A Log-Linear Model for a Poisson Process Change Point
Author(s): Clive R. Loader
Source: *The Annals of Statistics*, Vol. 20, No. 3 (Sep., 1992), pp. 1391-1411
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/2242017
Accessed: 24-01-2018 22:40 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *The Annals of Statistics*

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

The Annals of Statistics
1992, Vol. 20, No. 3, 1391–1411

A LOG-LINEAR MODEL FOR A POISSON PROCESS CHANGE POINT

BY CLIVE R. LOADER

AT&T Bell Laboratories

Many methods have been proposed for modelling nonhomogeneous Poisson processes, including change point models and log-linear models. In this paper, we use likelihood ratio tests to choose which of these models are necessary. Of particular interest is the test for the presence of a change point, for which standard asymptotic theory is not valid. Large deviation methods are applied to approximate the significance level, and power approximations are given. Confidence regions for the change point and other parameters in the model are also derived. A British coal mining accident data set is used to illustrate the methodology.

1. Introduction. Suppose we observe a nonhomogeneous Poisson process $X(t)$ on $[0, T]$ with rate $\lambda(t)$. The standard change point model considered, for example, by Akman and Raftery (1986a) assumes the rate switches between constant rates λ_0 and λ_1 at an unknown change point τ. Another commonly considered class of models is log-polynomial models, where $\log(\lambda(t))$ is assumed to be a polynomial with unknown coefficients.

When presented with a data set, it will rarely be clear in advance which of these classes of models is most appropriate, and some selection procedures are needed. Direct hypothesis testing seems inappropriate since the classes of models are nonnested, and the choice of null hypothesis is arbitrary. An alternative approach is to consider models in a larger parametric class, which includes both change point (cp) models and log-linear (ll) models. In this paper, we consider the model

$$(1) \qquad \lambda(t) = \begin{cases} e^{a+bt}, & 0 \le t \le \tau, \\ e^{a+\delta+bt}, & \tau < t \le T, \end{cases}$$

where a, b, δ and τ are unknown parameters, we refer to (1) as the llcp model. The cp model is the special case $b = 0$ and the ll model is the case $\delta = 0$. This model is a first step at mixing change points with gradually changing rate functions; at the end of this introduction we say more about possible extensions and the difficulties that arise. 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) will have some similarity.

Received August 1990; revised October 1991.

AMS 1980 *subject classifications*. Primary 62M99; secondary 60G55, 62F03.

Key words and phrases. Boundary crossing, change points, log-linear model, non-nested hypothesis, Poisson process.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

To select a suitable model, we propose a backward selection procedure. Beginning with the full model, likelihood ratio tests are used to determine whether b and δ are necessary. This is a multiple testing problem; adjustments to the individual significance levels are easily obtained using Bonferroni bounds. Backward selection is preferred to forward selection since the effects of a log-linear term and a change point may cancel if forward selection is used.

For a Poisson process with rate function $\lambda(t)$, the likelihood function is

$$L(\lambda) = \exp\left(-\int_0^T \lambda(t)\,dt\right)\prod_{i=1}^{n}\lambda(T_i), \tag{2}$$

where $n = X(T)$ and $T_1, \dots, T_n$ are the event times. See Cox and Lewis (1966) for a derivation of (2). For the llcp model, (2) becomes

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

For fixed τ, maximum likelihood estimates of a, b and δ are the solutions of

$$\begin{aligned} n &= \frac{e^a}{b}\left(e^{b\tau} - 1 + e^{\delta}\left(e^{bT} - e^{b\tau}\right)\right), \\ n - X(\tau) &= \frac{e^{a+\delta}}{b}\left(e^{bT} - e^{b\tau}\right), \end{aligned}$$

$$\sum_{i=1}^{n} T_i = e^a \frac{d}{db}\frac{e^{b\tau} - 1}{b} + e^{a+\delta}\frac{d}{db}\frac{e^{bT} - e^{b\tau}}{b}. \tag{3}$$

Rearranging the first two of these equations gives

$$\hat{a} = \log\left(\frac{\hat{b}X(\tau)}{e^{\hat{b}\tau} - 1}\right), \qquad \hat{\delta} = \log\left(\frac{(n - X(\tau))(e^{\hat{b}\tau} - 1)}{X(\tau)(e^{\hat{b}T} - e^{\hat{b}\tau})}\right).$$

Substituting these into (3) and rearranging leads to

$$\sum_{i=1}^{n} T_i = X(\tau)\frac{\tau e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1} + (n - X(\tau))\frac{Te^{\hat{b}T} - \tau e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}} - \frac{n}{\hat{b}}. \tag{4}$$

There is no closed form solution for $\hat{b}$, so (4) must be solved numerically. To establish the existence of $\hat{b}$, we note if $X(\tau) = m$, then $(n - m)\tau \le \sum_{i=1}^{n} T_i \le m\tau + (n - m)T$, and the right-hand side of (4) converges to these limits as $\hat{b} \to -\infty$ and $\hat{b} \to \infty$. The derivative of (4) with respect to $\hat{b}$ is $\operatorname{var}(\sum_{i=1}^{n} T_i | X(\tau), n) > 0$ and hence $\hat{b}$ is unique. Although (4) appears to take an indeterminate form as $\hat{b} \to 0$, the $1/\hat{b}$ terms cancel and the equation is defined at $\hat{b} = 0$ by continuity. Similar situations occur throughout this article.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

The maximum likelihood estimate $\hat{\tau}$ of τ is found by maximizing $L(\tau, \hat{a}, \hat{\delta}, \hat{b})$. We show in Lemma 2.4 that this maximum occurs at an event time. The likelihood is only right-continuous at these points, so the maximum may not be achieved; this can be overcome by considering left-continuous versions of the likelihood.

We will not discuss in detail the distribution of the estimators. Chernoff and Rubin (1956) show for a wide class of problems involving estimation of a discontinuity in a density that the maximum likelihood estimate will have error of $O_p(n^{-1})$. Similar results can be derived, at least heuristically, for our case. It follows that estimates of other parameters, in our case a, b and δ, will have the same asymptotic distributions as if τ were known, and the estimates are asymptotically independent of $\hat{\tau}$. See Yao (1986) for a rigorous development of these details for a similar problem.

