac{T(n-m)}{n(T-\tau)})}{\frac{n-m}{T-\tau} - \frac{m}{\tau}} \\
&= \left(\frac{a\hat{\lambda}_0}{\gamma}\right) + \log(\hat{\lambda}_0) - \log\left(\frac{n}{T}\right) \qquad (4.14)
\end{aligned}
$$

so $\hat{\lambda}_1$ and $\hat{\lambda}_0$ are conjugate values satisfying (2.9). Theorem 2.1 then gives (4.4). Similarly if $\rho > \frac{\tau}{T}$, we can use Theorem 2.2 to obtain (4.5). Finally, Theorem 2.3 shows $h(\delta, c) \sim \nu(\delta)e^{-c}$ as $c \to \infty$ which gives (4.6).

□

Theorem 4.1 provides a simple and easily computed approximation to the right hand side of (4.3). It is based on a local linearization of the likelihood ratio process and uses standard random walk theory. We will refer to the approximation as the random walk approximation. Since it depends only on $\hat{\delta}$ and not m, n or τ individually, we might expect similar results to hold for a wide range of problems involving discontinuities in densities and other locally Poisson processes; however, this generality means we cannot expect to always get good results. It can be shown that this approximation is exact if the Poisson process is observed on $(-\infty, \infty)$ and λ_1, λ_0 are known. Asymptotic values of c are given in Table 4.1. We note that c is not a smooth function of either δ or the confidence level, so one should be careful interpolating Table 4.1.

The alternative scaling is to let $c \to \infty$ as $n \to \infty$ with $\frac{c}{n}$ fixed. This large deviation scaling results in approximations similar to (3.21) and (3.22) with $\frac{1}{2}\eta^2 = \frac{1}{n}(l(\tau) + c)$ in the definition of p_t and q_t. We omit the full details.

The recursive formulae (2.28) and (2.29) can also be used to evaluate (4.3). The details are similar to the use of recursion for the significance level calculation. It is necessary to evaluate both probabilities in (4.3) separately. See Worsley (1986) for more details.

In Figures 4.1 and 4.2, we compare the three approximations to (4.3), for the case $T = 1$, $n = 100$, $\tau_0 = 1 - \tau_1 = 0.1$ and $c = 3$. In Figure 4.1 we take $\tau = 0.3$ and in Figure 4.2, $\tau = 0.5$. In both cases we are plotting $P_\tau^{m,n}(\sup_{\tau_0 \le t \le \tau_1} l(t) \ge l(\tau) + 3)$ as a function of $m = X(\tau)$. The large deviation approximation is generally performing much better than the random walk approximation, especially for moderate size changes. The random walk approximation tends to overestimate the coverage probability. For example, when $m = 65$ and $\tau = 0.5$, we have $l(0.5) = 4.57$ which is close to the 5% significance boundary, and we would include 0.5 in the confidence region if $\sup_{\tau_0 \le t \le \tau_1} l(t) \le 7.57$. The true coverage probability is about 88.8% while the random walk approximation gives about 91.1% coverage probability. Similar

δ	Confidence Level			
	90%	95%	98%	99%
0.2	2.9370	3.6434	4.5674	5.2630
0.4	2.9055	3.6118	4.5358	5.2315
0.6	2.8752	3.5815	4.5055	5.2012
0.8	2.8451	3.5525	4.4765	5.1722
1.0	2.8174	3.5229	4.4488	5.1446
1.2	2.7950	3.4964	4.4202	5.1182
1.4	2.7622	3.4749	4.3946	5.0897
1.6	2.7465	3.4506	4.3702	5.0658
1.8	2.7326	3.4178	4.3539	5.0414
2.0	2.7057	3.4109	4.3300	5.0266
2.2	2.6664	3.4010	4.2930	5.0093
2.4	2.6197	3.3782	4.2877	4.9774
2.6	2.5824	3.3433	4.2882	4.9477
2.8	2.6243	3.3006	4.2780	4.9526
3.0	2.6506	3.2549	4.2545	4.9566

Table 4.1: Asymptotic values of c for 1-way confidence regions for τ.

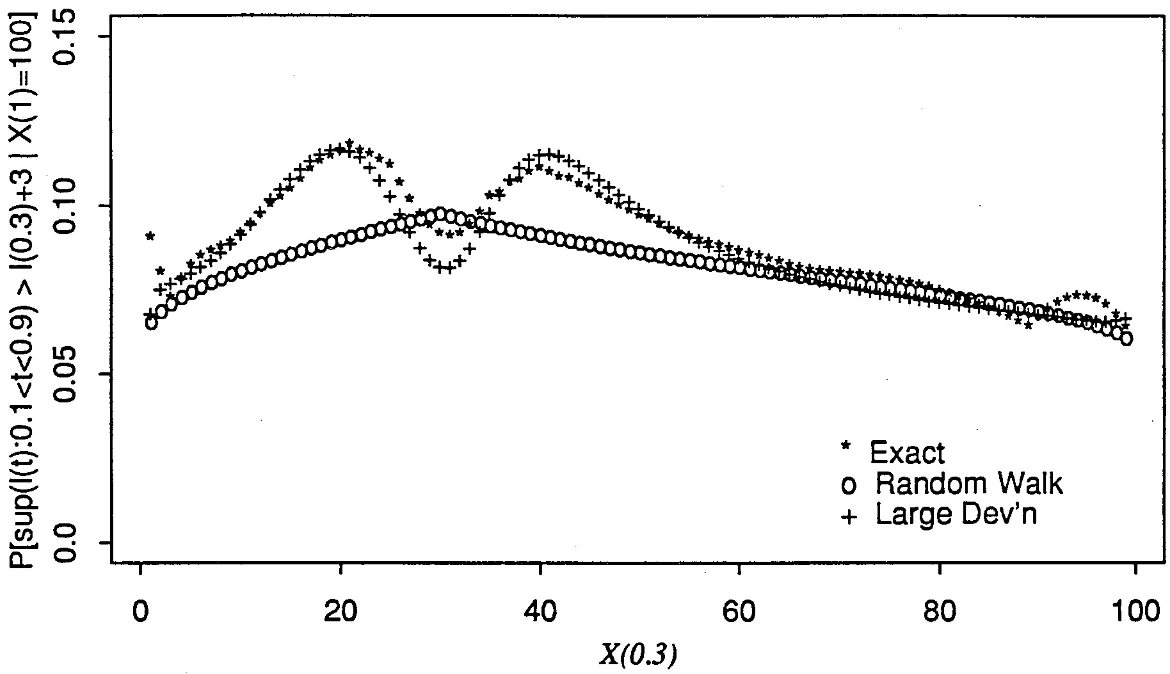


Figure 4.1: Confidence Level Approximations: $\tau = 0.3$

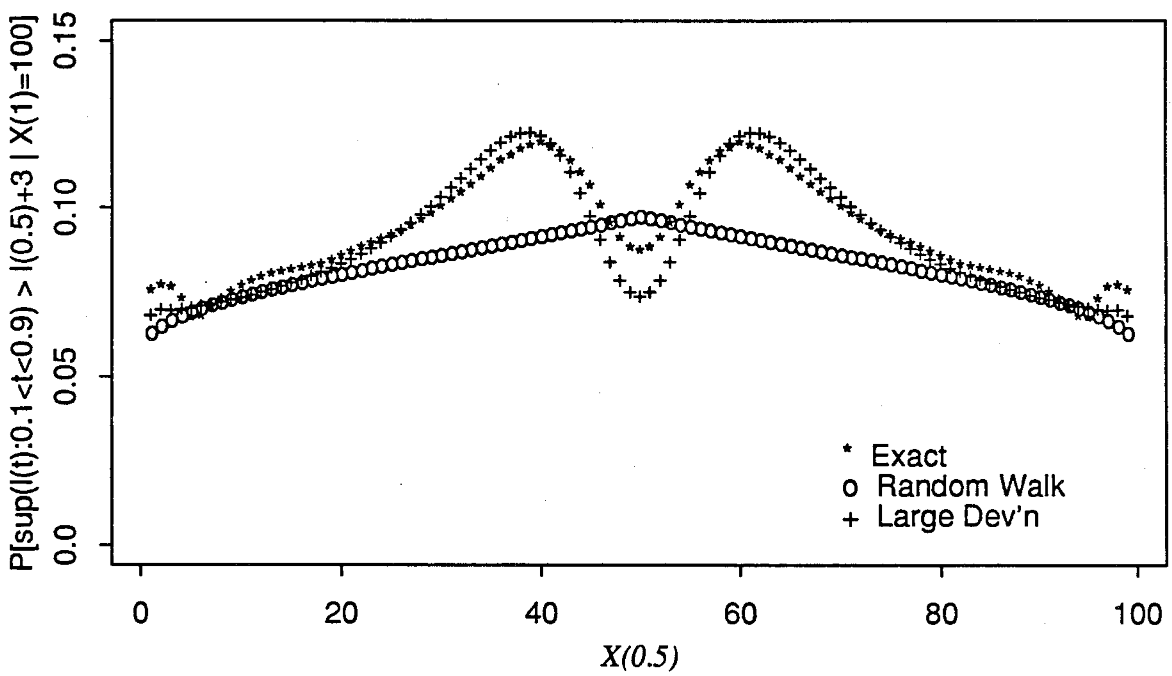


Figure 4.2: Confidence Level Approximations: $\tau = 0.5$

behaviour has been observed for other values of n and τ.

We show a 95% Confidence region for τ for Lucas' industrial accident data in Figure 4.3 and for the British Coal Mining data in Figure 4.4. The dotted lines represent $l(\hat{\tau}) - c$, where $c = c(t, X(t), n)$ is derived using the random walk approximation. The confidence region is those values of t for which $l(t)$ lies above the dotted lines.

The confidence sets are disconnected. However, in both cases the regions consist of one big segment and a few smaller segments which are only just in the interval. The total length of all segments in Figure 4.3 is 1.219 years (12.2% of the total observation period), and the total length in Figure 4.4 is 8.02 years (7.2% of the total).

	Random Walk		Exact		Large Dev'n	
1	–	–	74.041	74.044	74.041	74.050
2	74.055	74.303	74.055	74.321	74.055	74.325
3	74.353	74.359	74.353	74.376	74.353	74.381
4	74.386	74.415	74.386	74.432	74.386	74.437
5	74.460	74.471	74.460	74.488	74.460	74.493
6	–	–	74.759	74.772	74.759	74.777
7	74.795	74.811	74.795	74.829	74.795	–
8	74.833	74.869	74.833	–	–	–
9	74.885	75.757	–	75.776	–	75.781

Table 4.2: Confidence Region Comparisons for Lucas' Accident Data

We compare the approximate confidence region from Figure 4.3 with the more accurate large deviation method and the exact 95% confidence region in Table 4.2.

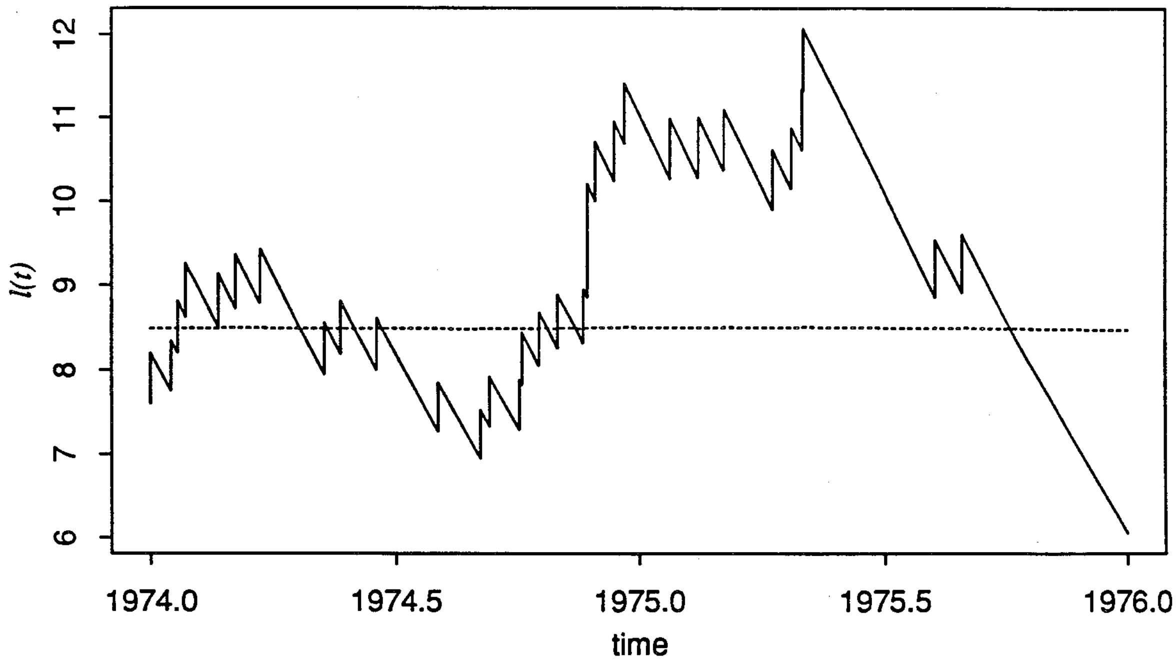


Figure 4.3: Industrial Accident Data: 95% Confidence Region for τ

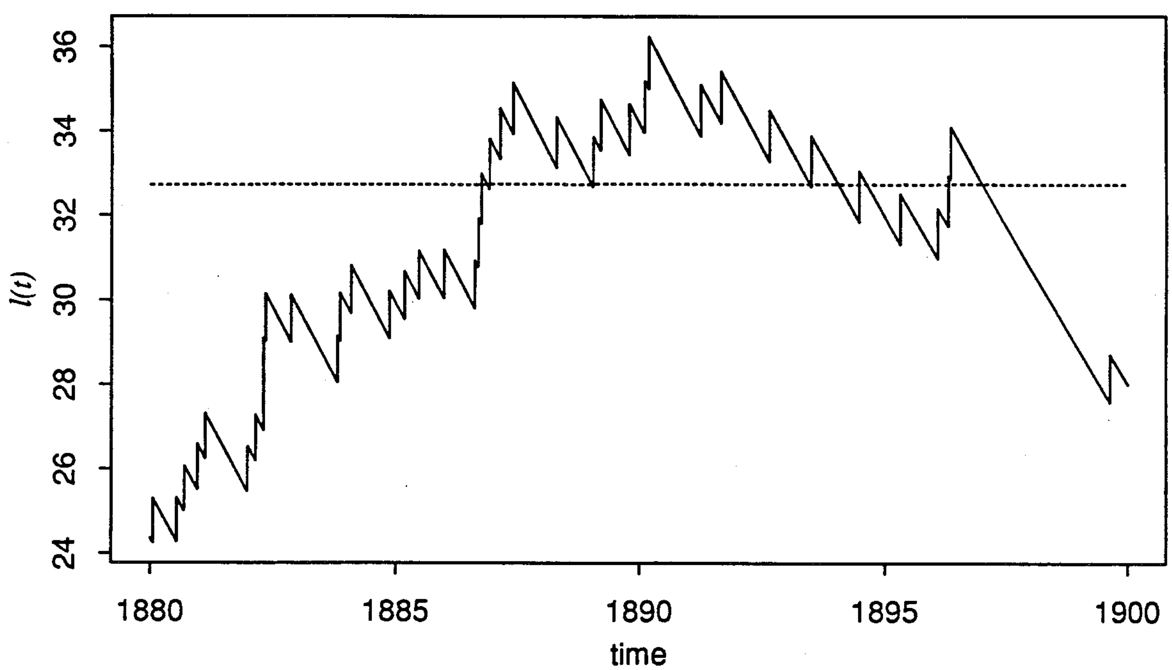


Figure 4.4: Coal Mining Data: 95% Confidence Region for τ

We have subtracted 1900 from all the years. Each line of the table represents one interval of the confidence region. The first interval in Line 1 includes just one point, and does not appear in the confidence region when the random walk approximation is used. Similarly, the interval in Line 6 is empty when the random walk approximation is used. The intervals in Lines 8 and 9 for the random walk approximation are merged into one interval when the exact method is used; the interval in Line 7 also merges with this when the large deviation approximation is used. The total lengths of all the intervals is 1.219, 1.357 and 1.395 years for the random walk, exact and large deviation method respectively.

4.2 Joint Confidence Regions

The likelihood ratio method can be extended to find joint confidence regions for τ and δ. Let $l(\tau|\delta)$ denote the likelihood ratio for testing $\mathcal{H}_0 : \lambda_0 = \lambda_1$ against the restricted alternative $\mathcal{H}_1 : \lambda_1 = e^{\delta}\lambda_0$. From (1.4) we get

$$l(\tau|\delta) = X(t)\log\left(\frac{T}{\tau + e^{\delta}(T-\tau)}\right) + (n - X(\tau))\log\left(\frac{Te^{\delta}}{\tau + e^{\delta}(T-\tau)}\right)$$

Then a joint confidence set may take the form

$$I_2 = \{(t,\delta) : l(t|\delta) \geq l(\hat{\tau}) - c\}.$$

Here, c will be chosen so that

$$P_{\tau,\delta}((\tau,\delta) \in I_2 | X(T) = n) = 1 - \alpha. \tag{4.15}$$

Conditioning on $X(\tau)$, we get

$$P^{(n)}_{\tau,\delta}\left\{\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau|\delta) + c\right\}$$

$$
\begin{aligned}
&= \sum_{m=0}^{n} P_{\tau}^{(m,n)} \left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \le l(\tau|\delta) + c \right\} P_{\tau,\delta}^{(n)}(X(\tau) = m) \\
&= \sum_{m=0}^{n} P_{\tau}^{(m,n)} \left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \le l(\tau) + c - (l(\tau) - l(\tau|\delta)) \right\} \\
&\quad \times P_{\tau,\delta}^{(n)}(X(\tau) = m). \qquad (4.16)
\end{aligned}
$$

The sum in (4.16) can be restricted to values of m such that $l(\tau) - l(\tau|\delta) \le c$. If $X(\tau) = nx$, we have

$$
\begin{aligned}
&\frac{1}{n}(l(\tau) - l(\tau|\delta)) \\
&= x \log\left(\frac{x(\tau + e^{\delta}(T-\tau))}{\tau}\right) + (1-x)\log\left(\frac{(1-x)(\tau + e^{\delta}(T-\tau))}{(T-\tau)e^{\delta}}\right). \qquad (4.17)
\end{aligned}
$$

The second derivative of (4.17) with respect to x is $1/x(1-x) \ge 4$ and hence

$$
l(\tau) - l(\tau|\delta) \ge 2n \left(\frac{X(\tau)}{n} - \frac{\tau}{\tau + e^{\delta}(T-\tau)} \right)^2 .
$$

Hence the range of summation in (4.16) can be restricted to a subset of

$$
\frac{\tau}{\tau + e^{\delta}(T-\tau)} - \sqrt{\frac{c}{2n}} \le \frac{m}{n} \le \frac{\tau}{\tau + e^{\delta}(T-\tau)} + \sqrt{\frac{c}{2n}}. \qquad (4.18)
$$

Applying Theorem 4.1 and compactness arguments gives, for any $0 < x_0 \le x_1 < 1$,

$$
\begin{aligned}
\sup_{nx_0 \le m \le nx_1} \Bigg| & P_{\tau}^{(m,n)} \left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \le l(\tau) + c - (l(\tau) - l(\tau|\delta)) \right\} \\
& -(1 - e^{-(c-x)}) \left(1 - h\left(\frac{|\hat{\delta}|}{e^{|\hat{\delta}|} - 1}, \frac{c-x}{e^{|\hat{\delta}|} - 1}\right) \right) \Bigg| \to 0 \qquad (4.19)
\end{aligned}
$$

as $n \to \infty$, where $x = l(\tau) - l(\tau|\delta)$. Moreover, we have $\sup |\hat{\delta} - \delta| \to 0$ where the supremum is taken over the range defined by (4.18), and hence (4.19) still holds if we replace $\hat{\delta}$ by δ. By standard likelihood theory, $2(l(\tau) - l(\tau|\delta))$ converges in law to a χ_1^2 random variable, and hence we can approximate (4.16) by

$$
\begin{aligned}
& P_{\tau,\delta}^{(n)} \left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \le l(\tau|\delta) + c \right\} \\
& \to \int_0^c (1 - e^{-(c-x)}) \left(1 - h\left(\frac{|\delta|}{e^{|\delta|} - 1}, \frac{c-x}{e^{|\delta|} - 1}\right) \right) \frac{1}{\sqrt{\pi x}} e^{-x} dx \qquad (4.20)
\end{aligned}
$$

as $n \longrightarrow \infty$.

In Table 4.3 we give values of c for various confidence levels and δ, derived from (4.20). The same points will hold asymptotically if we want to find confidence regions for τ and either λ_0 or λ_1.

If we pretend the approximation (4.6) is valid for all $c > 0$ instead of just as $c \rightarrow \infty$, we obtain the approximation

$$\begin{aligned} P^{(n)}_{\tau,\delta}((\tau,\delta) \in I_2) &\approx \int_0^c (1-e^{-(c-x)})(1-\nu(|\delta|)e^{-(c-x)})\frac{1}{\sqrt{\pi x}}e^{-x}dx \\ &= P(\chi_1^2 < 2c) - (1+\nu(|\delta|))e^{-c}\int_0^c \frac{1}{\sqrt{\pi x}}dx \\ &\quad +\nu(|\delta|)e^{-2c}\int_0^c \frac{1}{\sqrt{\pi x}}e^{x}dx \\ &\approx 1-2(1-\Phi(\sqrt{2c})) - \frac{2}{\sqrt{\pi}}(1+\nu(|\delta|))\sqrt{c}e^{-c} + \frac{\nu(|\delta|)}{\sqrt{\pi c}}e^{-c} \\ &\approx 1-\sqrt{c}e^{-c}\left(\frac{2}{\sqrt{\pi}}(1+\nu(|\delta|)) + \frac{1}{c\sqrt{\pi}}(1-\nu(|\delta|))\right). \end{aligned} \tag{4.21}$$

This is essentially equation (26) of Siegmund (1988a). It is not too difficult to establish the asymptotic relation

$$\begin{aligned} &\int_0^c (1-e^{-(c-x)})\left(1-h(\frac{|\delta|}{e^{|\delta|}-1}, \frac{c-x}{e^{|\delta|}-1})\right)\frac{1}{\sqrt{\pi x}}e^{-x} \\ &= 1-2\sqrt{\frac{c}{\pi}}e^{-c}(1+\nu(|\delta|))(1+o(1)) \end{aligned}$$

as $c \longrightarrow \infty$.

Large deviation methods may also be used to approximate the conditional boundary crossing probabilities. The calculation is similar to that for the confidence region for τ; however, because we sum over m, this is much more computational. Likewise, the recursive method may be used. For fixed c we must evaluate the summand of (4.16) for the $O(n^{1/2})$ values of m such that $l(\tau) - l(\tau|\delta) < c$, and therefore the exact evaluation of (4.15) is an $O(n^{5/2})$ computation.

δ	Confidence Level			
	90%	95%	98%	99%
0.2	3.7242	4.5204	5.5431	6.3024
0.4	3.6929	4.4890	5.5117	6.2709
0.6	3.6630	4.4589	5.4816	6.2408
0.8	3.6343	4.4303	5.4529	6.2121
1.0	3.6066	4.4027	5.4254	6.1846
1.2	3.5812	4.3760	5.3991	6.1584
1.4	3.5579	4.3517	5.3733	6.1332
1.6	3.5340	4.3292	5.3497	6.1084
1.8	3.5108	4.3079	5.3277	6.0860
2.0	3.4945	4.2847	5.3080	6.0651
2.2	3.4760	4.2646	5.2889	6.0460
2.4	3.4536	4.2514	5.2676	6.0291
2.6	3.4299	4.2360	5.2472	6.0101
2.8	3.4080	4.2165	5.2355	5.9896
3.0	3.3907	4.1946	5.2258	5.9724

Table 4.3: Asymptotic values of c for 2-way confidence regions for τ and δ.

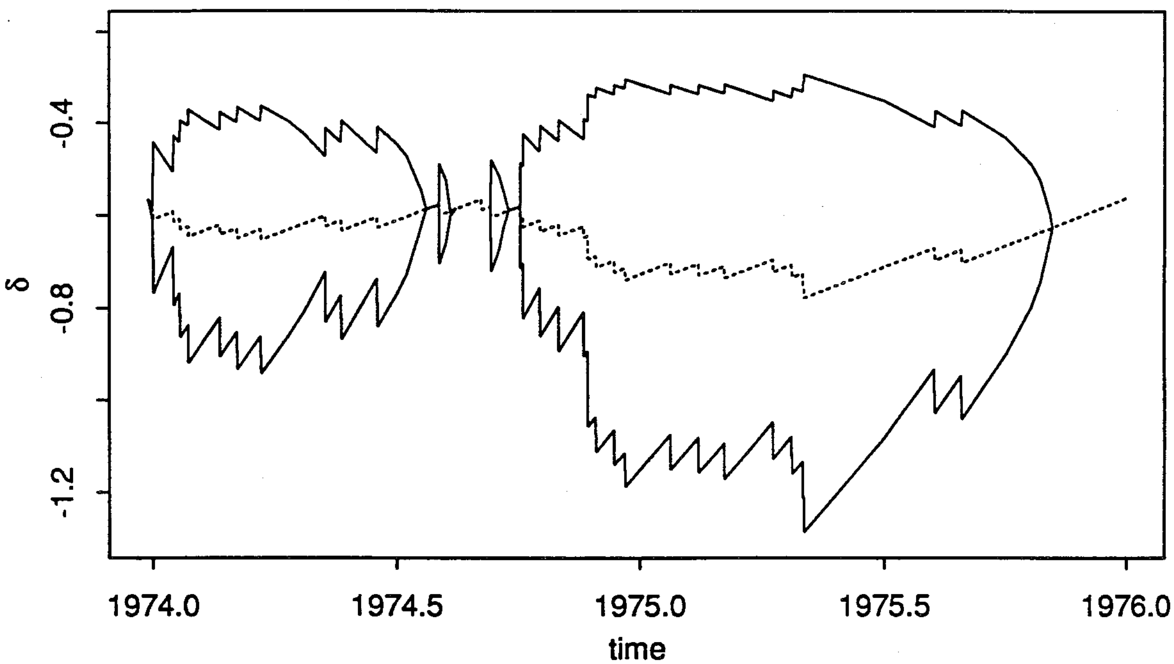


Figure 4.5: Industrial Accident Data: 95% Confidence Region for (τ, δ)

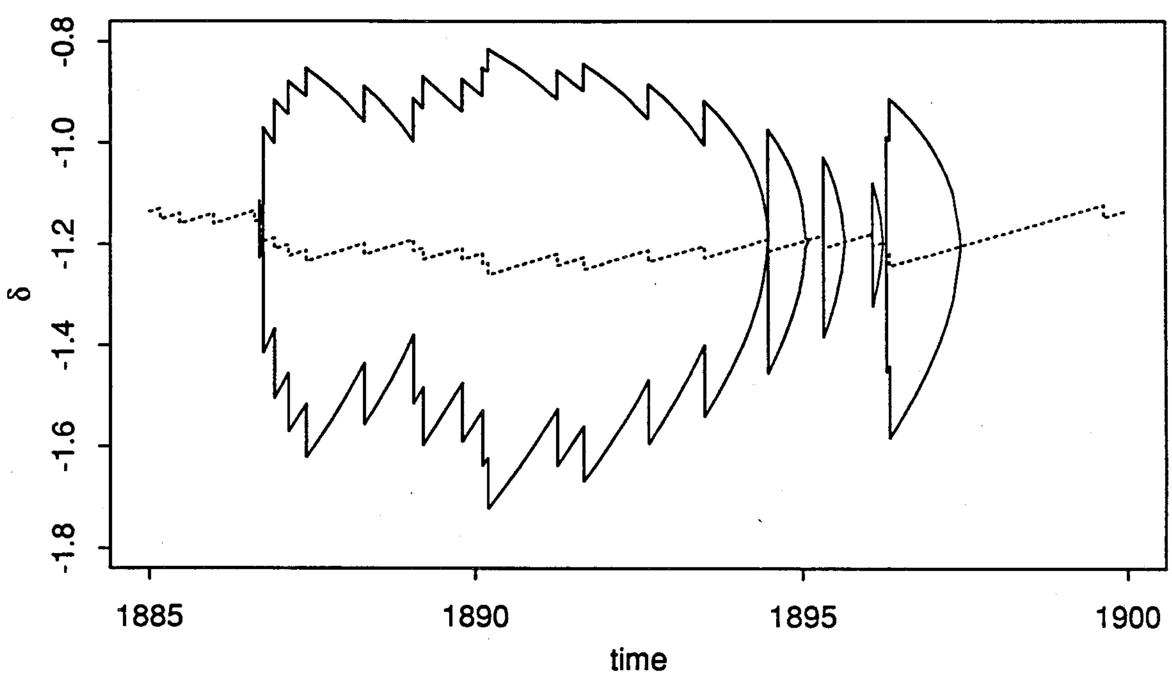


Figure 4.6: Coal Mining Data: 95% Confidence Region for (τ, δ)

Figure 4.5 displays an approximate 95% confidence region for (τ, δ) for Lucas' industrial accident data, computed using (4.20). The solid line outlines the confidence region and the dotted line is the maximum likelihood estimate of δ, as a function of t. A similar region for the British coal mining data is shown in Figure 4.6. As with the confidence regions for τ, these regions are disconnected, consisting of one big region and several smaller regions. In both cases, $\delta = 0$ is well outside of the confidence region so there is strong evidence of the presence of a change point.

4.3 Confidence Regions: Log-Linear Model

4.3.1 One and Two Way Confidence Sets

The methods described in the previous section and in particular Theorem 4.1 can be extended to the log-linear model. The confidence region I_1 for τ will again take the form given by (4.1). However, to evaluate the coverage probability we must condition on $S_n = \sum_{i=1}^{n} T_i$ to remove dependence on the unknown nuisance parameter b. We then require

$$\begin{aligned} 1-\alpha &= P_\tau^{(m,n)}(\tau \in I_1 | S_n = ny) \\ &= P_\tau^{(m,n)}\left(\sup_{\tau_0 \le \tau \le \tau_1} l(t) \le l(\tau) + c | S_n = ny\right). \end{aligned} \tag{4.22}$$

However, because of the conditioning on S_n, the processes $\{l(t); 0 \le t \le \tau\}$ and $\{l(t); \tau \le t \le T\}$ are not independent and therefore we cannot split (4.22) in the same manner as we split (4.2). To approximate (4.22) we must concentrate on the local properties of $l(t)$ for t close to τ.

Suppose we fix $S_n = ny$ as $n \to \infty$. As in the proof of Theorem 4.1,

$$l(\tau + \frac{t}{n}) = l(\tau) + a\left(X(\tau + \frac{t}{n}) - X(\tau)\right) + \gamma(t - \tau) + o_p(1)$$

where $o_p(1)$ holds uniformly in t for $0 \le t \le k$ and any $k < \infty$. We find a and γ by differentiating $l(t)$ with respect to $X(t)$ and t, holding y fixed:

$$\begin{aligned} a = a(\tau, \rho, y) &= -\hat{\delta} \\ \gamma = \gamma(\tau, \rho, y) &= \frac{(1-\rho)e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}} - \frac{\rho e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} \end{aligned}$$

where $\rho = \frac{m}{n}$. Moreover, we show in Appendix B that $\{X(\tau + \frac{t}{n}) - X(\tau); 0 < t < k\}$ and $\{X(\tau) - X(\tau - \frac{t}{n}); 0 < t < k\}$ for large n behave like independent Poisson processes with rates $\rho e^{\hat{b}\tau}/(e^{\hat{b}\tau} - 1)$ and $(1-\rho)e^{\hat{b}\tau}/(e^{\hat{b}T} - e^{\hat{b}\tau})$. As in (4.12) we have

$$-\hat{\delta}\frac{(1-\rho)e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}} < 0.$$

Using the same heuristic argument given in support of Theorem 4.1, we reach exactly the same conclusion in this case. As in (4.14), the local rates of $X(t)$ at $t = \tau$ are a conjugate pair satisfying (2.9). We will make this argument more rigorous in Appendix B.

The large deviation scaling, where we let $c \to \infty$ with $\frac{c}{n}$ fixed, can also be applied to this case. The details are similar to those for the power calculations for the log-linear model in Section 3.4, which will lead to integrals like (3.32) and (3.31) with $\eta^2 = \frac{2}{n}(l(\tau) + c)$. We omit the details. To apply this method, we must split (4.22) for $t > \tau$ and $t < \tau$. This cannot be justified using asymptotic independence since this scaling does not lead to asymptotic independence. However, the probabilities of exceeding the boundary on either side of τ will decrease exponentially in n and the probability of exceeding on both sides will decrease exponentially with a higher exponential order so will be asymptotically negligible.

We give 95% confidence regions for τ for Lucas' accident data in Figure 4.7 and for the coal mining data in Figure 4.8. For the accident data, the confidence region is very similar to that derived from the constant parameter model in 4.3. The main

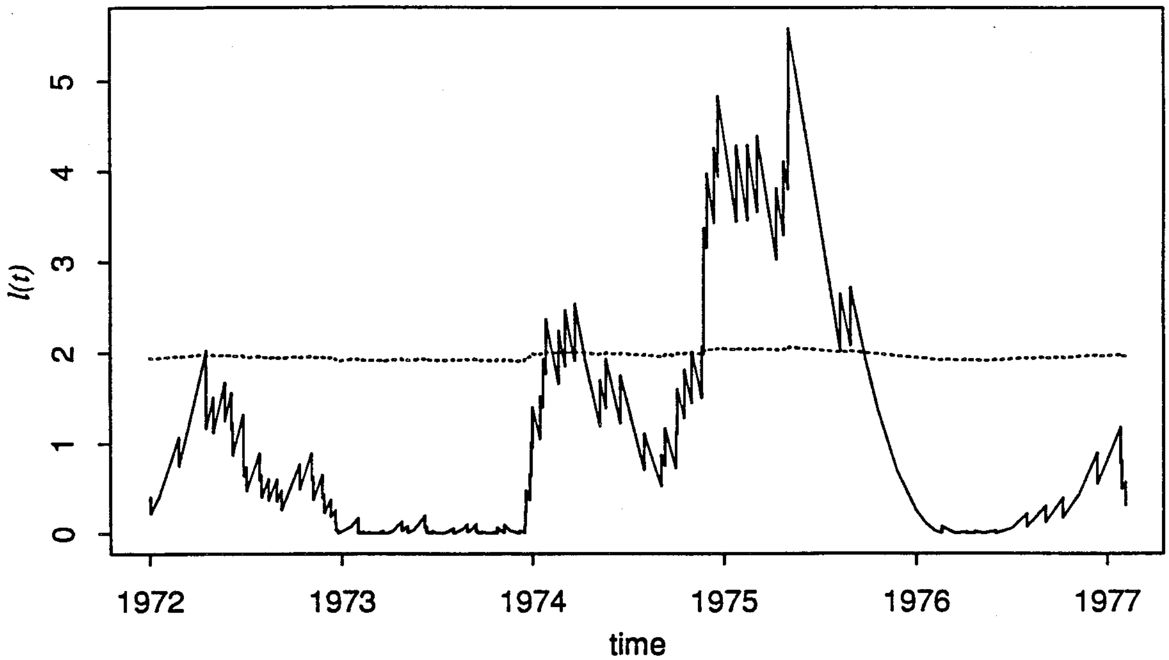


Figure 4.7: Industrial Accident Data: 95% Confidence Region for τ

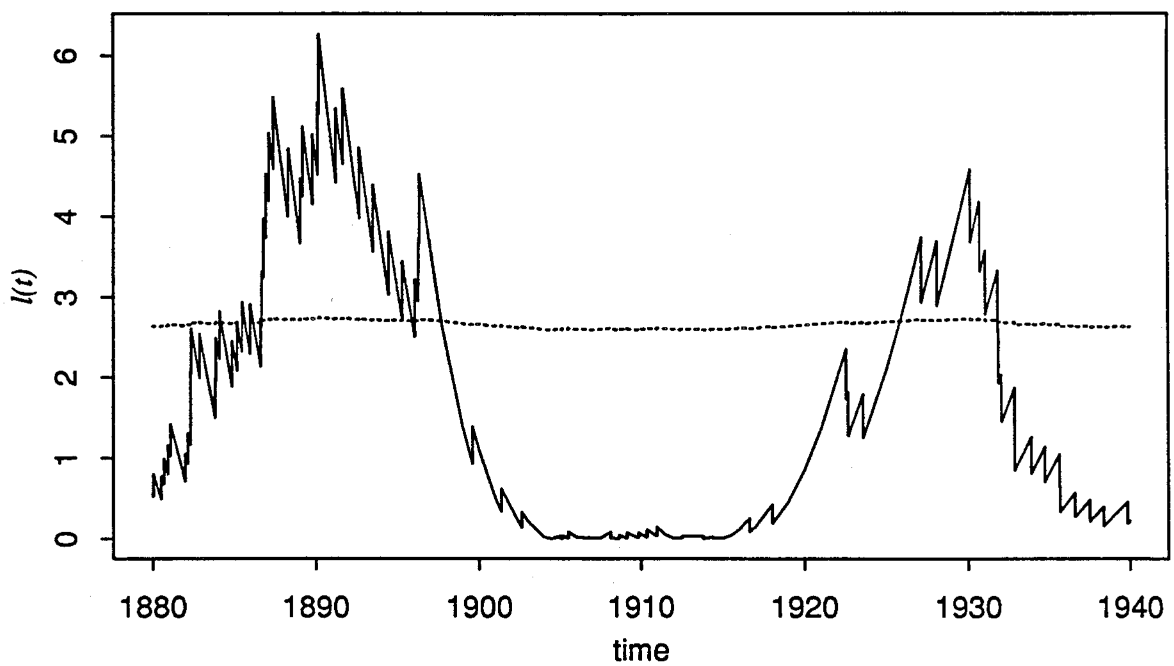


Figure 4.8: Coal Mining Data: 95% Confidence Region for τ

parts of the confidence region, between 1974.0 and 1976.0, have total length 0.98 years, which is slightly less than the corresponding size in Figure 4.3. However, one point in 1972 is just in the confidence region.

The main part of the confidence region for the coal mining data in Figure 4.8 occurs around 1890 has length 11.37 years, which is larger than the corresponding region of Figure 4.4. In addition, an interval around the second peak in 1930 is also included in this region. However, since we suspect a second change around this time, if we are interested in finding a confidence region for this change point we should split the data and form a separate region, rather than using the whole dataset. We should also note this second peak is crossing the lower boundary which Theorem 4.1 ignores.

We can also find joint confidence sets for τ and δ as before. The conditional likelihood is now

$$\begin{aligned} l(\tau|\delta) &= X(t)\log\left(\frac{\hat{b}_\delta(e^{\hat{b}_0 T}-1)}{b_0(e^{\hat{b}_\delta \tau}-1+e^\delta(e^{\hat{b}_\delta T}-e^{\hat{b}_\delta \tau}))}\right) \\ &\quad +(n-X(t))\log\left(\frac{\hat{b}_\delta e^\delta(e^{\hat{b}_0 T}-1)}{b_0(e^{\hat{b}_\delta \tau}-1+e^\delta(e^{\hat{b}_\delta T}-e^{\hat{b}_\delta \tau}))}\right)+(\hat{b}_\delta-b_0)\sum_{i=1}^n T_i \end{aligned}$$

where b_δ is the maximum likelihood estimator of b under the restricted alternative and is the solution of

$$\frac{1}{n}\sum_{i=1}^n T_i = \frac{\tau e^{b\tau}+e^\delta(Te^{bT}-\tau e^{b\tau})}{e^{b\tau}-1+e^\delta(e^{bT}-e^{b\tau})}-\frac{1}{b}.$$

The confidence level is

$$\begin{aligned} 1-\alpha &= P^{(n)}_{\tau,\delta}((\tau,\delta)\in I_2|S_n=ny) \\ &= \sum_{m=m_0}^{m_1} P^{(m,n)}_\tau\left\{\sup_{\tau_0\le t\le\tau_1}(l(t)-l(\tau))\le c-(l(\tau)-l(\tau|\delta))|S_n=ny\right\} \\ &\quad \times P^{(n)}_{\tau,\delta}(X(\tau)=m|S_n=ny) \end{aligned} \tag{4.23}$$

where m_0 and m_1 are chosen (dependent on y) so that $l(\tau) - l(\tau|\delta) < c$ if and only if $m_0 \leq X(\tau) \leq m_1$. Treating (4.23) in the same manner as we treated (4.17), we again get

$$\begin{aligned} & P_{\tau,\delta}^{(n)}\left\{\sup_{\tau_0\leq\tau\leq\tau_1} l(t) \leq l(\tau|\delta) + c|S_n = ny\right\} \\ = & \sum_{\{m:l(\tau)-l(\tau|\delta)<c\}} P_{\tau}^{(m,n)}\left\{\sup_{\tau_0\leq\tau\leq\tau_1} l(t) \leq l(\tau|\delta) + c|S_n = ny\right\} \\ \to & \int_0^c (1-e^{-(c-x)})\left(1 - h(\frac{|\delta|}{e^{|\delta|}-1}, \frac{c-x}{e^{|\delta|}-1})\right)\frac{1}{\sqrt{\pi x}}e^{-x}dx \end{aligned}$$

as $n \to \infty$ with y fixed. Using the approximation (4.8) again leads to the approximation (4.21).

The large deviation methods may also be used to approximate (4.23). The method is similar to that for the constant parameter model, and we omit the details.

We give a 95% joint confidence region for (τ, δ) for Lucas' Industrial accident data in Figure 4.9. The range of times covered is similar to the times covered in Figure 4.5, but the range of δ for fixed τ is slightly larger. For the coal mining data (Figure 4.10), the joint confidence region is substantially larger than in Figure 4.6, and the values of δ included are larger (closer to 0).