More importantly, consider the problem of testing $\mathscr{H}_0$: $b = 0$ against the llcp alternative. If $|\delta|$ is large, then $\hat{\tau}$ will be close to τ under both the null and alternative models, and under $\mathscr{H}_0$, the log-likelihood ratio will be distributed approximately as $\chi_1^2/2$. This approximation may not be very good if δ is small; however, we do not study this test further here.

For testing $\mathscr{H}_0$: $\delta = 0$ against the llcp alternative, standard asymptotics are not applicable. Under the null hypothesis, τ is meaningless and $\hat{\tau}$ does not satisfy regularity conditions required for standard asymptotic theory. If $\tau = t$ is known under the llcp alternative, the log-likelihood ratio statistic is

$$
\begin{aligned}
l(t) = X(t)\log&\left(\frac{\hat{b}X(t)(e^{\hat{b}_0 T} - 1)}{\hat{b}_0 n(e^{\hat{b}t} - 1)}\right) \\
&+ (n - X(t))\log\left(\frac{\hat{b}(n - X(t))(e^{\hat{b}_0 T} - 1)}{\hat{b}_0 n(e^{\hat{b}T} - e^{\hat{b}t})}\right) \\
&+ \left(\hat{b} - \hat{b}_0\right)\sum_{i=1}^{n} T_i,
\end{aligned} \tag{5}
$$

where $\hat{b}_0$ is the maximum likelihood estimate of b under $\mathscr{H}_0$ and satisfies

$$\frac{1}{n}\sum_{i=1}^{n} T_i = \frac{Te^{\hat{b}_0 T}}{e^{\hat{b}_0 T} - 1} - \frac{1}{\hat{b}_0}.$$

When τ is unknown, the likelihood ratio test will be based on the maximum of $l(t)$. To prevent a few events near an endpoint causing a spuriously large likelihood, we modify the test slightly by considering the maximum over a subinterval $[\tau_0, \tau_1]$ with $0 \le \tau_0 \le \tau_1 \le T$. The significance level is evaluated conditionally on $X(T)$ and $\Sigma_{i=1}^n T_i$, which by sufficiency removes dependence on a and b. The significance level is then

$$\alpha = \alpha(c, n, y) = P_0^{(n)}\left\{\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{c^2}{2} \,\middle|\, X(T) = n,\, S_n = ny\right\}, \tag{6}$$

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

where $S_n = \sum_{i=1}^{n} T_i$ and $P_0^{(n)}$ denotes a measure for which $\delta = 0$, conditioned on $X(t) = n$. We derive an approximation to the right-hand side of (6) in Section 2.

The problem of testing homogeneity against the cp alternative is very similar. In this case, the likelihood ratio process is

$$l(t) = X(t)\log\left(\frac{TX(t)}{tn}\right) + (n - X(t))\log\left(\frac{T(n - X(t))}{(T-t)n}\right)$$

and the significance level of the test is

$$\alpha = \alpha(c, n) = P_0^{(n)}\left\{\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \frac{c^2}{2}\right\}. \tag{7}$$

Exact methods have been applied to this problem by Worsley (1986). However, for computational reasons, approximations may be preferable when n is large.

Approximations to the power of the likelihood ratio test are derived in Section 3. Two approaches will be used; they differ in both accuracy and amount of computation required. One question of interest is how much power is lost by fitting the llcp model when $b = 0$ and the cp model could be used. Numerical examples are used to show this power loss is fairly substantial.

In Section 4, we derive confidence regions for the change point and joint confidence regions for the change point and other parameters. We follow the likelihood ratio method of Siegmund (1988), which gives rise to more boundary crossing problems. Again, several methods may be used to approximate these probabilities.

The methods are illustrated using British coal mining data in Section 5. This data set was published by Maguire, Pearson and Wynn (1952) and extended and corrected by Jarrett (1979). This version lists 191 accidents between 15 March 1851 and 22 March 1962.

There are several possible generalizations of (1), including higher-order log-polynomial terms, multiple change points and allowing the slope to change. Models with higher-order terms could in principle be analyzed by methods similar to those in this article, although sensitivity to differences between models will be further reduced. Formal analysis of multiple change point models leads to random field problems which we will not discuss here. An informal sequential procedure is to split the data into two sections when a change point is found and to analyze each section separately for further changes.

An alternative Bayesian approach to model selection for a nonhomogeneous Poisson process has been studied by Akman and Raftery (1986b) and applied to change point models in Raftery and Akman (1986). This approach has the advantage of allowing direct comparison of nonnested models and does not require any complicated distribution theory. The disadvantage is that posterior distributions may be difficult to compute, generally requiring multidimensional numerical integration.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

2. Significance level calculations. To approximate the significance level (6), we derive boundaries p_t and q_t such that the conditional likelihood ratio test is rejected if the process $X(t)$ exits the interval (p_t, q_t). The significance level can then be rewritten in terms of boundary crossing probabilities for $X(t)$, which are approximated by large deviation methods in Theorem 2.1.

LEMMA 2.1. *Let $b_0 = b_0(y)$ be defined by*

$$y = \frac{Te^{b_0T}}{e^{b_0T} - 1} - \frac{1}{b_0}.$$

Let $\eta = c/\sqrt{n}$ and define p_t and q_t to be the values of p obtained by simultaneously solving

$$y = p\frac{te^{bt}}{e^{bt} - 1} + (1-p)\frac{Te^{bT} - te^{bt}}{e^{bT} - e^{bt}} - \frac{1}{b}, \tag{8}$$

$$\begin{aligned}\frac{\eta^2}{2} &= p\log\left(\frac{pb(e^{b_0T} - 1)}{b_0(e^{bt} - 1)}\right)\\ &\quad + (1-p)\log\left(\frac{(1-p)b(e^{b_0T} - 1)}{b_0(e^{bT} - e^{bt})}\right) + (b - b_0)y,\end{aligned} \tag{9}$$

for p and b, subject to $p_t < (e^{b_0t} - 1)/(e^{b_0T} - 1) < q_t$. If $\sum_{i=1}^n T_i = ny$, then

$$l(t) \le \frac{c^2}{2} \quad \Leftrightarrow \quad np_t \le X(t) \le nq_t. \tag{10}$$

If p_t and q_t exist, then they are unique. If p_t does not exist, set $p_t = 0$, and if q_t does not exist, set $q_t = 1$, and (10) *still holds.*

PROOF. Let $f(t, p, b)$ denote the right-hand side of (9) and choose $b = b(t, p)$ to satisfy (8). If $p = X(t)/n$, then, by (4) and (5), $l(t) = nf(t, p, b(t, p))$; hence $l(t) = c^2/2$ only at solutions of (8) and (9).

In Lemma 2.4 we show for fixed t and varying p, $f(t, p, b(t, p))$ is decreasing, 0 or increasing when p is less than, equal to or greater than $(e^{b_0t} - 1)/(e^{b_0T} - 1)$, respectively. This implies uniqueness of p_t and q_t and implies that (10) holds. If p_t does not exist, then $f(t, p, b(t, p)) < \eta^2/2$ for all $p < (e^{b_0t} - 1)/(e^{b_0T} - 1)$, so we can take $p_t = 0$. The case where q_t does not exist is similar. □

Using Lemma 2.1, we can write

$$\begin{aligned}&P_0^{(n)}\left(\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \tfrac{1}{2}c^2 \,\middle|\, S_n = ny\right)\\ &\quad \approx P_0^{(n)}\left(\inf_{\tau_0 \le t \le \tau_1} (X(t) - np_t) \le 0 \,\middle|\, S_n = ny\right)\\ &\qquad + P_0^{(n)}\left(\sup_{\tau_0 \le t \le \tau_1} (X(t) - nq_t) \ge 0 \,\middle|\, S_n = ny\right).\end{aligned} \tag{11}$$

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

The approximation in (11) is in neglecting the probability of crossing both boundaries, which is generally small. The key to approximating these probabilities is the following lemmas. Lemma 2.2 is a special case of Lemma B.2 in Loader (1992) and Lemma 2.3 is Theorem 3.1 in Loader (1992).

LEMMA 2.2. *Fix τ, p and y. Let $Z_{1,n}(t) = X(\tau) - X(\tau - t/n)$ and $Z_{2,n}(t) = X(\tau + t/n) - X(\tau)$. Conditionally on $X(\tau) = [np]$ and $\sum_{i=1}^{n} T_i = ny$, $Z_{1,n}(t)$ and $Z_{2,n}(t)$ converge in law to a pair of independent Poisson processes with rates $\mu(\tau, p)$ and $\lambda(\tau, p)$, where*

$$\mu(\tau, p) = \frac{1}{n}\exp(\hat{a} + \hat{b}\tau) = \frac{p\hat{b}e^{\hat{b}\tau}}{e^{\hat{b}\tau} - 1},$$

$$\lambda(\tau, p) = \frac{1}{n}\exp(\hat{a} + \hat{\delta} + \hat{b}\tau) = \frac{(1-p)\hat{b}e^{\hat{b}\tau}}{e^{\hat{b}T} - e^{\hat{b}\tau}}.$$

LEMMA 2.3. *Suppose $\{X(t);\ 0 < t < T\}$ is a point process on a measurable space $(\Omega, \mathscr{F})$, and $P^{(n)}$ is a sequence of probability measures. Let $a(t)$ be a differentiable boundary, such that under $P^{(n)}$, conditionally on $X(\tau) = na(\tau)$, the local increments $X(\tau) - X(\tau - t/n)$ can be approximated by a Poisson process with rate $\mu(\tau, a(\tau)) < a'(\tau)$. Let t_k be the solution of $na(t) = k$. Then*

$$P^{(n)}\Big(\inf_{\tau_0 \le t \le \tau_1} (X(t) - na(t)) < 0\Big)$$
$$= \sum_{k=k_0}^{k_1}\left(1 - \frac{\mu(t_k, a(t_k))}{a'(t_k)}\right)P^{(n)}(X(t_k) = k)(1 + o(1))$$
$$= n\int_{\tau_0}^{\tau_1}(a'(t) - \mu(t, a(t)))g_n(t)\,dt(1 + o(1)),$$

where $g_n(t)$ is a suitable continuous approximation to $P^{(n)}(X(t_k) = k)$.

THEOREM 2.1. *For fixed t, let*

$$\psi(\delta, b) = \log\left(\frac{1}{b}\left(e^{bt} - 1 + e^{\delta}(e^{bT} - e^{bt})\right)\right). \tag{12}$$

Let $\psi''(\delta, b)$ be the second-derivative matrix and $\psi_0''(0, b)$ be the second derivative with respect to b when $\delta = 0$. Then

$$P_0^{(n)}\Big(\inf_{\tau_0 \le t \le \tau_1} (X(t) - np_t) \le 0 \Big| S_n = ny\Big)$$
$$= \frac{\sqrt{n}\,e^{-c^2/2}}{\sqrt{2\pi}}\int_{\tau_0}^{\tau_1}(p_t' - \mu(t, p_t))\sqrt{\frac{\psi_0''(0, \hat{b}_0)}{|\psi''(\hat{\delta}, \hat{b})|}}\,dt(1 + o(1)).$$

The probability of crossing q_t is approximated by time reversal; $\mu(t, p_t)$ is replaced by $\lambda(t, q_t)$.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

Theorem 2.1 is a direct application of Lemma 2.3. The large deviation approximation to the conditional distribution of $X(t)$ is derived in the Appendix; setting $\delta = 0$ in (24) gives

$$P_0^{(n)}(X(t) = np_t | S_n = ny) = \frac{\exp(-c^2/2)}{\sqrt{2\pi n}} \sqrt{\frac{\psi_0''(0, \hat{b}_0)}{|\psi''(\hat{\delta}, \hat{b})|}} (1 + o(1))$$

as $n \to \infty$, where $\hat{\delta}$ and $\hat{b}$ are all calculated under the assumption $\tau = t$ and $X(t) = np_t$.

Lemma 2.4 derives an expression for p_t' and proves two results concerning derivatives of the likelihood that have been claimed earlier.

LEMMA 2.4. (i) *$\hat{\tau}$ is either an event time or truncation point.*

(ii) *Uniqueness of p_t and q_t, as claimed in Lemma* 2.1.

(iii)
$$\frac{dp_t}{dt} = \frac{\lambda(t, p_t) - \mu(t, p_t)}{\log(\lambda(t, p_t)) - \log(\mu(t, p_t))}.$$

PROOF. Let $f(t, p, b)$ be as in Lemma 2.1. Differentiation shows

$$\frac{\partial f}{\partial t} = \frac{(1-p)be^{bt}}{e^{bT} - e^{bt}} - \frac{pbr^{bt}}{e^{bt} - 1},$$

$$\frac{\partial f}{\partial p} = \log\left(\frac{p(e^{bT} - e^{bt})}{(1-p)(e^{bt} - 1)}\right),$$

$$\frac{\partial f}{\partial b} = \frac{1}{b} - \frac{pte^{bt}}{e^{bt} - 1} - \frac{(1-p)(Te^{bT} - te^{bt})}{e^{bT} - e^{bt}} + y.$$

When $b = b(t, p)$ satisfies (8), these derivatives are $\lambda(t, p) - \mu(t, p)$, $\log(\mu(t, p)/\lambda(t, p))$ and 0, respectively.