4.3.2 Three Way Confidence Sets

We can extend previous methods in an obvious way to find joint confidence sets for τ, δ and b. We let $l(\tau|\delta, b)$ be the likelihood ratio statistic for the restricted alternative $\mathcal{H}_1 : \delta \neq 0$ with δ and b known,

$$\begin{aligned} l(\tau|\delta, b) &= X(\tau)\log\left(\frac{b(e^{b_0T}-1)}{b_0(e^{b\tau}-1+e^{\delta}(e^{bT}-e^{b\tau}))}\right) \\ &\quad +(n-X(\tau))\log\left(\frac{be^{\delta}(e^{b_0T}-1)}{b_0(e^{b\tau}-1+e^{\delta}(e^{bT}-e^{b\tau}))}\right) + (b-b_0)\sum_{i=1}^{n} T_i \end{aligned}$$

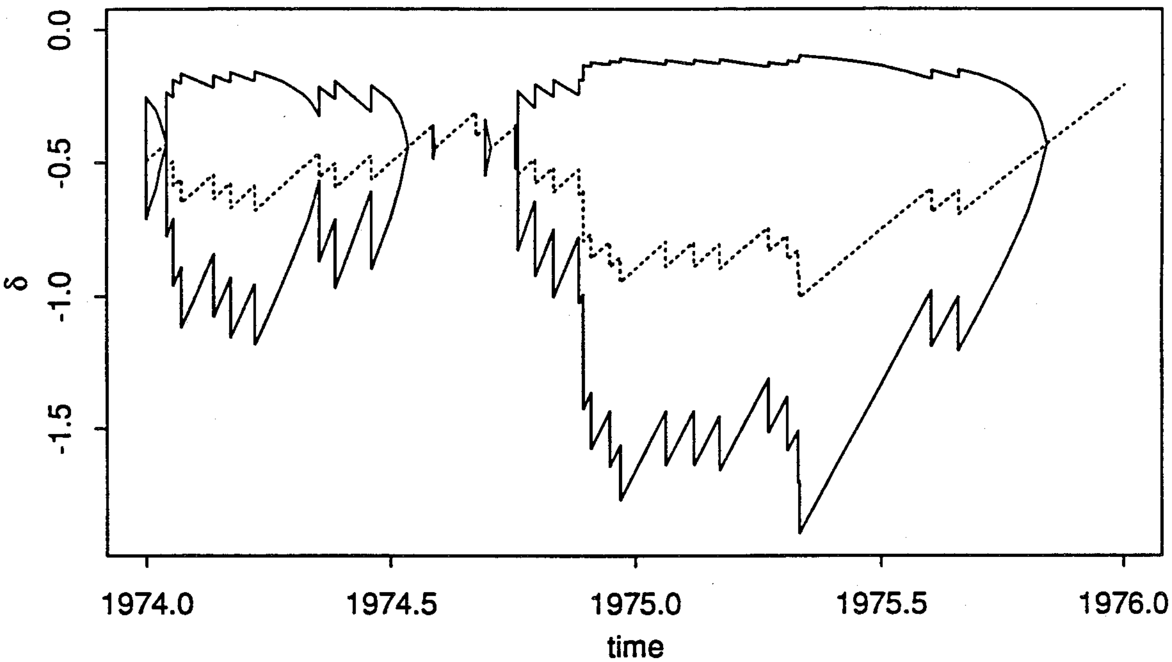


Figure 4.9: Industrial Accident Data: 95% Confidence Region for (τ, δ)

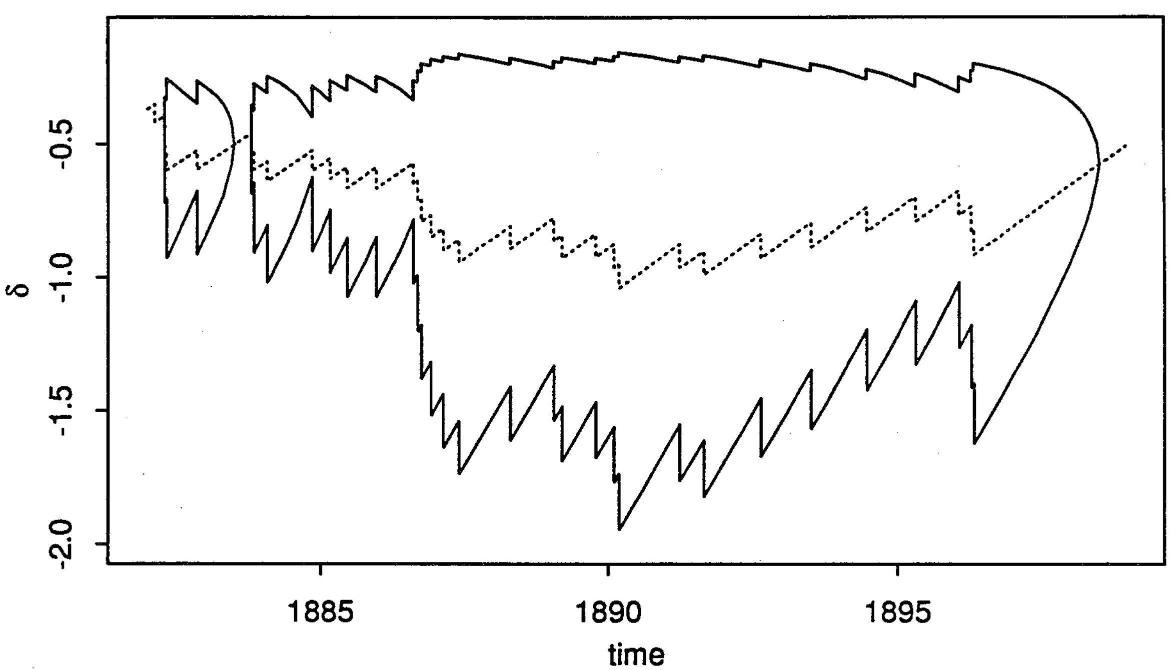


Figure 4.10: Coal Mining Data: 95% Confidence Region for (τ, δ)

where b_0 is the maximum likelihood estimate of b under $\mathcal{H}_0$. The confidence region will then be

$$I_3 = \left\{(\tau, \delta, b) : l(\tau|\delta, b) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c\right\}$$

where c is chosen so that

$$P^{(n)}_{\tau,\delta,b}\left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau|\delta, b) + c\right) = 1 - \alpha. \tag{4.24}$$

We can approximate (4.24) by conditioning on $X(\tau)$ and $\sum_{i=1}^{n} T_i$. We get

$$\begin{aligned}
&P^{(n)}_{\tau,\delta,b}\left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau|\delta, b) + c\right) \\
&= \sum_{m=0}^{n} \int_0^T P^{(m,n)}_{\tau}\left(\sup_{\tau_0 \leq t \tau_1} l(t) \leq l(\tau|\delta, b) + c | S_n = ny\right) \\
&\quad \times P^{(n)}_{\tau,\delta,b}(X(\tau) = m, S_n = ny)dy
\end{aligned}$$

Since $2(l(\tau) - l(\tau|\delta, b))$ has asymptotically a chi-square distribution with 2 degrees of freedom, we get the approximation

$$\begin{aligned}
&P^{(n)}_{\tau,\delta,b}((\tau, \delta, b) \in I_3) \\
&\approx \int_0^c (1 - e^{-(c-x)}) \left(1 - h(\frac{|\delta|}{e^{|\delta|}}, \frac{c - x}{e^{|\delta|} - 1})\right) e^{-x} dx \\
&= 1 - e^{-c} - ce^{-c} - \int_0^c (e^{-x} - e^{-c}) h(\delta, c - x) dx.
\end{aligned} \tag{4.25}$$

If we use the approximation (4.8) we get the approximation

$$\begin{aligned}
&P^{(n)}_{\tau,\delta,b}\left\{\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau|\delta, b) + c\right\} \\
&\approx 1 - ce^{-c}\left(1 + \nu(|\delta|) - \frac{1}{c}(1 - \nu(|\delta|))\right)
\end{aligned}$$

in place of (4.25). Asymptotic values of c for various confidence levels, derived from (4.25), are given in Table 4.4.

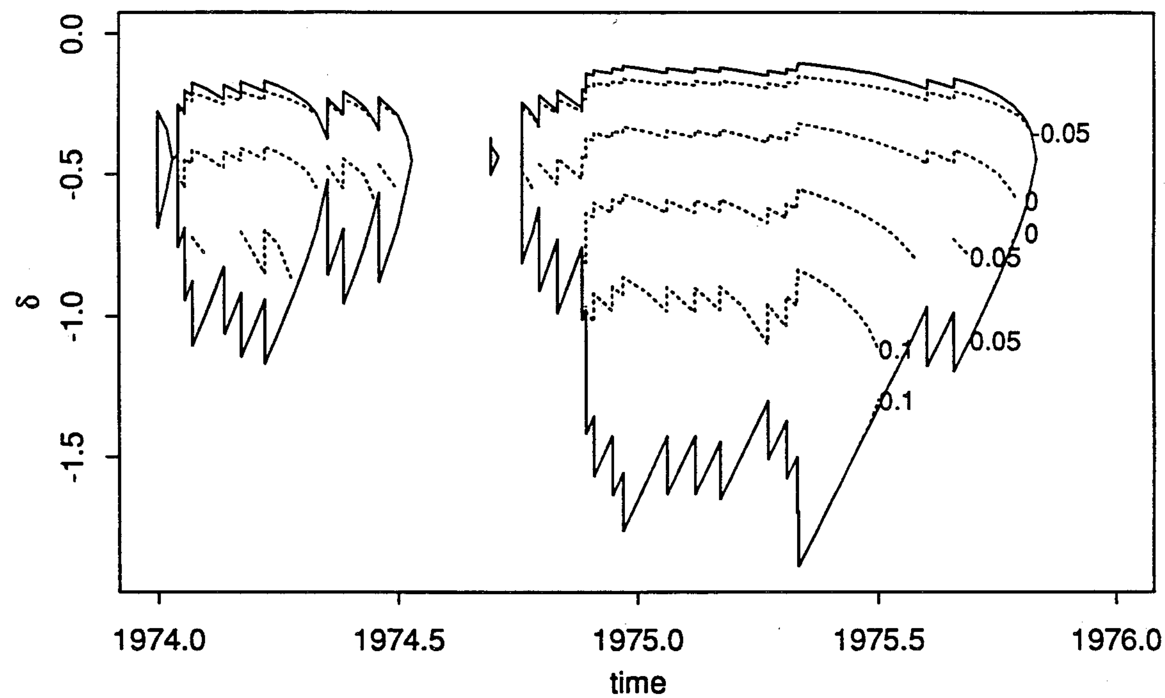
We show approximate 90% confidence regions for (τ, δ, b) for Lucas' accident data in Figure 4.11 and for the British coal mining accident data in Figure 4.12.

δ	Confidence Level			
	90%	95%	98%	99%
0.2	4.4687	5.3375	6.4402	7.2514
0.4	4.4375	5.3063	6.4089	7.2200
0.6	4.4078	5.2764	6.3790	7.1901
0.8	4.4379	5.2479	6.3504	7.1614
1.0	4.3524	5.2208	6.3232	7.1342
1.2	4.3268	5.1948	6.2972	7.1082
1.4	4.3030	5.1703	6.2724	7.0834
1.6	4.2806	5.1476	6.2488	7.0597
1.8	4.2588	5.1261	6.2269	7.0372
2.0	4.2379	5.1058	6.2064	7.0164
2.2	4.2197	5.0858	6.1872	6.9969
2.4	4.2025	5.0670	6.1689	6.9786
2.6	4.1844	5.0509	6.1509	6.9616
2.8	4.1660	5.0361	6.1336	6.9449
3.0	4.1487	5.0202	6.1184	6.9283

Table 4.4: Asymptotic values of c for 3-way confidence regions for (τ, δ, b).

Since representing a three dimensional region on a two dimensional page is difficult, some description of how to interpret the plots is necessary. For each value of δ and τ, there may be a range of values of b in the confidence region. The solid lines in Figure 4.11 represent the boundary of this region. Inside this region, we compute for each τ and δ the upper and lower extrema of the range of b such that (τ, δ, b) lie in the confidence region. The dotted lines then represent contour plots of the upper limits for b (top panel) and lower limits for b (bottom panel).

In Figure 4.11, for most values of t and δ, the upper limit for b is positive and the lower limit is negative. This is consistent with the hypothesis $b = 0$ and suggests the constant parameter model is adequate for this dataset. By contrast, in Figure 4.12, the upper limit for b is negative for many values of τ and δ, suggesting that the regression term is important for this dataset.



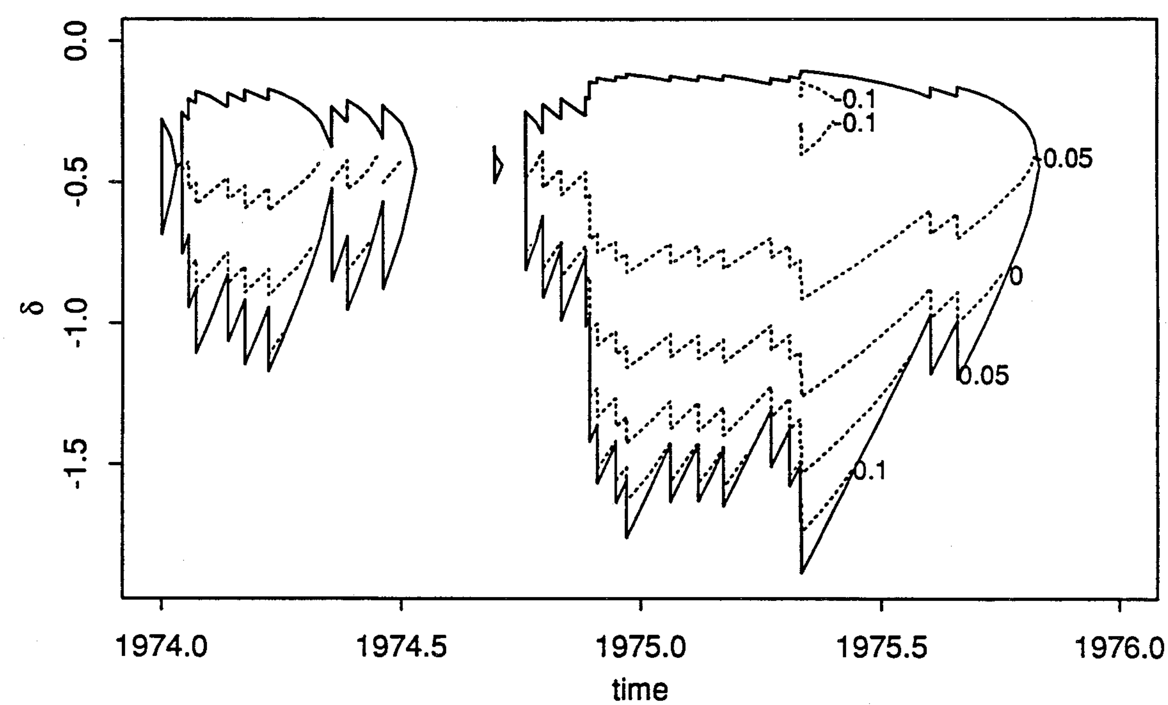


Figure 4.11: Lucas' Accident data: 90% Confidence region for (τ, δ, b)

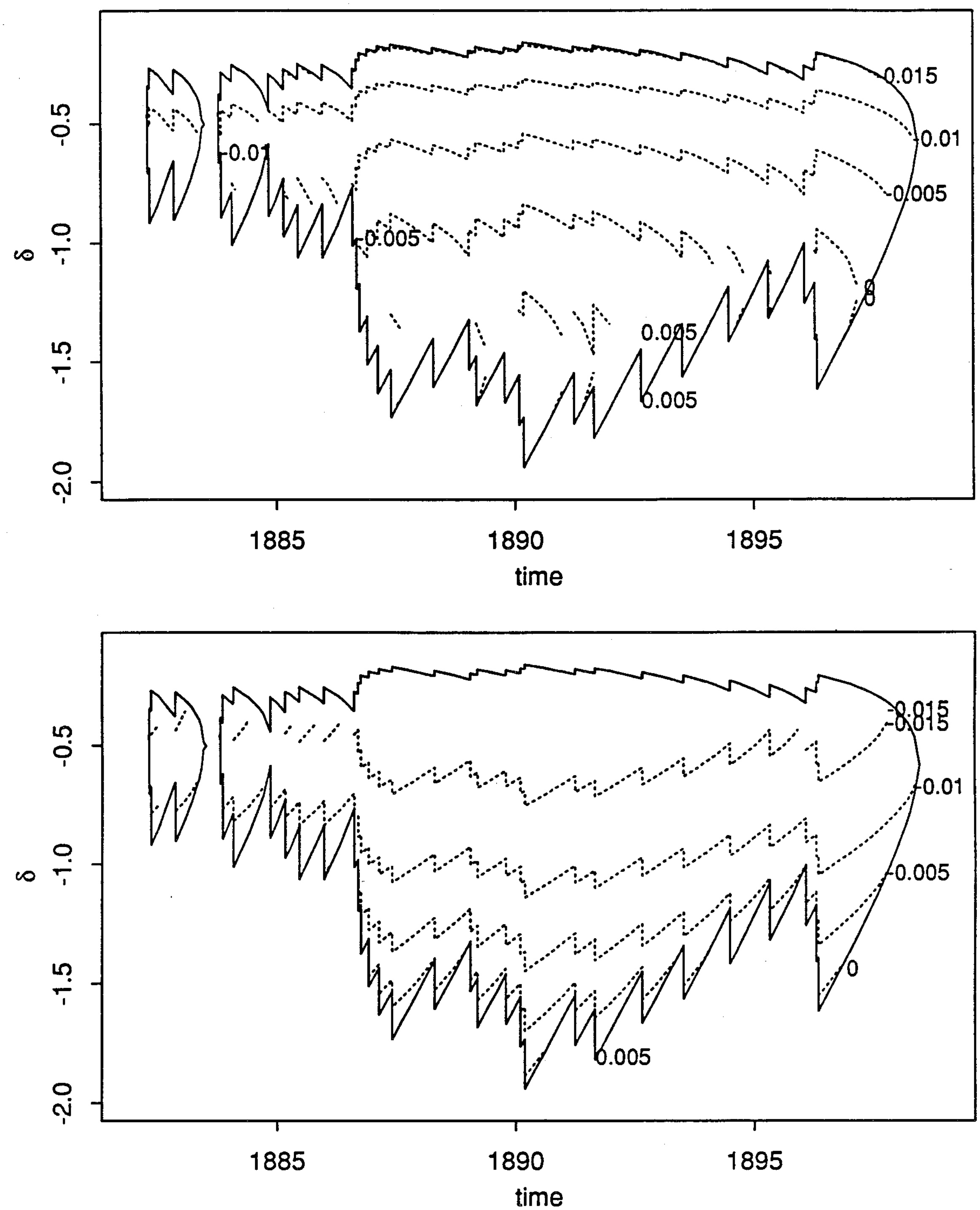


Figure 4.12: Coal mining data: 90% Confidence region for (τ, δ, b)

Chapter 5

Changes in Hazard Rates

In this chapter we depart from our discussion of Poisson processes and discuss a related model in Survival Analysis. We suppose independent subjects have a constant hazard of failing up to an unknown change point τ, at which time the hazard changes to a new level. This model was originally proposed by Matthews and Farewell (1982) who used simulations to derive the distribution of the likelihood ratio statistic for testing the null hypothesis of no change. The model has subsequently been studied by other authors.

5.1 The Hazard Rate Model

Suppose $T_1, \ldots, T_n$ denote independent identically distributed survival times of patients in a study. The hazard rate is defined by

$$h(t) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} P(T_1 \leq t + \epsilon | T_1 > t).$$

The change point model proposed by Matthews and Farewell (1982) is

$$h(t) = \begin{cases} \lambda_0 & 0 \le t < \tau \\ \lambda_1 & t \ge \tau \end{cases}. \tag{5.1}$$

The density $f(t)$ of T_1 can then be written

$$f(t) = h(t)\exp(-\int_0^t h(u)du) = \begin{cases} \lambda_0 \exp(-\lambda_0 t) & 0 \le t < \tau \\ \lambda_1 \exp(-\lambda_0\tau - \lambda_1(t-\tau)) & t \ge \tau \end{cases}.$$

The likelihood function based on observations $T_1, \ldots, T_n$ is

$$L(\lambda_0, \lambda_1, \tau) = \lambda_0^{X(\tau)}\lambda_1^{n-X(\tau)} \exp\left(-\lambda_0 \sum_{i=1}^{n}(T_i \wedge \tau) - \lambda_1 \sum_{i=1}^{n}(T_i - \tau)^+\right) \tag{5.2}$$

where $X(t) = \sum_{i=1}^{n} I(T_i \le t)$ is the number of deaths observed up to time t. For fixed τ, the maximum likelihood estimates of λ_0 and λ_1 can be found by differentiating (5.2), which gives

$$\hat{\lambda}_0 = \frac{X(\tau)}{\sum_{i=1}^{n}(T_i \wedge \tau)} \tag{5.3}$$

and

$$\hat{\lambda}_1 = \frac{n - X(\tau)}{\sum_{i=1}^{n}(T_i - \tau)^+}. \tag{5.4}$$

The maximum likelihood estimate $\hat{\tau}$ of τ is found by substituting $\hat{\lambda}_0$ and $\hat{\lambda}_1$ in (5.2) and maximizing over $\tau \in [\tau_0, \tau_1]$. The formal definition is similar to (1.7).

The use of (5.3) and (5.4) with an estimate of τ is common to most authors. However, the use of the maximum likelihood estimator $\hat{\tau}$ is not universal. One problem is that $L(\hat{\lambda}_0, \hat{\lambda}_1, t) \to \infty$ as $t \to T_{(n)}$. and hence it is necessary to restrict the range of the maximum likelihood estimator. Yao (1986) restricts to $\hat{\tau} \le T_{(n-1)}$ and then shows $\hat{\tau} - \tau = O_p(n^{-1})$. Nguyen, Rogers and Walker (1984) and Chang, Chen and Hsiung (1989) both propose alternative estimators of τ which they prove to be consistent, but do not establish a rate of convergence or provide any comparisons with other estimators.

The likelihood ratio statistic for testing $\mathcal{H}_0 : \lambda_0 = \lambda_1$ against the alternative $\mathcal{H}_1 : \lambda_0 \neq \lambda_1$ for known τ is given by

$$l(\tau) = X(\tau) \log \left(\frac{X(\tau) \sum_{i=1}^n T_i}{n \sum_{i=1}^n (T_i \wedge \tau)} \right) + (n - X(\tau)) \log \left(\frac{(n - X(\tau)) \sum_{i=1}^n T_i}{n \sum_{i=1}^n (T_i - (T_i \wedge \tau))} \right). \quad (5.5)$$

A likelihood ratio test for unknown τ can be based on the maximum of $l(t)$ over an interval $[\tau_0, \tau_1]$. The significance level of the likelihood ratio test is then

$$\alpha = P_0 \left\{ \sup_{\tau_0 < t < \tau_1} l(t) \geq \frac{1}{2} c^2 \right\}. \quad (5.6)$$

Unfortunately if τ_0 and τ_1 are constants, then (5.6) will depend on the unknown initial hazard rate λ_0. One possible solution to this problem is to condition on the sufficient statistic $\sum_{i=1}^n T_i$. However, the resulting crossing times cannot be discretized as we did for the Poisson process model in Chapter 3 and our methods cannot be applied in this case. Alternatively, we can choose τ_0 and τ_1 depending on the observed times $T_1, \ldots, T_n$ in such a way to make (5.6) independent of λ_0. This will be discussed in Theorem 5.1 below.

Tests based on the likelihood ratio have been discussed by several authors. Matthews, Farewell and Pyke (1985) show the normalized score process (essentially the signed square root of $l(t)$) converges in law to a time transformation of a normalized Brownian bridge. The significance level can then be approximated by (3.3) or tabulations of the probability. Worsley (1988) obtains results based on a time transformation which under $\mathcal{H}_0$ transforms $X(t)$ into a Poisson process. He then uses the recursive methods similar to those discussed in Section 2.4 to evaluate the significance level.

Theorem 5.1 *Define a random time transformation by* $s = \sum_{i=1}^n (T_i \wedge t)$ *and let* $Y(s) = X(t)$. *If there is no change in the hazard rate, then* $Y(s)$ *is a Poisson process with rate* λ_0, *observed up to the time of the* n*th event.*

Proof: Since the event times are exchangeable, it suffices to show the result conditional on $T_1 \leq \ldots \leq T_n$. Let $S_0 = 0$ and $S_j = \sum_{i=1}^n (T_i \wedge T_j)$ for $j = 1, \ldots, n$. It suffices to show $Y_j = S_j - S_{j-1}$ are independent exponential random variables with parameter λ_0. Note $S_j = \sum_{i=1}^{j-1} T_i + (n-j+1)T_j$ and therefore

$$\begin{aligned} Y_j &= T_{j-1} + (n-j+1)T_j - (n-j+2)T_{j-1} \\ &= (n-j+1)(T_j - T_{j-1}). \end{aligned}$$

Conditional on $T_1 \leq \ldots \leq T_n$, the joint density of $T_1, \ldots, T_n$ is

$$n!\lambda_0^n \exp(-\lambda_0 \sum_{i=1}^n T_i).$$

Since

$$\sum_{j=1}^n Y_j = S_n - S_0 = \sum_{j=1}^n T_j$$

and the transformation from T_j to Y_j is linear, the Jacobian is constant. Therefore, the joint density of $Y_1, \ldots, Y_n$ is proportional to $\exp(-\lambda_0 \sum_{i=1}^n Y_i)$ over the range $Y_j \geq 0$, $j = 1, \ldots, n$. It follows that $Y_j, j = 1, \ldots, n$ are *i.i.d.* exponential random variables with parameter λ_0 and hence $Y(s)$ is a Poisson process with rate λ_0, observed up to the time $S = \sum_{i=1}^n T_i$ of the nth event.

□

Under the time transformation of Theorem 5.1, the likelihood ratio statistic becomes

$$l_Y(s) = Y(s) \log\left(\frac{Y(s)S}{ns}\right) + (n - Y(s)) \log\left(\frac{(n - Y(s))S}{n(S-s)}\right). \tag{5.7}$$

Note that (5.7) is identical to the likelihood ratio statistic (1.8) for the constant parameter model. However, on the transformed scale, the change point τ is a random variable. Also, we know the nth event occurs at time S, so from (5.7) we see

$l_Y(s) \to \infty$ as $s \to S$. We set limits $u_0 S$ and $u_1 S$ for fixed $u_0 \geq 0$ and $u_1 < 1$. The significance level is then

$$\alpha = P_0 \left\{ \sup_{u_0 S \leq s \leq u_1 S} l_Y(s) \geq \frac{1}{2} c^2 \middle| S_n = S \right\}. \tag{5.8}$$

The probability (5.8) does not depend on λ_0. We see the relation between u_j and τ_j for $j = 1, 2$ is

$$u_j = \frac{\sum_{i=1}^n (T_i \wedge \tau_j)}{\sum_{i=1}^n T_i}.$$

A slightly different method of specifying the limits, used by Worsley (1988), is to truncate a fixed proportion of the events from each end. Our method of approximating (5.8) can be adapted for this modification.

We can approximate (5.8) using formulae similar to (3.6) and (3.7). We condition on S and let s_j be the solution of $l_Y(s) = \frac{1}{2}c^2$, subject to $s_j \leq \frac{j}{n} S$. We let

$$\mathcal{S} = \inf\{s \geq u_0 S : Y(s) \leq np_s\}$$

be the first passage time to the lower boundary, then use (2.23) to obtain

$$P_0(u_0 S \leq \mathcal{S} \leq u_1 S) \approx \sum_{j=m_0}^{m_1} \left(1 - \frac{a(s_j)}{s_j a'(s_j)}\right) \binom{n-1}{na(s_j)} \left(\frac{s_j}{S}\right)^{na(s_j)} \left(\frac{S - s_j}{S}\right)^{n(1-a(s_j))-1}$$

where m_0 and m_1 are chosen so that $u_0 S \leq s_j \leq u_1 S$ when $m_0 \leq j \leq m_1$. Using the relation $\binom{n-1}{np_s} = (1 - p_s)\binom{n}{np_s}$ we get the approximation

$$\begin{aligned} &P_0 \left\{ \inf_{u_0 S \leq s \leq u_1 S} (Y(s) - np_s) \leq 0 \right\} \\ &\quad \approx \; P_0(Y(u_0 S) \leq np_{u_0 S}) + c\phi(c) \int_{u_0 S}^{u_1 S} \frac{1 - p_s}{1 - \frac{s}{S}} \frac{p_s' - \frac{p_s}{s}}{\eta\sqrt{p_s(1 - p_s)}} ds. \end{aligned} \tag{5.9}$$

Since under $P_0^{(n)}$, $Y(u_0 S) \sim \mathcal{B}(n-1, u_0)$, the endpoint correction $P_0^{(n)}(Y(u_0 S) \leq np_{u_0 S})$ can be evaluated either by summing the Binomial probabilities or using an approximation such as (A.12). Note the additional factor of $\frac{1-p_s}{1-s/S}$ in (5.9) which does not appear in (3.6). Similarly, when we treat the upper boundary by time reversal, we get an additional factor of $\frac{1-q_s}{1-s/S}$ to include in (3.7), giving

$$
\begin{aligned}
& P_0\left\{\sup_{u_0 S \leq s \leq u_1 S}(Y(s) - nq_s) \geq 0\right\} \\
\approx \quad & P_0(Y(u_1 S) \geq nq_{u_1 S}) + c\phi(c)\int_{u_0 S}^{u_1 S} \frac{1-q_s}{1-\frac{s}{S}} \frac{q_s' - \frac{1-q_s}{S-s}}{\eta\sqrt{q_s(1-q_s)}} ds. \qquad (5.10)
\end{aligned}
$$

There are some additional differences for the second order terms which will be discussed in Appendix A.

We have fixed limits on the transformed time scale rather than the original time scale. Since the initial rate λ_0 is unknown, it may make more sense for limits to be specified here, since this automatically precludes detection of a change point based on a few observations. By comparison, Matthews, Farewell and Pyke (1985) fix limits on the original time scale, but use a transformation $s = \exp(-\lambda_0 t)$. Since λ_0 must be estimated from the data and then the limits are transformed, it would be more logical to have fixed limits on the transformed scale here also. Note that the Matthews, Farewell and Pyke transformation is asymptotically equivalent to a time reversal of our transformation.

So far we have assumed that all the times are uncensored. In medical applications of survival analysis, this is not normally the case. Typically, we have for the i^{th} individual a survival time X_i and a censoring time C_i. We observe (T_i, J_i) where $T_i = X_i \wedge C_i$ and $J_i = I(X_i \leq C_i)$.

The simplest stochastic model for censoring is to assume the C_i are *i.i.d.* from a distribution G, with density g. Moreover, we assume that $(X_1, \ldots, X_n)$ and $(C_1, \ldots, C_n)$ are independent. We then have

$$
\begin{aligned}
P(T_i \in (t, t+dt), J_i = 1) &= P(X_i \in (t, t+dt), C_i \geq t) \\
&= f(t)(1 - G(t))dt
\end{aligned}
$$

and similarly

$$P(T_i \in (t, t+dt), J_i = 0) = g(t)(1 - F(t))dt.$$

The likelihood is then given by

$$L(\tau, \lambda_0, \lambda_1, g) = \prod_{\substack{i=1 \\ J_i=1}}^{n} f(T_i)(1 - G(T_i)) \prod_{\substack{i=1 \\ J_i=0}}^{n} g(T_i)(1 - F(T_i)) \tag{5.11}$$

which is a function of g multiplied by (5.2), with $X(\tau)$ interpreted as the number of deaths observed up to time τ and $n - X(\tau)$ replaced by $N - X(\tau)$, where N is the total number of deaths observed.

Provided we do not specify a parametric family for g involving any of the existing parameters, we see from (5.11) that the maximum likelihood estimates derived without censoring are still valid with censoring. Moreover, the conclusion of Theorem 5.1 is still partially valid. The process $Y(s)$ will still evolve as a Poisson process. However, we do not observe the process up to the time of the final event, and the final time $S = \sum_{i=1}^{n} T_i$ does not seem to have any simple interpretation with respect to the Poisson process.

Given the partial extension of Theorem 5.1 to the case with censoring, we would expect the effect of censoring on the significance level approximations to be minor. However, if the last observation is censored, we might expect the process $\{Y(s); 0 \leq s \leq S\}$ to behave as if the time S were fixed rather than an event time. In this case, it would be appropriate to use (3.6) and (3.7) to approximate the significance level, rather than (5.9) and (5.10). To make more rigorous statements requires making distribution assumptions about the censoring times.

Suppose $G(t) = 1 - e^{-\mu t}$ so the C_i are exponentially distributed with parameter μ. Let $\tilde{X}(t)$ be the number of patients lost (either through death or censoring) up to time t, and $\tilde{Y}(s) = \tilde{X}(t(s))$. Then under $\mathcal{H}_0$, $\tilde{Y}$ is a Poisson process with rate $\lambda_0 + \mu$, observed up to the nth event. Each jump of $\tilde{Y}$ independently represents a death with probability $\frac{\lambda_0}{\lambda_0+\mu}$.

Suppose we have N uncensored observations, and the last observation is censored. Then $S_1, \ldots, S_N$ are equal in law to N *i.i.d.* uniform $[0, S]$ order statistics, and we may treat $Y(s)$ as a Poisson process observed up to a fixed time S, conditional on $Y(S) = N$. In this case, it is appropriate to use (3.6) and (3.7) to approximate the significance level.

Now suppose the last observation is uncensored. Then $S_1, \ldots, S_{N-1}$ are equal in law to $N-1$ *i.i.d.* uniform $[0, S]$ order statistics, and we may regard $Y(s)$ as a Poisson process observed up to S_N. In this case, it is appropriate to use (5.9) and (5.10) to approximate the significance level.

In both cases, we are calculating the significance level conditional on the number of uncensored observations and whether or not the last observation was censored. However, since conditional on N uncensored observations the probability that the last observation is uncensored is $\frac{N}{n}$, we can uncondition on this to obtain a significance level conditional only on the number of uncensored observations. The approximations may break down if there is heavy censoring and the exponential model for censoring times is poor.

5.2 Confidence Regions

The local approximations described in Chapter 4 can be used to find approximate confidence sets for the hazard rate problem. We will have a confidence region of the form

$$I_1 = \left\{ t : l(t) \geq \sup_{\tau_0 \leq t' \leq \tau_1} l(t') - c \right\}.$$

where c is chosen so that

$$1 - \alpha = P_\tau \left\{ \tau \in I_1 \mid X(\tau) = m, \sum_{i=1}^{n} (T_i \wedge \tau) = ny_1, \sum_{i=1}^{n} T_i = ny \right\}. \tag{5.12}$$

We condition on $X(\tau)$, $\sum_{i=1}^{n}(T_i \wedge \tau)$ and $\sum_{i=1}^{n} T_i$ in (5.12) to remove dependency on the unknown λ_0 and λ_1. Differentiating (5.5) shows that

$$l(\tau + \frac{t}{n}) - l(\tau) = a(X(\tau + \frac{t}{n}) - X(\tau)) + \gamma t + o_p(1)$$

as $n \to \infty$, where $o_p(1)$ holds uniformly for $t \in [-T, T]$ and any $T > 0$, and

$$\begin{aligned} a &= \log\left(\frac{my}{ny_1}\right) - \log\left(\frac{(n-m)y}{n(y-y_1)}\right) \\ &= -\hat{\delta} \\ \gamma &= (1 - \frac{m}{n})\left(\frac{n-m}{n(y-y_1)} - \frac{m}{ny_1}\right) \\ &= (1 - \frac{m}{n})(\hat{\lambda}_1 - \hat{\lambda}_0). \end{aligned}$$

Moreover, for $T > 0$, the process $\{X(\tau + \frac{t}{n}) - X(\tau); 0 < t < T\}$ converges in law to a Poisson process with rate $(1 - \frac{m}{n})\hat{\lambda}_1$ as $n \to \infty$ with y_1 and y fixed. Similarly, the local left process converges to a Poisson process with rate $(1 - \frac{m}{n})\hat{\lambda}_0$. A straightforward calculation, similar to (4.14), shows these are a conjugate pair satisfying (2.9), and therefore the result of Theorem 4.1 extends to this case. This is also true with censoring if we let $N \to \infty$ with $\frac{N}{n}$ fixed.

The use of Theorem 4.1 to approximate (5.12) will presumably have the same problem it had for our study of Poisson processes; namely that it is not always particularly accurate. We would like to be able to use recursive or large deviation methods to compare the results. However, the transformation given by Theorem 5.1 transforms τ to a random time and so we do not have any similar results under the alternative; in particular, $Y(s)$ is not a Poisson process with a change point.

We can also use the methods of Chapter 4 to find joint confidence regions for τ and $\delta = \log(\lambda_1/\lambda_0)$. Under the restricted alternative $\mathcal{H}_1 : \lambda_1 = e^{\delta}\lambda_0$ the maximum likelihood estimate of λ_0 is

$$\hat{\lambda}_0 = \frac{n}{\sum_{i=1}^n (T_i \wedge \tau) + e^{\delta} \sum_{i=1}^n (T_i - \tau)^+}$$

and the log-likelihood ratio statistic is

$$\begin{aligned} l(\tau|\delta) &= X(\tau)\log\left(\frac{\sum_{i=1}^n T_i}{\sum_{i=1}^n (T_i \wedge \tau) + e^{\delta}\sum_{i=1}^n (T_i - \tau)^+}\right) \\ &\quad +(n - X(\tau))\log\left(\frac{e^{\delta}\sum_{i=1}^n T_i}{\sum_{i=1}^n (T_i \wedge \tau) + e^{\delta}\sum_{i=1}^n (T_i - \tau)^+}\right). \end{aligned}$$

A joint confidence region for τ and δ will take the form

$$I_2 = \left\{(t,\delta) : l(t|\delta) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c\right\}$$

where c is chosen so that

$$P_{\tau,\delta}((\tau,\delta) \in I_2 \,|\, V_n = ny) = 1 - \alpha \tag{5.13}$$

where $V_n = \sum_{i=1}^n (T_i \wedge \tau) + e^{\delta}\sum_{i=1}^n (T_i - \tau)^+$ is the sufficient statistic for λ_0 when τ and δ are known. We can write

$$\begin{aligned} &P_{\tau,\delta}((\tau,\delta) \in I_2 \,|\, V_n = ny) \\ &= P_{\tau,\delta}\left\{\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau|\delta) + c \,|\, V_n = ny\right\} \\ &= P_{\tau,\delta}\left\{\sup_{\tau_0 \leq t \leq \tau_1} (l(t) - l(\tau)) \leq c - (l(\tau) - l(\tau|\delta)) \,|\, V_n = ny\right\}. \end{aligned}$$

As in Chapter 4, $2(l(\tau) - l(\tau|\delta))$ converges in law to a χ_1^2 random variable and is asymptotically independent of $\sup_{\tau_0 \leq t \leq \tau_1}(l(t) - l(\tau))$. We can therefore use Table 4.3 to approximate (5.13).

5.3 Example: Heart Transplant Data

To illustrate the methods developed in this chapter, we fit the change point model (5.1) to the Stanford heart transplant data, as published by Crowley and Hu (1977). Data is given for 103 patients; however, we consider only the 69 who actually received a transplant. Of these patients, 45 died during the followup period and 29 deaths were attributed to rejection of the transplanted heart (for 4 patients, cause of death was not recorded).

The likelihood ratio process is plotted on the original time scale in Figure 5.1 and on the transformed time scale in Figure 5.2. There is clear evidence of a change point early on in the time period. The maximum likelihood of 27.24 is attained at $\hat{\tau} = 68.01$ days. The sharply peaked likelihood indicates a change point model is preferable to a model with smoothly changing hazard rate. To check the validity of the assumption of exponential censoring times, we interchanged the censored and uncensored observations. The resulting log-likelihood ratio process was well behaved, with $l(t) < 3$ except for spurious fluctuations near the upper end point, indicating the assumption of exponential censoring times is reasonable. Of the 28 deaths recorded before $\hat{\tau}$, 17 were attributed to rejection and 2 were unspecified, which agrees closely with the overall proportions. If we consider a 'death' to be rejection of the heart and other patients to be censored, the pictures are similar but the maximum is reduced to 15.42.

We show approximations to the significance levels for various c and using various methods in Table 5.1. The observed $c = 7.4$ is highly significant regardless of which approximation is used. The exact method uses the recursive method of Section 2.4 and is a $O(n^2)$ computation; the amount of computation for other methods is independent of n. The Gaussian method is obtained by numerically solving an inte-

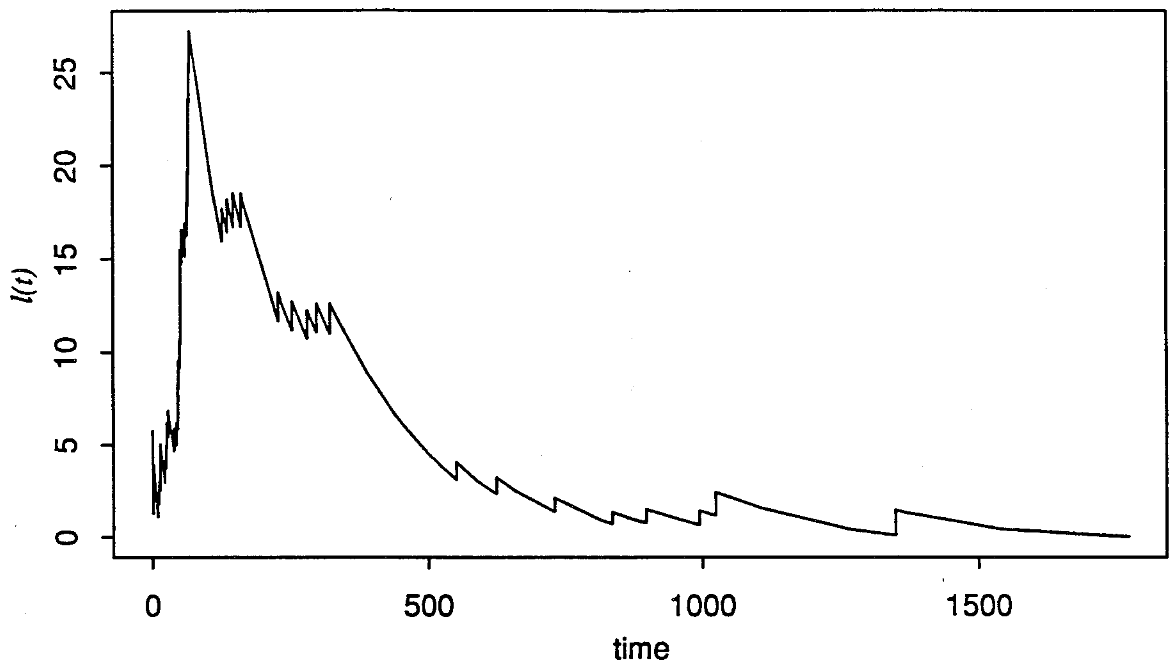


Figure 5.1: Heart Transplant Data: $l(t)$ on the Original Time Scale

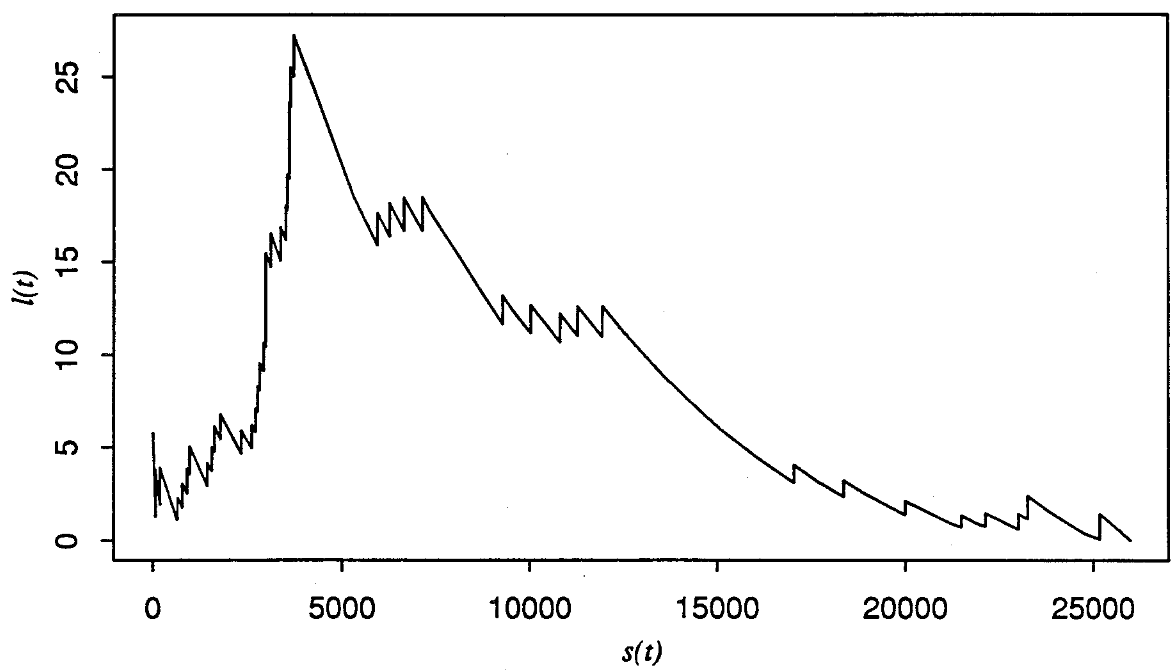


Figure 5.2: Heart Transplant Data: $l_Y(s)$ on the Transformed Time Scale

Method		c			
		2.5	3.0	3.5	4.0
Exact	a)	0.1707	0.0558	0.0130	0.00215
	b)	0.1763	0.0608	0.0139	0.00223
	c)	0.1726	0.0575	0.0133	0.00218
1st Order	a)	0.1898	0.0660	0.0152	0.00253
	b)	0.2021	0.0766	0.0175	0.00278
	c)	0.1941	0.0697	0.0160	0.00262
2nd Order	a)	0.1651	0.0557	0.0129	0.00211
	b)	0.1757	0.0623	0.0142	0.00227
	c)	0.1688	0.0580	0.0134	0.00217
Gaussian	c)	0.1896	0.0578	0.0134	0.00236

Table 5.1: Heart transplant data: Significance level approximations for the likelihood ratio test, with 45 uncensored observations, 24 censored observations and $u_0 = 1 - u_1 = 0.1$, a) When the largest observation is uncensored; b) Largest observation censored and c) Weighted combination. The Gaussian approximation does not distinguish between the three cases.

gral equation; see Appendix C. The Gaussian approximation performs surprisingly well, especially after taking the weighted average. Our large deviation approximations perform well for small tail probabilities. The second order approximation has excellent performance in all cases.

The Gaussian approximations must be used with some caution, since it is not method specific. For the lymphoma dataset considered by Matthews, Farewell and Pyke (1985), we obtain a maximum likelihood of 4.213, corresponding to $c = 2.90$, and with $u_0 = 1 - u_1 = 0.1$, the attained significance level is about 0.07. Matthews, Farewell and Pyke use a score statistic which has the same limiting Gaussian process but obtain $c = 6.35$, giving a very small attained significance level and vastly different conclusions if the Gaussian approximation is used. By contrast, our large deviation approximations are specific to the likelihood ratio statistic and this sort of discrepancy will not arise.

We show approximate 95% confidence regions for τ in Figure 5.3 and for (τ, δ) in Figure 5.4. The confidence region for τ is an interval and the joint region for (τ, δ) is a connected set; this is partially luck and partially due to the very peaked likelihood function. The confidence interval for τ is $(66, 82.4)$ and there is very strong evidence of an initial period following the transplant when there is a substantial risk of rejection or other side effects.