Between event times, $X(t)$ is constant. Letting $p = X(t)/n$,

$$\frac{1}{n}\frac{dl(t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial b}\frac{\partial b(t, p)}{\partial t} = \lambda(t, p) - \mu(t, p).$$

If $\lambda(t, p) = \mu(t, p)$, then $X(t) = n(e^{\hat{b}t} - 1)/(e^{\hat{b}T} - 1)$ and $\hat{b} = \hat{b}_0$. Therefore, $dl(t)/dt = 0 \Rightarrow l(t) = 0$ and hence $l(t)$ cannot have local maxima between event times. This gives (i).

When t is fixed and $b = b(t, p)$ we have

$$\frac{df}{dp} = \log(\mu(t, p)) - \log(\lambda(t, p)),$$

which is 0 only when $p = p_0 = (e^{b_0 t} - 1)/(e^{b_0 T} - 1)$. Hence, for fixed t, $f(t, p, b(t, p))$ is decreasing for $p < p_0$ and increasing for $p > p_0$, which implies (ii).

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

TABLE 1
Significance level approximations for testing the ll model against the llcp alternative, with $n = 100$ and $\tau_0 = 1 - \tau_1 = 0.1$

		c				
y	**Method**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0.5	Simulation	0.6728	0.3225	0.1054	0.0271	0.0040
	Approximation	1.0677	0.4254	0.1272	0.0289	0.0050
0.6	Simulation	0.6808	0.3264	0.1105	0.0253	0.0046
	Approximation	1.0723	0.4280	0.1284	0.0295	0.0051
0.7	Simulation	0.6903	0.3350	0.1107	0.0254	0.0044
	Approximation	1.0979	0.4399	0.1281	0.0283	0.0048
Simulation s.e.		0.005	0.005	0.003	0.002	0.0007

Setting $p = p_t$ gives

$$0 = \frac{df(t, p_t, b(t, p_t))}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial p}\frac{dp_t}{dt},$$

which leads to (iii). □

Table 1 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. When c is small the approximation does not perform particularly well. However, for larger values of c, which are of interest for significance testing, the approximating is more accurate.

The poor performance for small 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, 0.0174, 0.0015 and 0.0002 of crossing both boundaries. This accounts for most of the error of the approximation.

The method described in this section can be easily modified to approximate (7). Numerical examples and second-order corrections for this case may be found in Loader (1992).

3. Power calculations. Let $P_{\tau,\delta}^{(n)}$ be a measure under which the llcp model holds, conditioned on $X(T) = n$. The power of the likelihood ratio test is

$$\beta(\tau, \delta) = P_{\tau,\delta}^{(n)}(R_1 \cup R_2 | S_n = ny), \tag{13}$$

where

$$R_1 = \left\{ \sup_{\tau_0 \le t \le \tau} l(t) \ge \frac{c^2}{2} \right\}, \qquad R_2 = \left\{ \sup_{\tau \le t \le \tau_1} l(t) \ge \frac{c^2}{2} \right\}.$$

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

By sufficiency, (13) does not depend on a or b. Conditioning on $X(\tau)$,

$$\beta(\tau,\delta) = \sum_{j=0}^{n} P_{\tau}^{(n)}(R_1 \cup R_2 | X(\tau) = j,\, S_n = ny) P_{\tau,\delta}^{(n)}(X(\tau) = j | S_n = ny)$$

$$(14) \qquad = P_{\tau,\delta}^{(n)}\big(l(\tau) \geq \tfrac{1}{2}c^2 \big| S_n = ny\big) + \sum_{j=j_0}^{j_1} P_{\tau}^{(n)}(R_1 \cup R_2 | X(\tau) = j,\, S_n = ny) P_{\tau,\delta}^{(n)}(X(\tau) = j | S_n = ny),$$

where $j_0 = \lceil np_\tau \rceil$ and $j_1 = \lfloor nq_\tau \rfloor$. We can write

$$P_{\tau,\delta}^{(n)}\big(l(\tau) \geq \tfrac{1}{2}c^2 \big| S_n = ny\big) = P_{\tau,\delta}^{(n)}\big(X(\tau) < np_\tau \text{ or } X(\tau) > nq_\tau | S_n = ny\big).$$

For $\delta > 0$, the dominant term will be that arising from $X(\tau) < np_\tau$. These tail probabilities can be approximated by summing the approximation (24) or using the central limit theorem.

We approximate the terms

$$P_{\tau,\delta}^{(n)}(R_1 \cup R_2 | X(\tau) = j,\, S_n = ny)$$

by two methods. The first 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. The second is a large deviation method, applying the results of Lemma 2.3.

Both these methods have strengths and weaknesses. The first 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 large deviation for boundaries and sample sizes of interest. However, large deviation approximation requires much more computation.

Suppose $\delta > 0$ so the local expansion will be around (τ, p_τ); the case $\delta < 0$ involves a similar expansion around (τ, q_τ). The main result is contained in the following lemma.

LEMMA 3.1.

$$P_{\tau}^{(n)}(R_1 \cup R_2 | X(\tau) = np_\tau + x,\, S_n = ny) - e^{-\delta_0 x} - (1 - e^{-\delta_0 x}) h(\delta_0, \delta_0 x) \to 0$$

as $n \to \infty$ *with* $x = O(1)$, *where* $\delta_0 = \log(\lambda(\tau, p_\tau)/\mu(\tau, p_\tau))$ *and*

$$(15) \qquad h(\delta, z) = \left(1 - \frac{|\delta|}{e^{|\delta|} - 1}\right) \sum_{j=0}^{\infty} \left(\frac{|\delta|}{e^{|\delta|} - 1}\right)^{j} P\big(|\delta| U_j > z\big)$$

$$(16) \qquad \sim \nu(\delta) e^{-z},$$

where

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

In (15), U_j *is the sum of* j *independent* $\mathscr{U}[0,1]$ *random variables.*