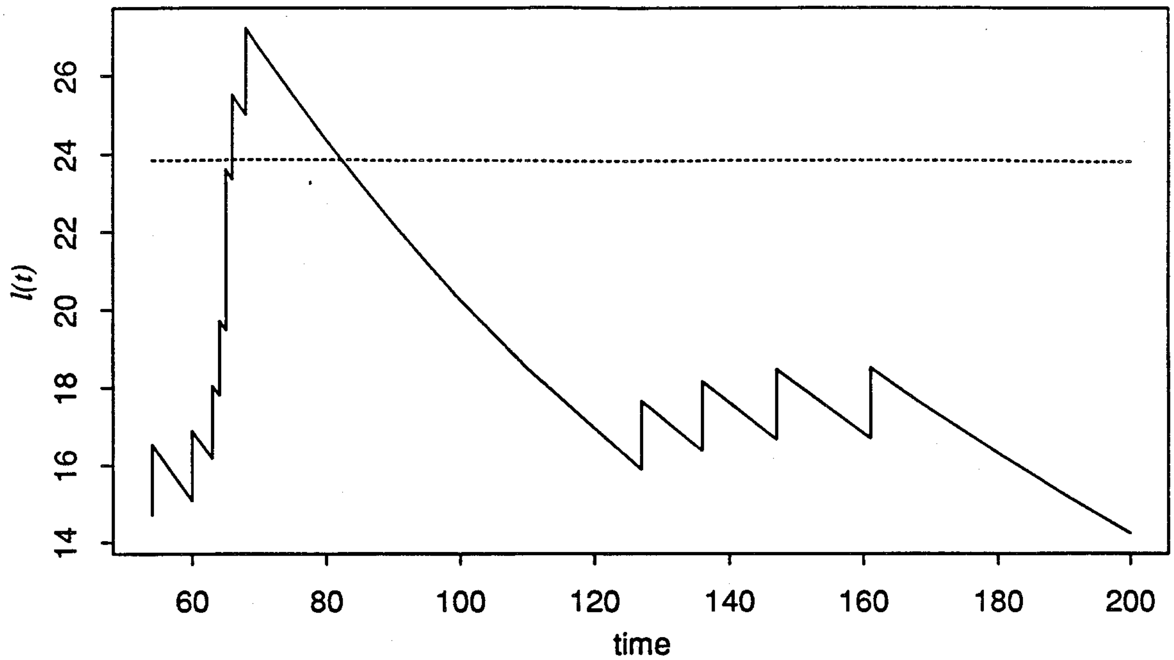


Figure 5.3: Heart Transplant Data: 95% Confidence Region for τ

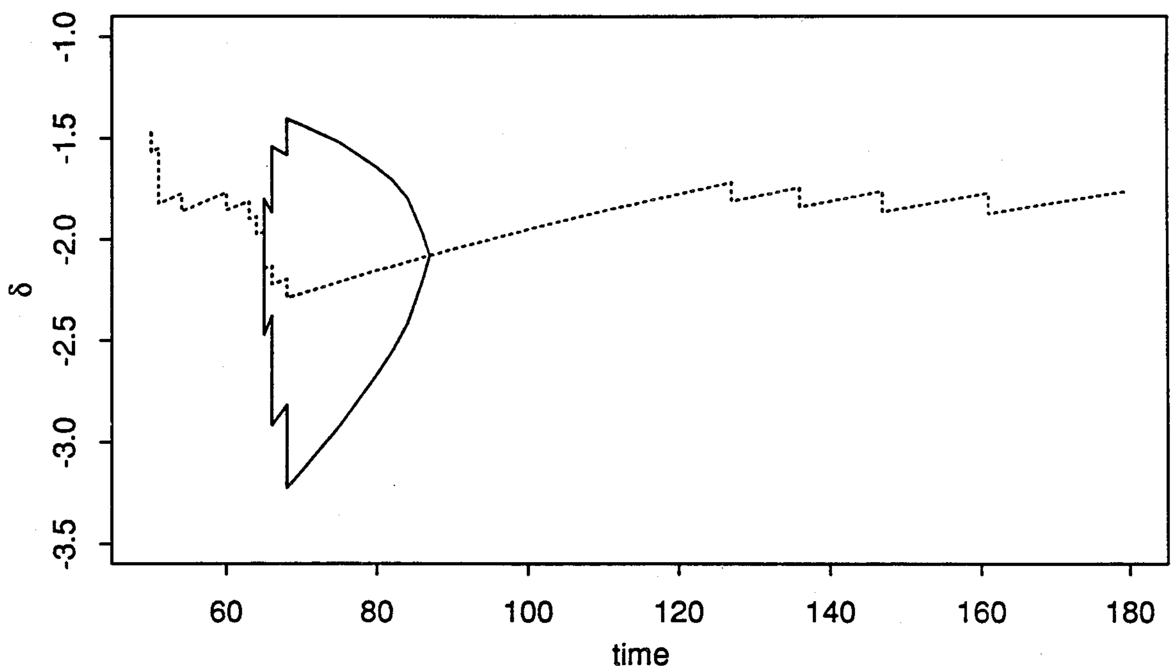


Figure 5.4: Heart Transplant Data: 95% Joint Confidence Region for (τ, δ)

Chapter 6

Kolmogorov-Smirnov Testing

The Kolmogorov-Smirnov test is designed to test the null hypothesis that independent observations come from a pre-specified distribution. The significance level is defined as a boundary crossing probability for the empirical process, for which we can apply the methods developed in Chapter 2. When the null distribution is completely specified, existing methods give comparable results and are less computational. However, the advantage of our method over existing methods occurs when the null hypothesis has unknown nuisance parameters which must be estimated from the data. It is well known that distributions for the point null hypothesis case are not valid in this case, even asymptotically. However, this case has received very little theoretical attention and our methods give better results than existing approximations.

6.1 The Known Parameter Case

Suppose $X_1, \ldots, X_n$ are *i.i.d.* from a continuous distribution F. We wish to test $\mathcal{H}_0 : F = F_0$ vs $\mathcal{H}_1 : F \neq F_0$. The Kolmogorov-Smirnov statistic is

$$D_n = \sup_x |\hat{F}(x) - F_0(x)| \tag{6.1}$$

where

$$\hat{F}(x) = \sum_{i=1}^{n} I(X_i \leq x)$$

is the empirical distribution function. By applying the probability integral transformation $X_i^* = F_0(X_i)$, we see the distribution of D_n under $\mathcal{H}_0$ does not depend on F_0, so we can without loss of generality assume $F_0(x) = x$ for $0 < x < 1$.

The Kolmogorov-Smirnov test rejects $\mathcal{H}_0$ if $D_n > \eta$ for some $\eta > 0$. The significance level is

$$\alpha = P_0\left(\sup_{0 \leq x \leq 1} |\hat{F}(x) - x| \geq \eta\right). \tag{6.2}$$

Since uniform order statistics are equal in law to the event times of a Poisson process $X(t)$ conditional on $X(1) = n$, we get

$$\alpha = P_0^{(n)}\left(\sup_{0 \leq t \leq 1} |X(t) - nt| \geq n\eta\right). \tag{6.3}$$

We can apply (2.24) directly to (6.3) with $a(t) = t - \eta$, giving

$$\alpha = 2\int_\eta^1 \frac{\eta}{t}\sqrt{\frac{n}{2\pi(t-\eta)(1-t+\eta)}}e^{-nl(t)}(1 + o(1)) \tag{6.4}$$

as $n \to \infty$, where

$$l(t) = (t - \eta)\log\left(1 - \frac{\eta}{t}\right) + (1 - t + \eta)\log\left(1 + \frac{\eta}{1-t}\right).$$

The leading factor of 2 in (6.4) arises from the upper boundary which is treated by time reversal. An application of Lemma A.1 shows the tangent approximation

(2.22) is exact in this case and therefore we would expect use of (2.24) to be fairly accurate.

The approximation (6.4) can be improved by using the second term of Stirling's formula when approximating the binomial coefficients, and some minor modifications at the lower limit of integration. We replace the lower limit by $\eta + \frac{1}{2n}$ and add $(1-\eta)^n$, which is the probability of hitting the boundary at η. These modifications make (6.4) a second order approximation; see Appendix A for more details.

Several other methods for approximating and evaluating (6.2) are available. These include combinatorial recursions and Gaussian approximations; see Durbin (1973a) for details. Note that for the one sided case, we can obtain exact $O(n)$ methods; for example, (2.23). A corrected diffusion approximation

$$P_0\left(\sup_{0\leq x\leq 1} |\hat{F}(x) - x| \geq \frac{c}{\sqrt{n}}\right) \approx 2e^{-2c(c+\frac{1}{3\sqrt{n}})}$$

is given by Smirnov (1944). The large deviation approximation given by Siegmund (1982) can be obtained by using a Laplace approximation to (6.4).

Lemma 6.1 (Laplace Approximation) *Suppose $f : [a,b] \to \mathcal{R}$ is continuous and twice continuously differentiable, and has a unique minimum at some point x^* with $a < x^* < b$. Further, suppose $\int_a^b |g(x)|dx < \infty$ and g is continuous at x^*. Then*

$$\int_a^b n^{1/2} g(x) e^{-nf(x)} dx = g(x^*)\sqrt{\frac{2\pi}{f''(x^*)}} e^{-nf(x^*)}(1 + o(1)) \quad \text{as } n \to \infty.$$

Proof: Follows by expanding f in a Taylor series around $x = x^*$.

□

We illustrate the accuracy of the various methods in Table 6.1. Here, $c = \eta\sqrt{n}$. We note the large deviation approximation is performing very well. The relative error is about 3% when $n = 20$ and much less for larger n. The second order approximation is inferior to the first order approximation for $n = 20$ but very good

		c				
n	Method	1.0	1.25	1.5	1.75	2.0
	Gaussian	0.2707	0.08787	0.02222	0.00437	0.000671
20	Exact	0.2326	0.07121	0.01651	0.00286	0.000363
	Large Dev'n	0.2376	0.07285	0.01693	0.00294	0.000374
	2nd Order	0.2196	0.06819	0.01590	0.00275	0.000347
	Smirnov	0.2332	0.07293	0.01777	0.00337	0.000498
50	Exact	0.2458	0.07742	0.01877	0.00349	0.000495
	Large Dev'n	0.2486	0.07808	0.01892	0.00352	0.000499
	2nd Order	0.2451	0.07728	0.01875	0.00349	0.000495
	Smirnov	0.2463	0.07810	0.01929	0.00371	0.000556
100	Exact	0.2527	0.08050	0.01984	0.00378	0.000555
	Large Dev'n	0.2542	0.08080	0.01991	0.00379	0.000557
	2nd Order	0.2531	0.08050	0.01984	0.00378	0.000555
	Smirnov	0.2532	0.08085	0.02010	0.00389	0.000587
200	Exact	0.2577	0.08267	0.02057	0.00397	0.000594
	Large Dev'n	0.2587	0.08281	0.02060	0.00398	0.000595
	2nd Order	0.2582	0.08267	0.02057	0.00397	0.000594
	Smirnov	0.2582	0.08285	0.02070	0.00403	0.000610

Table 6.1: Approximations to Significance Level of Kolmogorov-Smirnov Test

for $n = 100$ and $n = 200$. The error in the fourth decimal place when $c = 1.0$ in these cases can be attributed to the probability of crossing both boundaries. The Gaussian approximation performs poorly, with relative errors between 10 and 20%. Smirnov's corrected diffusion approximation performs slightly better than the large deviation approximation for small c and slightly worse for large c.

6.2 The Unknown Parameter Case

Let $\mathcal{F} = \{F(x,\theta), \theta \in \Theta\}$ be a one parameter exponential family, with densities given by

$$f(x,\theta) = \exp(\theta x - \psi(\theta))f(x)$$

for some non-negative function $f(x)$ and $\Theta = \{\theta : \int e^{\theta x} f(x)dx < \infty\}$. Our null hypothesis is now $\mathcal{H}_0 : F \in \mathcal{F}$. Since we have an unknown nuisance parameter, we can no longer use (6.1). It is natural to use an estimate of θ. We define

$$D_n = \sup_x |\hat{F}(x) - F(x, \hat{\theta})|$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ,

$$\sum_{i=1}^{n} X_i = n\psi'(\hat{\theta}).$$

In a few special cases, for example if θ is a location or scale parameter, invariance arguments can be used to show the distribution of D_n does not depend on θ. However, this is not true in general, and we must condition on $\sum_{i=1}^{n} X_i$. By sufficiency, the distribution of D_n is then independent of θ. We will reject $\mathcal{H}_0$ if $D_n \geq \eta$. The significance level of the test is then

$$\alpha = P_0(D_n \geq \eta | \sum_{i=1}^{n} X_i = ny) \tag{6.5}$$

where under P_0, $X_1, \dots, X_n$ are *i.i.d.* from $F(x, \theta)$ for some θ.

Durbin (1973b) shows the process

$$Y(t) = \sqrt{n}(\hat{F}(t) - F(t, \hat{\theta}))$$

converges weakly to a Gaussian process with mean 0 and covariance

$$\sigma(s,t) = F(s,\theta)(1 - F(t,\theta)) - \frac{g(s,\theta)g(t,\theta)}{I(\theta)}, s \leq t \tag{6.6}$$

where $g(t,\theta) = \frac{\partial}{\partial\theta}F(t,\theta)$ and $I(\theta) = -E(\frac{\partial^2}{\partial\theta^2}\log(f(X_1,\theta)))$. Although Durbin works with the unconditional distribution of $Y(t)$, the result also holds conditionally if y is fixed and $\theta = \hat{\theta}$ in (6.6). We can now use (2.3) to approximate $P(\sup_t Y(t) > c)$.

We can also use our large deviation methods to approximate the distribution of D_n. We let $X(t) = n\hat{F}(t)$. Then $X(t)$ can be approximated locally as a Poisson process and we can apply (2.26) to approximate the probability that $X(t)$ crosses the boundary $n(F(t,\hat{\theta}) - \eta)$, conditional on $\sum_{i=1}^{n} X_i = ny$. To achieve this, we need a large deviation approximation to the conditional distribution of $X(t)$ and an estimate of

$$n\mu(t, p_t) = \lim_{s \to t-} \frac{1}{t-s} E\left(X(t) - X(s) | X(t) = np_t, \sum_{i=1}^{n} X_i = ny\right)$$

where $p_t = F(t,\hat{\theta}) - \eta$.

These approximations can be obtained by embedding $\mathcal{F}$ in a two parameter exponential family $\mathcal{F}_t$ with densities given by

$$f_t(x,\theta,\delta) = \exp\left(\theta x + \delta I(x > t) - \psi_t(\theta,\delta)\right) f(x) \tag{6.7}$$

where

$$\psi_t(\theta,\delta) = \psi(\theta) + \log(F(t,\theta) + e^{\delta}(1 - F(t,\theta))).$$

The precise details of the approximations are now similar to those given in Section 3.2 and Appendix B for the log-linear model.

Let $\hat{\theta}_1$ and $\hat{\delta}_1$ be the maximum likelihood estimates of θ and δ when $X(t) = n(F(t,\hat{\theta}) - \eta)$. That is, $\hat{\theta}_1$ and $\hat{\delta}_1$ are solutions of

$$\begin{aligned} y &= \psi'(\theta) + \frac{(1-e^{\delta})\frac{\partial}{\partial\theta}F(t,\theta)}{F(t,\theta)+e^{\delta}(1-F(t,\theta))} && (6.8) \\ 1-p_t &= \frac{e^{\delta}(1-F(t,\theta))}{F(t,\theta)+e^{\delta}(1-F(t,\theta))}. && (6.9) \end{aligned}$$

We can then apply our approximation formula (2.26) for locally Poisson processes to get

$$\begin{aligned} & P\left(\inf_t (X(t) - np_t) \leq 0\right) \\ \approx & \int_{t_0}^{\infty} n\left(f(t,\hat{\theta}) - \mu(t,p_t)\right) \frac{1}{\sqrt{2\pi n}} \sqrt{\frac{\psi''(\hat{\theta})}{|\psi_t''(\hat{\theta}_1,\hat{\delta}_1)|}} e^{-l(t)} dt \qquad (6.10) \end{aligned}$$

where t_0 is the solution of $F(t,\hat{\theta}) = \eta$,

$$\psi_t''(\theta,\delta) = \begin{pmatrix} \frac{\partial y}{\partial\theta} & \frac{\partial y}{\partial\delta} \\ -\frac{\partial p_t}{\partial\theta} & -\frac{\partial p_t}{\partial\theta} \end{pmatrix}$$

and

$$\begin{aligned} l(t) = & (\hat{\theta}_1 - \hat{\theta})ny + \hat{\delta}_1 n(1-p_t) \\ & + n\left(\psi(\hat{\theta}) - \psi(\hat{\theta}_1) - \log(F(t,\hat{\theta}_1) + e^{\hat{\delta}_1}(1-F(t,\hat{\theta}_1)))\right) \end{aligned}$$

is the log-likelihood ratio for testing $\mathcal{H}_0 : \delta = 0$ when $X(t) = np_t$. By consideration of the density $f_t(x,\hat{\theta}_1,\hat{\delta}_1)$ and arguing informally as (3.14) or rigorously as in Appendix B, we obtain the local rate

$$\begin{aligned} \mu(t,p_t) &= f_t(t^-,\hat{\theta}_1,\hat{\delta}_1) \\ &= \frac{p_t e^{\hat{\theta}_1 t} f(t)}{\int_{-\infty}^{t} e^{\hat{\theta}_1 x} f(x)dx}. \end{aligned}$$

Finally, we can approximate the probability of crossing the upper boundary by time reversal.

Note that (6.10) is suitable for a Laplace approximation. If we choose t^* to minimize $l(t)$, the Laplace approximation is

$$\frac{e^{-l(t^*)}}{\sqrt{l''(t^*)/n}}\left\{f(t^*,\hat{\theta})-\frac{p_t^* e^{\hat{\theta}_1 t^*} f(t^*)}{\int_{-\infty}^{t^*} e^{\hat{\theta}_1 x} f(x)dx}\right\}\sqrt{\frac{\psi''(\hat{\theta})}{\psi''_{t^*}(\hat{\theta}_1,\hat{\delta}_1)}}. \tag{6.11}$$

Usually there is no closed form solution for t^* and so it must be found numerically. Also, $l(t)$ may have more than one local minimum so (6.11) should be used with caution.

We may consider replacing $F(t,\hat{\theta})$ by $P(X_1 < t|\sum_{i=1}^n X_i)$ in the definition of D_n. From the central limit viewpoint, both forms will have the same asymptotic Gaussian distribution. The large deviation method will give slightly different results. There is no strong theoretical preference for either form, although $F(t,\hat{\theta})$ is often easier to compute.

The method developed here can be extended to cases with more than one nuisance parameter; for example a normal distribution with unknown mean and variance. In this case, (6.8) will become a system of equations, one for each component of θ. Also, $\psi''(\theta)$ will be a matrix and so we replace $\psi''(\hat{\theta})$ by $|\psi''(\theta)|$ in (6.10). The main difference is that the large deviation approximation to the distribution of $X(t)$ will be slightly more complicated.

6.3 Examples: Exponential and Normal Distributions

The methods of the previous section will be applied to the exponential and normal distributions to compare the Gaussian and large deviation approximations to the significance level of the Kolmogorov-Smirnov test.

6.3.1 Exponential Distribution

For the exponential distribution, we have

$$\begin{aligned} f(x,\theta) &= \exp(\theta x + \log(-\theta))I(x \geq 0) \\ &= -\theta e^{\theta x} I(x \geq 0) \\ F(x,\theta) &= (1 - e^{\theta x})I(x \geq 0) \end{aligned}$$

where $\theta < 0$. If we let $y = \frac{1}{n}\sum_{i=1}^n X_i$ we have the simple expressions

$$\begin{aligned} \hat{\theta} &= -\frac{1}{y} \\ E\left\{X(t)\middle|\sum_{i=1}^n X_i = ny\right\} &= n\left\{1 - \left(1 - \frac{t}{ny}\right)^{n-1}\right\} \\ F(t,\hat{\theta}) &= 1 - \exp(-\frac{t}{y}). \end{aligned}$$

The covariance function (6.6) is given by

$$\sigma(s,t) = (1 - e^{-s/y})e^{-t/y} - \frac{ste^{-(s+t)/y}}{y^2}$$

which enables us to compute the Gaussian approximation to the significance level using (2.3) and numerical integration.

The embedded exponential family (6.7) has the form

$$f_t(x,\theta,\delta) = \frac{-\theta e^{\theta x + \delta I(x>t)}}{1 - e^{\theta t} + e^{\delta + \theta t}}$$

and

$$\psi_t(\theta,\delta) = -\log(-\theta) + \log(1 - e^{\theta t} + e^{\delta + \theta t}). \tag{6.12}$$

In this case (6.8) and (6.9) can be written as

$$y = -\frac{1}{\hat{\theta}_1} - \frac{(1 - e^{\hat{\delta}_1})te^{\hat{\theta}_1 t}}{1 - (1 - e^{\hat{\delta}_1})e^{\hat{\theta}_1 t}} \tag{6.13}$$

$$1 - p_t = \frac{e^{\hat{\delta}_1}e^{\hat{\theta}_1 t}}{1 - (1 - e^{\hat{\delta}_1})e^{\hat{\theta}_1 t}} \tag{6.14}$$

which can be simplified to

$$e^{\hat{\delta}_1} = \frac{(1-p_t)(1-e^{\hat{\theta}_1 t})}{p_t e^{\hat{\theta}_1 t}}$$

and hence

$$\frac{y}{t} = -\frac{1}{\hat{\theta}_1 t} - \frac{e^{\hat{\theta}_1 t} - 1 + p_t}{1 - e^{\hat{\theta}_1 t}} \tag{6.15}$$

There is no closed form solution for $\hat{\theta}_1$ so (6.15) must be solved numerically.

We find ψ_t'' by differentiating (6.13) and (6.14). This gives

$$\begin{aligned}
\frac{\partial y}{\partial \hat{\theta}_1} &= \frac{1}{\hat{\theta}_1^2} - \frac{t^2(1-e^{\hat{\delta}_1})e^{\hat{\theta}_1 t}}{(1-(1-e^{\hat{\delta}_1})e^{\hat{\theta}_1 t})^2} \\
\frac{\partial y}{\partial \hat{\delta}_1} = -\frac{\partial p_t}{\partial \hat{\theta}_1} &= \frac{t e^{\hat{\delta}_1 + 2\hat{\theta}_1 t}}{(1-(1-e^{\hat{\delta}_1})e^{\hat{\theta}_1 t})^2} \\
-\frac{\partial p_t}{\partial \hat{\delta}_1} &= \frac{e^{\hat{\delta}_1} e^{\hat{\theta}_1 t}(1-e^{\hat{\theta}_1 t})}{(1-(1-e^{\hat{\delta}_1})e^{\hat{\theta}_1 t})^2} \\
&= p_t(1-p_t).
\end{aligned}$$

A straightforward calculation shows

$$\psi''(\hat{\theta}) = \frac{1}{\hat{\theta}^2}.$$

When evaluating the integral (6.10) over the lower boundary, we get $t_0 = -y\log(1-\eta)$. Also, if $X(t) = np_t$, then $\sum_{i=1}^n X_i \geq ntp_t$ and hence we require $p_t \leq \frac{y}{t}$. Since $\frac{y}{t} \to 0$ and $p_t \to 1-\eta$ as $t \to \infty$, this gives a finite upper limit of integration. We have

$$f(t,\hat{\theta}) - \frac{p_t e^{\hat{\theta}_1 t}}{\int_0^t e^{\hat{\theta}_1 x} dx} = \frac{1}{y} e^{-t/y} + \frac{p_t \hat{\theta}_1 e^{\hat{\theta}_1 t}}{1 - e^{\hat{\theta}_1 t}}.$$

We now have all the components needed to evaluate the integrand of (6.10). It is straightforward to show the integral is independent of y so we can fix y to be an arbitrary (positive) number. We evaluate the probability of crossing the upper boundary $q_t = F(t,\hat{\theta}) + \eta$ by time reversal. The limits of integration are 0 and $y\log(\eta)$, and the derivative-local drift term is

$$f(t,\hat{\theta}) - \frac{(1-q_t)e^{\hat{\theta}_1 t}}{\int_t^\infty e^{\hat{\theta}_1 x}dx} = \frac{1}{y}e^{-t/y} - (1-q_t)\hat{\theta}_1.$$

We are unaware of any computationally feasible method for evaluating (6.5) exactly, or of any corrected diffusion approximations. We use simulations of size 10000 to check the performance of the Gaussian and Large Deviation approximations.

n	Method	c				
		0.6	0.8	1.0	1.2	1.4
	Gaussian	0.9511	0.3478	0.1004	0.0225	0.0039
20	Simulation	0.6574	0.2607	0.0719	0.0149	0.0019
	Large Dev'n	0.7829	0.2768	0.0742	0.0146	0.0021
50	Simulation	0.6844	0.2814	0.0815	0.0182	0.0028
	Large Dev'n	0.8415	0.3005	0.0834	0.0176	0.0028
100	Simulation	0.7059	0.2948	0.0865	0.0184	0.0023
	Large Dev'n	0.8702	0.3129	0.0881	0.0191	0.0032
200	Simulation	0.7059	0.3019	0.0864	0.0210	0.0033
	Large Dev'n	0.8917	0.3223	0.0916	0.0202	0.0034
Simulation s.e.		0.005	0.005	0.003	0.0015	0.0006

Table 6.2: Approximations to Significance Level of Kolmogorov-Smirnov Test for the Exponential Distribution

From Table 6.2 we see the tangent approximation is outperforming the Gaussian approximation, especially in the tail of the distribution. Neither approximation performs particularly well for small c, partly because there is a large probability of crossing both boundaries, which has not been allowed for. Using more accurate approximations for the Gaussian process boundary crossing probability does not substantially improve the approximation to the probability of interest for the Kolmogorov-Smirnov test. Durbin (1985) uses more accurate Gaussian approximations to obtain numerical results for the one-sided case. When the true asymptotic probability is 0.05, he obtains the first order approximation 0.0507. This will correspond to a value of c close to 1.0. We see from the appropriate column of Table 6.2 that the error of the Gaussian approximation is much more than the improvement obtained by using more accurate methods to evaluate the approximation.

We conclude by mentioning an alternative method of using the Kolmogorov-Smirnov test to test for exponentiality. By the strong law of large numbers,

$$\frac{\sum_{i=1}^{n}(X_i \wedge t)}{\sum_{i=1}^{n} X_i}$$

is a consistent estimator of $F(t;\theta)$. If we define a time transformation $s(t) = \sum_{i=1}^{n}(X_i \wedge t)$ as in Theorem 5.1 and $Y(s) = X(t(s))$, then $Y(s)$ is a Poisson process observed up to the time of the nth event. Therefore, $S_1/S_n, \ldots, S_{n-1}/S_n$ is distributed as the order statistics of a sample of size $n-1$ from a $\mathcal{U}[0,1]$ distribution. We can then use the Kolmogorov-Smirnov for the transformed sample,

$$\sup_{0<z<1} |Y(S_n z) - z| = \sup_{t>0} \left| X(t) - \frac{\sum_{i=1}^{n}(X_i \wedge t)}{\sum_{i=1}^{n} X_i} \right|$$

and use the standard percentage points for a sample of size $n-1$.

6.3.2 Normal Distribution

The evaluation of (6.10) for the normal distribution with unknown mean and variance is tedious and we omit most of the details. The density is parametrized as

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \exp\left\{\alpha y + \beta x + \frac{\beta^2}{4\alpha} - \frac{1}{2}\log(-\frac{\pi}{\alpha})\right\}$$

where $y = x^2$, $\alpha = -1/(2\sigma^2) < 0$ and $\beta = \mu/\sigma^2$. We then have a pair of equations from (6.8) to define the maximum likelihood estimates of α and β under the embedded change point family. Finally, ψ_t'' will be a 3×3 symmetric matrix.

We note that both limits of integration are finite. When $\sum_{1=1}^{n} X_i = 0$ and $\sum_{i=1}^{n} X_i^2 = n$, the lower limit will be the solution of $p_t = \Phi(t) - \eta = 0$. The upper limit is derived by noting if $t > 0$ and $X(t) = np_t$, then $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \geq t^2(1-p_t)/p_t$ and hence we require $t^2(1-p_t) \leq p_t$. Also note by symmetry the integral over the upper boundary will be the same as that over the lower boundary.

The Gaussian approximation is fairly straightforward. We can work in the original (μ, σ) parametrization. The covariance function is a generalization of (6.6); see Durbin (1973b). When $\mu = 0$ and $\sigma^2 = 1$, we get

$$\sigma(s,t) = \Phi(s)(1-\Phi(t)) - (1+\frac{st}{2})\phi(s)\phi(t)$$

for $s < t$, and we can then apply Durbin's approximation (2.3).

We give the two-boundary crossing approximations for some n and c in Table 6.3 and compare with the results of simulations of size 10000. The large deviation approximation tends to overestimate the true probability for small c but is performing well for the larger values of c which are of interest for significance testing. The Gaussian approximation is substantially overestimating the true probability. More extensive simulations may be found in Dallal and Wilkinson (1986).

n	Method	c			
		0.6	0.8	1.0	1.2
	Gaussian	0.7313	0.1495	0.0201	0.0018
20	Simulation	0.4476	0.0893	0.0096	0.0010
	Large Dev'n	0.5338	0.0976	0.0106	0.0006
50	Simulation	0.4887	0.1036	0.0133	0.0010
	Large Dev'n	0.5954	0.1149	0.0140	0.0011
100	Simulation	0.5058	0.1151	0.0154	0.0009
	Large Dev'n	0.6302	0.1243	0.0158	0.0013
200	Simulation	0.5078	0.1202	0.0167	0.0020
	Large Dev'n	0.6566	0.1313	0.0171	0.0014
Simulation s.e.		0.005	0.003	0.001	0.0004

Table 6.3: Approximations to Significance Level of Kolmogorov-Smirnov Test for the Normal Distribution

Chapter 7

Scan Statistics

In earlier chapters, we discussed detection of a single change point for Poisson processes defined on the line. In this chapter, we generalize this in two directions. First, we allow for the possibility of a second change point at which time the process reverts to its original rate. This leads to questions about the distribution of scan statistics, which we will define below. We can then formulate similar problems for multidimensional Poisson processes; we will discuss only the two dimensional case in detail.

Although we discuss alternatives where a region of higher event rate has a specific shape, the results may be useful more generally in detecting non-homogeneity. For 2 dimensional Poisson processes, we discuss detecting a rectangular region with a high event rate. In a clustering model we are unlikely to believe in advance that clusters form rectangles. However, many regions can be approximated as rectangles and our methods may still be useful for detecting clustering and non-homogeneity.

The problems discussed lead to questions about the distribution of the maxima of random fields. For problems involving the maxima of one dimensional stochastic processes, we exploited the relation between the maximum and first passage times. However, with multidimensional time, there is no natural definition of first passage times. Our technique is to place an artificial ordering on the indexing set, which allows us to define first passage times. We then exploit a special structure of the random fields we consider: that locally they behave like the superposition of independent one dimensional processes. This approach has been used before in random field problems, see for example Hogan and Siegmund (1986).

7.1 Scan Statistics on the Circle and Line

Suppose we observe a Poisson process on the unit circle, and let $X(\theta_1, \theta_2)$ be the number of events in $(\theta_1, \theta_2]$ for $0 \le \theta_1 < 2\pi$ and $0 < \theta_2 - \theta_1 \le 2\pi$. It is convenient to index the circle by $[0, 4\pi)$ with the convention $\theta = \theta - 2\pi$ if $\theta \ge 2\pi$. We model the rate as

$$\lambda(\theta) = \begin{cases} \lambda_1 & \theta_0 \le \theta < \theta_0 + \Delta \\ \lambda_0 & \text{otherwise} \end{cases}$$

where λ_0, λ_1 and θ_0 are unknown. We treat both the cases Δ known and unknown. Note that it suffices to consider one sided alternatives $\mathcal{H}_1 : \lambda_1 > \lambda_0$.

This problem is related to the square wave epidemic alternative, discussed by Levin and Kline (1985) and Siegmund (1988b), which considers two change points, with the rate returning to the original level at the second change point. Considering a two-sided alternative for this problem is exactly equivalent to a one-sided alternative for our problem.

We can also interpret our problem with respect to two dimensional problems. Suppose we observe a Poisson process on the unit disc in $\mathcal{R}^2$, and consider the projection of the events onto the circumference. This converts wedge shaped alternatives into our square wave alternative. A slightly more natural alternative in this setting may be to split the disc by an arbitrary straight line. However, we do not have any method that gives results for this problem.

The following lemma characterizes the local behaviour of $X(\theta_1, \theta_2)$ and will be useful for our following discussions.

Lemma 7.1 *Suppose for some θ, $X(\theta, \theta + \Delta) = m$. Define*

$$\begin{aligned} Z_1(t) &= X(\theta_1 - t, \theta_2) - m \\ Z_2(t) &= m - X(\theta_1 + t, \theta_2) \\ Z_3(t) &= m - X(\theta_1, \theta_2 - t) \\ Z_4(t) &= X(\theta_1, \theta_2 + t) - m. \end{aligned}$$

Let $n = X(0, 2\pi) \to \infty$ with $\frac{m}{n}$ fixed. Then for any $t > 0$ the process

$$\left\{Z_1(\frac{t_1}{n}), Z_2(\frac{t_2}{n}), Z_3(\frac{t_3}{n}), Z_4(\frac{t_4}{n}); 0 \le t_i \le T\right\}$$

converges in law to a vector of four independent Poisson processes on $[0, T]$, with rates $\lambda_1, \ldots, \lambda_4$ respectively, where

$$\lambda_1 = \lambda_4 = \frac{1 - \frac{m}{n}}{2\pi - \Delta}$$

and

$$\lambda_2 = \lambda_3 = \frac{m}{n\Delta}.$$

Proof: Suppose n is large enough so that $\theta_2 - \theta_1 \ge \frac{2T}{n}$ and $2\pi - (\theta_2 - \theta_1) \ge \frac{2T}{n}$. We can then piece together Z_1 and Z_4 to get a new process Z^*:

$$Z^*(t) = \begin{cases} Z_1(\frac{t}{n}) & 0 \le t \le T \\ Z_1(\frac{T}{n}) + Z_4(\frac{t-T}{n}) & T < t \le 2T \end{cases}.$$

Applying standard Poisson limit theorems to Z^* shows $\{Z_1(\frac{t_1}{n}), Z_4(\frac{t_4}{n})\}$ converge to a pair of independent Poisson processes. The convergence of $\{Z_2(\frac{t_2}{n}), Z_3(\frac{t_3}{n})\}$ is similar. Independence follows since we are conditioning on $X(\theta_1, \theta_2) = m$ and $\{Z_2, Z_3\}$ and $\{Z_1, Z_4\}$ depend only on events inside and outside $[\theta_1, \theta_2]$ respectively. We can determine the rates by considering expected values. For example,

$$\lambda_1 = E\left(Z_1(\frac{1}{n})|X(\theta, \theta + \Delta) = m\right) = \frac{1}{n}\frac{n-m}{2\pi - \Delta}$$

since $Z_1(\frac{1}{n})$ represents the number of points in an arc of length $\frac{1}{n}$, out of $n - m$ points in an arc of length $2\pi - \Delta$.

□

7.1.1 Known Δ

We first consider the simple but perhaps somewhat artificial case where Δ is known. Let $Y(\theta) = X(\theta, \theta + \Delta)$. We will reject $\mathcal{H}_0 : \lambda_0 = \lambda_1$ if

$$\sup_{0 \leq \theta < 2\pi} Y(\theta) \geq \frac{n\Delta}{2\pi}(1 + \epsilon) \tag{7.1}$$

for some $\epsilon > 0$ such that $\frac{n\Delta}{2\pi}(1 + \epsilon)$ is an integer, where $n = X(\theta_1, \theta_2)$ is the total number of events. The significance level of (7.1) is evaluated conditionally on n to remove dependency on the unknown λ_0 and λ_1. Since $Y(\theta)$ is stationary, it is easy to show the likelihood ratio test is equivalent to (7.1).

The statistic $\sup_{0 \leq \theta < 2\pi} Y(\theta)$ is known as the **circular scan statistic**. This and analagous statistics on the unit interval have been studied by several authors. Exact expressions for the distribution are discussed by Naus (1965) and Huntington and Naus (1975). Cressie (1980) obtains the asymptotic distribution of a standardized version of the scan statistic, which is related to the maximum of a Gaussian process with a triangular covariance function, discussed by Shepp (1971). Theorem 7.1 below gives a large deviation approximation to the significance level of (7.1), which is very accurate and much easier to evaluate than the exact expressions.

Theorem 7.1 *Suppose $\frac{\Delta}{2\pi}(1+\epsilon)$ is rational and $n \to \infty$ with $\frac{n\Delta}{2\pi}(1+\epsilon)$ integral. Then*

$$\begin{aligned} & P_0^{(n)}\left\{\sup_{0\le\theta<2\pi} Y(\theta) \ge \frac{n\Delta}{2\pi}(1+\epsilon)\right\} \\ = & \frac{2\pi n\epsilon}{2\pi-\Delta} P_0^{(n)}\left\{Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right\}(1+o(1)). \end{aligned} \tag{7.2}$$

Proof: Fix k large and let $\theta_j = \frac{2\pi j}{k}$ for $j = 0, \ldots, k-1$. Let $\Delta' = \Delta + \frac{2\pi}{k}$ so that for all θ, there exists j such that $(\theta, \theta+\Delta) \subset (\theta_j, \theta_j + \Delta')$. Let $Y'(\theta) = X(\theta, \theta+\Delta')$. Then

$$\begin{aligned} P_0^{(n)}\left(\sup_{0\le j<k} Y(\theta_j) \ge \frac{n\Delta}{2\pi}\right) & \le P_0^{(n)}\left(\sup_{0\le\theta\le 2\pi} Y(\theta) \ge \frac{n\Delta}{2\pi}\right) \\ & \le P_0^{(n)}\left(\sup_{0\le j<k} Y'(\theta_j) \ge \frac{n\Delta}{2\pi}\right). \end{aligned} \tag{7.3}$$

Let

$$\mathcal{T}_k = \inf\left\{\theta_j : Y(\theta_j) \ge \frac{n\Delta}{2\pi}(1+\epsilon)\right\}$$

with $\inf(\emptyset) = \infty$. We then have

$$\begin{aligned} & P_0^{(n)}\left(\sup_{0\le j<k} Y(\theta_j) \ge \frac{n\Delta}{2\pi}\right) \\ = & P_0^{(n)}(\mathcal{T}_k < \infty) \\ = & \sum_{j=0}^{k-1} P_0^{(n)}(\mathcal{T}_k = \theta_j) \\ \ge & \sum_{j=0}^{k-1} P_0^{(n)}\left(\mathcal{T}_k = \theta_j | Y(\theta_j) = \frac{n\Delta}{2\pi}(1+\epsilon)\right) P_0^{(n)}\left(Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right) \end{aligned} \tag{7.4}$$

Suppose now $Y(\theta_j) = \frac{n\Delta}{2\pi}(1+\epsilon)$ for some j. Let $Z(t) = Z_1(t) - Z_3(t)$ where $Z_1(t)$ and $Z_3(t)$ are defined in Lemma 7.1. Then

$$\begin{aligned}
&P_0^{(n)}\left(\mathcal{T}_k = \theta_j | Y(\theta_j) = \frac{n\Delta}{2\pi}(1+\epsilon)\right)\\
&= P_0^{(n)}\left(Z(\frac{2\pi i}{k}) < 0; i = 1,\ldots,j\right)\\
&\geq P_0^{(n)}\left(Z(\frac{2\pi i}{k}) < 0; i = 1,\ldots,J\right) - P_0^{(n)}\left(\cup_{i=J+1}^{j}\{Z(\frac{2\pi i}{k}) \geq 0\}\right). \qquad (7.5)
\end{aligned}$$

Suppose now $n \to \infty$ with $\frac{k}{n}$ fixed. Applying Lemma 7.1, the process $\{Z(\frac{t}{n}); 0 \leq t \leq \frac{2\pi n}{k}J\}$ converges in law to a process $\tilde{Z}(t)$, where $\tilde{Z}(t) = \tilde{Z}_1(t) - \tilde{Z}_3(t)$ and $\tilde{Z}_1(t)$ and $\tilde{Z}_3(t)$ are independent Poisson processes with rates $\lambda_1 = \frac{2\pi - \Delta(1+\epsilon)}{2\pi(2\pi-\Delta)}$ and $\lambda_3 = \frac{(1+\epsilon)}{2\pi}$ respectively. We therefore have

$$\begin{aligned}
&\lim_{n\to\infty} P_0^{(n)}\left(Z(\frac{2\pi i}{k}) < 0; i = 1,\ldots,J\right)\\
&= P\left(\tilde{Z}(\frac{2\pi i n}{k}) < 0; i = 1,\ldots,J\right)\\
&\geq P\left(\tilde{Z}(\frac{2\pi n}{k}) = -1\right) P\left(\tilde{Z}(t) \leq -1 \; \forall\, t > \frac{2\pi n}{k} | \tilde{Z}(\frac{2\pi n}{k}) = -1\right)\\
&\geq \frac{2\pi\lambda_3 n}{k}\exp\left(-2\pi\frac{n}{k}(\lambda_1+\lambda_3)\right)(1-\frac{\lambda_1}{\lambda_3})
\end{aligned}$$

where the last line is obtained using the Poisson distribution and (2.33). We wish to show the second term on the right of (7.5) is asymptotically negligible. This involves dividing the range of i into subintervals where different approximations are used; we omit some of the details. The most important values of i are those for which $\frac{2\pi i}{k}$ is small. Letting J_1 be the largest value of i such that $\frac{2\pi i}{k} < \Delta \wedge 2\pi - \Delta$, we get

$$\begin{aligned}
&P_0^{(n)}\left(\cup_{i=J+1}^{J_1}\{Z(\frac{2\pi i}{k}) \geq 0\}\right)\\
&\leq P_0^{(n)}\left(\cup_{i=J+1}^{J_1}\{Z_1(\frac{2\pi i}{k}) \geq \frac{ni}{k}\} \cup \cup_{i=J+1}^{j}\{Z_3(\frac{2\pi i}{k}) \leq \frac{ni}{k}\}\right)\\
&\leq P_0^{(n)}\left(Z_1(t) \geq \frac{nt}{2\pi} \text{ for some } t \geq \frac{2\pi J}{k}\right)\\
&\quad + P_0^{(n)}\left(Z_3(t) \leq \frac{nt}{2\pi} \text{ for some } t \geq \frac{2\pi J}{k}\right). \qquad (7.6)
\end{aligned}$$

Applying Corollary A.1, we get

$$\begin{aligned}&\limsup_{n\to\infty} P_0^{(n)}\left(Z_1(t) \geq \frac{nt}{2\pi} \text{ for some } t \geq \frac{2\pi J}{k}\right)\\ \leq\ & P_{\lambda_1}\left(Z_1(\frac{2\pi n}{k}J) \geq \frac{n}{k}J\right) + P_{\lambda_1'}\left(Z_1(\frac{2\pi n}{k}J) \leq \frac{n}{k}J\right)\end{aligned}$$

where λ_1' is the conjugate value for λ_1, defined in Lemma A.6 with $c = 1/2\pi$. Bounds for other ranges of i are similar. Letting $J \to \infty$, we have

$$\begin{aligned}&\liminf_{n\to\infty} P_0^{(n)}\left(\mathcal{T}_k = \theta_j | Y(\theta_j) = \frac{n\Delta}{2\pi}(1+\epsilon)\right)\\ \geq\ & 2\pi\frac{n}{k}\exp\left(-2\pi\frac{n}{k}(\lambda_1+\lambda_3)\right)(\lambda_3-\lambda_1).\end{aligned} \tag{7.7}$$

Moreover, (7.7) holds uniformly if we exclude small intervals around $\theta = 0$ and $\theta = 2\pi$. Neglecting the endpoint corrections, we get from (7.4),

$$\begin{aligned}&P_0^{(n)}\left\{\sup_{0\leq\theta\leq 2\pi} Y(\theta) \geq \frac{n\Delta}{2\pi}(1+\epsilon)\right\}\\ \geq\ & 2\pi n(\lambda_3-\lambda_1)\exp\left(-2\pi\frac{n}{k}(\lambda_1+\lambda_3)\right) P_0^{(n)}\left(Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right)(1+o(1)).\end{aligned}$$

Letting $\frac{n}{k} \to 0$ establishes a lower bound. A similar upper bound based on Y' completes the proof.