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

PROOF. Suppose $X(\tau) = np_\tau + x$ and $S_n = ny$. Using the derivatives calculated in Lemma 2.4 we have $l(t) = c^2/2 - \delta_0 x + o(1)$ and

$$l\left(\tau + \frac{t}{n}\right) = l(\tau) - \delta_0\left(X\left(\tau + \frac{t}{n}\right) - X(\tau)\right) + \left(\lambda(\tau, p_\tau) - \mu(\tau, p_\tau)\right)t + o_p(1).$$

Letting $Z_1(t)$ and $Z_2(t)$ be homogeneous Poisson processes with rates $\mu(\tau, p_\tau)$ and $\lambda(\tau, p_\tau)$, respectively, and applying Lemma 2.2,

$$P_\tau^{(n)}(R_1|X(\tau) = np_\tau + x, S_n = ny) - P\Big(\sup_{t>0}(\delta_0 Z_1(t) - \gamma t) \ge \delta_0 x\Big) \to 0,$$

$$P_\tau^{(n)}(R_2|X(\tau) = np_\tau + x, S_n = ny) - P\Big(\sup_{t>0}(-\delta_0 Z_2(t) + \gamma t) \ge \delta_0 x\Big) \to 0,$$

where $\gamma = \lambda(\tau, p_\tau) - \mu(\tau, p_\tau)$. Standard results for linear boundary crossing probabilities for Poisson processes [cf. Pyke (1959) or Loader (1992)] show

$$P\Big(\sup_{t>0}(\delta_0 Z_1(t) - \gamma t) \ge \delta_0 x\Big) = h(\delta_0, \delta_0 x),$$

$$P\Big(\sup_{t>0}(-\delta_0 Z_2(t) + \gamma t) \ge \delta_0 x\Big) = e^{-\delta_0 x}.$$

Using the local independence in Lemma 2.2 completes the result. A complete justification of the local approximation requires a truncation argument similar that used to prove Lemma 2.1 in Loader (1992). □

An approximation to the conditional distribution of $X(\tau)$ is derived in the appendix. From (24) (see the Appendix), we get

$$\frac{P_{\tau,\delta}^{(n)}(X(\tau) = np_\tau + x|S_n = ny)}{P_{\tau,\delta}^{(n)}(X(\tau) = np_\tau|S_n = ny)} - e^{(\delta_0 - \delta)x} \to 0$$

as $n \to \infty$ with $x = O(1)$. Using (14) gives

$$\begin{aligned}
&P_{\tau,\delta}^{(n)}\{R_1 \cup R_2;\, X(\tau) > np_\tau|S_n = ny\} \\
&\quad = \sqrt{\frac{\psi_\delta''(\delta, \hat{b}_\delta)}{2\pi n|\psi''(\hat{\delta}, \hat{b})|}}\exp(-nl_\delta(t)) \\
&\qquad \times \sum_{i=1}^{\infty} e^{(\delta_0 - \delta)x_i}\left(e^{-\delta_0 x_i} + (1 - e^{-\delta_0 x_i})h(\delta_0, \delta_0 x_i)\right)(1 + o(1))
\end{aligned}$$

as $n \to \infty$, where $i - 1 < x_i \le i$ and $np_\tau + x_i$ is an integer. Using the approximation (16) gives

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

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

The large deviation approach uses Lemma 2.3 to obtain more accurate approximations to the rejection probabilities in Lemma 3.1. Let $P_\tau^{(j,n)}$ denote conditioning on $X(\tau) = j$ and $X(T) = n$. Then

$$\begin{aligned}(17)\qquad & P_\tau^{(j,n)}\Big(\inf_{\tau_0 \le t \le \tau} (X(t) - np_t) \le 0 \Big| S_n = ny\Big) \\ &\approx \sum_{k=k_0}^{np_\tau} \left(1 - \frac{\mu_2(t_k)}{p'(t_k)}\right) P_\tau^{(j,n)}(X(t_k) = k | S_n = ny),\end{aligned}$$

where $\mu_2(t)$ is the local left rate at t conditioned on $X(t) = np_t$ and $X(\tau) = j$, and

$$\begin{aligned}(18)\qquad & P_\tau^{(j,n)}\Big(\sup_{\tau_0 \le t \le \tau} (X(t) - nq_t) \ge 0 \Big| S_n = ny\Big) \\ &\approx \sum_{k=k_1}^{j} \left(1 - \frac{\lambda_2(t_k)}{p'(t_k)}\right) P_\tau^{(j,n)}(X(t_k) = k | S_n = ny),\end{aligned}$$

where $\lambda_2(t)$ is the local right rate when $X(t) = nq_t$ and $X(\tau) = j$. To obtain expressions for $\mu_2(t)$ for $\lambda_2(t)$, we add a second change point at t. When $p = j/n$, the mle b_2 of b is the solution of

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

and the local rates are

$$\mu_2(t) = \frac{p_t b_2 e^{b_2 t}}{e^{b_2 t} - 1}, \qquad \lambda_2(t) = \frac{(p - q_t) b_2 e^{b_2 t}}{e^{b_2 \tau} - e^{b_2 t}}.$$

The conditional probabilities $P_\tau^{(j,n)}(X(t) = k | S_n = ny)$ can be approximated using Lemma 2.2 when t is close to τ and results similar to (24) otherwise. The summations (17) and (18) can be approximated by integrals; however, some care must be taken to avoid singularities.

Adding (17) and (18) approximates $P^{(j,n)}(R_1 | S_n = ny)$; we approximate $P^{(j,n)}(R_2 | S_n = ny)$ by time reversal. The power approximation is completed by substituting into (14) and using (24) to approximate $P_{\tau,\delta}^{(n)}(X(\tau) = j | S_n = ny)$.

Similar methods may be used to derive power approximations for the test of $\mathscr{H}_0$: $b = \delta = 0$ against the cp alternative. Alternatively, recursive formulae may be used for exact computations; see Worsley (1986). In Figure 1, we show the approximations to $P_{\tau,\delta}^{(n)}(R_1 \cup R_2;\ X(\tau) > np_\tau)$ for $T = 1$, $\tau = 0.5$ and $n = 100$, using $c = 3.0372$ and $\tau_0 = 1 - \tau_1 = 0.1$. The large deviation approximation performs very well. The local approximation slightly underestimates the power, but since precise power calculations are rarely important this is probably accurate enough for most purposes.