□

In proving Theorem 7.1, we have neglected the term $P(\mathcal{T}_k = 0)$, for which the local expansion is not valid. Since this endpoint is an artifact of our discretization and ordering, neglecting this term may seem reasonable; however, numerical results suggest it is better included. Also, as $\epsilon \to 0$ with n fixed, the right hand side of (7.2) tends to 0, while the left hand side tends to 1. Incorporating the endpoint correction $P_0^n(\mathcal{T}_k = 0)$ helps alleviate this discrepancy. Using Lemma A.3, we have

$$\begin{aligned}P_0^{(n)}(\mathcal{T}_k = 0) &= P_0^{(n)}\left(Y(0) \geq \frac{n\Delta}{2\pi}(1+\epsilon)\right)\\ &\leq \frac{(1+\epsilon)(2\pi-\Delta)}{2\pi\epsilon} P_0^{(n)}\left(Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right).\end{aligned}$$

The corresponding result for a process observed on the unit interval is derived similarly. We get

$$P_0^{(n)}\left\{\sup_{0<t<1-\Delta} Y(t) \geq n\Delta(1+\epsilon)\right\}$$
$$= \left(n\epsilon + \frac{(1+\epsilon)(1-\Delta)}{\epsilon}\right) P_0^{(n)}(Y(0) = n\Delta(1+\epsilon))(1+o(1)). \tag{7.8}$$

When $k = \frac{1}{\Delta}$ is an integer, Naus (1966) derives the exact expression

$$P_0^{(n)}\left\{\sup_{0<t<1-\Delta} Y(t) \geq m\right\} = 1 - \sum_{Q_0} n! k^{-n} \det(\frac{1}{c_{ij}!}) \tag{7.9}$$

where Q_0 is the set of all partitions of n into $m_1, \ldots, m_k$ such that $0 \leq m_i < m$ and $\sum_{i=1}^{k} m_i = n$. The elements c_{ij} are defined by

$$c_{ij} = \begin{cases} (j-i)m - \sum_{r=i+1}^{j-1} m_r & 1 \leq i < j \leq k \\ (j-i)m + \sum_{r=j}^{i} m_r & 1 \leq j \leq i \leq k \end{cases}.$$

A slightly more complicated expression when $\frac{1}{\Delta}$ is not an integer is given by Huntington and Naus (1975).

Evaluation of (7.9) involves computing $|Q_0|$ $k \times k$ determinants, and when n and m are large, the set Q_0 will also be large. Hence evaluation of (7.9) will often be difficult. In the special case $\Delta = \frac{1}{2}$, (7.9) reduces to

$$P_0^{(n)}\left\{\sup_{0<t<\frac{1}{2}} Y(t) \geq m\right\} = 1 - 2^{-n} \sum_{k=n-m+1}^{m-1} \left(\binom{n}{k} - \binom{n}{m}\right). \tag{7.10}$$

We compare the exact expression (7.10) with our approximate formula (7.8) in Table 7.1. We see the approximation is performing very well, even for large probabilities.

The approximations (7.2) and (7.8) do not appear to have been published previously, which is somewhat surprising given the simplicity of the approximations and the large amount of literature devoted to this problem. The most closely related

	m				
$n = 10$	6	7	8	9	10
Exact	0.9590	0.6953	0.3291	0.0898	0.0107
Approx.	1.0254	0.6738	0.3222	0.0891	0.0107
$n = 20$	11	13	15	17	19
Exact	0.9840	0.6328	0.1745	0.0167	0.00036
Approx	1.2013	0.6038	0.1700	0.0165	0.00036
$n = 50$	27	30	33	36	39
Exact	0.9597	0.5794	0.1640	0.0201	0.00099
Approx	1.0316	0.5442	0.1580	0.0197	0.00098

Table 7.1: Approximations to the Tail Probabilities for the 1-Dimensional Scan Statistic

result would seem to be that of Berman and Eagleson (1985) who work on the unit interval and show

$$P_0^{(n)}\left(\sup_{0\le t\le 1-\Delta} Y(t)\ge m\right)\le (n-m+1)P(\mathcal{B}(n,\Delta)\ge m-1).$$

Using the Binomial tail bound given by Lemma A.3 and a simple manipulation of the binomial probabilities, we can replace this bound by

$$P_0^{(n)}\left(\sup_{0\le t\le 1-\Delta} Y(t)\ge m\right)\le \frac{(m-1)(1-\Delta)^2 m}{\Delta(m-1-n\Delta)}P_0^{(n)}(Y(0)=m) \tag{7.11}$$

for $m > n\Delta+1$. This tends to substantially overestimate the true probability. For example, when $n=50$ and $\Delta=0.5$, (7.11) gives 33.7, 4.55, 0.660, 0.052 and 0.0019 for the values of m given in Table 7.1.

7.1.2 Unknown Δ

We now consider the case Δ unknown. We choose limits Δ^0 and Δ^1 such that $0\le\Delta^0<\Delta\le\Delta^1\le 2\pi$. Let $\psi=\theta+\Delta$. If θ and ψ are known, the log-likelihood ratio statistic for testing $\mathcal{H}_0:\lambda_0=\lambda_1$ vs $\mathcal{H}_1:\lambda_0<\lambda_1$ is

$$l(\theta,\psi)=$$
$$\begin{cases} X(\theta,\psi)\log\left(\frac{2\pi X(\theta,\psi)}{\Delta n}\right)+(n-X(\theta,\psi))\log\left(\frac{2\pi(n-X(\theta,\psi))}{(2\pi-\Delta)n}\right) & X(\theta,\psi)\ge\frac{n\Delta}{2\pi} \\ 0 & \text{otherwise}\end{cases}.$$

For unknown θ and Δ, we will reject $\mathcal{H}_0$ if

$$M_{\Delta^0,\Delta^1}=\sup_{\substack{0\le\theta<2\pi\\ \Delta^0\le\psi-\theta\le\Delta^1}} l(\theta,\psi)\ge\frac{1}{2}c^2. \tag{7.12}$$

If we define a function $h(\Delta)$ by

$$h(\Delta)\log\left(\frac{2\pi h(\Delta)}{\Delta}\right)+(1-h(\Delta))\log\left(\frac{2\pi(1-h(\Delta))}{2\pi-\Delta}\right)=\frac{c^2}{2n}$$

subject to $h(\Delta) \geq \frac{\Delta}{2\pi}$, then (7.12) is equivalent to

$$\sup_{0 \leq \theta < 2\pi} \sup_{\Delta^0 \leq \psi - \theta \leq \Delta^1} (X(\theta, \psi) - nh(\psi - \theta)) \geq 0. \tag{7.13}$$

The existence of $h(\Delta)$ for $\Delta \leq e^{-\pi\eta^2}$ is similar to the existence of q_t in (3.5). It is easily shown that $h(\Delta)$ is an increasing function of Δ.

Theorem 7.2 *Suppose $n \to \infty$, $c \to \infty$ with $\eta = \frac{c}{\sqrt{n}}$ fixed. Then*

$$\begin{aligned} & P_0^{(n)} \left\{ M_{\Delta^0, \Delta^1} \geq \frac{1}{2} c^2 \right\} \\ = \;& c^3 \phi(c) \int_{\Delta^0}^{\Delta^1} \frac{2\pi}{\eta^3 h'(\Delta)} \left(h'(\Delta) - \frac{1 - h(\Delta)}{2\pi - \Delta} \right)^2 \left(\frac{h(\Delta)}{\Delta} - \frac{1 - h(\Delta)}{2\pi - \Delta} \right) \\ & \frac{d\Delta}{\sqrt{h(\Delta)(1 - h(\Delta))}} (1 + o(1)). \end{aligned} \tag{7.14}$$

Proof: To facilitate the definition of a first passage time, we order the pairs (θ, ψ) according to

$$(\theta, \psi) < (\theta', \psi') \iff \begin{cases} \theta < \theta' \\ \theta = \theta', \psi > \psi' \end{cases}. \tag{7.15}$$

That is we think of starting with $(\theta, \psi) = (0, 2\pi)$, then decrease ψ until $\psi = 0$. We then increase θ a small amount, set $\psi = \theta + 2\pi$, and then decrease ψ until $\psi = \theta$. We repeat this process until $\theta = 2\pi$; see Figure 7.1.

Fix k large and let $\theta_i = \frac{2\pi i}{k}; i = 0, \ldots, k-1$. We define the first passage time τ_k by

$$\tau_k = \inf_{\substack{0 \leq i \leq k-1 \\ \theta_i + \Delta^0 \leq \psi \leq \theta_i + \Delta^1}} \{(\theta_i, \psi) : X(\theta_i, \psi) \geq nh(\psi - \theta_i)\} \tag{7.16}$$

Suppose we fix $\theta = \theta_i$ and increase (θ_i, ψ), which by (7.15) is equivalent to decreasing ψ. Then $X(\theta_i, \psi)$ is a decreasing and integer valued, and $h(\psi - \theta_i)$ is a decreasing function. From (7.16), we see τ_k can only occur when $nh(\psi - \theta_i)$ is an integer. We

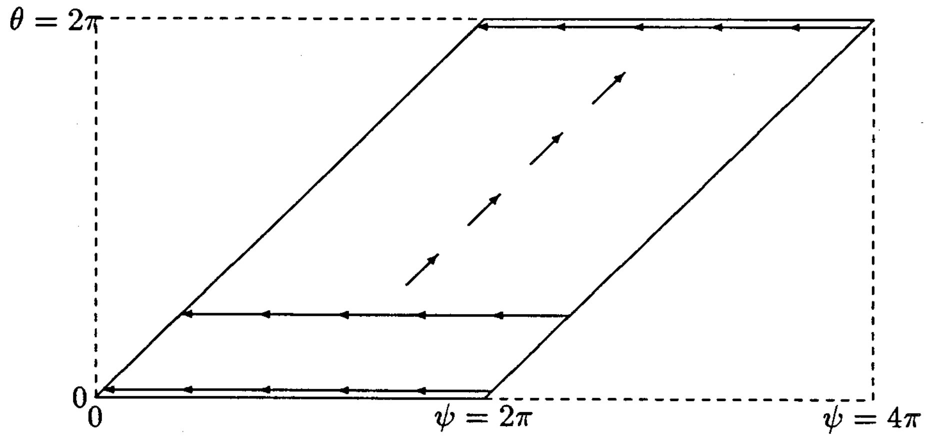


Figure 7.1: Ordering of (θ, ψ) for one dimensional scan statistic with unknown Δ. We begin at with $\psi = 2\pi$, $\theta = 0$ and first decrease ψ, then increase θ, finishing at $\psi = 2\pi$, $\theta = 2\pi$.

let Δ_j be the solution of $nh(\Delta) = j$ and $\psi_{i,j} = \theta_i + \Delta_j$. We can then rewrite (7.16) as

$$\tau_k = \inf_{\substack{0 \leq i \leq k-1 \\ j_0 \leq j \leq j_1}} \{(\theta_i, \psi_{i,j}) : X(\theta_i, \psi_{i,j}) \geq j\}$$

where the limits j_0 and j_1 are chosen so that $\Delta^0 \leq \Delta_j \leq \Delta^1$ for $j_0 \leq j \leq j_1$. As usual, we can write

$$P_0^{(n)}(\tau_k = (\theta_i, \psi_{i,j})) = P_0^{(n)}(\tau_k = (\theta_i, \psi_{i,j}) | X(\theta_i, \psi_{i,j}) = j) P_0^{(n)}(X(\theta_i, \psi_{i,j}) = j).$$

Suppose $\psi_{i,j} - \theta_i = \Delta_j$ and $X(\theta_i, \psi_{i,j}) = j$. We define processes $Z_k(t), k = 1, \ldots, 4$ as in Lemma 7.1.

For $\delta_1 \leq \theta_i$, $\delta_4 \leq 2\pi - \psi_{i,j}$ and $\delta_3 \leq \psi_{i,j} - \theta_i$, we can write

$$X(\theta_i - \delta_1, \psi_{i,j} + \delta_4) = X(\theta_i, \psi_{i,j}) + Z_1(\delta_1) + Z_4(\delta_4) \tag{7.17}$$

$$X(\theta_i - \delta_1, \psi_{i,j} - \delta_3) = X(\theta_i, \psi_{i,j}) + Z_1(\delta_1) - Z_3(\delta_3). \tag{7.18}$$

For $\tau_k = (\theta_i, \psi_{i,j})$ we require

$$X(\theta_i - \delta_1, \psi_{i,j} + \delta_4) \leq nh(\Delta_j + \delta_1 + \delta_4) \tag{7.19}$$

for $\delta_4 \geq 0$ and $\delta_1 = \frac{2\pi l}{k}, l = 0, 1, \ldots, i$ with equality only when $\delta_1 = \delta_4 = 0$, and

$$X(\theta_i - \delta_1, \psi_{i,j} - \delta_3) < nh(\Delta_j + \delta_1 - \delta_3) \tag{7.20}$$

for $\delta_3 \geq 0$ and $\delta_1 = \frac{2\pi l}{k}, l = 1, 2, \ldots, i$. We therefore have, for any $T > 0$,

$$P_0^{(j,n)}(\tau_k = (\theta_i, \psi_{i,j})) \leq P_0^{(j,n)}(U_1 \cap U_2) \tag{7.21}$$

where U_1 is the event that (7.19) holds for $n\delta_4 \leq T$ and $\delta_1 = \frac{2\pi l}{k}, l = 0, \ldots, \frac{T}{2\pi}\frac{k}{n}$ and U_2 is the event that (7.20) holds for $n\delta_3 \leq T$ and $\delta_1 = \frac{2\pi l}{k}, l = 1, \ldots, J$.

Let $\{\tilde{Z}_1(t_1), \tilde{Z}_2(t_2), \tilde{Z}_3(t_3), \tilde{Z}(t_4); t_j \geq 0\}$ be a vector of the limiting processes in Lemma 7.1. Letting $n \to \infty$ with $\frac{k}{n}$ fixed, and fixing θ_i and $\psi_{i,j}$, we use (7.17) and (7.18) to get

$$\begin{aligned}
&\limsup_{n\to\infty} P_0^{(j,n)}(U_1 \cap U_2) \\
&\leq P_0^{(j,n)} \Big\{ \tilde{Z}_1(\delta_1) + \tilde{Z}_4(\delta_4) < h'(\Delta_j)(\delta_1 + \delta_4), \delta_4 < T, \delta_1 = \frac{2\pi l n}{k}, l = 0, \ldots, \frac{Tk}{2\pi n}, \\
&\qquad \tilde{Z}_1(\delta_1) - \tilde{Z}_3(\delta_3) < h'(\Delta_j)(\delta_1 - \delta_3), \delta_3 < T, \delta_1 = \frac{2\pi l n}{k}, l = 1, \ldots, \frac{Tk}{2\pi n} \Big\} \\
&\leq P\left(\tilde{Z}_1(t) < h'(\Delta_j)t, t = \frac{2\pi n l}{k}, l = 1, \ldots, J\right) \\
&\qquad \times P\left(\tilde{Z}_4(t) \leq h'(\Delta_j)t, 0 < t < T\right) \\
&\qquad \times P\left(\tilde{Z}_3(t) \geq h'(\Delta_j)(t - \frac{2\pi}{k}), 0 < t < T\right).
\end{aligned} \tag{7.22}$$

Since T is arbitrary, we can use (7.21) and (7.22) to obtain, for any $\epsilon > 0$ and n sufficiently large,

$$P_0^{(n)}(\tau_k = (\theta_i, \psi_{i,j}) | X(\theta_i, \psi_{i,j}) = j)$$

$$\begin{aligned}
&\leq (1+\epsilon)P\left(\tilde{Z}_1(t) \leq h'(\Delta_j)t, t = \frac{2\pi ln}{k}, l = 1,2,\ldots\right) \\
&\quad \times P\left(\tilde{Z}_4(t) \leq h'(\Delta_j)t, t > 0\right) \\
&\quad \times P\left(\tilde{Z}_3(t) \geq h'(\Delta_j)(t - \frac{2\pi}{k}), t > 0\right).
\end{aligned}$$

We can apply Lemma 2.1 to obtain

$$P\left(\tilde{Z}_4(t) \leq h'(\Delta_j)t, t > 0\right) = 1 - \frac{(1-h(\Delta_j))}{h'(\Delta_j)(2\pi - \Delta_j)}$$

and if $\frac{n}{k}$ is sufficiently small,

$$P\left(\tilde{Z}_1(t) \leq h'(\Delta_j)t, t = \frac{2\pi nl}{k}, l = 1,2,\ldots\right) \leq (1+\epsilon)\left(1 - \frac{(1-h(\Delta_j))}{h'(\Delta_j)(2\pi - \Delta_j)}\right).$$

Similarly, we can use (2.8) to obtain

$$\begin{aligned}
&P\left(\tilde{Z}_3(t) \geq h'(\Delta_j)(t - \frac{2\pi}{k}), t > 0\right) \\
&= 1 - \exp\left(-\frac{2\pi n}{k}\left(\frac{h(\Delta_j)}{\Delta_j} - \frac{1-h(\Delta_j)}{2\pi - \Delta_j}\right)\right) \\
&\leq (1+\epsilon)\left(\frac{2\pi n}{k}\left(\frac{h(\Delta_j)}{\Delta_j} - \frac{1-h(\Delta_j)}{2\pi - \Delta_j}\right)\right) \qquad (7.23)
\end{aligned}$$

provided $\frac{n}{k}$ is sufficiently small. This gives, for $\frac{n}{k}$ sufficiently small and n sufficiently large,

$$\begin{aligned}
&P_0^{(n)}(\tau_k = (\theta_i, \psi_{i,j}) | X(\theta_i, \psi_{i,j}) = j) \\
&\leq (1+\epsilon)^3 \frac{2\pi n}{kh'(\Delta_j)^2}\left(h'(\Delta_j) - \frac{1-h(\Delta_j)}{2\pi - \Delta_j}\right)^2 \left(\frac{h(\Delta_j)}{\Delta_j} - \frac{1-h(\Delta_j)}{2\pi - \Delta_j}\right). \quad (7.24)
\end{aligned}$$

Moreover, (7.24) will hold uniformly if θ_i and ψ_{ij} are bounded away from the boundaries. We obtain a similar lower bound by subtracting from (7.21) the probability of either (7.17) or (7.18) failing outside the region covered by events U_1 and U_2. These probabilities are then approximated in a manner similar to that used in the proof of Lemma A.4 and Corollary A.1. We omit the details.

Assuming the boundary effects to be second order corrections, we sum (7.24) over i and j and obtain the approximation to the significance level

$$
\begin{aligned}
&P_0^{(n)}\left\{M_{\Delta^0,\Delta^1} \geq \frac{1}{2}c^2\right\} \\
&\approx \sum_{j=j_0}^{j_1}\sum_{i=0}^{k-1} \frac{2\pi n}{kh'(\Delta_j)^2}\left(h'(\Delta_j) - \frac{1-h(\Delta_j)}{2\pi-\Delta_j}\right)^2 \left(\frac{h(\Delta_j)}{\Delta_j} - \frac{1-h(\Delta_j)}{2\pi-\Delta_j}\right) \\
&\quad \times \binom{n}{nh(\Delta_j)}\left(\frac{2\pi h(\Delta_j)}{\Delta_j}\right)^{nh(\Delta_j)}\left(\frac{2\pi(1-h(\Delta_j))}{2\pi-\Delta_j}\right)^{n(1-h(\Delta_j))} \\
&\approx \sum_{j=j_0}^{j_1} \frac{2\pi n}{h'(\Delta_j)^2}\left(h'(\Delta_j) - \frac{1-h(\Delta_j)}{2\pi-\Delta_j}\right)^2 \left(\frac{h(\Delta_j)}{\Delta_j} - \frac{1-h(\Delta_j)}{2\pi-\Delta_j}\right) \\
&\quad \times \frac{1}{\sqrt{2\pi nh(\Delta_j)(1-h(\Delta_j))}} e^{-\frac{1}{2}c^2} \\
&\approx c^3\phi(c)\int_{\Delta^0}^{\Delta^1} \frac{2\pi}{\eta^3 h'(\Delta)}\left(h'(\Delta) - \frac{1-h(\Delta)}{2\pi-\Delta}\right)^2 \left(\frac{h(\Delta)}{\Delta} - \frac{1-h(\Delta)}{2\pi-\Delta}\right) \\
&\quad \frac{d\Delta}{\sqrt{h(\Delta)(1-h(\Delta))}}.
\end{aligned}
\tag{7.25}
$$

That (7.25) is an asymptotic relation follows by the uniformity of (7.24) and noting that the edge correction at $\Delta = \Delta^1$ is of a smaller order of magnitude. Setting $h(\Delta^1) = \frac{\Delta^1}{2\pi}(1+\epsilon)$, we get from (7.2) the approximation

$$
P_0^{(n)}\left\{\sup_{0\leq\theta\leq 2\pi} l(\theta, \theta+\Delta^1) \geq \frac{1}{2}c^2\right\} \approx \frac{c\phi(c)}{\eta}\frac{2\pi(2\pi h(\Delta^1) - \Delta^1)}{\Delta^1(2\pi-\Delta^1)\sqrt{h(\Delta^1)(1-h(\Delta^1))}}. \tag{7.26}
$$

The correction (7.26) is of a smaller order of magnitude than (7.25) but it seems reasonable to include it especially when $\Delta_1 - \Delta_0$ is small.

□

We can obtain an approximation for the limiting (as $n \to \infty$) Gaussian process by letting $\eta \to 0$. If $X(0,\Delta) = nh(\Delta) \approx \frac{n\Delta}{2\pi}$, we have

$$
l(0,\Delta) \approx \frac{1}{2}\frac{n(2\pi)^2}{\Delta(2\pi-\Delta)}\left(h(\Delta) - \frac{\Delta}{2\pi}\right)^2
$$

and hence

$$h(\Delta) = \frac{\Delta}{2\pi} + \frac{\eta}{2\pi}\sqrt{\Delta(2\pi - \Delta)} + o(\eta). \tag{7.27}$$

This gives

$$\begin{aligned}
\frac{h(\Delta)}{\Delta} &= \frac{1}{2\pi} + \eta\frac{2\pi - \Delta}{2\pi\sqrt{\Delta(2\pi - \Delta)}} + o(\eta) \\
\frac{1 - h(\Delta)}{2\pi - \Delta} &= \frac{1}{2\pi} - \eta\frac{\Delta}{2\pi\sqrt{\Delta(2\pi - \Delta)}} + o(\eta) \\
h'(\Delta) &= \frac{1}{2\pi} + \eta\frac{\pi - \Delta}{2\pi\sqrt{\Delta(2\pi - \Delta)}} + o(\eta)
\end{aligned}$$

and

$$h'(\Delta) - \frac{1 - h(\Delta)}{2\pi - \Delta} = \frac{\eta}{2\sqrt{\Delta(2\pi - \Delta)}} + o(\eta) \tag{7.28}$$

$$\frac{h(\Delta)}{\Delta} - \frac{1 - h(\Delta)}{2\pi - \Delta} = \frac{\eta}{\sqrt{\Delta(2\pi - \Delta)}} + o(\eta) \tag{7.29}$$

$$h(\Delta)(1 - h(\Delta)) = \frac{1}{(2\pi)^2}\Delta(2\pi - \Delta) + o(1). \tag{7.30}$$

Substituting (7.28), (7.29) and (7.30) into (7.25) gives

$$\begin{aligned}
& P_0^{(n)}\{M_{\Delta^0,\Delta^1} \geq \frac{1}{2}\eta^2\} \\
\approx\ & c^3\phi(c)\int_{\Delta^0}^{\Delta^1} \frac{(2\pi)^2}{\eta^3}\frac{\eta^2}{4\Delta(2\pi - \Delta)}\frac{2\pi\eta d\Delta}{\Delta(2\pi - \Delta)} \\
=\ & \frac{c^3\phi(c)}{4}\int_{\Delta^0}^{\Delta^1} \frac{8\pi^3}{(\Delta(2\pi - \Delta))^2}d\Delta \\
=\ & \frac{c^3\phi(c)}{4}\int_{u_0}^{u_1} \frac{1}{(u(1 - u))^2}du && (7.31) \\
=\ & \frac{c^3\phi(c)}{4}\left(2\log\left(\frac{u_1(1 - u_0)}{u_0(1 - u_1)}\right) + \frac{2u_1 - 1}{u_1(1 - u_1)} - \frac{2u_0 - 1}{u_0(1 - u_0)}\right) && (7.32)
\end{aligned}$$

where $u = \frac{\Delta}{2\pi}$. Siegmund (1988b) gives a version of (7.31) for a discrete time Gaussian process on the line. Dividing (7.31) by 2 gives Siegmund's result without the discrete time correction factor.

		c			
		3.0	3.5	4.0	4.5
	Gaussian	0.7947	0.2485	0.0569	0.0097
$n = 20$	Large Dev'n	0.4636	0.1173	0.0222	0.0032
	Simulation	0.3864	0.1083	0.0205	0.0041
$n = 50$	Large Dev'n	0.7113	0.2219	0.0421	0.0061
	Simulation	0.4717	0.1980	0.0396	0.0069
$n = 100$	Large Dev'n	0.6852	0.2144	0.0499	0.0091
	Simulation	0.4418	0.1637	0.0437	0.0096
Simulation s.e.		0.005	0.004	0.002	0.001

Table 7.2: Approximations to Distribution of Scan Statistic: 10% Truncation

		c			
		3.0	3.5	4.0	4.5
	Gaussian	0.3902	0.1220	0.0279	0.0048
$n = 20$	Large Dev'n	0.3764	0.0916	0.0168	0.0023
	Simulation	0.3400	0.0891	0.0189	0.0034
$n = 50$	Large Dev'n	0.3229	0.1006	0.0233	0.0042
	Simulation	0.2635	0.0856	0.0212	0.0040
$n = 100$	Large Dev'n	0.3328	0.1025	0.0232	0.0039
	Simulation	0.2763	0.0883	0.0234	0.0040
Simulation s.e.		0.005	0.003	0.002	0.0006

Table 7.3: Approximations to Distribution of Scan Statistic: 20% Truncation

We compare the Gaussian approximation (7.32) with the large deviation approximation (7.25) in Table 7.2 with $u_0 = 1 - u_1 = 0.9$ and in Table 7.3 with $u_0 = 1 - u_1 = 0.8$. The simulations, of size 10000, indicate the approximations are performing reasonably, even for small n. For small c, the approximations tend to overestimate the true probability, especially in Table 7.2 with 10% truncation. The large deviation approximation is in most cases substantially better than the Gaussian approximation. Both approximations are computed without the boundary corrections. The boundary corrections will make the errors larger.

We can define tests similar to (7.13) but with alternative choices of the function $h(\Delta)$. For example, Kuiper's (1960) test statistic is equivalent to the choice $h(\Delta) = \frac{\Delta}{2\pi} + \eta$. A similar derivation will give a large deviation approximation for this case, although we cannot simply substitute the new $h(\Delta)$ into (7.25). In particular, we must modify (7.23) to

$$P\left(\tilde{Z}_3(t) \geq nh'(\Delta_j)(t - \frac{2\pi}{k}), t > 0\right) \approx \frac{2\pi n}{k}\left(\frac{h(\Delta_j)}{\Delta_j} - \frac{h^*(\Delta_j)}{2\pi - \Delta_j}\right)$$

where

$$\frac{h(\Delta)}{\Delta h'(\Delta)} + \log\left(\frac{h(\Delta)}{\Delta}\right) = \frac{h^*(\Delta)}{(2\pi - \Delta)h'(\Delta)} + \log\left(\frac{h^*(\Delta)}{2\pi - \Delta}\right).$$

7.2 Change Regions on the Plane

Suppose we observe a Poisson Process on the unit square, and the rate is given by

$$\lambda(x,y) = \begin{cases} \lambda_1 & a_1 < x < a_2, b_1 < y < b_2 \\ \lambda_0 & \text{otherwise} \end{cases}$$

where $0 \leq a_1 < a_2 \leq 1$ and $0 \leq b_1 < b_2 \leq 1$. We consider both the cases $\Delta_1 = a_2 - a_1$ and $\Delta_2 = b_2 - b_1$ known and unknown. Our null hypothesis is $\mathcal{H}_0 : \lambda_0 = \lambda_1$. We will for simplicity consider the one-sided alternative $\mathcal{H}_1 : \lambda_1 > \lambda_0$. Similar results can be obtained for testing against the alternative $\mathcal{H}_2 : \lambda_1 < \lambda_0$ or for two sided alternatives. However, we believe $\mathcal{H}_1$ will often be the most interesting in practice.

For our development, the choice of the unit square for the domain is somewhat arbitrary but helps keep formulae tidy. However, the choice of a rectangular region of known orientation for λ_1 is crucial to our the success of our methods.

The problems we treat here lead to multidimensional versions of the scan statistics discussed in the previous section. These do not appear to have been analyzed in the literature previously. Other methods have been proposed to test the null hypothesis that a point process is generated by a homogeneous point process. Generally, these are ad hoc proposals and are formulated without specifying any precise alternative. Ripley (1981) has a discussion of several approaches.

Let $X(a_1, a_2, b_1, b_2)$ denote the number of events in the rectangle $[a_1, a_2] \times [b_1, b_2]$. The following Lemma characterizes the local behaviour of this process.

Lemma 7.2 *For fixed a_1, a_2, b_1 and b_2, we define processes $Z_1(u_1), \ldots, Z_8(u_8)$ to be the increments as each of the boundaries are moved:*

$$
\begin{aligned}
X(a_1, a_2, b_1, b_2 + u_1) &= m + Z_1(u_1) \\
X(a_1, a_2, b_1, b_2 - u_2) &= m - Z_2(u_2) \\
X(a_1, a_2, b_1 - u_3, b_2) &= m + Z_3(u_3) \\
X(a_1, a_2, b_1 + u_4, b_2) &= m - Z_4(u_4) \\
X(a_1, a_2 + u_5, b_1, b_2) &= m + Z_5(u_5) \\
X(a_1, a_2 - u_6, b_1, b_2) &= m - Z_6(u_6) \\
X(a_1 - u_7, a_2, b_1, b_2) &= m + Z_7(u_7) \\
X(a_1 + u_7, a_2, b_1, b_2) &= m - Z_7(u_7)
\end{aligned}
$$

where $m = X(a_1, a_2, b_1, b_2)$*; see Figure 7.2. Suppose* $n \to \infty$ *and* $m \to \infty$ *with* $\frac{m}{n}$ *fixed. Then*

$$\left\{Z_1(\frac{u_1}{n}), \ldots, Z_8(\frac{u_8}{n}); 0 \le u_j \le T\right\}$$

converges in law to a vector of 8 independent Poisson processes, with rates

$$\begin{aligned}
\lambda_1 &= \lambda_3 = (1 - \frac{m}{n})\frac{a_2 - a_1}{1 - \Delta_1\Delta_2} \\
\lambda_2 &= \lambda_4 = \frac{m}{n}\frac{a_2 - a_1}{\Delta_1\Delta_2} \\
\lambda_5 &= \lambda_7 = (1 - \frac{m}{n})\frac{b_2 - b_1}{1 - \Delta_1\Delta_2} \\
\lambda_6 &= \lambda_8 = \frac{m}{n}\frac{b_2 - b_1}{\Delta_1\Delta_2}.
\end{aligned}$$

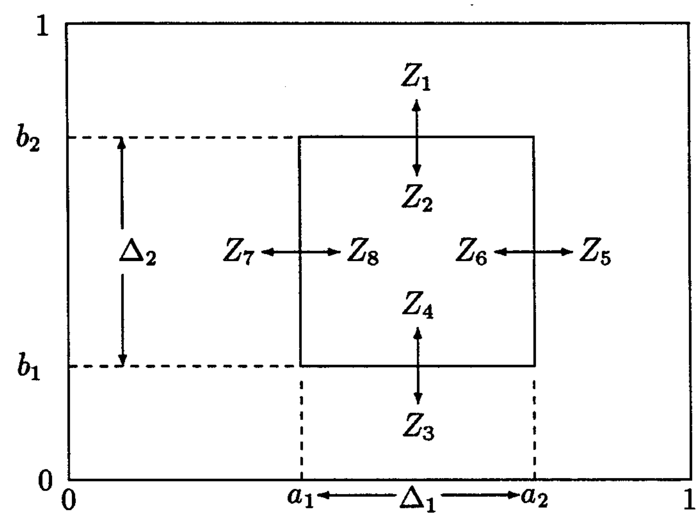


Figure 7.2: Local Expansion of $X(t)$. We fix a_1, a_2, b_1 and b_2, and define processes Z_j, $j = 1, \ldots, 8$ to be the increments of $X(t)$ as the boundaries are moved in the indicated directions.

Proof: First, we delete events in the corners $[x - \frac{T}{n}, x + \frac{T}{n}] \times [y - \frac{T}{n}, y + \frac{T}{n}]$ for $x \in \{a_1, a_2\}$ and $y \in \{b_1, b_2\}$, and construct processes $Z_j'(u_j); j = 1, \ldots, 8$ in the same manner as we constructed $Z_1, \ldots, Z_8$. Applying an argument similar to that of Lemma 7.1 gives the convergence of $\{Z'_j; j = 1, \ldots, 8\}$. Secondly, note the neglected corners have total area of $O(n^{-2})$ and hence the expected number of events in these corners is of $O(n^{-1})$. Therefore, the probability converging to 1, $Z_j(\delta)_j = Z'_j(\delta_j)$ for $0 \leq \delta_j \leq \frac{T}{n}$ and $1 \leq j \leq 8$, whcih completes the proof.

□

The importance of choosing a rectangular region for the hich rate λ_1 is apparent from Lemma 7.2. If we chose any other shaped region, we would not be able to decompose the resulting random field into independent one dimensional processes.

7.2.1 Known Δ_1 and Δ_2

Let $t = (t_1, t_2)$, $0 \leq t_1 \leq 1 - \Delta_1$ and $0 \leq t_2 \leq 1 - \Delta_2$. Define $Y(t) = X([t_1, t_1 + \Delta_1] \times [t_2, t_2 + \Delta_2])$. We will reject $\mathcal{H}_0$ if

$$\sup_t Y(t) \geq n\Delta_1\Delta_2(1 + \epsilon) \tag{7.33}$$

for some $\epsilon > 0$. The supremum is taken over the range $0 \leq t_1 \leq 1 - \Delta_1$ and $0 \leq t_2 \leq 1 - \Delta_2$. The significance level of (7.33) is then

$$P_0^{(n)}\left(\sup_t Y(t) \geq n\Delta_1\Delta_2(1 + \epsilon)\right). \tag{7.34}$$

Fix $k > 0$ and let

$$t_{i,j} = (\frac{i}{k}, \frac{j}{k}), \quad 0 \leq i < k(1 - \Delta_1) \quad 0 \leq j < k(1 - \Delta_2).$$

We define $Y'(t) = X([t_1, t_1 + \Delta_1 + \frac{1}{k}] \times [t_2, t_2 + \Delta_2 + \frac{1}{k}])$. As in the proof of Theorem 7.1, we have

$$P_0^{(n)}\left(\sup_{i,j} Y(t_{i,j}) \geq n\Delta_1\Delta_2(1 + \epsilon)\right) \leq P_0^{(n)}\left(\sup_t Y(t) \geq n\Delta_1\Delta_2(1 + \epsilon)\right)$$

$$\leq \; P_0^{(n)}\left(\sup_{i,j} Y'(t_{i,j}) \geq n\Delta_1\Delta_2(1+\epsilon)\right).$$

We define an ordering of the (i,j) pairs by

$$(i,j) < (i',j') \iff \begin{cases} j < j' \\ j = j', i < i' \end{cases}.$$

We define a first passage time by $\tau_k = \inf\{(i,j) : Y(t_{i,j}) \geq n\Delta_1\Delta_2(1+\epsilon)\}$. That is, we begin with a rectangle of size $\Delta_1 \times \Delta_2$ with lower left hand corner at $(0,0)$, and move right until the lower right corner is at $(1,0)$. Then the rectangle is moved up a small amount, and the process repeated.

We approximate (7.34) by $P_0^{(n)}(\tau_k \leq (k,k))$, which is the sum of the individual terms $P_0^{(n)}(\tau_k = (i,j))$. As usual, we

$$\begin{aligned} &P_0^{(n)}(\tau_k = (i,j)) \\ &= \; P_0^{(n)}\left(\tau_k = (i,j) | Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)\right) P_0^{(n)}(Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)). \end{aligned}$$

Suppose for some i,j, $Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)$. We define a process $\tilde{Y}$ by

$$\tilde{Y}(t_{i-u,j-v}) \;=\; Y(t_{i,j}) + \left(Z_7(\frac{u}{k}) - Z_6(\frac{u}{k})\right) + \left(Z_3(\frac{v}{k}) - Z_2(\frac{v}{k})\right) \tag{7.35}$$

$$\tilde{Y}(t_{i+u,j-v}) \;=\; Y(t_{i,j}) + \left(Z_5(\frac{u}{k}) - Z_8(\frac{u}{k})\right) + \left(Z_3(\frac{v}{k}) - Z_2(\frac{v}{k})\right) \tag{7.36}$$

where $0 \leq v \leq j$, $0 \leq u \leq i$ in (7.35) and $0 \leq u \leq k(1-\Delta_1) - i$ in (7.36), and $Z_j(u_j); j = 1,\ldots,8$ are defined in Lemma 7.2.

Fix $T > 0$. If there are no events in the corners defined in Lemma 7.2, then $\tilde{Y}(t_{i+u,j+v}) = Y(t_{i+u,j+v})$ for $-\frac{Tk}{n} \leq u \leq \frac{Tk}{n}$ and $0 \leq v \leq \frac{Tk}{n}$. Letting $J = \frac{Tk}{n}$, we can write

$$\begin{aligned} &P_0^{(n)}(\tau_k = t_{i,j} | Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)) \\ &\geq \; P_0^{(n)}\left((Z_7 - Z_6)(\frac{u}{k}) < 0, (Z_3 - Z_2)(\frac{v}{k}) < 0, (Z_5 - Z_8)(\frac{u}{k}) \leq 0 \;\forall\, 1 \leq u,v \leq J\right) \\ &\quad - P_0^{(n)}(A \cup B | Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)) \end{aligned}$$

where A is the event that there are any events in the corners and B is the event that $Y(t_{i+u,j+v}) \geq n\Delta_1\Delta_2(1+\epsilon)$ for some (u,v) with either $|u| > J$ or $|v| > J$. Letting $n \to \infty$ with $\frac{n}{k}$ and $t_{i,j}$ fixed, we get

$$\begin{aligned}
\limsup_{n\to\infty} P_0^{(n)} \Big(& (Z_7 - Z_6)(\frac{u}{k}) < 0, (Z_3 - Z_2)(\frac{v}{k}) < 0, \\
& (Z_5 - Z_8)(\frac{u}{k}) \leq 0 \;\forall\; 1 \leq u, v \leq J \Big) \\
= \; & P\left((\tilde{Z}_7 - \tilde{Z}_6)(\frac{nu}{k}) < 0, u = 1, \ldots, J\right) P\left((\tilde{Z}_3 - \tilde{Z}_2)(\frac{nv}{k}) < 0, v = 1, \ldots, J\right) \\
& \times P\left((\tilde{Z}_5 - \tilde{Z}_8)(\frac{nu}{k}) \leq 0, u = 1, \ldots, J\right) \\
\geq \; & P\left((\tilde{Z}_7 - \tilde{Z}_6)(\frac{nu}{k}) < 0, u = 1, \ldots\right) P\left((\tilde{Z}_3 - \tilde{Z}_2)(\frac{nv}{k}) < 0, v = 1, \ldots\right) \\
& \times P\left((\tilde{Z}_5 - \tilde{Z}_8)(\frac{nu}{k}) \leq 0, u = 1, \ldots\right)
\end{aligned}$$

where $\{\tilde{Z}_1, \ldots, \tilde{Z}_8\}$ is a vector of the limiting Poisson processes in Lemma 7.2. Also,

$$\limsup_{n\to\infty} P_0^{(n)}(A | Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)) = 0$$

and

$$\limsup_{n\to\infty} P_0^{(n)}(B | Y(t_{i,j}) = n\Delta_1\Delta_2(1+\epsilon)) = f(J)$$

where $f(J) \to 0$ as $J \to \infty$. Letting $J \to \infty$, we therefore have

$$\begin{aligned}
& \limsup_{n\to\infty} P_0^{(n)}(\tau_k = (i,j) | Y(t_{i,j}) \geq n\Delta_1\Delta_2(1+\epsilon)) \\
& \quad \geq \; P\left((\tilde{Z}_7 - \tilde{Z}_6)(\frac{nu}{k}) < 0, u = 1, \ldots\right) P\left((\tilde{Z}_3 - \tilde{Z}_2)(\frac{nv}{k}) < 0, v = 1, \ldots\right) \\
& \qquad \times P\left((\tilde{Z}_5 - \tilde{Z}_8)(\frac{nu}{k}) \leq 0, u = 1, \ldots\right) \\
& \quad \to \; \left(\frac{n}{k}\right)^2 (\lambda_2 - \lambda_3)(\lambda_6 - \lambda_7)(1 - \frac{\lambda_5}{\lambda_8}) \text{ as } \frac{n}{k} \to 0 \qquad (7.37) \\
& \quad = \; \left(\frac{n}{k}\right)^2 \frac{\epsilon\Delta_1}{1 - \Delta_1\Delta_2} \frac{\epsilon\Delta_2}{1 - \Delta_1\Delta_2} \frac{\epsilon}{(1+\epsilon)(1 - \Delta_1\Delta_2)} \qquad (7.38)
\end{aligned}$$

where we have used (2.34) and (2.35) to obtain (7.37). A lower bound similar to (7.38) can be derived similarly. Moreover, (7.38) will hold uniformly in $t_{i,j}$ provided this is bounded away from the edges. Neglecting these second order edge corrections, we sum over i and j to get an approximation to the significance level of (7.33),

$$\begin{aligned} & P_0^{(n)}\left(\sup_t Y(t) \geq n\Delta_1\Delta_2(1+\epsilon)\right) \\ = & \frac{n^2\Delta_1\Delta_2(1-\Delta_1)(1-\Delta_2)\epsilon^3}{(1-\Delta_1\Delta_2)^3(1+\epsilon)} P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1+\epsilon))(1+o(1)). \quad (7.39) \end{aligned}$$

We have edge corrections for three the boundaries $j = 0$, $i = 0$ and $i = \frac{1}{\delta}(1-\Delta_2)$. When $j = 0$ there is no Z_2 or Z_3, and using similar notation to before, we get

$$\begin{aligned} P_0^{(n)}(\tau_k = t_{i,0} | Y(t_{i,0}) = n\Delta_1\Delta_2(1+\epsilon)) & \approx \frac{n}{k}(\lambda_6 - \lambda_7) \\ & = \frac{n\Delta_2\epsilon}{k(1-\Delta_1\Delta_2)} \end{aligned}$$

and therefore

$$P_0^{(n)}(\tau = t_{i,0} \text{ for some } i) \approx \frac{n\Delta_2(1-\Delta_1)\epsilon}{1-\Delta_1\Delta_2} P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1+\epsilon)). \quad (7.40)$$

Similarly, the boundary correction when $i = 0$ is

$$P_0^{(n)}(\tau = t_{0,j} \text{ for some } j) \approx \frac{n\Delta_1(1-\Delta_2)\epsilon^2}{(1+\epsilon)(1-\Delta_1\Delta_2)^2} P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1+\epsilon)). \quad (7.41)$$

A simple calculation shows the third boundary correction is 0. The boundary corrections (7.40) and (7.41) are not symmetric in Δ_1 and Δ_2. There is also a corner correction when $t = (0,0)$,

$$P_0^{(n)}(Y(0) \geq n\Delta_1\Delta_2(1+\epsilon)) \approx \frac{(1+\epsilon)(1-\Delta_1\Delta_2)}{\epsilon} P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1+\epsilon)).$$

We will reject $\mathcal{H}_0$ in favour of the alternative $\mathcal{H}_2 : \lambda_1 < \lambda_0$ if

$$\inf_t \{Y(t) \leq n\Delta_1\Delta_2(1-\epsilon)\}$$

for some $\epsilon > 0$. We can approximate the null distribution of this statistic in an almost identical manner. The final result is

$$\begin{aligned} & P_0^{(n)}(\inf_t Y(t) \leq n\Delta_1\Delta_2(1-\epsilon)) \\ \approx & \frac{(1-\Delta_1)(1-\Delta_2)\Delta_1\Delta_2 n^2\epsilon^3}{(1-\Delta_1\Delta_2)^2(1-\Delta_1\Delta_2(1-\epsilon))} P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1-\epsilon)). \quad (7.42) \end{aligned}$$

We can take $\epsilon = 1$ and (7.42) is still valid; we are then looking for empty rectangles of known dimensions. In this case (7.42) becomes

$$\frac{(1-\Delta_1)(1-\Delta_2)\Delta_1\Delta_2 n^2}{(1-\Delta_1\Delta_2)^2}(1-\Delta_1\Delta_2)^n$$

which when $\Delta_1 = \Delta_2$ is derived by Aldous (1989), equation (H4d) by alternate means.

	m				
$n = 20$	6	8	10	12	14
Large Dev'n	1.1001	0.9742	0.4191	0.0644	0.0038
Simulation	1.0000	0.9258	0.4088	0.0659	0.0045
$n = 50$	18	20	22	24	26
Large Dev'n	0.9153	0.5129	0.1827	0.0431	0.0069
Simulation	0.8534	0.4919	0.1793	0.0431	0.0078
$n = 100$	33	36	39	42	45
Large Dev'n	0.9178	0.4678	0.1402	0.0260	0.0031
Simulation	0.8285	0.4374	0.1329	0.0249	0.0034

Table 7.4: Approximations to Distribution of 2 Dimensional Scan Statistic

The large deviation results are compared with simulation results in Table 7.4. We take $\Delta_1 = \Delta_2 = 0.5$, set $m = n\Delta_1\Delta_2(1+\epsilon)$ and show approximations to $P_0^{(n)}(\sup_t Y(t) > m)$. The approximations are performing reasonably in most cases, and slightly better for the large n.