One question of interest is how much power is lost if we test the ll model against the llcp alternative when the cp model is correct. In Figure 2, we compare the power of the two tests for $n = 100$ and $\tau = 0.3$. We also include

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

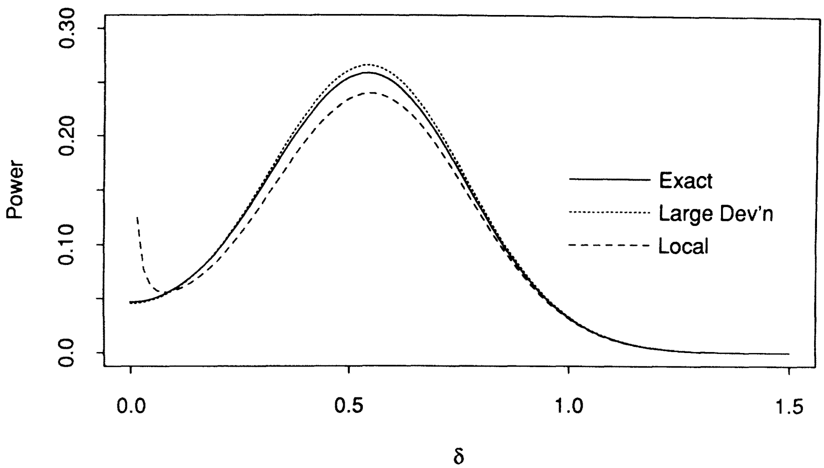

FIG. 1. *Comparing approximations to* $P^{(n)}_{\tau,\delta}(R_1 \cup R_2;\ X(\tau) > np_\tau)$ *using the cp model, with* $\tau = 0.5$.

the power of the two-sample test that would be appropriate for testing $\mathscr{H}_0$: $\delta = 0$ if τ were known. The power for the log-linear model is calculated for several values of y. We take $c = 3.0372$ when testing for the cp model and $c = 3.3301$, 3.3252 and 3.3301 for $y = 0.4$, 0.5 and 0.6, respectively, for the llcp 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 Figure 2.

Searching for the unknown change point results in only a small power loss. The test for the llcp model is substantially less powerful than the test for the

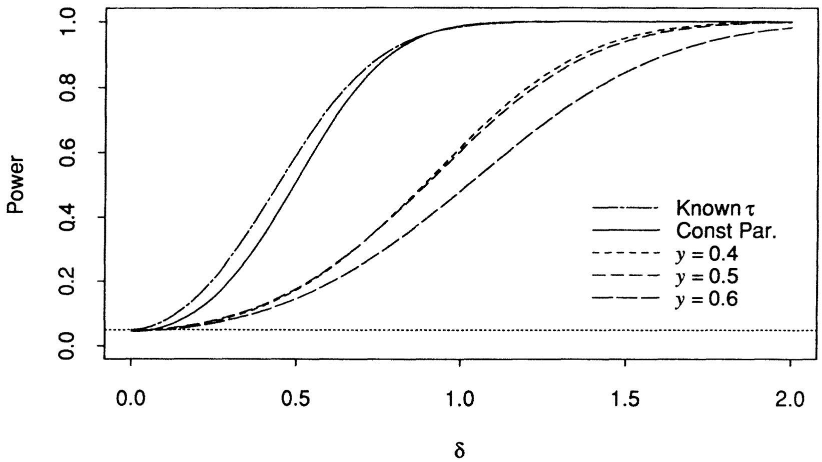

FIG. 2. *Power comparisons for* $n = 100$ *and* $\tau = 0.3$.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

cp model. The power of the llcp test shows some dependence on y; 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. This power loss implies that model selection may be difficult when only a small number of events are observed.

4. Confidence regions. In this section we derive confidence regions for τ, (τ, δ) and (τ, δ, b). The confidence regions are based on the likelihood ratio method of Siegmund (1988) and take the form

$$I_1 = \left\{\tau: l(\tau) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c_1\right\},$$

$$I_2 = \left\{(\tau, \delta): l(\tau|\delta) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c_2\right\},$$

$$I_3 = \left\{(\tau, \delta, b): l(\tau|\delta, b) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c_3\right\},$$

where c_1, c_2 and c_3 are chosen to obtain the required coverage probability and $l(\tau|\delta)$ and $l(\tau|\delta, b)$ are log-likelihood ratio statistics for $\mathscr{H}_0$: $\delta = 0$ against the llcp alternative with δ and (δ, b) specified, respectively.

We approximate the coverage probabilities using local expansions in lemmas 4.1–4.3. Large deviation approximations may also be used, but these become computationally expensive when finding joint confidence regions.

The methods in this section may also be used to find confidence regions for τ and (τ, δ) under the cp model; in particular, Lemmas 4.1 and 4.2 are unchanged. Worsley (1986) uses recursive methods to perform the computations exactly.

Let $P_\tau^{(m,n)}$ denote a measure under which the llcp model holds, conditioned on $X(\tau) = m$ and $X(t) = n$. For I_1 to have $(1 - \alpha)100\%$ coverage requires choosing $c_1 = c_1(\tau, m, n, y)$ such that $P_\tau^{(m,n)}(\tau \in I_1 | S_n = ny) = 1 - \alpha$. By sufficiency, the conditioning removes dependency on the unknown parameters a, δ and b.

LEMMA 4.1. *Let $m, n \to \infty$ with m/n convergent and τ and y fixed. Then*

$$P_\tau^{(m,n)}(\tau \in I_1 | S_n = ny) \to e^{-c_1} + (1 - e^{-c_1}) h\big(\hat{\delta}, c_1\big)$$
$$\sim \big(1 + \nu(\hat{\delta})\big) e^{-c_1} \quad \textit{as } c_1 \to \infty.$$

PROOF. We note $\tau \in I_1 \Leftrightarrow l(\tau) \geq \sup_{\tau_0 \leq t \leq \tau_1} l(t) - c_1$. Therefore,

$$1 - \alpha = P_\tau^{(m,n)}(\tau \in I_1 | S_n = ny)$$
$$= P_\tau^{(m,n)}\left(\sup_{\tau_0 \leq t \leq \tau_1} l(t) \leq l(\tau) + c_1 \,\middle|\, S_n = ny\right).$$

The result then follows from Lemma 3.1 with p_τ replaced by $p = m/n$. □

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

The coverage probability of I_2 is evaluated conditionally on n and S_n, giving

$$(19)\qquad \begin{aligned} 1-\alpha &= P_{\tau,\delta}^{(n)}\big((\tau,\delta)\in I_2 \big| S_n = ny\big) \\ &= P_{\tau,\delta}^{(n)}\Big(\sup_{\tau_0\le t\le\tau_1} l(t) \le l(\tau|\delta) + c_2 \Big| S_n = ny\Big). \end{aligned}$$