7.2.2 Unknown Δ_1 and Δ_2

Suppose $0 \leq a_1 < a_2 \leq 1$ and $0 \leq b_1 < b_2 \leq 1$. We let $t = (a_1, a_2, b_1, b_2)$ and $X(t)$ be the number of events in $[a_1, a_2] \times [b_1, b_2]$. The likelihood ratio statistic for known a_1, a_2, b_1, b_2 is

$$l(t) = \begin{cases} X(t)\log\left(\frac{X(t)}{n\Delta_1\Delta_2}\right) + (n - X(t))\log\left(\frac{n-X(t)}{n(1-\Delta_1\Delta_2)}\right) & X(t) \geq n\Delta_1\Delta_2 \\ 0 & \text{otherwise} \end{cases}$$

where $\Delta_1 = a_2 - a_1$ and $\Delta_2 = b_2 - b_1$. We fix limits u_0 and u_1 such that $0 < u_0 < u_1 < 1$ and will reject $\mathcal{H}_0$ if

$$M_{u_0,u_1} \geq \frac{1}{2}c^2 \tag{7.43}$$

where

$$M_{u_0,u_1} = \sup_{u_0 \leq \Delta_1\Delta_2 \leq u_1} \sup_{\substack{0 \leq a_1 \leq 1-\Delta_1 \\ 0 \leq b_1 \leq 1-\Delta_2}} l(t).$$

The significance level is then

$$\alpha = P_0^{(n)}\left\{M_{u_0,u_1} \geq \frac{1}{2}c^2\right\}.$$

We may wish to place further restrictions on Δ_1 and Δ_2 to exclude long thin strips, for example by placing restrictions on the ratio Δ_1/Δ_2. This presents no real difficulty and we will discuss the necessary modification later. If we define $h(u)$ by

$$h(u)\log\left(\frac{h(u)}{u}\right) + (1 - h(u))\log\left(\frac{1-h(u)}{1-u}\right) = \frac{\eta^2}{2}$$

subject to $h(u) \geq u$, where $\eta = \frac{c}{\sqrt{n}}$, we can rewrite (7.43) as

$$\sup_t(X(t) - nh(\Delta_1\Delta_2)) \geq 0.$$

We order points $t = (a_1, a_2, b_1, b_2)$ by

$$t \leq t' \iff \begin{cases} a_1 < a_1' \\ a_1 = a_1', a_2 > a_2' \\ a_1 = a_1', a_2 = a_2', b_1 < b_1' \\ a_1 = a_1', a_2 = a_2', b_1 = b_1', b_2 > b_2' \end{cases} . \tag{7.44}$$

This ordering begins by considering $t = (0, 1, 0, 1)$ (i.e. the whole square) and decreasing b_2 until $b_1 = b_2$; increasing b_1 a small amount and resetting b_2 to 1, then repeating until $b_1 = 1$. Then a_2 is decreased, and the process repeated until $a_2 = a_1$. Then a_1 is increased, and the process repeated until $a_1 = 1$.

Fix m_1 and m_2 large and let $t = (\frac{i}{m_1}, \frac{i+j}{m_1}, \frac{k}{m_2}, b_2)$ for $i = 0, \ldots, m_1$, $j = 1, \ldots, m_1 - i$, $k = 1, \ldots, m_2$ and $\frac{k}{m_2} < b_2 \leq 1$ and

$$\tau_{m_1, m_2} = \inf \left\{ t : X(t) \geq nh(\frac{j(m_2 b_2 - k)}{m_1 m_2}) \right\} .$$

The ordering (7.44) is designed so that τ_{m_1,m_2} can be discretized in a manner similar to the one dimensional scan statistic with unknown Δ. We define u_l to be the solution of $nh(u) = l$, and choose l_0, l_1 such that $u_0 \leq u_j \leq u_1$ for $l_0 \leq l \leq l_1$. We then define

$$t_{i,j,k,l} = (\frac{i}{m_1}, \frac{i+j}{m_1}, \frac{k}{m_2}, \frac{k}{m_2} + \frac{m_1 u_l}{j})$$

for appropriate values of (i, j, k, l). We then have

$$\begin{aligned} & P_0^{(n)} \left(\tau_{m_1,m_2} = t_{i,j,k,l} \right) \\ = \;& P_0^{(n)} \left(\tau_{m_1,m_2} = t_{i,j,k,l} | X(t_{i,j,k,l}) = nh(u_l) \right) P_0^{(n)} \left(X(t_{i,j,k,l}) = nh(u_l) \right) . \end{aligned} \tag{7.45}$$

The second term on the right of (7.45) is a Binomial probability. The first term can be approximated using a local expansion. Provided there are no events in the corners defined in Lemma 7.2 we can write, for fixed $T > 0$ and $0 \leq \delta_j \leq \frac{T}{n}$,

$$X(a_1 - \delta_7, a_2 + \delta_5, b_1 + \delta_4, b_2 + \delta_1)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) + Z_5(\delta_5) - Z_4(\delta_4) + Z_1(\delta_1)$$

$$X(a_1 - \delta_7, a_2 + \delta_5, b_1 + \delta_4, b_2 - \delta_2)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) + Z_5(\delta_5) - Z_4(\delta_4) - Z_2(\delta_2)$$

$$X(a_1 - \delta_7, a_2 + \delta_5, b_1 - \delta_3, b_2 + \delta_1)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) + Z_5(\delta_5) + Z_3(\delta_3) + Z_1(\delta_1)$$

$$X(a_1 - \delta_7, a_2 + \delta_5, b_1 - \delta_3, b_2 - \delta_2)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) + Z_5(\delta_5) + Z_3(\delta_3) - Z_2(\delta_2)$$

$$X(a_1 - \delta_7, a_2 - \delta_6, b_1 + \delta_4, b_2 + \delta_1)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) - Z_6(\delta_6) - Z_4(\delta_4) + Z_1(\delta_1)$$

$$X(a_1 - \delta_7, a_2 - \delta_6, b_1 + \delta_4, b_2 - \delta_2)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) - Z_6(\delta_6) - Z_4(\delta_4) - Z_2(\delta_2)$$

$$X(a_1 - \delta_7, a_2 - \delta_6, b_1 - \delta_3, b_2 + \delta_1)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) - Z_6(\delta_6) + Z_3(\delta_3) + Z_1(\delta_1)$$

$$X(a_1 - \delta_7, a_2 - \delta_6, b_1 - \delta_3, b_2 - \delta_2)$$

$$= X(a_1, a_2, b_1, b_2) + Z_7(\delta_7) - Z_6(\delta_6) + Z_3(\delta_3) - Z_2(\delta_2).$$

This leads to the local conditions fo $\tau_{m_1,m_2} = t$ as

$$\begin{aligned}
Z_7(\delta_7) + Z_5(\delta_5) - Z_4(\delta_4) + Z_1(\delta_1) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 + \delta_5) - \Delta_1(\delta_4 - \delta_1)) \\
Z_7(\delta_7) + Z_5(\delta_5) - Z_4(\delta_4) - Z_2(\delta_4) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 + \delta_5) - \Delta_1(\delta_4 + \delta_2)) \\
Z_7(\delta_7) + Z_5(\delta_5) + Z_3(\delta_3) + Z_1(\delta_4) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 + \delta_5) + \Delta_1(\delta_3 + \delta_1)) \\
Z_7(\delta_7) + Z_5(\delta_5) + Z_3(\delta_3) - Z_2(\delta_2) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 + \delta_5) + \Delta_1(\delta_3 - \delta_2)) \\
Z_7(\delta_7) - Z_6(\delta_6) - Z_4(\delta_4) + Z_1(\delta_1) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 - \delta_6) - \Delta_1(\delta_4 - \delta_1)) \\
Z_7(\delta_7) - Z_6(\delta_6) - Z_4(\delta_4) - Z_2(\delta_2) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 - \delta_6) - \Delta_1(\delta_4 + \delta_2)) \\
Z_7(\delta_7) - Z_6(\delta_6) + Z_3(\delta_3) + Z_1(\delta_1) &\leq nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7 - \delta_6) + \Delta_1(\delta_3 + \delta_1))
\end{aligned}$$

$$Z_7(\delta_7) - Z_6(\delta_6) + Z_3(\delta_3) - Z_2(\delta_2) \;\le\; nh'(\Delta_1\Delta_2)(\Delta_2(\delta_7-\delta_6)+\Delta_1(\delta_3-\delta_2)).$$

for appropriate values of $\delta_1,\ldots,\delta_8$ arising from (7.44).

These conditions can be simplified to marginal conditions on $Z_1,\ldots,Z_7$. For example, if $\delta_1 > 0$ then $t' = (a_1, a_2, b_1, b_2 + \delta_1) < t$, and we get the condition $Z_1(u) \le nh'(\Delta_1\Delta_2)\Delta_1 u$ for all $u > 0$. If we take $\delta_2 > 0$, we must have either $\delta_3 > 0$, $\delta_5 > 0$ or $\delta_7 > 0$ to make $t' = (a_1 - \delta_7, a_2 + \delta_5, b_1 - \delta_3, b_2 - \delta_2) < t$. If we take $\delta_3 = \frac{1}{m_2}$ (its smallest positive value) and $\delta_5 = \delta_7 = 0$, we get the condition $Z_2(u) \ge nh'(\Delta_1\Delta_2)(\Delta_1 u - \Delta_1/m_2)$ and similar conditions if we take $\delta_5 > 0$ or $\delta_7 > 0$. Continuing in this manner gives the conditions

$$\begin{aligned}
Z_1(u) &\le k_1 u,\ \forall\, u > 0\\
Z_2(u) &\ge k_2 u - \min(c_2^3, c_2^5, c_2^7),\ \forall\, u > 0\\
Z_3(u) &\le k_3 u,\ u = \frac{l}{m_2}, l = 0,1,\ldots\\
Z_4(u) &\ge k_4 u - \min(c_4^5, c_4^7),\ u = \frac{l}{m_2}, l = 0,1,\ldots\\
Z_5(u) &\le k_5 u,\ u = \frac{l}{m_1}, l = 0,1,\ldots\\
Z_6(u) &\ge k_6 u - c_6^7,\ u = \frac{l}{m_1}, l = 0,1,\ldots\\
Z_7(u) &\le k_7 u,\ u = \frac{k}{m_1}, l = 0,1,\ldots
\end{aligned}$$

where

$$\begin{aligned}
k_1 = k_2 = k_3 = k_4 &= nh'(\Delta_1\Delta_2)\Delta_1\\
k_7 = k_6 = k_7 &= nh'(\Delta_1\Delta_2)\Delta_2\\
c_2^3 &= nh'(\Delta_1\Delta_2)\frac{\Delta_1}{m_2}\\
c_2^5 = c_4^5 = c_2^7 = c_4^7 = c_6^7 &= nh'(\Delta_1\Delta_2)\frac{\Delta_2}{m_1}.
\end{aligned}$$

If we take m_2 larger than m_1 so $c_2^3 < c_2^5 = c_2^7$, we use Theorem 2.1 and Lemma 2.1 to obtain

$$\begin{aligned}
&P_0^{(n)}(\tau_{m_1,m_2} = t | X(t) = nh(\Delta_1\Delta_2)) \\
&\approx \left(1 - \frac{1-h(\Delta_1\Delta_2)}{h'(\Delta_1\Delta_2)(1-\Delta_1\Delta_2)}\right)^4 \left(1 - \exp(\frac{1}{m_2}(\lambda_2' - \lambda_2))\right) \\
&\quad \times \left(1 - \exp(\frac{\Delta_2}{m_2\Delta_1}(\lambda_4' - \lambda_4))\right) \left(1 - \exp(\frac{1}{m_1}(\lambda_6' - \lambda_6))\right) \\
&\approx \left(1 - \frac{1-h(\Delta_1\Delta_2)}{h'(\Delta_1\Delta_2)(1-\Delta_1\Delta_2)}\right)^4 \frac{n^3\Delta_2^2\Delta_1}{m_2 m_1^2} \left(\frac{1-h(\Delta_1\Delta_2)}{1-\Delta_1\Delta_2} - \frac{h(\Delta_1\Delta_2)}{\Delta_1\Delta_2}\right)^3
\end{aligned}$$

where the conjugate values λ_2', λ_4' and λ_6' are defined using (2.9). This gives

$$\begin{aligned}
&P_0^{(n)}\left\{M_{u_0,u_1} \geq \frac{1}{2}c^2\right\} \\
&\approx \sum_{\Delta_1}\sum_{\Delta_2}\sum_{i=0}^{m_1(1-\Delta_1)}\sum_{k=0}^{m_2(1-\Delta_2)} \left(1 - \frac{1-h(\Delta_1\Delta_2)}{h'(\Delta_1\Delta_2)(1-\Delta_1\Delta_2)}\right)^4 \frac{n^3\Delta_2^2\Delta_1}{m_2 m_1^2} \\
&\quad \times \left(\frac{1-h(\Delta_1\Delta_2)}{1-\Delta_1\Delta_2} - \frac{h(\Delta_1\Delta_2)}{\Delta_1\Delta_2}\right)^3 \frac{e^{-\frac{1}{2}c^2}}{\sqrt{2\pi n h(\Delta_1\Delta_2)(1-h(\Delta_1\Delta_2))}} \\
&\approx \sum_{\Delta_1}\sum_{\Delta_2} \left(1 - \frac{1-h(\Delta_1\Delta_2)}{h'(\Delta_1\Delta_2)(1-\Delta_1\Delta_2)}\right)^4 \frac{n^3(1-\Delta_1)(1-\Delta_2)\Delta_2^2\Delta_1}{m_1} \\
&\quad \times \left(\frac{1-h(\Delta_1\Delta_2)}{1-\Delta_1\Delta_2} - \frac{h(\Delta_1\Delta_2)}{\Delta_1\Delta_2}\right)^3 \frac{e^{-\frac{1}{2}c^2}}{\sqrt{2\pi n h(\Delta_1\Delta_2)(1-h(\Delta_1\Delta_2))}}
\end{aligned} \tag{7.46}$$

where the sum is over $\Delta_1 = 0, \frac{1}{m_1}, \frac{2}{m_1}, \ldots, 1$ and $u_0 \leq \Delta_1\Delta_2 \leq u_1$ with $nh(\Delta_1\Delta_2)$ an integer. Approximating the sums in (7.46) by integrals with differentials $d\Delta_2 = 1/(n\Delta_1 h'(\Delta_1\Delta_2))$ and $d\Delta_1 = 1/m_1$ gives

$$\begin{aligned}
&P_0^{(n)}\left\{M_{u_0,u_1} \geq \frac{1}{2}c^2\right\} \\
&\approx c^7\phi(c)\int_{\Delta_1}\int_{\Delta_2} \frac{\Delta_2^2\Delta_1^2(1-\Delta_1)(1-\Delta_2)}{\eta^7 h'(\Delta_1\Delta_2)^3}\left(h'(\Delta_1\Delta_2) - \frac{1-h(\Delta_1\Delta_2)}{1-\Delta_1\Delta_2}\right)^4 \\
&\quad \times \left(\frac{1-h(\Delta_1\Delta_2)}{1-\Delta_1\Delta_2} - \frac{h(\Delta_1\Delta_2)}{\Delta_1\Delta_2}\right)^3 \frac{d\Delta_2 d\Delta_1}{\sqrt{h(\Delta_1\Delta_2)(1-h(\Delta_1\Delta_2))}}.
\end{aligned} \tag{7.47}$$

Making the change of variables $u = \Delta_1\Delta_2$ and $v = \Delta_1$, (7.47) becomes

$$\begin{aligned} & P_0^{(n)}(M_{u_0,u_1} \geq \frac{1}{2}c^2) \\ \approx \; & c^7\phi(c) \int_{u_0}^{u_1} \int_u^1 \frac{u^2}{\eta^7 h'(u)^3} \left(h'(u) - \frac{1-h(u)}{1-u} \right)^4 \left(\frac{1-h(u)}{1-u} - \frac{h(u)}{u} \right)^3 \\ & \times \frac{|J|(1-\Delta_1)(1-\Delta_2)}{\sqrt{h(u)(1-h(u))}} dv du \end{aligned} \tag{7.48}$$

where J is the Jacobian of the transformation and

$$\begin{aligned} |J|(1-\Delta_1)(1-\Delta_2) &= \begin{vmatrix} \Delta_2 & \Delta_1 \\ 1 & 0 \end{vmatrix}^{-1} (1-\Delta_1)(1-\Delta_2) \\ &= \frac{1}{v}(1-v)(1-\frac{u}{v}) \\ &= \frac{1+u}{v} - 1 - \frac{u}{v^2} \end{aligned}$$

and therefore

$$\int_u^1 |J|(1-\Delta_1)(1-\Delta_2) dv = -(1+u)\log(u) - 2(1-u) \tag{7.49}$$

which reduces (7.48) to a one dimensional integral. If, as suggested earlier, we place additional restrictions on the values Δ_1 and Δ_2 may take, we must adjust the range of integration in (7.49) appropriately. If we let $\eta \to 0$ and expand $h(u)$ as in (7.27) we get the Gaussian approximation

$$\frac{c^7\phi(c)}{16} \int_{u_0}^{u_1} \frac{u^2(-(1+u)\log(u) - 2(1-u))}{(u(1-u))^4} du.$$

The highest order boundary correction occurs at $u = u_1$. To evaluate the boundary correction, we need approximations to the maximum number of points in a rectangle of fixed area. The corresponding random field cannot be decomposed into the sums of asymptotically independent processes and so our methods cannot handle this problem. Moreover, there are other terms of the same order as these boundary effects and the numerical results below suggest the approximation is tending to overestimate the true probability and adding boundary effects without other higher order terms will not improve the approximation.

If instead we placed limits on Δ_1 and Δ_2 individually, boundary corrections would be correspond to the case Δ_1 known and Δ_2 unknown. This leads to a boundary crossing problem for a three dimensional random field which can be decomposed as independent one dimensional fields so our methods should yield results; however we will not treat this case here. Note also that the range of integration in (7.48) will be more complicated in this case.

We compare the approximation (7.48) with simulation results, for $u_0 = 1 - u_1 = 0.1$ in Table 7.5 and for $u_0 = 1 - u_1 = 0.2$ in Table 7.6. In most cases the performance of the large deviation approximation in the tails is adequate. However, the Gaussian approximation is extremely poor, giving about twice the true probability even when $n = 100$.

		c				
		3.0	3.5	4.0	4.5	5.0
	Gaussian	4.6660	2.7030	1.0556	0.2875	0.0559
$n = 20$	Large Dev'n	1.3157	0.6485	0.2168	0.0509	0.0085
	Simulation	0.8387	0.4845	0.1822	0.0465	0.0074
$n = 50$	Large Dev'n	1.9071	1.0059	0.3499	0.0854	0.0149
	Simulation	0.8846	0.5839	0.2315	0.0649	0.0103
$n = 100$	Large Dev'n	2.3138	1.2317	0.4459	0.1141	0.0206
	Simulation	0.9208	0.6402	0.2802	0.0801	0.0167
Simulation s.e.		0.004	0.005	0.004	0.002	0.001

Table 7.5: Tail probabilities for the 2-Dimensional Scan Statistic: 10% Truncation

		c				
		3.0	3.5	4.0	4.5	5.0
	Gaussian	1.6124	0.9340	0.3648	0.0994	0.0193
$n = 20$	Large Dev'n	0.7305	0.3621	0.1210	0.0282	0.0047
	Simulation	0.6614	0.3289	0.1189	0.0262	0.0057
$n = 50$	Large Dev'n	0.8584	0.4632	0.1705	0.0446	0.0083
	Simulation	0.7032	0.3921	0.1489	0.0435	0.0102
$n = 100$	Large Dev'n	0.9946	0.5412	0.1994	0.0515	0.0095
	Simulation	0.7498	0.4118	0.1616	0.0418	0.0087
Simulation s.e.		0.004	0.005	0.004	0.002	0.001

Table 7.6: Tail probabilities for the 2-Dimensional Scan Statistic: 20% Truncation

Chapter 8

Discussions and Conclusions

The change point methodology we have discussed provides a way of analyzing the relationship between a response variable (in our case, a Poisson process) and a covariate (in our case, time). Many other methods of analyzing such relationships are available. For example, log-linear models (cf. McCullagh and Nelder (1983)) model the rate as

$$\log(\lambda(t)) = \beta' f(t) \tag{8.1}$$

where $f(t)$ is a vector of known functions of t and β is a vector of unknown constants. Another method, discussed by Tibshirani and Hastie (1987), does not assume a parametric form for $\lambda(t)$ but instead assumes only that $\lambda(t)$ is a smooth function of t and uses local estimation. Inference about (8.1) can be made using standard likelihood theory, and similar techniques can be used to make approximate inferences for local likelihood fits. By contrast, the introduction of a change point leads to complicated inference problems. The natural question is then 'Why use the change point models?'.

The first and most obvious answer is that in some situations the change point model may be the correct one. Of course, in practice, when a statistician is presented with a set of data, he does not know the correct model in advance but must fit several models and select the best. In this sense, the change point methodology provides an addition to the collection of possible models. In some cases, the change point model may perform better than other methods. For example, with the British coal mining accident data, Raftery and Akman (1986) have concluded, using Bayesian methods, that the constant parameter model with a change point (1.1) provides a much better fit than the linear model given by (8.1) with $f(t) = (1, t)'$ fitted by Cox and Lewis (1966) and other authors. However, the general topic of choosing between non-nested models is an area which needs further study.

Another important area in which the change point models may have advantage over other models is interpretability. To illustrate this point, we consider Lucas' industrial accident data. Suppose at the end of the time period (1970,1980), it was noticed that there had been a substantial drop in the accident rate. In an industrial environment, there will have been many procedural changes over this period, many of which might have contributed to safety. In the interests of further reductions in the accident rate, we may be interested in knowing which changes contributed most to the reduction. Fitting a model which results in a smooth estimate of $\lambda(t)$ is unlikely to help answer this sort of question; however the change point models are designed to answer precisely this type of question.

Our discussion of inference in change point methods is centered around the likelihood function. Alternative methods for deriving tests and confidence regions have been proposed. Several tests amount to selecting different boundaries for $X(t)$ or the process under consideration in a particular problem. The choice between tests seems somewhat arbitrary. Different boundaries will trade off power against various alternatives; for example a test with linear boundaries will have more power to detect changes in the middle of $[0, T]$ and less power to detect changes near the end points. Classical optimality results seem an unrealistic goal in this setting.

We have not discussed estimation in any detail. The use of point estimates such as (1.5) and (1.6) with an estimate of τ has been shown to be efficient for fairly general change point problems. However, since the estimate $\hat{\tau}$ is in some sense seeking to maximize the difference between $\hat{\lambda}_0$ and $\hat{\lambda}_1$, some form of shrinkage estimation may be plausible. This question does not explicitly appear in any literature, but presumably Bayes estimates will have some shrinkage.

Less is known about the efficiency of the maximum likelihood estimate $\hat{\tau}$. Asymptotic results for various change point problems are given by Chernoff and Rubin (1955) and Hinkley (1970). For detecting a change in the drift of Brownian motion, Ibragimov and Khasminski (1981) show the maximum likelihood estimate is less efficient than a Bayes estimate. For our constant parameter model, the Bayesian approach has been studied by Raftery and Akman (1986).

Several methods have been proposed for confidence region problems. For example, we could define

$$I_1' = \{t : |\hat{\tau} - t| < c\} \tag{8.2}$$

where c is chosen so that

$$P_\tau^{(m,n)}(|\hat{\tau} - \tau| < c) = 1 - \alpha.$$

In the case of normally distributed random variables, Siegmund (1988a) shows for large n the interval (8.2) is on average twice as large as the interval (4.1). The inefficiency of (8.2) as a confidence region may be related to the inefficiency of $\hat{\tau}$ as a point estimate of τ.

A method used by Akman and Raftery (1986) is to test separately for a change point in $[0, t]$ and $[t, T]$, and include t in the confidence region if both tests accept the null hypothesis at level $1 - \sqrt{1-\alpha}$. However, this method seems very susceptible to model specification: If there is a large change but also some additional structure that cannot be explained by a change point model, we may get an empty confidence set. This suggests we must begin with a model which accounts for all the structure in the data. A more common data analytic practice is to look for the most significant structure in a dataset, remove this structure and continue looking for further structure. In addition, we are testing for change-points near the end points of an interval, and therefore truncation and specification of the boundaries may have a substantial effect on the confidence region.

Siegmund (1988a) suggests some other methods which have asymptotic performance similar to that of the likelihood ratio method. However, one big advantage for the likelihood ratio method is the ease with which it extends to find joint confidence sets for τ and other parameters. We are not aware of any other methods which have been used to find such joint confidence regions.

8.1 Some Other Statistical Applications

The change point methods may be applicable to some statistical problems that we have not yet discussed. We suggest some potential areas in this section. Although we formulate the problems in terms of Poisson processes, similar ideas may be applicable for for other problems, such as sequences of *i.i.d.* random variables. We do not try to be completely rigorous, but leave the details as an open research problem.

8.1.1 Local Fitting

The question of choosing between a change point model and a smooth (either parametric or non-parametric) model needs additional attention. The most intuitive approach is to fit a model incorporating both possibilities, and then assessing the contribution of each component. We suppose the rate function $\lambda(t)$ can be written

$$\log(\lambda(t)) = a(t) + H(t) \tag{8.3}$$

where $a(t)$ is smooth and $H(t)$ is a step function. Most modeling takes the form (8.3) with $H(t) = 0$. By contrast, our constant parameter model (1.1) is of the form (8.3) with $a(t) = 0$ and the log-linear model (1.2) has $a(t) = bt$.

If the points at which $H(t)$ is discontinuous are known, then local likelihood procedures may be used to estimate $a(t)$. By fitting both with a change at $t = \tau$ and without a change, we can form a local likelihood ratio process to assess at what t a smooth fit is inadequate.

We will use local likelihood estimates of the smooth part of $\lambda(t)$. Suppose we fix $h > 0$ and consider a window $[t - h, t + h]$ around a point t. We may assume the rate is locally constant round t and the local likelihood estimate of $\lambda(t)$ is

$$\hat{\lambda}(t) = \frac{X(t+h) - X(t-h)}{2h}. \tag{8.4}$$

If $t - h < 0$ or $t + h > T$ we must make small modifications to (8.4).

Other methods may be used to define the local windows; for example, we may choose h, dependent on the data, to include a fixed number of points. Also, we may wish to use some form of weighted estimate in place of (8.4), where the density of events close to t is more important than the density near the end-points of the interval.

Now suppose we assume there is a change point at time t. We can estimate rates on each side of t by

$$\hat{\lambda}_0(t) = \frac{X(t) - X(t-h)}{h}$$

and

$$\hat{\lambda}_1(t) = \frac{X(t+h) - X(t)}{h}.$$

We can define a local likelihood ratio process to be

$$\begin{aligned} l(t) &= (X(t) - X(t-h)) \log\left(\frac{2(X(t) - X(t-h))}{X(t+h) - X(t-h)}\right) \\ &\quad + (X(t+h) - X(t)) \log\left(\frac{2(X(t+h) - X(t))}{X(t+h) - X(t-h)}\right) \end{aligned} \tag{8.5}$$

for $h \leq t \leq T - h$ and a slightly modified form for other t. Values of t for which $l(t)$ is large may be considered possible change points.

Inference based on (8.5) seems difficult. The methods developed in Chapter 3 do not work for this process. If $\lambda(t)$ is constant, we would expect the distribution of $\sup_{\tau_0 \leq t \leq \tau_1} l(t)$ to be stochastically larger than the distribution of (1.9). Moreover, smooth variation of $\lambda(t)$ will inflate $l(t)$.

The choice of h needs some study. In more common smoothing problems, varying the smoothing parameter may be regarded as a trade off between variance and bias. However, for the change point problems, there are also power conditions and indeed even the definition of what counts as a change will also influence the selection.

8.1.2 Recursive Partitioning

Regression Trees (Breiman, Friedman, Olshen and Stone (1984)) are a method used to analyse data where each point consists of a single response variable and many independent covariates. The covariate domain is partitioned into regions and the mean response is taken to be a constant over each of the regions. There is an obvious analogy with change point problems: selection of a partition is equivalent to selecting a point at which the mean response level changes.

In the usual setting, choice of where to split is made by assigning a 'value' to each possible split and finding a split to maximize this value. One possible criterion is the reduction in residual sum of squares. A decision must be made on the basis of the maximum value whether to stop the algorithm or to make the split and continue with further iterations. This question is related to the significance of the split; however, in practice the choice of stopping rule is somewhat arbitrary.

For the special case of a Poisson process observed on a d-dimensional unit hypercube, we can make use of our change point methods to assess the significance of splits. Suppose we observe a total of n events and let the event times be $T_1, \ldots, T_n$ with $T_j = (T_{j1}, \ldots, T_{jd})$. Define $X_i(t) = \sum_{j=1}^{n} I(T_{ji} \leq t)$. Then $X_i(t), i = 1, \ldots, d$ are d independent 1 dimensional Poisson processes, conditional on $X_i(1) = n$ for all i. We can then make a split in any one of the d dimensions. We place limits $0 < \tau_0 \leq t \leq \tau_1$ on the location of a split. We then choose a split to maximize

$$l_i(t) = X_i(t) \log\left(\frac{X_i(t)}{nt}\right) + (n - X_i(t)) \log\left(\frac{n - X_i(t)}{n(1-t)}\right)$$

over the range $1 \leq i \leq d$ and $\tau_0 \leq t \leq \tau_1$. We will carry out the split if and only if

$$\max_{1\leq i\leq d} \sup_{\tau_0 \leq t \leq \tau_1} l_i(t) \geq \frac{1}{2}c^2 \tag{8.6}$$

for some $c > 0$. We are then interested in the significance of a split, which is

$$\alpha = P_0^{(n)}\left(\max_{1\leq i\leq d} \sup_{\tau_0 \leq t \leq \tau_1} l_i(t) \geq \frac{1}{2}c^2\right). \tag{8.7}$$

Since the projections $X_i(t)$ are independent under $P_0^{(n)}$, we can write (8.7) as

$$1 - \alpha \quad = \quad \prod_{i=1}^{d} P_0^{(n)}\left(\sup_{\tau_0 \leq t \leq \tau_1} l_i(t) \leq \frac{1}{2}c^2\right)$$

$$
\begin{aligned}
&= \left(P_0^{(n)} \left(\sup_{\tau_0 \le t \le \tau_1} l_1(t) \ge \frac{1}{2} c^2 \right) \right)^d \\
&\approx 1 - dP_0^{(n)} \left(\sup_{\tau_0 \le t \le \tau_1} l_1(t) \ge \frac{1}{2} c^2 \right) \qquad (8.8)
\end{aligned}
$$

if c is sufficiently large. We can approximate (8.8) using (3.6) and (3.7).

After selecting a split, we will then want to continue looking for further splits. The obvious (although not strictly valid) approach is to then treat each segment as an independent Poisson process on a hypercube and search within each region for the best split. The maximization (8.6) is then taken over all regions and the significance level approximation (8.8) is then adjusted to take the multiple regions into account.

An open question is how to modify (8.8) when the process is observed on a non-rectangular domain $\mathcal{V}$. If we let $\mathcal{A}_i(t)$ be the cross sectional area when we fix $t_i = t$, then under the null hypothesis of homogeneity, we would expect $X_i(t)$ to have rate proportional to $\mathcal{A}_i(t)$. This can then be simply transformed to a homogeneous Poisson process. We can then assess the significance of the marginal likelihood ratio processes as before. However, it is not clear whether or in what circumstances adding marginal tail probabilities as in (8.8) will be a good approximation.

8.1.3 Permutation Tests

The Poisson process assumes that the times between events have an exponential distribution. In some situations we may not want to assume this, but instead assume the event times come from a renewal process. Let $Y_i = T_{i+1} - T_i$ for $i = 1, \ldots, n-1$ and F_i be the (unknown) distribution of Y_i given T_i. Our model is

$$
F_i = \begin{cases} F & T_i \le \tau \\ G & T_i \ge \tau \end{cases}. \qquad (8.9)
$$

The model isn't completely specified by (8.9) since for some (random) i we will have $T_i \leq \tau < T_{i+1}$. To be completely rigorous, (8.9) would be specified in terms of hazard rates depending on the time since the last event.

The null hypothesis is $\mathcal{H}_0 : F = G$. We would like to use the likelihood ratio statistic to test $\mathcal{H}_0$. When $X(t)$ is a Poisson process, F is an exponential distribution which has coefficient of variation equal to 1. If the inter-event times have a coefficient of variation less than 1, we would expect the likelihood ratio statistic to be less variable than for a Poisson process. Likewise, if F has a coefficient of variation greater than 1, we expect the likelihood ratio statistic to be more variable. We cannot attempt to find the distribution of the maximum of $l(t)$, since it depends upon the unknown F.

Under $\mathcal{H}_0$, the $n-1$ increments Y_i are exchangeable and we can use a permutation test to approximate the distribution of the likelihood ratio statistic. Note that conditional on $X(T) = n$, the differences $Y_1, \ldots, Y_{n-1}$ are not independent. For each of the $(n-1)!$ permutations of $Y_1, \ldots, Y_n$ we calculate the supremum of $l(t)$. If the observed supremum is among the top 5%, we will reject $\mathcal{H}_0$. If n is even moderately large, it will be unfeasible to calculate the maximum of the likelihood ratio statistic for all $(n-1)!$ permutations and it will be necessary to use a Monte Carlo approximation.

In the context of *i.i.d.* observations, bootstrap and randomization tests have been proposed by Romano (1989). However, the performance of non-parametric tests; for example, their efficiency relative to parametric methods, has not been studied.

A permutation method can also be used to find confidence regions for the change point. Suppose $\tau = T_j$. We then want to approximate $P_\tau(\sup_t l(t) \geq l(\tau) + c)$. We can split into the sets $t < \tau$ and $t > \tau$ as before, and then use permutations of the interevent times on each side of τ to estimate $P_\tau(\sup_{t<\tau} l(t) \geq l(\tau) + c)$

and $P_\tau(\sup_{t>\tau} l(t) \geq l(\tau) + c)$. Usually we will have to use random sampling of permutations. Since for each τ our main interest is whether or not to include τ in the confidence region, we should use a sequential stopping rule that will enable us to decide quickly for most τ and use more permutations for a few marginal values of τ.

8.2 Conclusions

We have discussed inference about change points in Poisson processes for two different models. The tests and confidence intervals considered give rise to non-linear boundary crossing problems for counting processes. Various methods to approximate these probabilities have been discussed and we have illustrated the trade-off between accuracy and simplicity of the approximations.

The constant parameter model has a parameter space that is a subset of that for the log-linear model and therefore the loglinear model may be preferable on the grounds of being more general and guarding against detecting change points where in fact the rate is changing smoothly over time. However, there is a price to pay: The likelihood ratio test for the log-linear model is much less powerful than the test for the constant parameter model, and therefore important structure may be lost by fitting this model unnecessarily. On the other hand, structure may be found in data that is hidden by the constant parameter model. For the British coal mining data, fitting the log-linear model raised the possibility of a second change point which was completely hidden in the constant parameter model.

Our methodology extends to other problems. For example, we have discussed Kolmogorov-Smirnov testing in the presence of unknown nuisance parameters and obtained approximations to significance levels that are much more accurate than existing methods.

In medical trials, there may be an initial period following treatment during which patients are at a considerable risk of side effects. Fitting a change point model to survival times as in Chapter 5 will be helpful in detecting such initial periods of high risk. By contrast, fitting more commonly used survival distributions will not help detect such periods.

In Chapter 7 we discussed the approximation of various scan statistics. The large deviation methods we used give very good approximations to the tail of the distributions. The results here appear to be new; the only case for which exact results are known is the one dimensional scan statistic on the line with fixed width. Even in this case, our result is much simpler to evaluate than the exact formula and gives very good results.

Appendix A

More on Boundary Crossing Probabilities

This Appendix will be concerned with further development of the boundary crossing approximations used in this thesis, and will also introduce some related methods. The results will be applied to make rigorous some of the results derived in Chapters 3 and 4. We begin by studying the tangent approximation (2.24) in Section A.1, establishing that under fairly general conditions this provides a first order asymptotic approximation and indicating how to obtain the second order corrections. Section A.2 studies the local boundary crossing approximations used in the power and confidence region calculations. In Section A.3 we will discuss some extensions to locally Poisson process. Section A.4 will discuss an alternative approach to boundary crossing probabilities and Section A.5 discusses upper boundary approximations.

A.1 The Tangent Approximation

Our object is to show that under fairly general circumstances, the tangent approximation (2.24) provides a first order asymptotic approximation to boundary crossing probabilities. We begin with some preliminary lemmas. The main result is presented in Theorem A.1 below.

Lemma A.1 *Let $X(t)$ be a Poisson process on $[0,T]$ with $X(T) = n$. Let c be a constant with $1 - cT < 0$. Then*

$$P^{(n)}\left(\exists t < T \ :\ X(t) - n(1 - c(T - t)) \leq 0\right) = \frac{1}{cT}. \tag{A.1}$$

Proof: This result is due to Daniels (1945). The present proof was suggested by D. Siegmund. Let

$$Y(t) = \frac{X(t) - \frac{nt}{T}}{T - t}. \tag{A.2}$$

Then if $0 \leq s < t < T$,

$$\begin{aligned} E^{(n)}(Y(t)|Y(s)) &= \frac{1}{T-t}\left(X(s) + \frac{t-s}{T-s}(n - X(s)) - \frac{nt}{T}\right) \\ &= \frac{1}{T-t}\left(X(s)\frac{T-t}{T-s} - \frac{n}{T(T-s)}(t(T-s) - T(t-s))\right) \\ &= Y(s). \end{aligned}$$

Hence $Y(t)$ is a martingale under $P^{(n)}$, with mean $E(Y(t)) = 0$. Let

$$\mathcal{T} = \inf\{t : X(t) - n(1 - c(T - t)) \leq 0\}$$

with $\inf\{\phi\} = T$. Now choose t sufficiently large so that $n(1-c(T-t)) > n-1$. We note that on the set $\{\mathcal{T} > t\}$, $X(t) > n(1-c(T-t)) > n-1$ and therefore $X(t) = n$.

Also, on $\{\mathcal{T} \le t\}$, $X(\mathcal{T}) = n(1 - c(T - \mathcal{T}))$. Using the martingale stopping theorem (Shiryayev (1984) p459), $E^{(n)}(Y(\mathcal{T} \wedge t)) = 0$. Therefore,

$$\begin{aligned} 0 &= E^{(n)}(Y(\mathcal{T} \wedge t); \mathcal{T} \le t) + E^{(n)}(Y(\mathcal{T} \wedge t); \mathcal{T} > t) \\ &= \frac{n(1 - c(T-t) - \frac{t}{T})}{T - t} P^{(n)}(\mathcal{T} \le t) + \frac{n}{T}(1 - P^{(n)}(\mathcal{T} \le t)) \\ &= -ncP^{(n)}(\mathcal{T} \le t) + \frac{n}{T} \end{aligned} \tag{A.3}$$

Solving (A.3) for $P^{(n)}(\mathcal{T} \le t)$ and letting $t \to T$ gives (A.1).

□

Using a time reversal argument, we can rewrite (A.1) as

$$P^{(n)}\left(\exists t < T \, : \, X(t) - nct > 0\right) = \frac{\hat{\lambda}}{nc}$$

where $\hat{\lambda} = \frac{n}{T}$. Thus Lemma A.1 can be regarded as a conditional form of Lemma 2.1.

Lemma A.2 *Let $a(t)$ be defined for $t_0 \le t \le T$, with $a(t_0) = 0$, $a(T) = 1$, $a'(t_0) > 0$ and $a''(t) > 0$. Choose ϵ sufficiently small that*

$$Ta(T - \epsilon) + t_0 - T + \epsilon \ge 0. \tag{A.4}$$

Then for all $n \ge 1$,

$$\begin{aligned} \frac{1}{Ta'(T)} &\le P^{(n)}\left(\exists t < T \, : \, X(t) - na(t) < 0\right) \\ &\le \frac{\epsilon}{T(1 - a(T - \epsilon))} + Ce^{-2n(\frac{\epsilon}{T} - 1 + a(T-\epsilon))^2} &\text{(A.5)} \\ &\le \frac{1}{Ta'(T)}\left(1 + \frac{\epsilon M}{a'(T)}\right) + Ce^{-2n(\frac{\epsilon}{T} - 1 + a(T-\epsilon))^2}. &\text{(A.6)} \end{aligned}$$

where $M = \sup_{0<t<T} a''(t)$. We require the additional conditions $M < \infty$ and $\epsilon M < a'(T)$ for (A.6). Letting $n \to \infty$ then $\epsilon \to 0$ in (A.5) gives

$$\lim_{n\to\infty} P^{(n)}\left(\exists t < T \, : \, X(t) - na(t) < 0\right) = \frac{1}{Ta'(T)}. \tag{A.7}$$

Proof: Since $a''(t) > 0$, we have $a(t) \geq 1 - (T-t)a'(T)$ for all t. Therefore by Lemma A.1,

$$P^{(n)}\left(\exists t < T \,:\, X(t) - na(t) < 0\right) \geq \frac{1}{Ta'(T)}. \tag{A.8}$$

Since $a(t)$ is convex, we have for $T - \epsilon \leq t \leq T$,

$$\begin{aligned} a(t) &\leq \frac{T-t}{\epsilon}a(T-\epsilon) + \frac{t-T+\epsilon}{\epsilon}a(T) \\ &= 1 - (T-t)\frac{1-a(T-\epsilon)}{\epsilon} \end{aligned}$$

and for $t_0 \leq t \leq T - \epsilon$,

$$\begin{aligned} a(t) &\leq \frac{T-\epsilon-t}{T-\epsilon-t_0}a(t_0) + \frac{t-t_0}{T-\epsilon-t_0}a(T-\epsilon) \\ &= a(T-\epsilon) - \frac{T-\epsilon-t}{T-\epsilon-t_0}a(T-\epsilon) \qquad \text{(A.9)} \\ &\leq \frac{t}{T} - \frac{T-\epsilon}{T} + a(T-\epsilon). \qquad \text{(A.10)} \end{aligned}$$

Here, (A.10) follows from (A.9) by (A.4). Therefore,

$$\begin{aligned} &P^{(n)}\left(\exists t < T \,:\, X(t) - na(t) < 0\right) \\ &\leq P^{(n)}\left(\exists t < T \,:\, X(t) - n(T-t)\frac{1-a(T-\epsilon)}{\epsilon} < 0\right) \\ &\quad + P^{(n)}\left(\exists t < T \,:\, X(t) - \frac{t}{T} < a(T-\epsilon) - \frac{T-\epsilon}{T}\right) \\ &\leq \frac{\epsilon}{T(1-a(T-\epsilon))} + Ce^{-2n(\frac{\epsilon}{T}-1+a(T-\epsilon))^2} \qquad \text{(A.11)} \end{aligned}$$

using Lemma A.1 and a uniform bound for Kolmogorov-Smirnov statistics. Hu (1985) gives the explicit bound $C = 2\sqrt{2}$. Combining (A.8) and (A.11) gives (A.5). A straightforward Taylor series expansion gives (A.6).

□

Remark: The condition $a''(t) > 0$ simplifies the proof of Lemma A.2. The limiting relation (A.7) will hold when this condition is weakened to $Ta'(T) > 1$. We will not establish this rigorously here; see Lemma A.4 below for a more detailed approach.

Lemma A.3 *(See Feller (1968) p151) Suppose $X \sim \mathcal{B}(n,p)$ and $m > np$. Then*

$$P(X \geq m) \leq \frac{m(1-p)}{m-np} P(X = m). \tag{A.12}$$

Proof: If $n \geq j > m$, then

$$\begin{aligned} P(X = j) &= \frac{n-j+1}{j}\frac{p}{1-p}P(X = j-1) \\ &\leq \frac{n-m}{m}\frac{p}{1-p}P(X = j-1) \end{aligned}$$

and by induction,

$$P(X = j) \leq \left(\frac{(n-m)p}{m(1-p)}\right)^{j-m} P(X = m)$$

for $j \geq m$. Hence

$$\begin{aligned} P(X \geq m) &= \sum_{j=m}^{n} P(X = j) \\ &\leq P(X = m) \sum_{j=m}^{\infty} \left(\frac{(n-m)p}{m(1-p)}\right)^{j-m} \\ &= \frac{1}{1 - \frac{(n-m)p}{m(1-p)}} P(X = m) \end{aligned}$$

which gives (A.12).

□

Theorem A.1 *Suppose the boundary $a(t)$ is continuous, differentiable and satisfies:*

1. *$a(t) < \frac{t}{T}$ for all t.*

2. *$ta'(t) - a(t) > 0$ for all $t > 0$.*

3. For some $t_0 > 0$, $a(t_0) = 0$.

. Let $\mathcal{T} = \inf\{t : X(t) < na(t)\}$. We define

$$l(t) = a(t)\log\left(\frac{Ta(t)}{t}\right) + (1 - a(t))\log\left(\frac{T(1 - a(t))}{T - t}\right).$$

Fix τ_0 and τ_1 such that $t_0 \le \tau_0 < \tau_1 \le T$. Let t^* be a value of t that minimizes $l(t)$ over the interval $[\tau_0, \tau_1]$. Suppose also either $l(t) > l(t^*)$ or $l'(t) = 0$ for $t = \tau_0$ and $t = \tau_1$. Then

$$P^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1) = I_n(\tau_0, \tau_1)(1 + o(1)) \tag{A.13}$$

as $n \to \infty$, where

$$I_n(\tau_0, \tau_1) = \sqrt{n}\int_{\tau_0}^{\tau_1}\left(a'(t) - \frac{a(t)}{t}\right)\frac{1}{\sqrt{2\pi a(t)(1 - a(t))}}e^{-nl(t)}dt \tag{A.14}$$

and under $P_0^{(n)}$, $X(t)$ is a homogeneous Poisson process conditioned on $X(T) = n$.

Proof: We follow the informal derivation of (2.24) and use notation from there. We concentrate first on the case $l(t) = \frac{1}{2}\eta^2$ is constant, $\tau_0 > t_0$ and $\tau_1 < T$. Using Lemma A.2 and the condition $ta'(t) - a(t) > 0$, we get

$$P_0^{(n)}(\mathcal{T} = t_j | X(t_j) = j) = \left(1 - \frac{a(t_j)}{t_j a'(t_j)}\right)(1 + o(1)) \tag{A.15}$$

and by Stirling's formula,

$$\binom{n}{na(t_j)} = \frac{1}{\sqrt{2\pi n a(t_j)(1 - a(t_j))}}\frac{1}{a(t_j)^{na(t_j)}(1 - a(t_j))^{n(1-a(t_j))}}(1 + o(1)). \tag{A.16}$$

Moreover, (A.15) and (A.16) hold uniformly in t_j for $\tau_0 \le t_j \le \tau_1$. uniformly. Fix $\epsilon > 0$. For n sufficiently large, multiplying (A.15) and (A.16) and applying the summation formula (2.21) gives

$$\begin{aligned} &P_0^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1) \\ &\ge (1 - \epsilon)\sqrt{n}e^{-\frac{1}{2}c^2}\sum_{j=j_0}^{j_1}\left(1 - \frac{a(t_j)}{t_j a'(t_j)}\right)\frac{1}{n\sqrt{2\pi a(t_j)(1 - a(t_j))}} \end{aligned} \tag{A.17}$$

where $j_0 = \lceil na(\tau_0) \rceil$, $j_1 = \lfloor na(\tau_1) \rfloor$ and $c = \eta\sqrt{n}$. But the sum in (A.17) is simply an approximating Reimann sum to an integral, and we therefore get

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{e^{\frac{1}{2}c^2}}{\sqrt{n}} P_0^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1) \\
\ge\ & (1-\epsilon)\int_{a(\tau_0)}^{a(\tau_1)} \left(1 - \frac{a(t)}{ta'(t)}\right) \frac{1}{\sqrt{2\pi a(t)(1-a(t))}} da(t) \\
=\ & (1-\epsilon)\int_{\tau_0}^{\tau_1} \left(a'(t) - \frac{a(t)}{t}\right) \frac{1}{\sqrt{2\pi a(t)(1-a(t))}} dt.
\end{aligned}$$

Letting $\epsilon \to 0$ and an identical upper bound complete the proof.

It remains to show (A.13) holds in the cases $\tau_0 = t_0$ and $\tau_1 = T$. Since $P_0^{(n)}(t_0 \le \mathcal{T} \le \tau_1) \ge P_0^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1)$ for all $\tau_0 > t_0$, we obtain the lower bound by letting $\tau_0 \to t_0$. When $l(t) = \frac{1}{2}\eta^2$, $a(t)$ is convex and hence by Lemma A.2, the tangent-sum formula (2.23) is an upper bound for $P_0^{(n)}(t_0 < \mathcal{T} \le \tau_1)$. Fix $\tau_0 > t_0$ and choose n large enough so that $P^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1) \le (1+\epsilon)I_n(\tau_0, \tau_1)$. Then

$$\begin{aligned}
&P^{(n)}(t_0 \le \mathcal{T} \le \tau_1) \\
=\ & P^{(n)}(\tau_0 \le \mathcal{T} \le \tau_1) + P^{(n)}(\mathcal{T} < t_0) \\
\le\ & (1+\epsilon)I_n(\tau_0,\tau_1) + \sum_{j=0}^{na(\tau_0)} \left(1 - \frac{a(t_j)}{t_j a'(t_j)}\right) \binom{n}{j} \left(\frac{t_j}{T}\right)^{na(t_j)} \left(\frac{T-t_j}{T}\right)^{n(1-a(t_j))} \\
\le\ & (1+\epsilon)I_n(\tau_0,\tau_1) + \left(\frac{T-t_0}{T}\right)^n \\
&+ \sum_{j=1}^{na(\tau_0)} \frac{e^{1/4}}{\sqrt{2\pi j(1-\frac{j}{n})}} \left(\frac{\tau_0}{Ta(\tau_0)}\right)^{na(\tau_0)} \left(\frac{T-\tau_0}{T(1-a(\tau_0))}\right)^{n(1-a(\tau_0))} \\
\le\ & (1+\epsilon)I_n(\tau_0,\tau_1) + e^{-\frac{1}{2}c^2}\left\{1 + \frac{e^{1/4}}{\sqrt{2\pi(1-a(\tau_0))}} \sum_{j=1}^{[na(\tau_0)]} \frac{1}{\sqrt{j}}\right\}. \qquad \text{(A.18)}
\end{aligned}$$

We can rewrite (A.18) as

$$\frac{e^{-\frac{1}{2}c^2}}{\sqrt{n}} P^{(n)}(t_0 \le \mathcal{T} \le \tau_1)$$

$$\leq \quad (1+\epsilon)\int_{\tau_0}^{\tau_1} \frac{a'(t) - \frac{a(t)}{t}}{\sqrt{2\pi a(t)(1-a(t))}}dt + \frac{1}{\sqrt{n}} + \frac{2e^{1/4}\sqrt{a(\tau_0)}}{\sqrt{2\pi}}$$

and letting $n \to \infty$, $\epsilon \to 0$ and $\tau_0 \to t_0$ gives the result. The case $\tau_1 = T$ can be treated similarly.

Now suppose $l(t)$ is not constant. We can then restrict the sum to small neighbourhoods of the point(s) (or intervals) at which $l(t)$ attains its minimum, and use Laplace approximations to both (2.23) and (A.14) to obtain the result.

$\square$

For our applications, we are interested probabilities of the form

$$P_0^{(n)}\left(\inf_{\tau_0 \leq t \leq \tau_1}(X(t) - na(t)) < 0\right) = P_0^{(n)}(\mathcal{T}' \leq \tau_1)$$

where $\mathcal{T}' = \inf\{t \geq \tau_0 : X(t) \leq na(t)\}$. We can write

$$\begin{aligned} P^{(n)}(\mathcal{T}' \leq \tau_1) \quad &= \quad P^{(n)}(\tau_0 < \mathcal{T} \leq \tau_1) + P^{(n)}(X(\tau_0) \leq na(\tau_0)) \\ &\quad + P^{(n)}(\mathcal{T} < \tau_0, X(\tau_0) > na(\tau_0), \mathcal{T}' \leq \tau_1). \qquad \text{(A.19)} \end{aligned}$$

Under the conditions of Theorem A.1, both correction terms will be second or higher order effects. To prove this, and obtain second order corrections, we must examine the four sources of approximation in Theorem A.1:

1. The tangent approximation.
2. Truncation and endpoint effects.
3. Use of Stirling's formula.
4. Approximation of the sum by the integral.

In the case of Brownian motion, second order corrections to first exit times through curved boundaries have been considered by Jennen (1985). Our results are similar but easier to derive due to the discreteness of our first passage times. We begin by deriving second order corrections to the tangent approximation, which uses the following Lemma.

Lemma A.4 *Under the conditions of Lemma A.2,*

$$P^{(n)}(\mathcal{T} < T) = \frac{1}{Ta'(T)} + \frac{a''(T)E^{(n)}(T-\mathcal{T})}{2a'(T)} + o(n^{-1}) \tag{A.20}$$

$$= \frac{1}{Ta'(T)} + \frac{Ta''(T)}{2na'(T)(Ta'(T)-1)^2} + o(n^{-1}). \tag{A.21}$$

and hence, in the notation of Theorem A.1,

$$P_0^{(n)}(\mathcal{T} = t_j | X(t_j) = j) = 1 - \frac{a(t_j)}{t_j a'(t_j)} - \frac{t_j a(t_j) a''(t_j)}{2na'(t_j)(t_j a'(t_j) - a(t_j))^2} + o(n^{-1}) \tag{A.22}$$

uniformly for $\tau_0 \le t_j \le \tau_1$*, if* $\tau_0 > t_0$.

Proof: The proof involves a more careful analysis of the proof of Lemma A.1. Let $\mathcal{T} = \inf\{t : X(t) - na(t) \le 0\}$ and $Y(t)$ be defined by (A.2). Using the Martingale stopping theorem we have

$$\begin{aligned} 0 &= E^{(n)}(Y(\mathcal{T} \wedge t)) \\ &= E^{(n)}\left\{\frac{na(\mathcal{T}) - \frac{n\mathcal{T}}{T}}{T - \mathcal{T}}; \mathcal{T} < t\right\} + \frac{n}{T}P(\mathcal{T} = T). \end{aligned} \tag{A.23}$$

Using the conditions $a(t) < t$ and $ta'(t) > a(t)$ for all t, we can find a constant m such that

$$\frac{ta(T)}{T} > a(T) - m(T-t) \ge a(t)\ \forall\, t < T$$

and since $a(t)$ is twice differentiable, there exists v such that

$$a(t) \ge a(T) - (T-t)a'(T) + \frac{v}{2}(T-t)^2\ \forall\, t < T.$$

Moreover, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$a(t) \ge a(T) - (T-t)a'(T) + \frac{(a''(T)-\epsilon)}{2}(T-t)^2\ \forall\, T - \delta < t < T.$$

Fix $k > 0$ and choose n large enough so that $k/n < \delta$. From (A.23) we get

$$\begin{aligned} 0 \;\geq\; & \frac{a(T)}{T}P(\tau = T) + \left\{\frac{a(T)}{T} - a'(T)\right\}P(\mathcal{T} < T) \\ & \frac{v}{2}E(T-\mathcal{T};\mathcal{T} \leq T - \frac{k}{n}) + \frac{a''(T)-\epsilon}{2}E(T-\mathcal{T};T-\frac{k}{n} < \mathcal{T} < T) \end{aligned} \tag{A.24}$$

Rearranging (A.24) gives

$$\begin{aligned} P(\mathcal{T} < T) \;\geq\; & \frac{a(T)}{Ta'(T)} + \frac{v}{2a'(T)}E(T-\mathcal{T};\mathcal{T} \leq T - \frac{k}{n}) \\ & + \frac{a''(T)-\epsilon}{2a'(T)}E(T-\mathcal{T};T-\frac{k}{n} < \mathcal{T} < T). \end{aligned} \tag{A.25}$$

We suppose $v < 0$; otherwise, the term involving v can be omitted. We then want an upper bound for $E(T-\mathcal{T};\mathcal{T} \leq T - \frac{k}{n})$. We define $\mathcal{T}' = \inf\{t : X(t) \leq n(a(T) - m(T-t))\}$ so $\mathcal{T}' < \mathcal{T}$ and hence

$$E(T-\mathcal{T};\mathcal{T} \leq T - \frac{k}{n}) \leq E(T-\mathcal{T}';\mathcal{T}' \leq T - \frac{k}{n}).$$

We can write

$$\begin{aligned} E(T-\mathcal{T}';\mathcal{T}' \leq T - \frac{k}{n}) \;=\; & \sum_{j=j_1}^{na(T)} E(\frac{j}{nm};\mathcal{T}' = j) \\ \leq\; & \sum_{j=j_1}^{na(T)} \frac{j}{nm}P(X(t_j) = na(T) - j) \end{aligned} \tag{A.26}$$

where $t_j = T - \frac{j}{nm}$ and $j_1 = \lceil mk \rceil$. We split the sum into $j \leq \sqrt{n}$ and $j > \sqrt{n}$. On the former, we have

$$\begin{aligned} P(X(t_j) = na(T) - j) \;=\; & \binom{na(T)}{j}\left(\frac{t_j}{T}\right)^{na(T)-j}\left(1 - \frac{t_j}{T}\right)^{j} \\ \leq\; & \frac{(na(T))^j}{j!}\left(\frac{j}{nmT}\right)^{j}\left(1 - \frac{j}{nmT}\right)^{na(T)-j} \\ \leq\; & \left(\frac{ea(T)}{mT}\right)^{j}\left(1 - \frac{j}{nmT}\right)^{na(T)}\left(1 - \frac{a(T)}{\sqrt{n}mT}\right)^{\sqrt{n}a(T)} \\ \leq\; & \left(1 - \frac{a(T)}{\sqrt{n}mT}\right)^{\sqrt{n}a(T)}\left(\frac{a(T)e^{1-a(T)/(mT)}}{mT}\right)^{j}. \end{aligned}$$

Summing (A.26) over $j < \sqrt{n}$ gives

$$\begin{aligned}\sum_{j=j_1}^{\sqrt{n}} \frac{j}{nm} P(X(t_j) = na(T) - j) &\leq \left(1 - \frac{a(T)}{\sqrt{n}mT}\right)^{\sqrt{n}} \sum_{j=j_1}^{\infty} \frac{j}{nm} \left(\frac{a(T)e^{1-a(T)/(mT)}}{mT}\right)^j \\ &\sim \frac{1}{n} f(j_1)\end{aligned}$$

as $n \to \infty$, where $f(j_1) \to 0$ as $n \to \infty$. Note $m > a(T)/T$ and hence the relation $\log(x) < x - 1$ with $x = a(T)/mT$ shows $a(T)e^{1-a(T)/(mT)}/mT < 1$ and hence the sum is convergent.

For $na(T) > j > \sqrt{n}$, we use standard bounds on factorials to obtain

$$\begin{aligned}P(X(t_j) = na(T) - j) &\leq e^{\frac{1}{4}} \sqrt{\frac{na(T)}{2\pi j(na(T) - j)}} \left(\frac{a(T)}{mT}\right)^j \left(\frac{1 - j/(nmT)}{1 - j/(na(T))}\right)^{na(T)-j} \\ &= e^{\frac{1}{4}} \sqrt{\frac{na(T)}{2\pi j(na(T) - j)}} \exp\left(-nf(\frac{j}{n})\right)\end{aligned}$$

where

$$f(x) = x \log\left(\frac{mT}{a(T)}\right) + (a(T) - x) \log\left(\frac{1 - x/a(T)}{1 - x/(mT)}\right).$$

We note $f(x)$ is an increasing function over $x > 0$ and hence $f(j/n)$ is minimized (over $j > \sqrt{n}$) at $j = \sqrt{n}$. Hence

$$\begin{aligned}&\sum_{j=\sqrt{n}}^{n} P(X(t_j) = na(T) - j) \\ &\leq \exp(-nf(\frac{1}{\sqrt{n}})) \left\{1 + \sum_{j=\sqrt{n}}^{na(T)-1} e^{\frac{1}{4}} \sqrt{\frac{na(T)}{2\pi j(na(T) - j)}}\right\} \\ &= o(n^{-1})\end{aligned}$$

as $n \to \infty$ since the first factor converges exponentially to 0. We can hence rewrite (A.25) as

$$P(\mathcal{T} < T) \geq \frac{a(T)}{Ta'(T)} + \frac{a''(T) - \epsilon}{2a'(T)} E(T - \mathcal{T}; T - \frac{k}{n} < \mathcal{T} < T) + \frac{v}{2n} f(mk) + o(n^{-1}). \tag{A.27}$$

A similar limiting argument as $n \to \infty$ and an application of Lemma A.5 below shows

$$\lim_{k\to\infty}\lim_{n\to\infty} nE(T - \mathcal{T}; T - \frac{k}{n} < \mathcal{T} < T) = \frac{Ta(T)}{(Ta'(T) - a(T))^2}$$

which we substitute into (A.27), then let $\epsilon \to 0$ to obtain

$$P(\mathcal{T} < T) \geq \frac{a(T)}{Ta'(T)} + \frac{Ta(T)a''(T)}{2na'(T)(Ta'(T) - a(T))^2} + o(n^{-1}).$$

A similar lower bound completes the proof.

The approximation (A.22) follows by a scale transformation, and continuity and compactness arguments give the uniformity.

□

Lemma A.5 *Let $X(t)$ be a Poisson process with rate $\lambda < \gamma$, and $\tau = \sup\{t : X(t) \geq \gamma t\}$. Then*

$$E(\tau) = \frac{\lambda}{(\gamma - \lambda)^2}. \tag{A.28}$$

Proof: We assume $\gamma = 1$; the general case follows by a straightforward time transformation. Since $X(t)$ is non-decreasing, we must have $X(\tau) = \tau$ and therefore τ is an integer valued random variable. We have

$$\begin{aligned} P(\tau = j) &= P(\tau = j, X(\tau) = j) \\ &= P(\tau = j | X(j) = j) P(X(j) = j) \\ &= (1 - \lambda)\frac{(\lambda j)^j}{j!} e^{-\lambda j}, \end{aligned}$$

using Lemma 2.1. Summing over j, we get

$$\begin{aligned} E(\tau) &= (1 - \lambda)\sum_{j=1}^{\infty} \frac{(\lambda j)^j}{(j-1)!} e^{-\lambda j} \\ &= (1 - \lambda)\sum_{j=1}^{\infty}\sum_{k=0}^{\infty} (-1)^k \frac{(\lambda j)^{j+k}}{(j-1)!k!} \end{aligned}$$

$$= \sum_{l=1}^{\infty} \lambda^l \sum_{k=0}^{l-1} (-1)^k \frac{(l-k)^l}{(l-k-1)!k!} \tag{A.29}$$

$$= (1-\lambda) \sum_{l=1}^{\infty} \lambda^l \frac{l(l+1)}{2}. \tag{A.30}$$

Here, we have used a combinatorial identity given on page 60 of Feller (1968) to evaluate the inner sum of (A.29). Evaluating (A.30) gives (A.28).

□

We now analyse the endpoint corrections in (A.19). Since under $P_0^{(n)}$, $X(\tau_0)$ has a $\mathcal{B}(n, \frac{\tau_0}{T})$ distribution, we have

$$\begin{aligned}
&P_0^{(n)}(X(\tau_0) \leq na(\tau_0)) \\
&= \sum_{j=0}^{\lfloor na(\tau_0) \rfloor} \binom{n}{j} \left(\frac{\tau_0}{T}\right)^j \left(1 - \frac{\tau_0}{T}\right)^{n-j} \\
&= \frac{\exp(-nl(\tau_0))}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}} \frac{(\mu/\lambda)^{na(\tau_0) - \lfloor na(\tau_0) \rfloor)}}{1 - \mu/\lambda}(1 + o(1))
\end{aligned}$$

as $n \to \infty$, where $\mu = a(\tau_0)/\tau_0$ and $\lambda = (1 - a(\tau_0))/(1 - \tau_0)$. This approximation is derived from the tail approximation (B.18).

The second correction term in (A.19) is slightly more complicated. We can write

$$\begin{aligned}
&P_0^{(n)}(\mathcal{T} \leq \tau_0, X(\tau_0) \geq na(\tau_0), \mathcal{T}' \leq \tau_1) \\
&\quad = \sum_x P_0^{(n)}(\mathcal{T} \leq \tau_0, \mathcal{T}' \leq \tau_1 | X(\tau_0) = na(\tau_0) + x) P_0^{(n)}(X(\tau_0) = na(\tau_0) + x)
\end{aligned}$$

where the sum is taken over those values of x for which $na(\tau_0) + x$ is an integer. We can write

$$\begin{aligned}
&P_0^{(n)}(\mathcal{T} \leq \tau_0, \mathcal{T}' \leq \tau_1 | X(\tau_0) = na(\tau_0) + x) \\
&= P_0^{(n)} \left(\inf_{t < \tau_0} (X(t) - na(t)) < 0 | X(\tau_0) = na(\tau_0) + x \right) \\
&\quad \times P_0^{(n)} \left(\inf_{t > \tau_0} (X(t) - na(t)) < 0 | X(\tau_0) = na(\tau_0) + x \right) \qquad \text{(A.31)}
\end{aligned}$$

which can be approximated using the same methods used for the local power approximations in Chapter 3 and confidence region approximations in Chapter 4. A rigorous justification of this local approximation follows in a manner similar to that used in the proof of Lemma A.4. We get

$$P_0^{(n)}\left(\inf_{t<\tau_0}(X(t)-na(t))<0|X(\tau_0)=na(\tau_0)+x\right) \rightarrow h\left(\frac{\mu}{a'(\tau_0)},\frac{\mu x}{a'(\tau_0)}\right)$$
$$P_0^{(n)}\left(\inf_{t>\tau_0}(X(t)-na(t))<0|X(\tau_0)=na(\tau_0)+x\right) \rightarrow \exp\left(\frac{x}{a'(\tau_0)}(\lambda'-\lambda)\right)$$

as $n\to\infty$, where $h(y,z)$ is defined by (2.13) and λ' is defined by (2.9) with $a=-1$ and $\gamma=a'(\tau_0)$. We also define a conjugate $\mu'>\mu$ as the solution of (2.9) with $a=1$ and $\gamma=-a'(\tau_0)$. Expanding the Binomial probabilities shows

$$\begin{aligned} P_0^{(n)}(X(\tau_0)=na(\tau_0)+x) &= P_0^{(n)}(X(\tau_0)=na(\tau_0))\left(\frac{\lambda}{\mu}\right)^x(1+o(1)) \\ &= \frac{\exp(-nl(\tau_0))}{\sqrt{na(\tau_0)(1-a(\tau_0))}}\left(\frac{\lambda}{\mu}\right)^x(1+o(1)) \end{aligned}$$

as $n\to\infty$ with x fixed, which in conjunction with (A.31) gives the approximation

$$\begin{aligned} P_0^{(n)}(\mathcal{T}\le\tau_0, X(\tau_0)>na(\tau_0),\mathcal{T}'\le\tau_1) &= \frac{e^{-nl(\tau_0)}}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}} \\ \times\sum_x \exp\left\{x\left(\frac{\lambda'-\lambda}{a'(\tau_0)}+\log(\frac{\lambda}{\mu})\right)\right\} & h\left(\frac{\mu}{a'(\tau_0)},\frac{\mu x}{a'(\tau_0)}\right)(1+o(1)) \end{aligned}$$

where the sum is over the values of x such that $na(\tau_0)+x$ is a positive integer. If we use the approximation (2.13) to $h(y,z)$ we get

$$\begin{aligned} & P_0^{(n)}(\mathcal{T}\le\tau_0, X(\tau_0)>na(\tau_0),\mathcal{T}'\le\tau_1) \\ \approx\; & \frac{e^{-nl(\tau_0)}}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}}\frac{a'(\tau_0)-\mu}{\mu'-a'(\tau_0)}\sum_x\left(\frac{\lambda'}{\mu'}\right)^x \\ \approx\; & \frac{e^{-nl(\tau_0)}}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}}\frac{a'(\tau_0)-\mu}{\mu'-a'(\tau_0)}\frac{(\lambda'/\mu')^{1-na(\tau_0)+\lfloor na(\tau_0)\rfloor}}{1-\lambda'/\mu'}. \end{aligned} \quad (A.32)$$

The conditions $\lambda' < a'(\tau_0) < \mu'$ ensure the sum is convergent.

The second order correction to Stirling's formula is

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n} + o(\frac{1}{n})\right).$$

This leads to

$$\binom{n}{na(t)} = \frac{1}{\sqrt{2\pi na(t)(1-a(t))}} \times \frac{1}{a(t)^{na(t)}}\frac{1}{(1-a(t))^{n(1-a(t))}}\left(1 - \frac{1-a(t)+a(t)^2}{12na(t)(1-a(t))} + o(\frac{1}{n})\right).$$

Making the approximation of summation by integration accurate to the second order just requires some care with the limits. The simple end-point approximations used in the first order approximation will have error of order $\frac{1}{n}$. If we instead choose limits $\tau_0' = a^{-1}([na(\tau_0)] + 0.5)$ and $\tau_1' = a^{-1}([na(\tau_1)] + 0.5)$, then the summation (2.23) will be a midpoint approximation to 2.24 and hence we will get the required second order accuracy.

The final second order approximation to the one-sided boundary crossing probability is

$$\begin{aligned} &P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1}(X(t) - na(t)) < 0\right) \\ &= \sqrt{n}\Bigg\{\int_{\tau_0'}^{\tau_1'}\left(a'(t) - \frac{a(t)}{t} - \frac{ta(t)a''(t)}{2n(ta'(t)-a(t))^2}\right) \\ &\qquad \times\left(1 - \frac{1-a(t)+a(t)^2}{12na(t)(1-a(t))}\right)\frac{\exp(-nl(t))dt}{\sqrt{2\pi a(t)(1-a(t))}} \\ &\quad + \left(\frac{(\mu/\lambda)^{na(\tau_0)-\lfloor na(\tau_0)\rfloor}}{1-\mu/\lambda} + \sum_x \left(\frac{\lambda'}{\mu}\right)^x h\left(\frac{\mu}{a'(\tau_0)}, \frac{\mu x}{a'(\tau_0)}\right)\right) \\ &\qquad \times \frac{\exp(-nl(\tau_0))}{n\sqrt{2\pi a(\tau_0)(1-a(\tau_0))}}\Bigg\}(1 + o(\frac{1}{n})) \end{aligned} \tag{A.33}$$

where for numerical purposes it will often be convenient to approximate the sum in (A.33) by (A.32). We note that the end-point corrections are only necessary for the validity of (A.33) when either $l(t)$ is constant or is minimized at $t = \tau_0$ with $l'(\tau_0) = 0$. In this special case, it is easily shown that $\mu' = \lambda$ and $\lambda' = \mu$. When $l(t)$ attains a unique minimum at some point $t_0 \in (\tau_0, \tau_1)$, it is possible to use a Laplace approximation to (A.33) and obtain an expression with the same order error term. However, (A.33) is generally much more accurate for problems of interest.

For the hazard rate model discussed in Chapter 5, second order corrections to (5.8) are slightly different since we are conditioning on the nth event occurring at time S; moreover, the upper and lower boundaries must also be treated separately. For the lower boundary the correct binomial coefficient if $\binom{n-1}{np_t}$. For the upper boundary, being treated by time reversal, the binomial coefficient is $\binom{n-1}{nq_t-1}$ and the local drift is $(nq_t - 1)/t$ instead of nq_t/t. This can be derived by noting the martingale (A.2) now has mean $1/T$ instead of 0.

A.2 Small Deviation Boundary Crossings

The results of Section A.1 are scaled to obtain asymptotic approximations when the boundary becomes remote as $n \to \infty$. The examples in earlier chapters show the final approximation given by Theorem A.1 usually gives good results. However, the integral must usually be evaluated numerically which can be computationally heavy, especially in applications such as the joint confidence regions of Chapter 4, where the integral must be evaluated repeatedly.

To reduce computation, and also provide more analytic insight into the results, we derived approximations based on local linearizations of the likelihood ratio process, and application of the results in Section 2.2. Although not as accurate, the results are much simpler. We give more a rigorous derivation of the results here.

After a preliminary Lemma we derive conditional and non-linear versions of Theorems 2.1 and 2.2.

Lemma A.6 *Suppose $X(t)$ is a Poisson process with rate $\lambda < c$ and $t_0 > 0$. Then*

$$P_\lambda(\exists t > t_0 : X(t) > ct) \leq P_\lambda(X(t_0) \geq ct_0) + P_{\lambda'}(X(t_0) < ct_0) \tag{A.34}$$

where $c\log(\lambda) - \lambda = c\log(\lambda') - \lambda'$ and $\lambda' > c$. If $\lambda > c$, then

$$P_\lambda(\exists t > t_0 : X(t) < ct) = P_\lambda(X(t_0) \leq ct_0) + P_{\lambda'}(X(t_0) > ct_0) \tag{A.35}$$

where λ' is as before except $\lambda' > c$.

Proof: Consider first $\lambda < c$. Let $\mathcal{T} = \inf\{t \geq t_0 : X(t) > ct\}$. By Wald's likelihood ratio identity,

$$P_\lambda(\mathcal{T} < \infty) = P_\lambda(X(t_0) \geq ct_0) + E_{\lambda'}\left\{\left(\frac{\lambda}{\lambda'}\right)^{X(\mathcal{T})-c\mathcal{T}}; t_0 < \mathcal{T} < \infty\right\}$$

from which (A.34) since $\frac{\lambda}{\lambda'} < 1$ and $X(\mathcal{T}) - c\mathcal{T} \geq 0$. To prove (A.35) we set $\mathcal{T} = \inf\{t \geq t_0 : X(t) \leq ct\}$ and proceed similarly.

□

Corollary A.1 *Let $X(t)$ be a Poisson process on $[0,T]$ and $cT > 1$. Then*

$$\limsup_{n\to\infty} P^{(n)}\left(\exists t > \frac{t_0}{n} : X(t) > nct\right) \leq P_{\frac{1}{T}}(X(t_0) \geq ct_0) + P_{(\frac{1}{T})'}(X(t_0) < ct_0) \tag{A.36}$$

and hence

$$\lim_{t_0\to\infty} \limsup_{n\to\infty} P^{(n)}\left(\exists t > \frac{t_0}{n} : X(t) > nct\right) = 0. \tag{A.37}$$

Similar results hold for $X(t) < nct$ when $cT < 1$.

Proof: Suppose first that $cT > 1$. Let $\mathcal{T} = \sup\{t : X(t) > nct\}$ and $j_0 = \lceil nct_0 \rceil$. Then for any $j_1 > j_0$ and n large enough so that $ncT > j_1$,

$$\begin{aligned} P^{(n)}\left(\mathcal{T} \geq \frac{t_0}{n}\right) &= \sum_{j=j_0}^{[ncT]} P^{(n)}\left(\mathcal{T} = \frac{j}{nc}\right) \\ &\leq \sum_{j=j_0}^{j_1} P^{(n)}\left(\mathcal{T} = \frac{j}{nc} | X(\frac{j}{nc}) = j\right) P^{(n)}\left(X(\frac{j}{nc}) = j\right) \\ &\quad + \sum_{j=j_1+1}^{[ncT]} P^{(n)}\left(X(\frac{j}{nc}) = j\right). \end{aligned} \tag{A.38}$$

Using standard Poisson limit theorems we have

$$\lim_{n\to\infty} \sum_{j=j_0}^{j_1} P^{(n)}\left(\mathcal{T} = \frac{j}{nc} | X(\frac{j}{nc}) = j\right) P^{(n)}\left(X(\frac{j}{nc}) = j\right) = P_{\frac{1}{T}}\left(\mathcal{T} \leq \frac{j_1}{c}\right). \tag{A.39}$$

We split the remaining terms according to $j < \sqrt{n}$ and $j \geq \sqrt{n}$. If $j_1 \leq j < \sqrt{n}$, then

$$\begin{aligned} P^{(n)}\left(X(\frac{j}{nc}) = j\right) &\leq \frac{n^j}{j!}\left(\frac{j}{ncT}\right)^j \left(1 - \frac{j}{ncT}\right)^{n-j} \\ &\leq \left(\frac{e}{cT}\right)^j \left(1 - \frac{j}{ncT}\right)^n \left(1 - \frac{1}{\sqrt{n}cT}\right)^{-\sqrt{n}} \\ &\leq \left(1 - \frac{1}{\sqrt{n}cT}\right)^{-\sqrt{n}} \left(\frac{e^{1-1/(cT)}}{cT}\right)^j. \end{aligned}$$

Since $cT \neq 1$, the relation $\log(x) < x - 1$ with $x = \frac{1}{cT}$ gives $\frac{e^{1-1/(cT)}}{cT} < 1$ and hence

$$\begin{aligned} \limsup_{n\to\infty} \sum_{j=j_1}^{\sqrt{n}} P^{(n)}\left(X(\frac{j}{nc}) = j\right) &\leq \sum_{j=j_1}^{\infty} \left(1 - \frac{1}{\sqrt{n}cT}\right)^{-\sqrt{n}} \left(\frac{e^{1-1/cT}}{cT}\right)^j \\ &= \left(1 - \frac{1}{\sqrt{n}cT}\right)^{-\sqrt{n}} \frac{(\frac{1}{cT}e^{1-1/cT})^{j_1}}{1 - \frac{1}{cT}e^{1-1/cT}}. \end{aligned} \tag{A.40}$$

For $n > j > \sqrt{n}$, we use standard bounds for factorials to obtain

$$P^{(n)}\left(X(\frac{j}{nc}) = j\right)$$

$$
\begin{aligned}
&\leq e^{\frac{1}{4}}\sqrt{\frac{n}{2\pi j(n-j)}}\left(\frac{1}{cT}\right)^{j}\left(\frac{1-\frac{j}{ncT}}{1-\frac{j}{n}}\right)^{n-j} \\
&\leq \frac{e^{1/4}}{\sqrt{2\pi j_1(1-\frac{j}{n})}}\exp\left(-nf(\frac{j}{n})\right)
\end{aligned}
$$

where $f(x) = x\log(cT) - (1-x)\log(\frac{1-x/cT}{1-x})$. Differentiating shows

$$
f'(x) = \log(cT) + \log\left(\frac{1-x/cT}{1-x}\right) + \frac{1-x}{1-x/cT} - 1 > 0
$$

since $cT > 1$ and $\log(x) \geq 1 - \frac{1}{x}$. We therefore have

$$
\begin{aligned}
\sum_{j=\sqrt{n}}^{n} P^{(n)}\left(X(\frac{j}{nc}) = j\right) &\leq \exp(-nf(\frac{1}{\sqrt{n}}))\left\{1 + \sum_{j=\sqrt{n}}^{n-1}\frac{e^{1/4}}{\sqrt{2\pi j_1(1-\frac{j}{n})}}\right\} \\
&\to 0
\end{aligned}
\tag{A.41}
$$

as $n \to \infty$ since the first factor converges exponentially to 0. Substituting (A.39), (A.40) and (A.41) into (A.38), letting $j_1 \to \infty$ and applying Lemma A.6 completes the proof of (A.36). Using tail bounds for the Poisson distribution gives (A.37). The case $cT < 1$ is similar, but using $\mathcal{T} = \inf\{t : X(t) \leq nct\}$.

□

Theorem A.2 *Suppose $n \to \infty$ and a and γ are constants such that $a + \gamma T < 0$. Then*

$$
\begin{aligned}
&\lim_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T}(aX(t) + \gamma nt) > c\right\} \\
&= \begin{cases} \exp(\frac{c}{\gamma n}(\hat{\lambda}' - \hat{\lambda})) & a < 0 \\ 1 - (1 - \frac{\hat{\lambda}a}{n|\gamma|})\sum_{j=0}^{\infty}(\frac{\hat{\lambda}a}{n|\gamma|})^j P(aU_j < c) & a > 0 \end{cases}
\end{aligned}
$$

where $\hat{\lambda} = \frac{n}{T}$ and $\hat{\lambda}'$ are a conjugate pair satisfying (2.9) and U_j is the convolution of j independent $\mathcal{U}[0,1]$ random variables.

Proof: Fix $k > 0$, $\alpha > 0$ and $\delta > 0$. Let $Y(t) = aX(t) + \gamma nt$. Suppose n is large enough so that $(1 - \frac{k}{nT})^n \geq (1-\delta)\exp(-\frac{k}{T})$. Then

$$\begin{aligned}
&P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} \\
&\geq P^{(n)}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c\right\} \\
&\geq \sum_{m=0}^{\alpha} P^{(n)}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c \,\middle|\, X(\frac{k}{n}) = m\right\}\binom{n}{m}\left(\frac{k}{nT}\right)^m\left(1-\frac{k}{nT}\right)^{n-m} \qquad (A.42)\\
&\geq \sum_{m=0}^{\alpha} P_{\hat{\lambda}}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c \,\middle|\, X(\frac{k}{n}) = m\right\} \\
&\qquad \times\frac{1}{m!}\left(1-\frac{\alpha}{n}\right)^{\alpha}\left(\frac{\hat{\lambda}k}{n}\right)^m(1-\delta)e^{-\frac{\hat{\lambda}k}{n}} \qquad (A.43)\\
&= \left(1-\frac{\alpha}{n}\right)^{\alpha}(1-\delta)P_{\hat{\lambda}}\left\{\sup_{0<t<\frac{k}{n}} Y(t); X(\frac{k}{n}) \leq \alpha\right\}.
\end{aligned}$$

Here, (A.43) follows from (A.42) since $\frac{1}{n^m}\binom{n}{m} \geq \frac{1}{m!}(1-\frac{m}{n})^{\alpha} \geq \frac{1}{m!}(1-\frac{\alpha}{n})^{\alpha}$ for $m \leq \alpha$. Letting $n \to \infty$ and $\delta \to 0$, we get

$$\begin{aligned}
\liminf_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} &\geq \liminf_{n\to\infty} P_{\hat{\lambda}}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c; X(\frac{k}{n}) \leq \alpha\right\} \\
&\geq \liminf_{n\to\infty} P_{\hat{\lambda}}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c\right\} \\
&\quad -\limsup_{n\to\infty} P_{\hat{\lambda}}\left\{X(\frac{k}{n}) > \alpha\right\}. \qquad (A.44)
\end{aligned}$$

Under $P_{\hat{\lambda}}$, $X(\frac{k}{n})$ has a Poisson distribution with parameter $\frac{n}{T}\frac{k}{n} = \frac{k}{T}$ and therefore $P_{\hat{\lambda}}(X(\frac{k}{n}) > \alpha)$ does not depend on n. Letting $\alpha \to \infty$, (A.44) becomes

$$\liminf_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} \geq \liminf_{n\to\infty} P_{\hat{\lambda}}\left\{\sup_{0<t<\frac{k}{n}} Y(t) > c\right\}. \qquad (A.45)$$

But a straightforward time rescaling shows the probability on the right of (A.45) is independent of n, so letting $k \to \infty$ gives

$$\liminf_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} \geq P_{\hat\lambda}\left\{\sup_{t>0} Y(t) > c\right\}.$$

Derivation of an upper bound is similar. We have

$$P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} \leq P^{(n)}\left\{\sup_{0<t\leq \frac{k}{n}} Y(t) > c\right\} + P^{(n)}\left\{\sup_{\frac{k}{n}<t<T} Y(t) > 0\right\} \quad \text{(A.46)}$$

and similarly to before, we get

$$\limsup_{n\to\infty} P^{(n)}\left\{\sup_{0<t\leq \frac{k}{n}} Y(t) > c\right\} \leq P_{\hat\lambda}\left\{\sup_{0<t\leq \frac{k}{n}} Y(t) > c\right\}. \quad \text{(A.47)}$$

By an application of Corollary (A.1) we have

$$\begin{aligned}\limsup_{n\to\infty} P^{(n)}\left\{\sup_{\frac{k}{n}<t<T} Y(t) > 0\right\} &\leq P_{\hat\lambda}\left\{Y(\frac{k}{n}) \geq 0\right\} + P_{\hat\lambda'}\left\{Y(\frac{k}{n}) > 0\right\} \\ &\to 0 \text{ as } k \to \infty.\end{aligned}$$

Hence from (A.46) and (A.47) we have

$$\limsup_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T} Y(t) > c\right\} \leq P_{\hat\lambda}\left\{\sup_{0<t<T} Y(t) > c\right\}$$

which completes the proof.

□

Corollary A.2 *Suppose $g(t,p)$ is twice continuously differentiable and $g(0,0) = 0$. Let $a = \frac{\partial}{\partial p} g(t,p)|_{0,0}$ and $\gamma = \frac{\partial}{\partial t} g(t,p)|_{0,0}$. We suppose $a + \gamma T < 0$. Moreover, the solution p_t of $g(t,p) = 0$, where it exists, satisfies $p_t > \frac{t}{T}$ for $t > 0$ if $a > 0$ and $p_t < \frac{t}{T}$ if $a < 0$. Also, $p'_t|_{t=0} \neq \frac{1}{T}$. Then*

$$\begin{aligned}&\lim_{n\to\infty} P^{(n)}\left\{\sup_{0<t<T} ng(t, \frac{X(t)}{n}) > c\right\} \\ &= \begin{cases} \exp(\frac{c}{n\gamma}(\hat\lambda' - \hat\lambda)) & a < 0 \\ 1 - (1 - \frac{a}{T|\gamma|}) \sum_{j=0}^{\infty} (\frac{a}{T|\gamma|})^j P(aU_j < c) & a > 0 \end{cases}.\end{aligned}$$

Proof: We denote by g_1, g_2, g_{11}, g_{12} and g_{22} the first and second order partial derivatives of g. We assume all the second derivatives are bounded by M so

$$|g(t,p) - tg_1(0,0) - pg_2(0,0)| \leq \frac{M}{2}(t+p)^2.$$

For $X(t) < \alpha$ and $t < \frac{k}{n}$ we have

$$ng(t, \frac{X(t)}{n}) \leq n\left\{\gamma t + a\frac{X(t)}{n} + \frac{M}{2}(\frac{k}{n} + \frac{\alpha}{n})^2\right\}$$

and hence

$$\begin{aligned} &P^{(n)}\left\{\sup_{0<t<T} ng(t, \frac{X(t)}{n}) \geq c\right\} \\ \geq \;& P^{(n)}\left\{\sup_{0<t<\frac{k}{n}} (aX(t) + \gamma nt) \geq c + \frac{M}{2n}(\alpha + k)^2\right\}\end{aligned}$$

where $a = g_2(0,0)$ and $\gamma = g_1(0,0)$. Proceeding as in the proof of Theorem A.2 gives the lower bound.

Derivation of an upper bound is similar. By the conditions on p_t we can choose β such that $\beta T > a + \gamma T$ and $\beta t \leq ap_t + \gamma t$ for all $t > 0$. We then have

$$\begin{aligned} &P^{(n)}\left\{\sup_{0<t<T} ng(t, \frac{X(t)}{n}) > c\right\} \\ \leq \;& P^{(n)}\left\{\sup_{0<t<\frac{k}{n}} (aX(t) + \gamma nt) \geq c - \frac{M}{2}(\alpha + k)^2\right\} + P^{(n)}\left\{X(\frac{k}{n}) \geq \alpha\right\} \\ &+ P^{(n)}\left\{\sup_{\frac{k}{n}<t<T} Y(t) \geq \beta nt\right\}\end{aligned}$$

and the proof then follows similarly to the upper bound in Theorem A.2.

□

Our main use for Corollary A.2 is to prove Theorem 4.1. There are two boundaries to consider, with $np_t \leq \frac{t}{T} \leq nq_t$ such that $l(t) = l(\tau) + c$ when $X(t) = np_t$ or $X(t) = nq_t$. We shift the origin to $(\tau, l(\tau))$ and apply Corollary A.2 to one of the boundaries: nq_t if $\rho > \frac{\tau}{T}$ and np_t if $\rho < \frac{\tau}{T}$. It remains to show the probability of crossing the other boundary converges to 0; this is a trivial consequence of Corollary A.1: if $\rho > \frac{\tau}{T}$, the probability under $P_\tau^{(m,n)}$ that $X(t) < \frac{t}{T}$ for any $t \in [\tau_0, \tau_1]$ converges to 0, and similarly when $\rho < \frac{\tau}{T}$.

A.3 Locally Poisson Processes

Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{I}[0,T]$ be the class of integer valued, non-decreasing right continuous functions f on $[0,T]$ with $f(0) \geq 0$ and $f(T) < \infty$. A counting process on $[0,T]$ is a function $X : \Omega \to \mathcal{I}[0,T]$ such that the random variable $X(A) = \int_A dX$ is measurable for all Borel subsets A of $[0,T]$. That is,

$$\{\omega \in \Omega : X_\omega(A) = k\} \in \mathcal{F}$$

for all $k \geq 0$ and Borel sets A. If P_λ is a probability measure such that (X, P_λ) satisfy conditions 1 and 2 on page 2 with $\lambda(t) = \lambda$ for all t, then X is a Poisson process on $[0,T]$ with respect to P_λ and rate λ.

Definition A.1 *Suppose $\{X_n\}_{n=1}^\infty$ be a sequence of counting processes on Ω and $\{P^{(n)}\}_{n=1}^\infty$ is a sequence of probability measures on $\mathcal{F}$. We say $(X_n, P^{(n)})$* **converges weakly** *to a Poisson process with rate λ if for all finite sequences $A_1, \ldots, A_m$ of disjoint Borel subsets of $[0,T]$,*

$$P^{(n)}(\cap_{i=1}^m X_n(A_i) = k_i) \to \prod_{i=1}^m \left(\frac{(\lambda|A_i|)^{k_i}}{k_i!} e^{-\lambda|A_i|}\right) \tag{A.48}$$

as $n \to \infty$, where $|A|$ denotes the Lebesgue measure of A.