LEMMA 4.2. *Let $n\to\infty$ with τ, δ and y fixed. Then*

$$P_{\tau,\delta}^{(n)}\big((\tau,\delta)\in I_2 | S_n = ny\big)$$

$$(20)\qquad \to \int_0^{c_2}\big(1-e^{-(c_2-z)}\big)\big(1-h(\delta,c_2-z)\big)\frac{1}{\sqrt{\pi z}}e^{-z}\,dz$$

$$(21)\qquad \approx 1-\sqrt{\frac{c_2}{\pi}}\,e^{-c_2}\big(1(1+\nu(\delta)) + c_2^{-1}(1-\nu(\delta))\big).$$

PROOF. Conditioning on $X(\tau)$, we can rewrite (19) as

$$(22)\qquad \begin{aligned} 1-\alpha = \sum_{m=0}^{n} P_\tau^{(m,n)}\Big\{ &\sup_{\tau_0\le t\le\tau_1}(l(t)-l(\tau)) \\ &\le c_2 - \big(l(\tau)-l(\tau|\delta)\big)\Big| S_n = ny\Big\} \\ &\times P_{\tau,\delta}^{(n)}\big(X(\tau)=m|S_n=ny\big). \end{aligned}$$

The sum in (22) can be restricted to values of m such that $l(\tau)-l(\tau|\delta)\le c_2$. The conditional likelihood is given by

$$\begin{aligned} l(\tau|\delta) = X(\tau)&\log\left(\frac{\hat b_\delta\big(e^{\hat b_0 T}-1\big)}{\hat b_0\big(e^{\hat b_\sigma\tau}-1+e^{\delta}\big(e^{\hat b_\delta T}-e^{\hat b_\delta\tau}\big)\big)}\right) \\ &+(n-X(\tau))\log\left(\frac{\hat b_\delta e^{\delta}\big(e^{\hat b_0 T}-1\big)}{\hat b_0\big(e^{\hat b_0\tau}-1+e^{\delta}\big(e^{\hat b_\delta T}-e^{\hat b_\delta\tau}\big)\big)}\right) \\ &+\big(\hat b_\delta - b_0\big)\sum_{i=1}^{n} T_i, \end{aligned}$$

where $\hat b_\delta$ is the maximum likelihood estimator of b under the restricted alternative and is the solution of

$$(23)\qquad \frac{1}{n}\sum_{i=1}^{n} T_i = \frac{\tau e^{b\tau}+e^{\delta}\big(Te^{bT}-\tau e^{b\tau}\big)}{e^{b\tau}-1+e^{\delta}\big(e^{bT}-e^{b\tau}\big)} - \frac{1}{b}.$$

Let

$$x_0 = \frac{e^{\hat b_\delta\tau}-1}{e^{\hat b_\delta\tau}-1+e^{\delta}\big(e^{\hat b_\delta T}-e^{\hat b_\delta\tau}\big)}.$$

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

If $X(\tau) = nx_0$, then $\hat{b} = \hat{b}_\delta$, $\hat{\delta} = \delta$ and $l(\tau) - l(\tau|\delta) = 0$. A proof similar to Lemma 2.4 shows $(l(\tau) - l(\tau|\delta))/n$ is a convex function of $X(\tau)/n$ minimized at x_0, and hence the range of $X(\tau)$ such that $l(\tau) - l(\tau|\delta) < c_2$ will be an interval $[m_0, \mathrm{m}_1]$, where

$$m_0 = nx_0 - O(\sqrt{n}) \quad \text{and} \quad m_1 = nx_1 + O(\sqrt{n}).$$

When $X(\tau)$ is restricted to this range, $\hat{\delta} - \delta = O(n^{-1/2})$. Applying Lemma 4.1 and compactness arguments shows

$$P_\tau^{(m,n)}\left\{ \sup_{\tau_0 \le t \le \tau_1} l(t) \le l(\tau|\delta) + c_2 \middle| S_n = ny \right\}$$
$$-(1 - e^{-(c_2 - z)})(1 - h(\delta, c_2 - z)) \to 0$$

uniformly over $m_0 < m < m_1$ as $n \to \infty$, where $z = l(\tau) - l(\tau|\delta)$. By standard likelihood theory, $2(l(\tau) - l(\tau|\delta))$ converges in law to a χ_1^2 random variable, which leads to (20). Substituting (16) and expanding the integral around $z = c_2$ leads to (21); this is a continuous time version of equation (26) of Siegmund (1988). □

LEMMA 4.3. *Let $n \to \infty$ with τ, δ and b fixed. Then*

$$P_{\tau,\delta,b}^{(n)}((\tau, \delta, b) \in I_3) \to \int_0^{c_3} (1 - e^{-(c_3 - z)})(1 - h(\delta, c_3 - z))e^{-z}\, dz$$
$$\approx 1 - c_3 e^{-c_3}\left(1 + \nu(\delta) - \frac{1}{c_3}(1 - \nu(\delta))\right).$$

PROOF. This involves conditioning on $X(\tau)$ and $\sum_{i=1}^n T_i$ and following the proof of Lemma 4.2. The conditional likelihood ratio is given by

$$l(\tau|\delta, b) = X(\tau)\log\left(\frac{b(e^{\hat{b}_0 T} - 1)}{\hat{b}_0(e^{b\tau} - 1 + e^{\delta}(e^{bT} - e^{b\tau}))}\right)$$
$$+(n - X(\tau))\log\left(\frac{be^{\delta}(e^{\hat{b}_0 T} - 1)}{\hat{b}_0(e^{b\tau} - 1 + e^{\delta}(e^{bT} - e^{b\tau}))}\right) + (b - \hat{b}_0)\sum_{i=1}^n T_i,$$

and $2(l(\tau) - l(\tau|\delta, b))$ has asymptotically a chi-square distribution with two degrees of freedom, which leads to the result. □

5. Example: British coal mining data. We apply the methods of the preceding sections to British coal mining accident data as published by Jarrett (1979). This version lists 191 accidents resulting in 10 or more deaths between 15 March 1851 and 22 March 1962. The cumulative accident count is shown in Figure 3.