Definition A.2 *Suppose X is a counting process on Ω, and $\{P^{(n)}\}_{n=1}^\infty$ is a sequence of probability measures on $\mathcal{F}$. For constants $\{c_n\}_{n=1}^\infty$ with $c_n > 0$ and $c_n \to 0$, and* *fixed $t_0 < T$, we define*

$$X_n(t) = X(t_0 + c_n t) - X(t_0).$$

If, for all $S > 0$, the pair $(X_n, P^{(n)})$ restricted to $0 \le t \le S$ converges weakly to a Poisson Poisson process with rate μ^+ on $[0, S]$ then X is **locally right-Poisson** *at t_0 under $P^{(n)}$, and X has local rate μ^+/c_n under $P^{(n)}$. If $t_0 > 0$ and $\{X(t_0) - X((t_0 - c_n t)^-), P^{(n)}\}$ converges to a Poisson process with rate μ^- on $[0, S]$, then X is* **locally left-Poisson**. *If the pair $\{X(t_0) - X(t_0 - c_n t), X(t_0 + c_n t) - X(t_0)\}$ converges to independent Poisson processes, then X is* **locally Poisson** *at t_0.*

For our applications, n will represent the total number of events and we take $c_n = \frac{1}{n}$. The most common example of a locally Poisson process is the empirical process: if $X_1, \ldots, X_n$ are *i.i.d.* $\mathcal{U}[0,1]$ random variables, then $X(t) = \sum_{i=1}^n I(X_i \le t)$ is locally Poisson for all t, with local left and right rates both n under $P^{(n)}$. Despite the immense literature on empirical processes, very little attention has been given to local approximation by Poisson processes, with most authors preferring to approximate $X(t)$ by a Brownian bridge when n is large. A treatment of local Poisson approximations for the empirical process may be found in Major (1990), where explicit bounds for the difference between the two processes are considered. The formulation used by Major allows $S \to \infty$ as $n \to \infty$, which would simplify the handling of error terms, for example in Corollary A.1, but formulating similar results for the more general class of processes we wish to consider appears difficult.

Establishing results similar to Theorem A.1 for the boundary crossing approximation (2.26) will depend on the choice of the approximation $f_c(t)$, which will be discussed in Appendix B, and an extension of Lemma A.2 to locally Poisson processes. We can write

$$P^{(n)}\left(\exists t \,:\, T - \frac{k}{n} \le t < T, X(t) - na(t) < 0\right)$$

$$
\begin{aligned}
&\leq P^{(n)}\left(\exists t < T \,:\, X(t) - na(t) < 0\right) \\
&\leq P^{(n)}\left(\exists t \,:\, T - \frac{k}{n} \leq t < T, X(t) - na(t) < 0\right) \\
&\quad + P^{(n)}\left(\exists t < T - \frac{k}{n} \,:\, X(t) - na(t) < 0\right).
\end{aligned}
$$

Letting $n \to \infty$ and $k \to \infty$, using standard Poisson limit theorems in conjunction with Lemma 2.1, gives

$$
\begin{aligned}
&\frac{\mu(T, a(T))}{a'(T)} \\
&\leq \lim_{n\to\infty} P^{(n)}\left(\exists t < T \,:\, X(t) - na(t) < 0\right) \\
&\leq \frac{\mu(T, a(T))}{a'(T)} + \lim_{k\to\infty}\lim_{n\to\infty} P^{(n)}\left(\exists t < T - \frac{k}{n} \,:\, X(t) - na(t) < 0\right) \quad \text{(A.49)}
\end{aligned}
$$

where $\mu(T, a(T))$ denotes the local left rate at T, and we assume $\mu(T, a(T)) < a'(T)$. We are also assuming the limits in (A.49) exist; otherwise, similar relations hold with lim sup's and lim inf's. The extension of Lemma A.2 to locally Poisson processes now requires showing the last term of (A.49) is 0. This must be carried out in special cases using the structure of $X(t)$ and the boundary $a(t)$, but given the large deviation scaling, the necessary conditions will be mild. The main requirement will be

$$
\liminf_{n\to\infty} \frac{1}{n} E^{(n)}(X(t) | X(T) = na(T)) > a(t)
$$

for all $t < T$ which generalizes the condition $Ta(t) - t < 0$ in Lemma A.2 which ensures boundary crossing is a rare event and justifies the local approximation.

Extension of the second order correction to the log-linear model appears difficult; moreover, the probability of crossing both boundaries would appear to be as important as obtaining the second order approximations. One of the problems encountered include obtaining second order corrections to the conditional distribution of $X(t)$ given by (B.28). Although the next term is given by Borokov and Rogozin (1965), evaluation appears difficult. We also need to correct the approximation to the local rate $\mu(t, a(t))$ given by (B.38) and the curvature of $E^{(n)}(X(t')|X(t) = na(t))$. We will not pursue this topic here.

A.4 Aldous' Poisson Clumping Heuristic

Aldous (1989) presents an interesting method which can be used to derive heuristic approximations to the probabilities of rare events. The method, called the Poisson clumping heuristic, is used by Aldous for several problems involving boundary crossings by Gaussian processes. We show here how similar methods can be applied to locally Poisson processes; in particular, we use this method to obtain (2.24). The heuristic can also be used to study the multidimensional fields discussed in Chapter 7.

A.4.1 One Dimensional Processes

Since the heuristic is more naturally suited to the study of stationary processes, we begin by studying the one dimensional scan stattistic with known Δ, and derive the significance level approximation (7.2). The idea of the heuristic is that values of θ for which $Y(\theta) = \frac{n\Delta}{2\pi}(1+\epsilon)$ will occur in 'clumps'. That is, if $Y(\theta) = \frac{n\Delta}{2\pi}(1+\epsilon)$ for some θ, then the process will stay around this level for only a short time, then drift back towards the stationary level $\frac{n\Delta}{2\pi}$. In Figure A.1, we illustrate this by plotting $Y(t)$ vs $t = \frac{\theta}{2\pi}$ with $n = 50$, $\Delta = 0.2$ and $\epsilon = \frac{1}{2}$. The process $Y(t)$ first reaches 15 near $t = 0.3$, climbs higher then returns to 15 at about 0.5. These values of t represent one clump. The process then returns to lower levels, but increases to reach 15 again between 0.8 and 0.9. This represents a second clump.

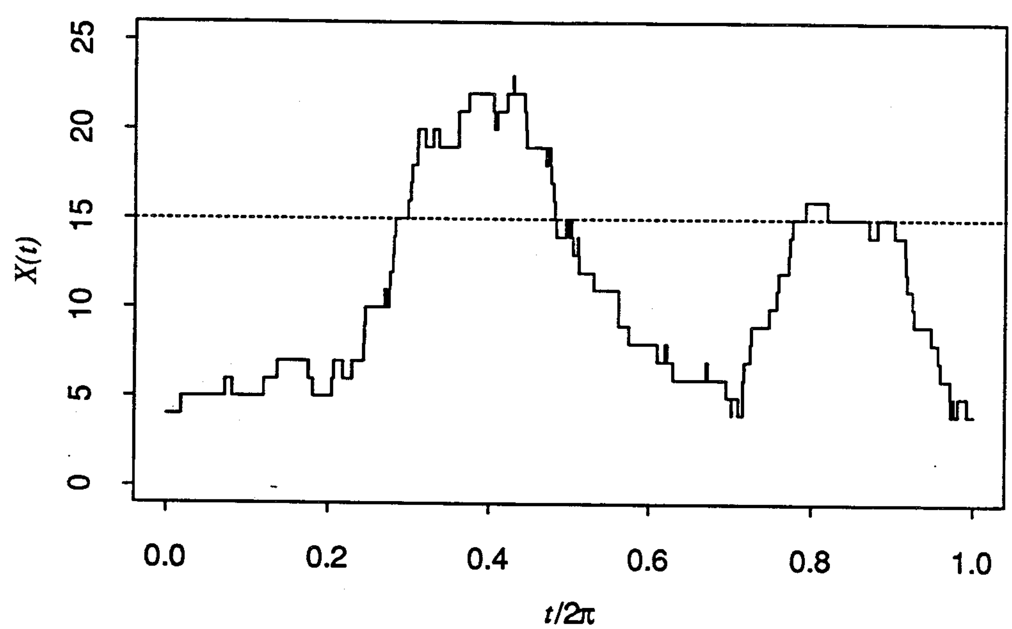


Figure A.1: Diagram of the Poisson Clumping. In this example, we have 50 points and a scan width of 0.2, so the mean level is 10. A clump is taken to be a set of values of t for which $X(t) = 15$, indicated by the dotted line. In this example, we have two clumps.

The probability of exceeding the boundary is the probability of one or more clumps. Moreover, if the boundary is remote, in the sense that there is only a small probability of ever exceeding the boundary, then there will usually be only 0 or 1 clumps, and the probability of crossing the boundary is just the expected number of clumps, which can be found as the ratio of the expected total size of all clumps to the mean clump size. By approximating the mean clump size and the total size of all clumps, we can obtain an approximation to the probability of exceeding the boundary.

Suppose $Y(\theta) = \frac{n\Delta}{2\pi}(1+\epsilon)$. Then for small δ, $Y(\theta+\delta) - Y(\theta)$ behaves like the difference of two homogeneous Poisson processes $Z_1(\delta)$ and $Z_2(\delta)$, where $Z_1(\delta)$ has rate $\lambda = \frac{n}{2\pi}(1 - \frac{\Delta\epsilon}{2\pi-\Delta})$ and $Z_2(\delta)$ has rate $\mu = \frac{n}{2\pi}(1+\epsilon)$. We can now use (2.36) to obtain the mean clump size,

$$\begin{aligned} \int_0^\infty P(Z_1(\delta) = Z_2(\delta))d\delta &= \frac{1}{|\lambda - \mu|} \\ &= \frac{2\pi - \Delta}{\epsilon n}. \end{aligned} \tag{A.50}$$

The expected total size of all clumps is

$$\int_0^{2\pi} P_0^{(n)}\left(Y(\theta) = \frac{n\Delta}{2\pi}(1+\epsilon)\right) d\theta = 2\pi P_0^{(n)}\left(Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right). \tag{A.51}$$

From (A.50) and (A.51), we get the approximation

$$P_0^{(n)}\left(\sup_{0<\theta<2\pi} Y(\theta) \geq \frac{n\Delta}{2\pi}(1+\epsilon)\right) \approx \frac{2\pi\epsilon n}{2\pi - \Delta} P_0^{(n)}\left(Y(0) = \frac{n\Delta}{2\pi}(1+\epsilon)\right)$$

which agrees with (7.2).

The use of the heuristic for non-stationary processes and non-linear boundaries is slightly more complicated, since the mean clump size will depend on the location of the clump. The following Lemma is used to estimate the clump size.

Lemma A.7 *Suppose $X(t)$ is a Poisson process with rate $\lambda > 0$, and $a > 0$ and $\gamma < 0$ are constants such that $a\lambda + \gamma < 0$. Define an occupancy measure U by*

$$U(A) = \int_0^\infty I(aX(t) + \gamma t \in A)dt.$$

Then

$$\lim_{\delta \to 0^-} \frac{1}{|\delta|} E(U[\delta, 0]) = \frac{1}{|a\lambda + \gamma|}.$$

Proof: Each time the process $aX(t) + \gamma t$ returns to 0, we get a contribution of $|\frac{\delta}{\gamma}|$ to $U[\delta, 0]$. By Lemma 2.1, the probability of returning to 0 is $\frac{\lambda a}{|\gamma|}$ and so the expected number of visits to 0 is

$$\frac{1}{1 + \frac{a\lambda}{\gamma}} = \frac{|\gamma|}{|a\lambda + \gamma|}$$

and therefore

$$E(U[\delta, 0]) = \frac{|\delta|}{|a\lambda + \gamma|}(1 + o(\delta)).$$

□

Suppose a locally Poisson process $X(t)$ hits a boundary $na(t)$ and consider the set $\{t : X(t) \in [na(t), na(t) + \delta]\}$ for small δ. The local event rate for $s < t$ given by

$$n\mu(t, a(t)) = \lim_{s \to t^-} \frac{P(X(t) - X(s) > 0 | X(t) = na(t))}{t - s}.$$

We assume $\mu(t, a(t)) < a'(t)$. By Lemma A.7, the mean clump size at time t is given by

$$\frac{\delta}{n(a'(t) - \mu(t, a(t)))}.$$

The clump rate is the probability that $X(t)$ lies in the interval $[na(t), na(t)+\delta]$, which approximately $\delta f_c(t)$, where $f(t)$ is a continuous approximation to the distribution $f(t)$ of $X(t)$. We therefore get the approximation

$$P_0^{(n)}\left(\inf_{t_0 < t < t_1} (X(t) - na(t)) \leq 0\right) \approx \int_{t_0}^{t_1} n(a'(t) - \mu(t, a(t))) f_c(t) dt$$

which agrees with (2.26).

A.4.2 Random Fields

Although we have defined a clump to be times at which the process $X(t)$ is at a specified value, we can also define a clump to be the times at which $X(t)$ exceeds a specified level. For the one-dimensional scan statistic, we then approximate the mean clump size using (2.37). The expected size of all clumps is then $2\pi P_0^{(n)}(Y(0) \geq \frac{n\Delta}{2\pi}(1+\epsilon))$, and if we use Lemma A.3, this gives an alternative derivation of (7.2).

The advantage of this technique is that it can be extended to some random field problems. If the random field can be locally decomposed as the sum of independent one dimensional processes, then the product rule discussed in Chapter J of Aldous (1989) for Gaussian random fields may be applicable.

We derive the significance level approximation (7.39) for the two dimensional scan statistic with known Δ_1 and Δ_2. If $Y(t) = n\Delta_1\Delta_2(1+\epsilon)$, the increments of $Y(t)$ parallel to the ith axis behave like the difference of two independent Poisson processes, with rates given by

$$\mu_i = n(1+\epsilon)\frac{\Delta_1\Delta_2}{\Delta_i}$$

and

$$\lambda_i = n\frac{(1-\Delta_1\Delta_2(1+\epsilon))}{1-\Delta_1\Delta_2}\frac{\Delta_1\Delta_2}{\Delta_i}$$

for $i = 1, 2$. We then use (2.37) to get the mean clump size parallel to the ith axis,

$$\frac{\mu_i}{(\lambda_i - \mu_i)^2} = \frac{1}{n}\frac{\Delta_i}{\Delta_1\Delta_2}\frac{(1+\epsilon)(1-\Delta_1\Delta_2)^2}{\epsilon^2}.$$

The product rule then gives the mean clump size,

$$\frac{(1+\epsilon)^2(1-\Delta_1\Delta_2)^4}{n^2\Delta_1\Delta_2\epsilon^4}. \tag{A.52}$$

The mean total size of all clumps is $(1-\Delta_1)(1-\Delta_2)P_0^{(n)}(Y(0) \geq n\Delta_1\Delta_2(1+\epsilon))$. The probability of exceeding the level $n\Delta_1\Delta_2(1+\epsilon)$ is approximated by

$$\begin{aligned} &P_0^{(n)}\left(\sup_t Y(t) \geq n\Delta_1\Delta_2(1+\epsilon)\right) \\ &\quad\approx (1-\Delta_1)(1-\Delta_2)\frac{n^2\Delta_1\Delta_2\epsilon^4}{(1-\Delta_1\Delta_2)^4(1+\epsilon)^2}P_0^{(n)}(Y(0) > n\Delta_1\Delta_2(1+\epsilon)) \\ &\approx \quad \frac{n^2(1-\Delta_1)(1-\Delta_2)\Delta_1\Delta_2\epsilon^3}{(1-\Delta_1\Delta_2)^3(1+\epsilon)}P_0^{(n)}(Y(0) = n\Delta_1\Delta_2(1+\epsilon)) \end{aligned}$$

using (B.18). The clumping heuristic has led to an alternative, and simpler, derivation of (7.39). One advantage of the use of the clumping heuristic is that it immediately generalizes to $d \geq 2$ dimensions. Using similar notation, the one-dimensional mean clump size parallel to the jth axis is

$$\frac{(1+\epsilon)\Delta_j(1-\prod_{i=1}^d \Delta_i)^2}{n\epsilon^2\prod_{i=1}^d \Delta_i}$$

and therefore

$$\begin{aligned} &P_0^{(n)}\left(\sup_t Y(t) \geq n(1+\epsilon)\prod_{i=1}^d \Delta_i\right) \\ &\approx \quad \frac{n^d\epsilon^{2d-1}\prod_{i=1}^d \Delta_i^{d-1}(1-\Delta_i)}{(1-\prod_{i=1}^d \Delta_i)^{2d-1}(1+\epsilon)^{d-1}}P_0^{(n)}\left(Y(0) = n(1+\epsilon)\prod_{i=1}^d \Delta_i\right). \end{aligned}$$

To conclude this section, we use the clumping heuristic to derive an approximation to the significance level for the variable width scan statistic studied in Theorem 7.2. However, we are unable to exactly reproduce (7.14).

Again, we use the product rule to write the clump size as the product of one dimensional clump sizes. However, deriving the one dimensional clump sizes involves a little trickery. For one dimensional processes, we can either apply the clumping heuristic to sets $X(t) \in [na(t), na(t)+\delta]$ or $X(t) \geq na(t)$. We chose the former since it is easiest to derive the clump size for these sets. However, the ratio of clump size to clump rate must be the same for both sets, so if we consider the later, we have the clump size

$$\frac{1}{n(a'(t)-\mu(t,a(t)))}\frac{P_0^{(n)}(X(t)\geq na(t))}{P_0^{(n)}(X(t)=na(t))}$$
$$\approx \frac{a(t)(1-\frac{t}{T})}{n(a'(t)-\mu(t,a(t)))(a(t)-\frac{t}{T})}.$$

Here, we are considering upper boundaries so $\mu(t,a(t))=\frac{1-a(t)}{T-t}$.

For the circular scan statistic with unknown Δ, we get the one dimensional clump sizes

$$\frac{h(\Delta)(1-\frac{\Delta}{2\pi})}{n(h'(\Delta)-\frac{1-h(\Delta)}{2\pi-\Delta})(h(\Delta)-\frac{\Delta}{2\pi})}$$

and therefore we obtain the boundary crossing approximation

$$P_0^{(n)}\left(M_{\Delta^0,\Delta^1}\geq \frac{1}{2}c^2\right)$$
$$\approx 2\pi\int_{\Delta^0}^{\Delta^1}\left(\frac{n(h'(\Delta)-\frac{1-h(\Delta)}{2\pi-\Delta})(h(\Delta)-\frac{\Delta}{2\pi})}{h(\Delta)(1-\frac{\Delta}{2\pi})}\right)^2 P_0^{(n)}(X(0,\Delta)\geq nh(\Delta))d\Delta$$

where the leading 2π arises from the integral over θ with Δ fixed. If we use the tail approximation given by Lemma A.3, this gives the approximation

$$c^3\phi(c)\int_{\Delta^0}^{\Delta^1}\frac{2\pi\Delta}{\eta^3 h(\Delta)}\left(h'(\Delta)-\frac{1-h(\Delta)}{2\pi-\Delta}\right)^2\left(\frac{h(\Delta)}{\Delta}-\frac{1-h(\Delta)}{2\pi-\Delta}\right)\frac{d\Delta}{\sqrt{h(\Delta)(1-h(\Delta))}}$$

which differs from (7.14) by a factor of $\Delta h'(\Delta)/h(\Delta)$ in the integrand.

A.5 Upper Boundary Tangent Approximations

When deriving the tangent approximation (2.24), we made the assumption that the continuous boundary $a(t)$ satisfies $a(0) < 0$. For the processes discussed in this thesis, this caused no difficulty since the upper boundary could always be treated by time reversal. However, it is possible to use (2.24) with trivial modifications to evaluate boundary crossing probabilities for unconditioned Poisson processes. In this case, time reversal cannot be used for upper boundaries. We show here how the result of Mallows and Nair (1989) can be used in this case, and also to generalize the results when we allow random jump sizes.

Theorem A.3 *Suppose $X(t)$ be a Poisson process with rate λ and $Y(t) = \sum_{i=1}^{X(t)} Y_i$ where Y_i are a sequence of i.i.d. non-negative random variables with moment generating function $Ee^{\theta Y_1} = e^{\psi(\theta)}$. Let $\tau = \inf\{t > 0 : Y(t) - \gamma t = 0\}$ where $\gamma > 0$. Then*

$$P(\tau < \infty) = \begin{cases} \frac{\lambda}{\gamma}\psi'(0) & \lambda\psi'(0) < \gamma \\ \frac{\lambda'}{\gamma}\psi'(\frac{\lambda'-\lambda}{\gamma}) & \lambda\psi'(0) > \gamma \end{cases}$$

where

$$\log(\lambda') - \log(\lambda) = \psi(\frac{\lambda' - \lambda}{\gamma}) \tag{A.53}$$

and $\lambda' < \lambda$.

Proof: This result is given by Mallows and Nair (1989) but we give an alternative proof here. The case $\lambda\psi'(0) < 1$ is a straightforward generalization of Lemma 2.1. Using the same notation but working in real variables, we get

$$G_+(\theta) = \frac{\lambda\psi'(0)}{\gamma}\frac{e^{\psi(\theta)} - 1}{\theta\psi'(0)}$$

and letting $\theta \to 0$ gives the result.

To treat the case $\lambda\psi'(0) > 1$, we denote by $P_{\lambda,\theta}$ the measure under which $X(t)$ has rate λ and Y_1 has density $e^{\theta y-\psi(\theta)}f_0(y)$. If $\tau = t$ and $X(\tau) = n$ the likelihood ratio is

$$\frac{dP_{\lambda',\theta}}{dP_{\lambda,0}} = e^{\theta\gamma t - n\psi(\theta)}\left(\frac{\lambda'}{\lambda}\right)^n e^{(\lambda-\lambda')t} \tag{A.54}$$

and if we choose $\theta = \frac{\lambda'-\lambda}{\gamma}$ and λ' to satisfy (A.53), then (A.54) is equal to 1. By Wald's likelihood ratio identity, we have

$$P_{\lambda,0}(\tau < \infty) = P_{\lambda',\theta}(\tau < \infty) = \frac{\lambda'}{\gamma}\psi'(\frac{\lambda'-\lambda}{\gamma}).$$

□

Now suppose $X(t)$ is a Poisson process with rate $\lambda > 0$ and $a(t)$ is a continuous differentiable boundary with $a(0) > 0$ and $a'(t) > \lambda$ for all t. Let $\mathcal{T} = \inf\{t : X(t) = a(t)\}$ and t_j be the solution of $a(t) = j$ for $j \geq a(0)$. We have

$$\begin{aligned} P(\mathcal{T} < T) &= \sum_{j:t_j<T} P(\mathcal{T} = t_j | X(t_j) = j)P(X(t_j) = j) \\ &\approx \sum_{j:t_j<T}\left(1 - \frac{a^*(t)}{ta'(t)}\right)\frac{(\lambda t_j)^j}{j!}e^{-\lambda t_j}. \end{aligned}$$

where

$$\log(a(t)) - \log(a^*(t)) = \frac{a^*(t) - a(t)}{ta'(t)}.$$

Using Stirling's formula and approximating the sum by an integral gives

$$P(\mathcal{T} < T) \approx \int_0^T (a'(t) - \frac{a^*(t)}{t})\frac{1}{\sqrt{2\pi a(t)}}\exp\left(-\lambda t + a(t)\left(1 + \log(\frac{\lambda t}{a(t)})\right)\right). \tag{A.55}$$

An example where this may be useful is a repeated likelihood ratio test of $\mathcal{H}_0 : \lambda = \lambda_0$ vs $\mathcal{H}_1 : \lambda > \lambda_0$ for a Poisson process $X(t)$. The maximum likelihood estimate of λ at time t is

$$\hat{\lambda} = \frac{X(t)}{t}$$

and the log-likelihood ratio process is

$$l(t) = X(t)\log\left(\frac{X(t)}{\lambda_0 t}\right) - (X(t) - \lambda_0 t).$$

If we stop the test and reject $\mathcal{H}_0$ at $\tau = \inf\{t \geq \tau_0 : l(t) \geq \frac{1}{2}c^2\}$, then from (A.55) we obtain the approximation

$$P_{\lambda_0}(\tau_0 \leq \tau \leq \tau_1) \approx P_{\lambda_0}(l(\tau_0) \geq \frac{1}{2}c^2) + \phi(c)\int_{\tau_0}^{\tau_1}(a'(t) - \frac{a^*(t)}{t})\frac{dt}{\sqrt{a(t)}}$$

where $a(t)$ is the solution of $a(t)\log(\frac{a(t)}{\lambda_0 t}) - (a(t) - \lambda_0 t) = \frac{1}{2}c^2$.

Appendix B

Large Deviation Approximations

B.1 Conditional Distributions

Suppose we have a d-parameter exponential family with densities

$$f_{\theta,\lambda}(x,y) = \exp(\theta x + <\lambda, y> - \psi(\theta,\lambda)) f_{0,0}(x,y) \tag{B.1}$$

where $x \in \mathcal{R}$, $\theta \in \mathcal{R}$, $y \in \mathcal{R}^{d-1}$ and $\lambda \in \mathcal{R}^{d-1}$. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an independent sample of size n from $f_{0,\lambda}(x,y)$ for some unknown λ. Let $S_X = \sum_{i=1}^n X_i$ and $S_Y = \sum_{i=1}^n Y_i$. Our aim is to find an approximation to the conditional distribution of S_X given S_Y.

The simplest approximation is based on the multivariate central limit theorem. This states, if the true parameters are θ and λ,

$$\frac{1}{\sqrt{n}} \begin{pmatrix} S_X - n\psi_1(\theta,\lambda) \\ S_Y - n\psi_2(\theta,\lambda) \end{pmatrix} \Rightarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \psi_{11}(\theta,\lambda) & \psi_{12}(\theta,\lambda) \\ \psi_{21}(\theta,\lambda) & \psi_{22}(\theta,\lambda) \end{pmatrix} \right) \tag{B.2}$$

as $n \to \infty$. Applying standard results for the conditional distributions of normally distributed random variables, we get the approximate conditional distribution

$$S_X \sim \mathcal{N}(n\mu_{X|y}, n\sigma^2_{X|y})$$

where

$$\mu_{X|y} = \psi_1(\theta,\lambda) + \psi_{12}(\theta,\lambda)\psi_{22}^{-1}(\theta,\lambda)(y - \psi_2(\theta,\lambda))$$

and

$$\sigma^2_{X|y} = \psi_{11}(\theta,\lambda) - \psi_{12}(\theta,\lambda)\psi_{22}^{-1}(\theta,\lambda)\psi_{21}(\theta,\lambda).$$

Here, subscripts of ψ denote derivatives; for example

$$\psi_{12}(\theta,\lambda) = \frac{\partial^2}{\partial\theta\partial\lambda}\psi(\theta,\lambda).$$

However, this central limit approach is somewhat inadequate. Firstly, by sufficiency, the distribution of S_X given S_Y does not depend on λ; however, the central limit approximation will depend on λ. This apparently contradictory situation arises because the approximation holds uniformly only for $y = \psi_2(\theta,\lambda)O_p(\frac{1}{\sqrt{n}})$ rather than all values of y. Secondly, likelihood ratios are not preserved; for example, the likelihood ratio statistic for testing $\mathcal{H}_0 : \theta = 0$ against the alternative $\mathcal{H}_1 : \theta \neq 0$ obtained using the central limit approximation to the conditional distribution is not correct. Thirdly, the approximation only takes into account mean and variance, and not the shape of the distributions, so we cannot expect good approximations in all circumstances and in particular we will not get good approximations to the tails of the conditional distribution.

Let $\hat{\theta}$ and $\hat{\lambda}$ be the maximum likelihood estimates of θ and λ, so if $S_X = nx$ and $S_Y = ny$,

$$x = \psi_1(\hat{\theta},\hat{\lambda}) \tag{B.3}$$

$$y = \psi_2(\hat{\theta},\hat{\lambda}). \tag{B.4}$$

Since we know the likelihood ratios and the central limit approximation is best when $(\hat{\theta}, \hat{\lambda})$ is close to (θ, λ), it makes sense to try to build this into the approximation. We can write

$$f_{0,0}^{(n)}(nx, ny) = \frac{f_{0,0}^{(n)}(nx, ny)}{f_{\hat{\theta},\hat{\lambda}}^{(n)}(nx, ny)} f_{\hat{\theta},\hat{\lambda}}^{(n)}(nx, ny) \tag{B.5}$$

where $f_{\theta,\lambda}^{(n)}$ denotes the n-fold convolution of $f_{\theta,\lambda}$. From (B.1) we have

$$\begin{aligned} \frac{f_{0,0}^{(n)}(nx, ny)}{f_{\hat{\theta},\hat{\lambda}}^{(n)}(nx, ny)} &= \exp\left(-n(\hat{\theta}x + < \hat{\lambda}, y > -\psi(\hat{\theta}, \hat{\lambda}))\right) \\ &= \exp(-nH_1(x, y)) \end{aligned} \tag{B.6}$$

where

$$\begin{aligned} H_1(x, y) &= \sup_{\theta,\lambda}(\theta x + \lambda y - \psi(\theta, \lambda)) \\ &= \hat{\theta}_1 x + \hat{\lambda}_1 y - \psi(\hat{\theta}_1, \hat{\lambda}_1). \end{aligned} \tag{B.7}$$

Applying (B.2) gives the approximation

$$\begin{aligned} f_{\hat{\theta},\hat{\lambda}}^{(n)}(nx, ny) &\approx \frac{1}{(2\pi n)^{d/2} |\psi''(\hat{\theta}, \hat{\lambda})|} \\ &= \frac{\sqrt{|H_1''(x, y)|}}{(2\pi n)^{d/2}} \end{aligned} \tag{B.8}$$

where $\psi''(\hat{\theta}, \hat{\lambda})$ is the second derivative matrix of ψ and $H_1''(x, y)$ is the second derivative matrix of H_1. The identity $\psi'' = (H_1'')^{-1}$ follows by differentiating (B.7) to get

$$H_1''(x, y) = \begin{pmatrix} \frac{\partial \hat{\theta}}{\partial x} & \frac{\partial \hat{\lambda}}{\partial x} \\ \frac{\partial \hat{\theta}}{\partial y} & \frac{\partial \hat{\lambda}}{\partial y} \end{pmatrix} \tag{B.9}$$

and differentiating (B.3) and (B.4) to get

$$\psi''(\hat{\theta}, \hat{\lambda}) = \begin{pmatrix} \frac{\partial x}{\partial \hat{\theta}} & \frac{\partial y}{\partial \hat{\theta}} \\ \frac{\partial x}{\partial \hat{\lambda}} & \frac{\partial y}{\partial \hat{\lambda}} \end{pmatrix}.$$

By substituting (B.6) and (B.8) into (B.5), we get the approximation

$$f_{0,0}^{(n)}(nx, ny) \sim \frac{\sqrt{|H_1''(x,y)|}}{(2\pi n)^{d/2}} \exp(-nH_1(x,y)) \tag{B.10}$$

Moreover, Borokov and Rogozin (1965) show the ratio of the two sides of (B.10) converges to 1 uniformly in x and y. They consider only the case where the support of $f_{0,0}$ is the whole plane; however, similar results hold much more generally. If the support of $f_{0,0}$ is a subset Ξ of $\mathcal{R}^2$, then (B.10) holds on the interior of Ξ, with the uniform convergence holding on closed subsets of the interior of Ξ. There is also a similar result for discrete distributions. For example, if X_1 has a Bernoulli distribution, then S_X has a binomial distribution and (B.10) is equivalent to using Stirling's formula to approximate the binomial coefficient.

Applying the same argument to the marginal distribution of S_Y, we get the approximation

$$f_Y^{(n)}(ny) \sim \frac{\sqrt{|H_0''(y)|}}{(2\pi n)^{(d-1)/2}} \exp(-nH_0(y))$$

where

$$\begin{aligned} H_0(y) &= \sup_\lambda(\lambda y - \psi(0,\lambda)) \\ &= \hat{\lambda}_0 y - \psi(0,\hat{\lambda}_0) \\ y &= \psi_2(0,\hat{\lambda}_0) \end{aligned} \tag{B.11}$$

The large deviation approximation to the density of S_X given S_Y is then

$$\begin{aligned} f_{X|Y}^{(n)}(nx|ny) &= \frac{f^{(n)}(nx,ny)}{f_Y^{(n)}(ny)} \\ &\sim \frac{1}{\sqrt{2\pi n}}\sqrt{\frac{|H_1''(x,y)|}{H_0''(y)}} \exp\left(-n(H_1(x,y) - H_0(y))\right). \end{aligned} \tag{B.12}$$

We note the term $n(H_1(x,y) - H_0(y))$ is simply the log-likelihood ratio for testing $\mathcal{H}_0 : \theta = 0$ vs $\mathcal{H}_1 : \theta \neq 0$. Since (B.10) holds uniformly in x and y, so does (B.12).

Lemma B.1 *In the case $d = 2$, so λ and y are scalars,*

$$\frac{|H_1''(x,y)|}{H_0''(y)} = \frac{\psi_{22}(0,\hat{\lambda}_0)}{\psi_{22}(\hat{\theta},\hat{\lambda})}\frac{\partial\hat{\theta}}{\partial x} \tag{B.13}$$

Proof: From the identity $\psi''(\hat{\theta},\hat{\lambda}) = H_1''(x,y)$ and a similar relation for H_0, we have

$$\frac{|H_1''(x,y)|}{H_0''(y)} = \frac{\psi_{22}(0,\hat{\lambda}_0)}{|\psi''(\hat{\theta},\hat{\lambda})|}. \tag{B.14}$$

Differentiating (B.3) and (B.4) with respect to x gives

$$\begin{aligned} 1 &= \psi_{11}(\hat{\theta},\hat{\lambda})\frac{\partial\hat{\theta}}{\partial x} + \psi_{12}(\hat{\theta},\hat{\lambda})\frac{\partial\hat{\lambda}}{\partial x} \\ 0 &= \psi_{21}(\hat{\theta},\hat{\lambda})\frac{\partial\hat{\theta}}{\partial x} + \psi_{22}(\hat{\theta},\hat{\lambda})\frac{\partial\hat{\lambda}}{\partial y} \end{aligned}$$

which gives

$$\frac{\partial\hat{\theta}}{\partial x} = \frac{\psi_{22}(\hat{\theta},\hat{\lambda})}{|\psi''(\hat{\theta},\hat{\lambda})|}. \tag{B.15}$$

Substituting (B.15) into (B.14) gives the result.

□

B.2 Tail Probabilities

A Laplace approximation coupled with large deviation theory gives simple approximations to $P(S_Y \geq ny)$ and $P(S_X \geq nx|S_Y = ny)$.

Lemma B.2 *Suppose $\hat{\lambda}_0 > 0$. Then if Y_1 has a continuous distribution,*

$$P(S_Y \geq ny) = \frac{1}{\hat{\lambda}_0}\sqrt{\frac{H_0''(y)}{2\pi n}}e^{-nH_0(y)}(1+o(1)) \quad \text{as } n \to \infty$$

and if Y_1 has an arithmetic distribution of span 1,

$$P(S_Y \geq ny) = \frac{e^{-\hat{\lambda}_0(\lceil ny\rceil - ny)}}{1-e^{-\hat{\lambda}_0}}\sqrt{\frac{H_0''(y)}{2\pi n}}e^{-nH_0(y)}(1+o(1)) \quad \text{as } n \to \infty.$$

If $\hat{\lambda}_0 < 0$, we get similar approximations for $P(S_Y \leq ny)$.

Proof: We choose n large enough so that

$$f_n(ny) \geq (1-\epsilon)\sqrt{\frac{H_0''(y)}{2\pi n}} e^{-nH_0(y)}$$

for all y. Choosing y such that $\hat{\lambda}_0 > 0$ and $y' > y$, we have

$$\begin{aligned}
P(S_Y \geq ny) &\geq P(ny' \geq S_Y \geq ny) \\
&= \int_{ny}^{ny'} f_Y^{(n)}(x)dx \\
&= n\int_y^{y'} f_Y^{(n)}(nu)du \\
&\geq (1-\epsilon)n\int_y^{y'} \sqrt{\frac{H_0''(u)}{2\pi n}} \exp(-nH_0(u))du \\
&\geq (1-\epsilon)\sqrt{\frac{\inf_{y<u<y'} H_0''(u)}{2\pi n}} \int_y^{y'} e^{-nH_0(u)}du. \qquad \text{(B.16)}
\end{aligned}$$

Since H_0 is a convex function, we have

$$H_0(u) \leq \frac{u-y}{y'-y}H_0(y') + \frac{y'-u}{y'-y}H_0(y)$$

and therefore

$$\int_y^{y'} e^{-nH_0(u)}du \leq \frac{y'-y}{H_0(y')-H_0(y)}(e^{-nH_0(y)} - e^{-nH_0(y')}). \qquad \text{(B.17)}$$

Substituting (B.17) into (B.16) and letting $y' \to y$ slowly enough so that $n(H_0(y') - H_0(y)) \to \infty$ gives a lower bound. Since $H_0(u) \geq H_0(y) + (u-y)H_0'(y)$, we have for n sufficiently large,

$$P(S_Y \geq ny) \leq (1+\epsilon)n\int_Y^{\infty} \sqrt{\frac{H_0''(u)}{2\pi n}} e^{-n(H_0(y)+(u-y)H_0'(y))}du.$$

The discrete case is similar, but with summations instead of integrals.

□

The conditional version of Lemma B.2 says

$$P(S_X \geq nx | S_Y = ny) = \frac{1}{\hat{\theta}_1}\sqrt{\frac{|H_1''(x,y)|}{2\pi n H_0''(y)}} e^{-n(H_1(x,y)-H_0(y))}(1+o(1))$$

in the continuous case and a similar result in the discrete case.

Examples: Suppose Y_1 is distributed $\mathcal{N}(0,1)$ so $\psi(0,\lambda) = \frac{1}{2}\lambda^2$, $\hat{\lambda}_0 = y$ and $H_0(y) = \frac{1}{2}y^2$. By B.2, we get

$$P(S_Y \geq ny) \approx \frac{1}{y\sqrt{2\pi n}} e^{-\frac{1}{2}ny^2}.$$

Letting $c = \sqrt{n}y$, this becomes the familiar approximation

$$1 - \Phi(c) \approx \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2}c^2}$$

where $\Phi(c)$ is the standard normal distribution function.

Now suppose $p = P(Y_1 = 1) = 1 - P(Y_1 = 0)$ so S_Y is distributed $\mathcal{B}(n,p)$. Then under P_λ, S_Y is distributed $\mathcal{B}(n,p_1)$ where

$$\lambda = \log\left(\frac{p_1(1-p)}{p(1-p_1)}\right)$$

and

$$\psi(\lambda) = \log\left(\frac{1-p}{1-p_1}\right).$$

The maximum likelihood estimate of λ when $S_Y = ny$ is

$$\hat{\lambda} = \log\left(\frac{y(1-p)}{p(1-y)}\right).$$

This gives

$$H(y) = y\log\left(\frac{y}{p}\right) + (1-y)\log\left(\frac{1-y}{1-p}\right)$$

and hence if $y > p$ and ny is an integer,

$$P(S_Y \geq ny) \approx \frac{y(1-p)}{y-p}\frac{1}{\sqrt{2\pi ny(1-y)}}\left(\frac{p}{y}\right)^{ny}\left(\frac{1-p}{1-y}\right)^{n(1-y)}. \tag{B.18}$$

We see that (B.18) is very similar to the upper bound given in Lemma A.3.

B.3 Example: Log-Linear Model

Our interest in (B.10) arises from its applicability to finding the conditional distribution of $X(\tau)$ given $X(T) = n$ and $S_n = ny$ for the log-linear model (1.2). We obtain an exact expression for this distribution in Lemma B.3 below; however the form is not particularly useful either for theoretical or numerical purposes. We compute the approximation (B.10) for this special case and compare the results.

Lemma B.3 *Let $X(t)$ be a Poisson process with rate $\lambda(t)$ given by (1.2) with $\delta = 0$. Denote by $f_n(x)$ the density of the sum of n i.i.d uniform $[0, 1]$ random variables,*

$$f_n(x) = \frac{1}{(n-1)!}\sum_{k=0}^{n}\binom{n}{k}(-1)^k(x-k)_+^{n-1} \tag{B.19}$$

for $0 < x < n$. Note that (B.19) holds for all x, including $x < 0$ and $x > n$. Then

$$\begin{aligned} &P_0^{(n)}(X(t) = j|S_n = ny) \\ &= \binom{n}{j}\frac{t^{j-1}(T-t)^{n-j-1}}{T^{n-1}f_n(\frac{ny}{T})}\int_0^{ny} f_j(\frac{x}{t})f_{n-j}(\frac{ny-x-(n-j)t}{T-t})dx \end{aligned} \tag{B.20}$$

for $n(1-\frac{y}{t}) \vee 1 \le j \le n\frac{T-y}{T-t} \wedge (n-1)$. If $y > t$ then

$$P_0^{(n)}(X(t) = 0|S_n = ny) = \left(1-\frac{t}{T}\right)^{n-1}\frac{f_n(\frac{n(y-t)}{T-t})}{f_n(\frac{ny}{T})} \tag{B.21}$$

and if $y < t$ then

$$P_0^{(n)}(X(t) = n|S_n = ny) = \left(\frac{t}{T}\right)^{n-1}\frac{f_n(\frac{ny}{t})}{f_n(\frac{ny}{T})}. \tag{B.22}$$

Proof: Conditional on $X(T) = n$, the joint density of the event times is

$$f(t_1, \ldots, t_n) = n!\left(\frac{b}{e^{bT}-1}\right)^n e^{b\sum_{i=1}^n t_i},\ 0 \le t_1 \le \ldots \le t_n \le T.$$

If $b = 0$, then S_n has the same distribution as the sum of n independent uniform $[0, T]$ random variables. By a likelihood ratio argument, the density of S_n is

$$f_{S_n}(ny|X(T) = n) = \left(\frac{bT}{e^{bT} - 1}\right)^n e^{bny} \frac{1}{T} f_n(\frac{ny}{T}). \tag{B.23}$$

For $0 < t < T$,

$$P_0^{(n)}(X(t) = j|S_n = ny) = \frac{f_{S_n,X(t)}(ny, j|X(T) = n)}{f_{S_n}(ny|X(T) = n)}.$$

Letting $S_n = Y_1 + Y_2$ where $Y_1 = \sum_{i=1}^{j} T_i$ and $Y_2 = Y - Y_1$, we get

$$f_{S_n,X(t)}(ny, j|X(T) = n) = f_{S_n}(ny|X(t) = j, X(T) = n)P_0^{(n)}(X(t) = j)$$

and

$$\begin{aligned} & f_{S_n}(ny|X(t) = j, X(T) = n) \\ = & \int_0^{ny} f_{Y_1}^{(n)}(x|X(t) = j) f_{Y_2}^{(n)}(ny - x|X(t) = j). \end{aligned}$$

By appropriate transformations of (B.23) we get

$$\begin{aligned} f_{Y_1}^{(n)}(x|X(t) = j) &= \left(\frac{bt}{e^{bt} - 1}\right)^j e^{bx} \frac{1}{t} f_j(\frac{x}{t}) \\ f_{Y_2}^{(n)}(ny - x|X(t) = j) &= \left(\frac{b(T - t)}{e^{bT} - e^{bt}}\right)^{n-j} e^{b(ny-x)} \frac{1}{T - t} f_{n-j}(\frac{ny - x - (n - j)t}{T - t}) \end{aligned}$$

and therefore

$$\begin{aligned} P_0^{(n)}(X(t) = j|S_n = ny) = & \frac{\binom{n}{j} \left(\frac{e^{bt}-1}{e^{bT}-1}\right)^j \left(\frac{e^{bT}-e^{bt}}{e^{bT}-1}\right)^{n-j}}{\left(\frac{bT}{e^{bT}-1}\right)^n e^{nby} \frac{1}{T} f_n(\frac{ny}{T})} \\ & \times \int_0^{ny} \frac{e^{bny} b^n t^j (T - t)^{n-j}}{(e^{bt} - 1)^j (e^{bT} - e^{bt})^{n-j}} \frac{f_j(\frac{x}{t}) f_{n-j}(\frac{ny-x-(n-j)t}{T-t})}{t(T - t)} dx \end{aligned}$$

which simplifies to (B.20). The special cases for $j = 0$ and $j = n$ are similar.

□

We apply the results of the previous section to approximate the conditional distribution of $X(t)$ given $\sum_{i=1}^{n} T_i$ for the log-linear model (1.2). We represent $T_1, \ldots, T_n$ as the order statistics of a sample of size n from a density $f_{0,b}(t)$, where

$$f_{\delta,b}(t) = \frac{be^{\delta I(t>\tau)+bt}}{e^{\delta}(e^{bT} - e^{b\tau}) + e^{b\tau} - 1} I_{[0,T]}(t)$$

and hence

$$\psi(\delta, b) = \log(e^{b\tau} - 1 + e^{\delta}(e^{bT} - e^{b\tau})) - \log(b). \tag{B.24}$$

Differentiating (B.24), we get

$$\psi_{22}(\delta, b) = \frac{1}{b^2} - \frac{\tau^2 e^{b\tau} - e^{\delta}((T-\tau)^2 e^{b(T+\tau)} - T^2 e^{bT} + \tau^2 e^{b\tau}) + e^{2\delta}(T-\tau)^2 e^{b(T+\tau)}}{(e^{b\tau} - 1 + e^{\delta}(e^{bT} - e^{b\tau}))^2}.$$

This gives

$$\psi_{22}(0, \hat{b}_0) = \frac{1}{\hat{b}_0^2} - \frac{T^2 e^{\hat{b}_0 T}}{(e^{\hat{b}_0 T} - 1)^2} \tag{B.25}$$

and

$$\begin{aligned} \psi_{22}(\hat{\delta}, \hat{b}) &= \frac{1}{\hat{b}^2} + \frac{p\tau^2 e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} + \frac{(1-p)(T^2 e^{\hat{b}T} - \tau^2 e^{\hat{b}\tau})}{e^{\hat{b}T} - e^{\hat{b}\tau}} \\ &\quad - \left(\frac{p\tau e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} + \frac{(1-p)(Te^{\hat{b}T} - \tau e^{\hat{b}\tau})}{e^{\hat{b}T} - e^{\hat{b}\tau}} \right)^2 \end{aligned} \tag{B.26}$$

where $p = \frac{X(\tau)}{n}$ and

$$e^{\hat{\delta}} = \frac{1-p}{p} \frac{e^{\hat{b}\tau} - 1}{e^{\hat{b}T} - e^{\hat{b}\tau}}.$$

Differentiating (1.11), we get

$$-\frac{\partial \hat{\delta}}{\partial p} = \frac{1}{p(1-p)} + \frac{\left(\frac{\tau e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} - \frac{Te^{\hat{b}T} - \tau e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}} \right)^2}{\frac{1}{\hat{b}^2} - \frac{p\tau^2 e^{\hat{b}\tau}}{(e^{\hat{b}\tau} - 1)^2} - \frac{(1-p)(T-\tau)^2 e^{\hat{b}(T+\tau)}}{(e^{\hat{b}T} - e^{\hat{b}\tau})^2}}. \tag{B.27}$$

Substituting (B.25), (B.26) and (B.27) into (B.12) leads to the approximation

$$\begin{aligned} P_0^{(n)}(X(\tau) = np | S_n = ny) &= P_0^{(n)}(n - X(\tau) = n(1-p) | S_n = ny) \\ &\approx \frac{1}{\sqrt{2\pi n}} \sqrt{-\frac{\psi_{11}(0, \hat{b}_0)}{\psi_{11}(\hat{\delta}, \hat{b})} \frac{\partial \hat{\delta}}{\partial p}} e^{-l(\tau)}. \end{aligned} \tag{B.28}$$

For the special cases $p = 0$ and $p = 1$ we can use large deviation approximations to f_n in conjunction with (B.21) and (B.22).

	$y = 0.5, t = 0.5$			$y = \frac{2}{3}, t = \frac{1}{3}$		
j	(B.20)	(B.28)	Ratio	(B.20)	(B.28)	Ratio
0	0.0000	NA	NA	0.1315	NA	NA
1	1.221(-7)	1.350(-7)	0.9043	0.4838	0.5249	0.9218
2	2.825(-4)	2.964(-4)	0.9531	0.3449	0.3608	0.9559
3	2.193(-2)	2.268(-2)	0.9670	3.967(-2)	4.100(-2)	0.9674
4	0.2318	0.2381	0.9737	1.761(-4)	1.791(-4)	0.9832
5	0.4923	0.5046	0.9756	0.0000	NA	NA
6	0.2318	0.2381	0.9737	0.0000	NA	NA

Table B.1: Exact and Approximate Conditional Distributions of $X(t)$ with $n = 10$. We evaluate $P_0^{(n)}(X(t) = j | S_n = ny)$ using the exact formula (B.20) and the large deviation approximation (B.28). Numbers in parenthesis represent powers of 10. For example, $1.221(-7) = 1.221 \times 10^{-7}$.

We show the performance of the approximation (B.28) in Table B.1. Even for the small sample size ($n = 10$), the approximation is performing acceptably, with relative errors less than 10%.

We illustrate the uniform convergence of the large deviation approximation on intervals excluding 0 and 1 in Table (B.2). We have fixed $y = 0.5$ and for several n and ϵ calculated the supremum and infimum over $\epsilon \leq p \leq 1 - \epsilon$ of the ratio of the true probability (B.20) to the large deviation approximation (B.28). For uniform convergence, both the supremum and infimum should converge to 1 as $n \to \infty$. The uniform convergence seems reasonably clear when $\epsilon = 0.2$ and $\epsilon = 0.3$ but less so

ϵ		n				
		10	20	30	50	70
0.3	sup	0.976	0.988	0.992	0.995	0.999
	inf	0.967	0.983	0.989	0.993	0.998
0.2	sup	0.976	0.988	0.992	0.995	0.999
	inf	0.953	0.980	0.983	0.991	0.996
0.1	sup	0.976	1.098	0.992	1.017	1.036
	inf	0.904	0.936	0.910	0.975	0.986

Table B.2: Uniform Convergence of Large Deviation Approximations

when $\epsilon = 0.1$. The exact formula becomes unstable for $n > 70$. By contrast, a central limit approximation will only converge uniformly for $\frac{i}{n}$ in intervals which shrink to the point $\frac{1}{n}E(X(t)|S_n = ny)$ as $n \to \infty$.

To establish the validity of the significance level and power approximations in Chapter 3 and the confidence level approximations in Chapter 4 for the log-linear model, we need to show that conditionally on $S_n = ny$, $X(t)$ is locally Poisson. See Section A.3 for the appropriate definitions.

Theorem B.1 *Suppose $X(t) = m = np$, $S_n = ny$ and $n \to \infty$ with p and y fixed. Then under $P_0^{(m,n)}(\,\cdot\,|S_n = ny)$, X is locally Poisson at t with rates*

$$\begin{aligned}\mu^-(t,p) &= \frac{p\hat{b}e^{\hat{b}t}}{e^{\hat{b}t}-1}\\ \mu^+(t,p) &= \frac{(1-p)\hat{b}e^{\hat{b}t}}{e^{\hat{b}T}-e^{\hat{b}t}}\end{aligned}$$

on the left and right respectively, and the estimate $\hat{b}$ is calculated with a change point at t..

Proof: To simplify the notation, we will only show for each $S > 0$, $X(t) - X(t - \frac{S}{n})$ converges in law to a Poisson random variable with mean $\mu^-(t,p)S$. Establishing that (A.48) holds more generally and the independence of the left and right processes is similar. We will not be completely rigorous; in particular, we will omit some uniformity conditions.

We have

$$\begin{aligned}
&P_0^{(m,n)}\left(X(t) - X(t - \frac{S}{n}) = k | S_n = ny\right) \\
&= \frac{P_0^{(m,n)}(X(t) - X(t - \frac{S}{n}) = k, S_n = ny)}{P_0^{(m,n)}(S_n = ny)} \\
&= \frac{P_0^{(m,n)}(X(t) - X(t - \frac{S}{n}) = k) P_0^{(m-k,n-k)}(\tilde{S}_{n-k} = ny - kt)}{P_0^{(m,n)}(S_n = ny)} \qquad \text{(B.29)}
\end{aligned}$$

where $\tilde{S}_{n-k}$ denotes the sum of $n-k$ random variables $\tilde{X}_1, \ldots, \tilde{X}_{n-k}$ from a density

$$\tilde{f}(x) = \frac{be^{bx}}{e^{bT} - e^{bt} + e^{bt - S/n} - 1} I\left([0, t - \frac{S}{n}] \cup [t, T]\right).$$

We treat each of the terms in (B.29) separately. Under $P_0^{(m,n)}$, $X(t) - X(t - \frac{S}{n})$ has a $\mathcal{B}(m, (e^{bt} - e^{b(t - S/n)})/(e^{bt} - 1))$ distribution and the standard Poisson limit theorem gives

$$P_0^{(m,n)}\left(X(t) - X(t - \frac{S}{n}) = k\right) \to \frac{1}{k!}\left(\frac{pbSe^{bt}}{e^{bt} - 1}\right)^k \exp\left(-\frac{pbSe^{bt}}{e^{bt} - 1}\right) \qquad \text{(B.30)}$$

as $n \to \infty$.

We write the second term from the numerator of (B.29) as

$$P_0^{(m-k,n-k)}(\tilde{S}_{n-k} = ny - kt) = \frac{P_0^{(n-k)}(\tilde{S}_{n-k} = ny - kt, \tilde{X}(t) = m - k)}{P_0^{(n-k)}(\tilde{X}(t) = m - k)}. \qquad \text{(B.31)}$$

Embedding $\tilde{f}(x)$ in a two parameter exponential family with a change point of size δ at t, we can approximate the numerator of (B.31) by

$$P_0^{(n-k)}(\tilde{S}_{n-k} = ny - kt, \tilde{X}(t) = m - k) \sim \frac{|\tilde{H}''(\frac{m-k}{n-k}, \frac{ny-kt}{n-k})|^{1/2}}{2\pi n}$$

$$\times \exp\left(-(n-k)((\tilde{b}-b)\frac{ny-kt}{n-k}+\tilde{\delta}\frac{n-m}{n-k}-(\tilde{\psi}(\tilde{\delta},\tilde{b})-\tilde{\psi}(0,b)))\right)$$

as $n \to \infty$, where

$$\tilde{\psi}(\delta,b)=\log\left(\frac{e^{b(t-S/n)}-1+e^{\delta}(e^{bT}-e^{bt})}{b}\right),$$

$\tilde{b}$ and $\tilde{\delta}$ are the maximum likelihood estimates of b and δ and $\tilde{H}(p,y)=\tilde{b}y+\tilde{\delta}(1-p)-\tilde{\psi}(\tilde{\delta},\tilde{b})$. A straightforward calculation shows

$$\tilde{\psi}(\delta,b)=\psi(\delta,b)-\frac{bSe^{bt}}{n(e^{bt}-1+e^{\delta}(e^{bT}-e^{bt}))}+o(\frac{1}{n})$$

and hence

$$\begin{aligned}
P_0^{(n-k)}(\tilde{S}_{n-k}=ny-kt,\tilde{X}(t)=m-k) &\sim \frac{|\tilde{H}''(\frac{m-k}{n-k},\frac{ny-kt}{n-k})|^{1/2}}{2\pi n}\\
&\times \exp\left(-(\tilde{b}-b)(ny-kt)-\tilde{\delta}(n-m)+(n-k)(\psi(\tilde{\delta},\tilde{b})-\psi(0,b))\right)\\
&\times \exp\left(-\frac{\tilde{b}Se^{\tilde{b}t}}{e^{\tilde{b}t}-1+e^{\tilde{\delta}}(e^{\tilde{b}T}-e^{\tilde{b}t})}+\frac{bSe^{bT}}{e^{bT}-1}\right).
\end{aligned} \tag{B.32}$$

Under $P_0^{(n-k)}$, $\tilde{X}(t)$ has a binomial distribution with success probability

$$\frac{e^{b(t-s/N)}-1}{e^{b(t-S/n)}-1+e^{bT}-e^{bt}}.$$

This gives

$$\begin{aligned}
&P_0^{(n-k)}(\tilde{X}(t)=m-k)\\
&= \binom{n-k}{m-k}\left(\frac{e^{b(t-S/n)}-1}{e^{b(t-S/n)}-1+e^{bT}-e^{bt}}\right)^{m-k}\left(\frac{e^{bT}-e^{bt}}{e^{b(t-S/n)}-1+e^{bT}-e^{bt}}\right)^{n-m}\\
&\sim \binom{n-k}{m-k}\left(\frac{e^{bt}-1}{e^{bT}-1}\right)^{m-k}\left(\frac{e^{bT}-e^{bt}}{e^{bT}-1}\right)^{n-m}\\
&\quad\times \exp\left(\frac{bSe^{bt}}{e^{bT}-1}\right)\exp\left(-\frac{bpSe^{bt}}{e^{bt}-1}\right).
\end{aligned} \tag{B.33}$$

as $n \to \infty$.

We can write the denominator of (B.29) as

$$P_0^{(m,n)}(S_n = ny) = \frac{P_0^{(n)}(S_n = ny, X(t) = m)}{P_0^{(n)}(X(t) = m)}$$

and using the large deviation approximation,

$$\begin{aligned} &P_0^{(n)}(S_n = ny, X(t) = m) \\ &\sim \quad \frac{|H''(1-p,y)|}{2\pi n} \exp\left(-n((\hat{b}-b)y + \hat{\delta}(1-p) - (\psi(\hat{\delta},\hat{b}) - \psi(0,b)))\right) \end{aligned} \tag{B.34}$$

as $n \to \infty$, and

$$P_0^{(n)}(X(t) = m) = \binom{n}{m} \left(\frac{e^{bt}-1}{e^{bT}-e^{bt}}\right)^m \left(\frac{e^{bT}-e^{bt}}{e^{bT}-1}\right)^{n-m}. \tag{B.35}$$

Substituting (B.30), (B.32), (B.33), (B.34) and (B.35) into (B.29) gives

$$\begin{aligned} &P_0^{(m,n)}\left(X(t) - X(t - \frac{S}{n}) = k | S_n = ny\right) \\ &\sim \quad \frac{1}{k!}\left(\frac{pbSe^{bt}}{e^{bt}-1}\right)^k \exp\left(-\frac{pbSe^{bt}}{e^{bt}-1}\right) \\ &\quad \times \frac{n!(m-k)!}{m!(n-k)!}\left(\frac{e^{bt}-1}{e^{bT}-1}\right)^k \exp\left(-\frac{bSe^{bt}}{e^{bT}-1}\right) \exp\left(\frac{pbSe^{bt}}{e^{bt}-1}\right) \\ &\quad \times \exp\left((\hat{b}-b)ny - (\tilde{b}-b)(ny-kt) + n(1-p)(\hat{\delta}-\tilde{\delta})\right) \\ &\quad \times \exp\left((n-k)(\psi(\tilde{\delta},\tilde{b}) - \psi(0,b)) - n(\psi(\hat{\delta},\hat{b}) - \psi(0,b))\right) \\ &\quad \times \exp\left(-\frac{S\tilde{b}e^{\tilde{b}t}}{e^{\tilde{b}t} - 1 + e^{\tilde{\delta}}(e^{\tilde{b}T} - e^{\tilde{b}t})} + \frac{sbe^{bt}}{e^{bT}-1}\right) \end{aligned} \tag{B.36}$$

$$\begin{aligned} &\sim \quad \frac{1}{k!}\left(\frac{pbSe^{bt}}{e^{bT}-1}\right)^k \left(\frac{n}{m}\right)^k \exp\left(k(\psi(0,b) - bt)\right) \\ &\quad \times \exp\left(k(\tilde{b}t - \psi(\tilde{\delta},\tilde{b})) - \frac{S\tilde{b}e^{\tilde{b}t}}{e^{\tilde{b}t} - 1 + e^{\tilde{\delta}}(e^{\tilde{b}T} - e^{\tilde{b}t})}\right) \\ &\quad \times \exp\left((\hat{b}-\tilde{b})ny + n(1-p)(\hat{\delta}-\tilde{\delta}) + n(\psi(\tilde{\delta},\tilde{b}) - \psi(\hat{\delta},\hat{b}))\right). \end{aligned} \tag{B.37}$$

The first line of (B.36) arises from (B.30), the second line comes from (B.33) and (B.35) and the remaining lines come from (B.32) and (B.34). It is easy to show

$\tilde{b} - \hat{b} = O(\frac{1}{n})$ and $\tilde{\delta} - \hat{\delta} = O(\frac{1}{n})$ as $n \to \infty$. This implies

$$\psi(\tilde{\delta}, \tilde{b}) = \psi(\hat{\delta}, \hat{b}) + (\tilde{\delta} - \hat{\delta})\psi_1(\hat{\delta}, \hat{b}) + (\tilde{b} - \hat{b})\psi_2(\hat{\delta}, \hat{b}) + O(\frac{1}{n})$$

and hence we can rewrite (B.37) as

$$\begin{aligned} & P_0^{(m,n)}\left(X(t) - X(t - \frac{S}{n}) = k | S_n = ny\right) \\ \sim \; & \frac{1}{k!}\left(\frac{S\tilde{b}e^{\tilde{b}t}}{e^{\tilde{b}t} - 1 + e^{\tilde{\delta}}(e^{\tilde{b}T} - e^{\tilde{b}t})}\right)^k \exp\left(-\frac{S\tilde{b}e^{\tilde{b}t}}{e^{\tilde{b}t} - 1 + e^{\tilde{\delta}}(e^{\tilde{b}T} - e^{\tilde{b}t})}\right) \\ & \times \exp\left(n(\hat{b} - \tilde{b})(y - \psi_2(\hat{\delta}, \hat{b})) + n(\hat{\delta} - \tilde{\delta})(1 - p - \psi_1(\hat{\delta}, \hat{b}))\right) \\ \to \; & \frac{1}{k!}\left(\frac{S\hat{b}e^{\hat{b}t}}{e^{\hat{b}t} - 1 + e^{\hat{\delta}}(e^{\hat{b}T} - e^{\hat{b}t})}\right)^k \exp\left(-\frac{S\hat{b}e^{\hat{b}t}}{e^{\hat{b}t} - 1 + e^{\hat{\delta}}(e^{\hat{b}T} - e^{\hat{b}t})}\right) \end{aligned}$$

as $n \to \infty$. The result follows by noting

$$e^{\hat{b}t} - 1 + e^{\hat{\delta}}(e^{\hat{b}T} - e^{\hat{b}t}) = \frac{1}{p}(e^{\hat{b}t} - 1).$$

□

If we are prepared to assume the process $X(t)$ is locally Poisson, it is much easier to derive the local rate. We can write

$$\mu(t, p_t) = \lim_{n \to \infty} \frac{1}{n} \lim_{s \to t^-} \frac{1}{t - s} P_0^{(m,n)}(X(s) = m - 1 | S_n = ny)$$

where $m = np_t$. A straightforward calculation then shows

$$\begin{aligned} n\mu(t, p_t) &= \lim_{s \to t^-} \frac{1}{t - s} \frac{f_{S_n}^{(m,n)}(ny | X(s) = m - 1) P_0^{(m,n)}(X(s) = m - 1)}{f_{S_n}^{(m,n)}(ny)} \\ &= \frac{f_{S_n}^{(m-1,n-1)}(ny - t)}{f_{S_n}^{(m,n)}(ny)} \frac{m(e^{bt} - 1)^{n-1}}{(e^{bt} - 1)^n} \lim_{s \to t^-} \frac{e^{bt} - e^{bs}}{t - s} \\ &= \frac{f_{S_n}^{(n-1)}(ny - t, X(t) = m - 1)}{f_{S_n}^{(n)}(ny, X(t) = m)} \frac{P^{(n)}(X(t) = m)}{P^{(n-1)}(X(t) = m - 1)} \frac{mbe^{bt}}{e^{bt} - 1}. \end{aligned}$$

We use the large deviation approximation to the distribution of S_n to get

$$\begin{aligned} n\mu(t,p_t) &= \frac{\sqrt{|H_1''(\frac{m-1}{n-1},\frac{ny-t}{n-1})|}\exp(-(n-1)H_1(\frac{m-1}{n-1},\frac{ny-t}{n-1}))}{\sqrt{|H_1''(\frac{m}{n},y)|}\exp(-nH_1(\frac{m}{n},y))}(1+o(1)) \\ &\quad \times\frac{\exp(b(ny-t)-(n-1)\psi(0,b))}{\exp(n(by-\psi(0,b)))}\frac{n(e^{bt}-1)}{m(e^{bT}-1)}\frac{mbe^{bt}}{e^{bt}-1} \\ &\sim n\exp\left(H_1(p_t,y)-(1-p_t)\hat{\delta}+(t-y)\hat{b}\right) \end{aligned} \tag{B.38}$$

as $n \to \infty$, using the continuity of H and its derivatives, and using $\frac{\partial H_1(p_t,y)}{\partial p_t} = -\hat{\delta}$ and $\frac{\partial H_1(p_t,y)}{\partial y} = \hat{b}$. We have $H_1(p_t,y) = (1-p_t)\hat{\delta} + \hat{b}y - \psi(\hat{\delta},\hat{b})$ and therefore (B.38) becomes

$$\mu(t,p_t) = e^{\hat{b}t-\psi(\hat{\delta},\hat{b})}.$$

From (B.24) and (1.11) we get

$$\psi(\hat{\delta},\hat{b}) = -\log\left(\frac{bp_t}{e^{bt}-1}\right)$$

and hence

$$\mu(t,p_t) = p_t\frac{be^{bt}}{e^{bt}-1}.$$

Appendix C

Computational Details

Implementing many of the procedures described in this thesis requires some numerical computation. For the most part, the numerical methods used are fairly standard and are described in many introductory numerical analysis books; see for example Conte and De Boor (1980). We provide some additional details of the computing here.

All the computing was performed on Sun 3/50 work stations. The New S Language (Becker, Chambers and Wilks, 1988) was used to produce the graphics and for some of the data manipulation. The C programming language was used for numerical procedures such as solving non-linear equations and numerical integration. Double precision arithmetic with an MC68881 floating point processor was used for all floating point computations.

C.1 Significance and Power Calculations

The evaluation of (3.6) is straightforward. The values of p_t can be found very easily from (3.5) using Newton's method. Starting values can be obtained using the relation

$$f(t,p) \approx \frac{1}{2}\frac{(p-t)^2}{t(1-t)}$$

if p is close to t, together with the fact that $p_t < t$. Since $f(t,p)$ is convex, Newton's method will always converge to the correct solution, provided we make sure at each iteration we have $p > 0$. If some step leads to $p < 0$, we simply reduce the step size. Convergence is usually very fast. The numerical integration is very straightforward. Since there is a singularity in the case $\tau_0 = t_0$, we use the transformation $u = \sqrt{t - t_0}$. Using Simpson's rule with 50 intervals to do the numerical integration, computing the second order approximation for fixed n and all 5 values of c takes about 2.2 seconds of cpu time and is independent of n. The values in Tables 3.1 and 3.2 can be reproduced with as few as 20 intervals. By contrast, performing the exact calculation for one value of c takes 1.5, 8.8 and 35.4 seconds for $n = 20$, 50 and 100, and increases quadratically in n.

The significance level calculation (3.16) is only slightly more complicated to evaluate. We first evaluate $\hat{b}_0$ by solving (1.13). For each t we can find b and p_t by solving (3.9) and (3.10) numerically. Since (3.9) is linear in p we could write a single non-linear equation for b; however, applying the bivariate Gauss-Newton method directly works reasonably well. A starting value for p can be found using the relation

$$f(t,p,b) \approx \frac{1}{2}\frac{(p - \frac{e^{b_0 t}-1}{e^{b_0 T}-1})^2}{\sigma(t,t)}$$

for b close to b_0 and p close to $\frac{e^{b_0 t}-1}{e^{b_0 T}-1}$. Here, $f(t,p,b)$ is as defined in the proof of

Lemma 3.1. A Taylor series expansion suggests a starting value for b:

$$b \approx b_0 + \left(\frac{\partial f(t,b,p)}{\partial b}\bigg|_{b=b_0} \right)^{-1} (\frac{\eta^2}{2} - f(t, b_0, p)).$$

The simulation of the significance levels presents some difficulty since we are evaluating the significance level conditionally on $\sum_{i=1}^{n} T_i = ny$. This requires conditional sampling of the T_i which is a difficult problem. However, we circumvented this problem by sampling unconditionally with $b = b_0$ and rejecting the sample if $|\frac{1}{n}\sum_{i=1}^{n} T_i - y| > \epsilon$. For the results in Table 3.3, we chose $\epsilon = 0.05$, which resulted in about 10% rejection. The results indicate that for fixed c the dependence of the significance level on y is small and differences between approximations is much larger than any error that may be caused by the sampling scheme.

We briefly describe the numerical technique we used to evaluate the exact Gaussian probabilities given in Tables 3.1 and 3.2. We assume $X(t)$ is a continuous, mean 0, Gaussian Markov process with covariance function $\sigma(s,t)$, and $\mathcal{T} = \inf\{t \geq \tau_0 : |X(t)| \geq c(t)\}$. Letting $G(t)$ be the distribution function of $\mathcal{T}$, we have

$$\frac{1}{2}G(t) = P(\mathcal{T} < t, X(\mathcal{T}) = c(\mathcal{T})) = P(\mathcal{T} < t, X(\mathcal{T}) = -c(\mathcal{T}))$$

and a simple conditioning argument in conjunction with the strong Markov property of $X(t)$ gives

$$\begin{aligned}
&P(X(t) \geq c(t)) \\
&= \int_{c(\tau_0)}^{\infty} (P(X(t) > c(t)|X(\tau_0) = x) + P(X(t) > c(t)|X(\tau_0) = -x))dF_{\tau_0}(x) \\
&\quad + \frac{1}{2}\int_{\tau_0}^{t} (P(X(t) > c(t)|X(s) = c(s)) + P(X(t) > c(t)|X(s) = -c(s)))dG(t) \\
&= \int_{c(\tau_0)}^{\infty} P(|X(t)| > c(t)|X(\tau_0) = x)dF_{\tau_0}(x) \\
&\quad + \frac{1}{2}\int_{\tau_0}^{t} P(|X(t)| > c(t)|X(s) = c(s))dG(s) \\
&= \int_{c(\tau_0)}^{\infty} K(t, c(t); \tau_0, u)dF_{\tau_0}(x) + \frac{1}{2}\int_{\tau_0}^{t} K(t, c(t); s, c(s))dG(s) \qquad \text{(C.1)}
\end{aligned}$$

where $F_{\tau_0}(x)$ is the distribution of $X(\tau_0)$ and

$$\begin{aligned} K(t, c(t); s, u) &= P(|X(t)| > c(t)|X(s) = u) \\ &= 1 - \Phi\left(\frac{c(t) - \mu(t|s, u)}{\sigma(t|s)}\right) + \Phi\left(\frac{-c(t) - \mu(t|s, u)}{\sigma(t|s)}\right). \end{aligned}$$

Equations of the form (C.1) have been studied by Park and Schuurmann (1976). Numerical solutions of (C.1) are constructed using midpoint approximations. Suppose $\tau_0 = t_0 < \ldots < t_n$, $u_i = \frac{1}{2}(t_i + t_{i-1})$ and $\Delta_i = G(t_i) - G(t_{i-1})$. Then we approximate (C.1) by

$$\begin{aligned} &P(X(t_n) \geq c(t_n)) \\ &\approx \int_{c(\tau_0)}^{\infty} K(t_n, c(t_n); \tau_0, x) dF_{\tau_0}(x) + \frac{1}{2}\sum_{i=1}^{n} K(t_n, c(t_n); u_i, c(u_i))\Delta_i. \end{aligned} \tag{C.2}$$

Here, $\mu(t|s, u) = E(X(t)|X(s) = u)$ and $\sigma(t|s)$ is the conditional standard deviation of $X(t)$ given $X(s)$. The integral on the right of (C.2) can be evaluated numerically, and the triangular system of equations can be solved successively for $\Delta_1, \Delta_2, \ldots$.

It remains to choose the points t_i. The most obvious choice is to choose equally spaced points. However, even without the truncation this is inefficient; it is better to sample more densely at times where the first exit is most likely to occur. With truncation at τ_0 this is even more important since the density has a singularity at τ_0. We choose upper limits m and h such that $\Delta_i \leq m$ and $h_i = t_i - t_{i-1} \leq h$ for all i. At the ith step, we initially set $h_i = (0.9\frac{m}{\Delta_{i-1}}h_{i-1}) \wedge h$. If this leads to $\Delta_i > m$, we repeat the step with a smaller h_i.

The computation time is then quadratic in the number of steps. For $\tau_0 = 0.2$ and $c = 2$, we used $h = 0.01$ and $m = 0.001$. This resulted in 172 steps, taking 139.4 seconds cpu time. This gives the probability 0.3479 to 4 decimal places and the 4th place may not be accurate. For moderate values of n, it is faster and easier to evaluate the exact probability for the Poisson process than for the limiting Gaussian process.

Power calculations for the constant parameter model are performed by three methods. The simplest is the local expansion (3.26). If we wish to evaluate the power for m values of δ, this is an $O(m)$ algorithm. We have to approximate the marginal probability $P_{\tau,\delta}(X(\tau) \le np_\tau)$. As δ increases, np_τ moves through the lower tail, middle and upper tail of the distribution of $X(\tau)$ and so (B.18) and the central limit approximation are appropriate for different values of δ. For the most interesting values of δ, the central limit approximation is the most appropriate.

The tangent approximations using (3.21) and (3.22) in conjunction with (3.18) is an $O(k) + O(m)$ algorithm, where k is the number of points used to approximate the integrals (3.21) and (3.22). For numerical purposes it is convenient to use a slightly modified form of (3.22). The expansion of the binomial coefficients using Stirling's formula introduces a singularity at $t = t'$. For values of j close to nq_t, the probability of crossing the boundary at t' is not negligible, and therefore we calculate this probability exactly and replace t' by the solution of $nq_t = j - 0.5$ in (3.22). Since we need to evaluate the terms $P^{(n)}_{\tau,\delta}(X(\tau) = j)$, we can keep track of the sum of these to evaluate $P^{(n)}_{\tau,\delta}(l(\tau) \ge \frac{1}{2}c^2)$.

As noted in Section 3.3, the exact computation may be performed in two ways. Directly evaluating (3.17) for each δ is an $O(mn^2)$ calculation, while evaluating (3.19) for each j then using (3.18) for each δ is a $O(n^3) + O(mn)$ calculation.

Evaluating at 150 values of δ with $n = 100$ took 0.7, 30.3 and 5802.1 seconds of cpu time for the local, tangent (with $k = 50$) and exact methods respectively.

C.2 Confidence Regions

Finding the constants c to make (4.3) equal to $1 - \alpha$ involves numerically solving an equation; repeating this for all τ will be a considerable chore. However, to construct the confidence set we do not need to evaluate c explicitly. For each possible value of τ, need to test whether or not $\tau \in I_1$. For fixed τ let $c_{obs} = \sup_{\tau_0 \leq t \leq \tau_1} l(t) - l(\tau)$ and

$$\alpha_{obs} = P_\tau^{(m,n)} \left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) > l(\tau) + c_{obs} \right). \tag{C.3}$$

Then we have $\tau \in I_1$ if and only if $\alpha_{obs} \geq \alpha$, and therefore we need only evaluate (C.3) for a suitable set of values of τ to determine the confidence region.

Since the confidence region will be made up of a small number of disjoint intervals, we can concentrate on finding the endpoints of these intervals. Moreover, these intervals will usually have one endpoint at an observed event and the other endpoint will be in between two events. Therefore, we can first test the times immediately before and immediately after each event for inclusion or exclusion in the confidence set. If from this we conclude there is an endpoint between successive events, we can then use the secant method to find this accurately. This is not completely foolproof, but the method will only fail when points for which $l(t)$ is very small are included in the confidence region; i.e. when there is only weak evidence of a change.

Depending on the method used to approximate (C.3), even this calculation can be laborious. For example, if the exact calculation is used, we have an $O(n^2)$ calculation for each t for which we evaluate α_{obs}. However, the computation can be reduced by first approximating (C.3) by a simpler and less computational approximation, then using the more complicated methods only in marginal cases. From Table 4.1, we see the values of c will usually be about 3.5 for $1-\alpha = 0.95$. We take $\alpha_{obs} = 0$ if $c_{obs} > 5$ and $\alpha_{obs} = 1$ if $c_{obs} < 2$. For Lucas' Accident data with $n = 178$, computation of the 95% confidence regions in Table 4.2 took 0.7, 72.5 and 3149.1 seconds by the local, tangent and exact methods respectively.

The computation of two way confidence regions for (τ, δ) is similar. For each τ, we first determine whether the set $\{\delta : (\tau, \delta) \in I_2\}$ is empty or non-empty. To a good approximation, we only need to test whether $(\tau, \hat{\delta}) \in I_2$. If the set is non-empty, we then find the upper and lower limits for δ such that $(\tau, \delta) \in I_2$. This method is not exact since for fixed α, c depends on δ. Occasionally the set $\{\delta : (\tau, \delta) \in I_2\}$ will not be an interval. However, this situation is rare since $l(\tau) - l(\tau|\delta)$ is a convex function of δ and the dependence of c on δ is small.

C.3 Scan Statistics

The evaluation of the approximations to the distribution of scan statistics presented in Chapter 7 is fairly straightforward. However, there are efficient and inefficient ways to do the simulations, so we describe the methods used.

We first consider the one dimensional scan statistic with unknown Δ, defined by (7.12). For fixed $\Delta = \Delta_j$ with $nh(\Delta_j) = j$, we have $\sup_{0 \leq \theta < 2\pi} l(\theta, \theta + \Delta_j) \geq \frac{1}{2}c^2$ if and only if

$$\inf_{1 \leq i \leq n} (\theta_{i+j-1} - \theta_i) \leq \Delta_j. \tag{C.4}$$

We can therefore rewrite (7.12) as

$$M_{\Delta_0, \Delta_1} = \sup_j nh(m_j, \frac{j}{n})$$

where m_j is defined as the left hand side of (C.4), subject to $m_j \geq \Delta_0$ and if $m_j > \Delta_1$, then $m_j = 2\pi$. This enables us to evaluate M_{Δ_0, Δ_1} in an $O(n^2)$ calculation. The computation can be further reduced by noting $m_j \geq m_{j-1}$ for all j. Hence if $\sup_{i<j} nh(m_i, \frac{i}{n}) \geq nh(m_{j-1}, \frac{j}{n})$, we do not need to evaluate m_j and can skip directly to $j+1$.

The two dimensional scan statistics are more difficult to evaluate. We first study the case Δ_1 and Δ_2 are known. We first note that if $\sup_t Y(t) = m$, then we can place the rectangle to include m points, then move it as far right and as far up as we can, so there are points on both the left and bottom edges. Therefore, we can evaluate $\sup_t Y(t)$ by evaluating $Y(x_1, y_2)$ for each pair of points (x_1, y_1), (x_2, y_2) with $x_1 \leq x_2 \leq x_1 + \Delta_1$ and $y_2 \leq y_1 \leq y_2 + \Delta_2$. In its crudest form this is an $O(n^3)$ algorithm. However, we can reduce the computation to $O(n^2)$ by fixing the left point (x_1, y_1) and moving a rectangle of size $\Delta_1 \times \Delta_2$ vertically and keeping count of the number of points in the window.

For $\Delta_0 = \Delta_1 = 0.5$, evaluating $\sup_t Y(t)$ takes 0.0563, 0.247 and 0.836 seconds for $n = 20$, 50 and 100 respectively. These times are based on the average of 1000 simulations and include the time to generate and sort the points.

When Δ_1 and Δ_2 are unknown, the maximum of $l(t)$ will occur at a value of t for which there are events on all 4 boundaries (adjacent boundaries may be represented by one point in a corner). To exploit this in evaluating M_{u_0,u_1}, we choose in turn each pair of points to represent the left and right boundaries and slide a window of fixed width but variable height to consider all possible top and bottom boundaries. This is a $O(n^4)$ computation. For $u_0 = 1 - u_1 = 0.1$, computation times, based on the average of 1000 simulations, is 0.412, 9.233 and 121.5 seconds for $n = 20$, 50 and 100 respectively.

Bibliography

[1] Akman, V. E. and Raftery, A. E. (1986). Asymptotic Inference for a Change-Point Poisson Process. *Ann. Stat.* **14**, 1583-1590.

[2] Aldous, David (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, New York.

[3] Becker, R. A.; Chambers, J. M. and Wilks, A. R. (1988). *The New S Language. A Programming Environment for Data Analysis and Graphics.* Wadsworth & Brooks/Cole, Moneterey, CA.

[4] Berman, M. and Eagleson, G. K. (1985). A Useful Upper Bound for the Tail Probabilities of the Scan Statistic When the Sample Size is Large. *Jour. Amer. Statist. Assn.* **80**, 886-889.

[5] Billingsley, Patrick (1968). *The Convergence of Probability Measures.* John Wiley & Sons, New York.

[6] Borokov, A. A. and Rogozin, R. A. (1965). On the Multidimensional Central Limit Theorem. *Theory Prob. Applic.* **10**, 55-62.

[7] Breiman, Leo; Freidman; Jerome H; Olshen, Richard A. and Stone, Charles J. (1984). *Classification and Regression Trees.* Wadsworth, Monterey, CA.

[8] Chang, I. S.; Chen, C. H. and Hsiung, C. A. (1989). Asymptotic Inference for a Change-point Hazard Rate Model with Random Censorship. Preprint.

[9] Chernoff, H. and Rubin, H. (1955). The Estimation of the Location of a Discontinuity in Density. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, 19-37.

[10] Chung, Kai Lai (1974). *A Course in Probability Theory*, Second Edition. Academic Press, New York.

[11] Conte, S. D. and De Boor, C. (1980). *Elementary Numerical Analysis. An Algorithmic Approach.* McGraw-Hill, New York.

[12] Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Methuen & Co. Ltd, London.

[13] Cressie, Noel (1980). *The Asymptotic Distribution of the Scan Statistic Under Uniformity. Ann. Prob.* **8**, 828-840.

[14] Crowley, John and Hu, Marie (1977). Covariance Analysis of Heart Transplant Survival Data. *Jour. Amer. Statist. Assn* **72**, 27-36.

[15] Dallal, Gerald E. and Wilkinson, L. (1986). An Analytic Approximation to the Distribution of Lilliefors's Test Statistic for Normality. *The American Statistician* **40**, 294-296.

[16] Daniels, H. E. (1945). The Statistical Theory of the Strength of Bundles of Thread. *Proc. Royal Society, London Ser. A* **183**, 405-435.

[17] Durbin, J. (1973a). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

[18] Durbin, J. (1973b). Weak Convergence of the Sample Distribution Function when the Parameters are Estimated. *Ann. Stat.* **1**, 279-290.

[19] Durbin, J. (1985). The First Passage Density of a Gaussian Process to a General Boundary. *Journal of Applied Probability* **22**, 99-122.

[20] Dvoretzky, A; Kiefer, J. and Wolfowitz, J. (1953). Sequential Decision Problems for Processes with Continuous Time Parameter. Testing Hypothesis. *Ann. Math. Stat.* **24**, 254-264.

[21] Feller, W. (1966). *An Introduction to Probability Theory and its Applications.* Vol 2. John Wiley & Sons, New York.

[22] Feller, W. (1968). *An Introduction to Probability Theory and its Applications.* Vol 1, 3rd ed. John Wiley & Sons, New York.

[23] Hinkley (1970). Inference About the Change Point in a Sequence of Random Variables. *Biometrika* **57**, 1-17.

[24] Hogan, M. L. and Siegmund, D. (1986). Large Deviations for the Maxima of Some Random Fields. *Advances in Applied Mathematics* **7**, 2-22.

[25] Hu, Inchi (1985). A Uniform Bound for the Tail Probability of Kolmogorov-Smirnov Statistics. *Ann. Stat.* **13**, 821-826.

[26] Huntington, Raymond J. and Naus, Joseph I. (1975). A Simpler Expression for kth Nearest Neighbor Coincidence Probabilities. *Ann. Prob.* **3**, 894-896.

[27] Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[28] James, Barry; James, Kang Ling and Siegmund, David (1987). Tests for a Change-Point. *Biometrika* **74**, 71-83.

[29] Jarrett, R. G. (1979). A Note on the Intervals Between Coal-Mining Disasters. *Biometrika* **66**, 191-193.

[30] Jennen, C. and Lerche, H. R. (1981). First Exit Densities of Brownian Motion through One-Sided Moving Boundaries. *Ziet. Wahrs. und Verwandte Gebiete* **55**, 133-148.

[31] Jennen, Christel (1985). Second-order Approximations to the Density, Mean and Variance of Brownian First-Exit Times. *Ann. Prob.* **13**, 126-144.

[32] Kim, Hyune-Ju (1988). *Change Point Problems in Regression.* Ph.D Dissertation, Stanford Univesity.

[33] Kim, H. J. and Siegmund, D. (1989). The Likelihood Ratio Test for a Change-Point in Simple Linear Regression. *Biometrika* **76**, 409-423.

[34] Knowles, M. and Siegmund, D. (1988). On Hotelling's Geometric Approach to Testing for a Non-Linear Parameter in Regression. *Technical Report* **235**, Dept. Of Statistics, Stanford University.

[35] Kuiper, N. H. (1960). Tests Concerning Random Points on a Circle. *Nederl. Akad. Wetensch. Proc. Series A* **63**, 38-47.

[36] Levin, B. and Kline, J. (1985). The CUSUM test of Homogeneity with an Application to Spontaneous Abortion Epidemiology. *Statistics in Medicine* **4**, 469-488.

[37] Lucas, James M. (1985). Counted Data CUSUM's. *Technometrics* **27**, 129-144.

[38] Maguire, E. A.; Pearson, E. S. and Wynn, A. H. A. (1952). The Time Intervals Between Industrial Accidents. *Biometrika* **38**, 168-180.

[39] Major, P. (1990). A Note on the Approximation of the Uniform Empirical Process. *Ann. Prob.* **18**, 129-139.

[40] Mallows, Colin L. and Nair, Vijayan N. (1989). On a Zero-Crossing Probability. *Ann. Inst. Statist. Math.* **41**, 1-8.

[41] Matthews, David E. and Farewell, Vernon T. (1982). On Testing for a Constant Hazard Rate against a Change-Point Alternative. *Biometrics 38*, 463-468.

[42] Matthews, D. E.; Farewell, V. T. and Pyke, R. (1985). Asymptotic Score-Statistic Processes and Tests for Constant Hazard Rate Against a Change-Point Alternative. *Ann. Stat.* **13**, 583-591.

[43] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

[44] Naus, J. I. (1965). The Distribution of the Size of the Maximum Cluster of Points on a Line. *Jour. Amer. Statist. Assn.* **60**, 532-538.

[45] Naus, J. I. (1966). Some Probabilities, Expectations and Variances for the Size of the Largest Clusters and Smallest Intervals. *Jour. Amer. Statist. Assn.* **61**, 1191-1199.

[46] Nguyen, H. T.; Rogers, G. S. and Walker, E. A. (1984). Estimation in Change-point Hazard Rate Models. *Biometrika* **71**, 299-304.

[47] Noé, Marc (1972). The Calculation of Two-Sided Kolmogorov-Smirnov Type Distributions. *Ann. Math. Stat.* **43**, 58-64.

[48] Park, C. and Schuurmann, F. S. (1976). Evaluation of Barrier-Crossing Probabilities of Wiener Paths. *Journal of Applied Probability* **13**, 267-275.

[49] Pyke, R. (1959). The Supremum and Infimum of the Poisson Process. *Ann. Math. Stat.* **30**, 568-576.

[50] Raftery, A. E. and Akman, V. E. (1986). Bayesian Analysis of a Poisson Process with a Change-Point. *Biometrika* **73**, 85-89.

[51] Ripley, Brian D. (1981). *Spatial Statistics.* John Wiley & Sons, New York.

[52] Romano, Joseph P. (1989). Bootstrap and Randomization Tests of Some Nonparametric Hypothesis. *Ann. Stat.* **17**, 141-159.

[53] Shepp, L. (1971). First Passage Time for a Particular Gaussian Process. *Ann. Math. Stat.* **42**, 946-951.

[54] Shiryayev, A. N. (1963). On optimal methods in Earliest Detection Problems. *Theory Prob. Applic.* **8**, 26-51.

[55] Shiryayev, A. N. (1984). *Probability.* Springer-Verlag, New York.

[56] Shorack, Galen R. and Wellner, Jon A. (1986). *Empirical Processes with Applications to Statistics.* John Wiley & Sons, New York.

[57] Siegmund, D. (1982). Large Deviations for Boundary Crossing Probabilities. *Ann. Prob.* **10**, 581-588.

[58] Siegmund, David (1985). *Sequential Analysis. Tests and Confidence Intervals.* Springer-Verlag, New York.

[59] Siegmund, David (1988a). Confidence Sets in Change-Point Problems. *International Statistical Review* **56**, 31-48.

[60] Siegmund, David (1988b). Approximate Tail Probabilities for the Maxima of some Random Fields. *Ann. Prob.* **16**, 487-501.

[61] Smirnov, N. V. (1944). An Approximation to the Distribution Laws of Random Quantities Determined by Empirical Data. *Usephi Mat. Nauk.* **10**, 179-206.

[62] Tibshirani, R. and Hastie, T. (1987). Local Likelihood Estimation. *Jour. Amer. Stat. Assn.* **82**, 559-568.

[63] Woodroofe, M. (1976). Frequentist Properties of Bayesian Sequential Tests. *Biometrika* **63**, 101-110.

[64] Worsley, K. J. (1986). Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables. *Biometrika* **73**, 91-104.

[65] Worsley, K. J. (1988). Exact Percentage points of the Likelihood Ratio Test for a Change-Point Hazard-Rate Model. *Biometrics* **44**, 259-263.

[66] Yao, Yi-Ching (1986). Maximum Likelihood Estimation in Hazard Rate Models with a Change-point. *Commun. Statist. Theor. Meth.*, **15**, 2455-2466.