Several analyses of this data set appear in the literature. Cox and Lewis (1966) fit a log-quadratic model to the data original data. Change point models are fitted in Akman and Raftery (1986a) and Raftery and Akman (1986).

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

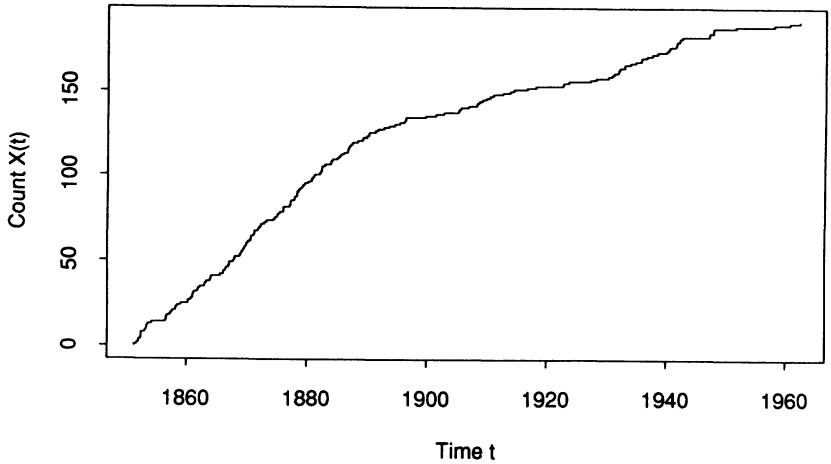

FIG. 3. *British coal mining accident data.*

We plot $l(t)$ using the cp model in Figure 4 and using the llcp model in Figure 5. Since endpoints for the observation interval are not precisely defined, we take the first and last observations to be the endpoints and do not count these events.

With both models, $l(t)$ is maximized at $t = 1890 \cdot 19$. The values of four test statistics are summarized in Table 2, in the 1851–1962 column. For testing $\mathscr{H}_0$: $\delta = 0$ against the llcp alternative, the maximum of 6.27 has an attained significance level of 0.027 (using 10% truncation). Hence we conclude the change point is necessary. For testing $\mathscr{H}_0$: $b = 0$ against the llcp alternative, the log-likelihood ratio is 0.36, which is not significant. This suggests the

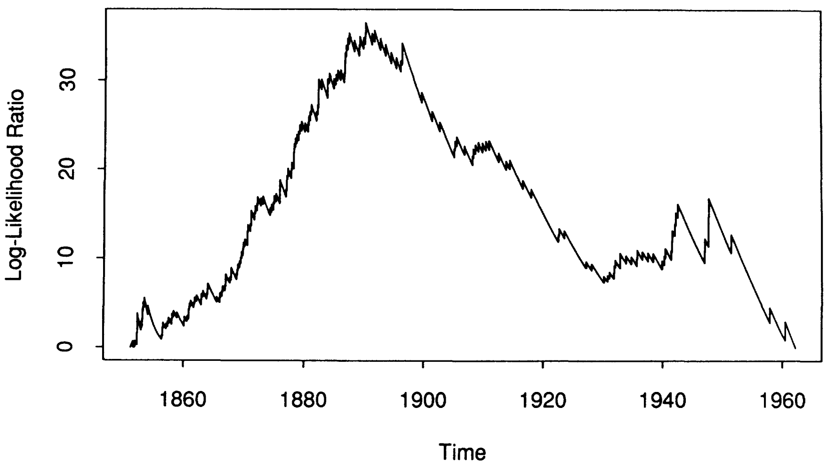

FIG. 4. $l(t)$ *for the cp model.*

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

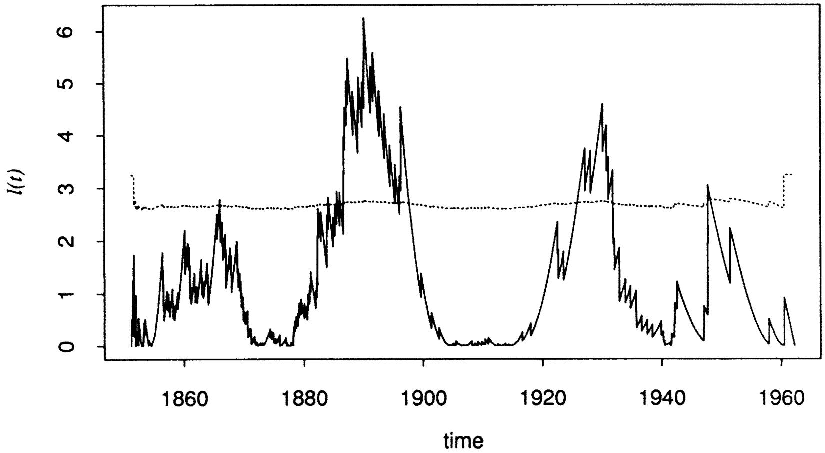

FIG. 5. $l(t)$ *for the llcp model. The dotted line defines an approximate* 95% *confidence region for* τ.

cp model provides a much better fit than the llcp model; similar conclusions have been reached by Raftery and Akman (1986) using Bayesian methods.

However, the cp model does not appear to be a perfect fit. In Figure 5, the likelihood ratio process also has a second peak around 1930, raising the possibility of a second change. To analyze this more thoroughly, we consider the subset of 67 accidents from 10 March 1890 to 22 March 1962. The results of applying various tests are shown in Table 2. Due to the smaller number of events, truncation has been increased to 30%. The first two tests, testing homogeneity against the cp and ll alternatives, respectively, are both not significant. However, the final two tests with the llcp alternative are both significant. The point estimates suggest a gradual decrease in the accident rate ($\hat{b} = -0.042$), and a sudden increase ($\hat{\delta} = 1.76$) at $\hat{\tau} = 1930.16$. This example shows the advantage of backward selection over forward selection: The forward selection method would conclude the data is homogeneous, while the backward selection shows evidence of both a change point and log-linear term.

TABLE 2
Log-likelihood ratio statistics for various tests with the coal mining data

		1851–1962		1890–1962	
Null	**Alternative**	**llr**	$\hat{\tau}$	**Sllr**	$\hat{\tau}$
$b = \delta = 0$	$b = 0$	36.24	1890.19	1.25	1940.43
$b = \delta = 0$	$\delta = 0$	30.33	—	0.58	—
$b = 0$	llcp	0.36	—	5.09	—
$\delta = 0$	llcp	6.27	1890.19	5.76	1930.15

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

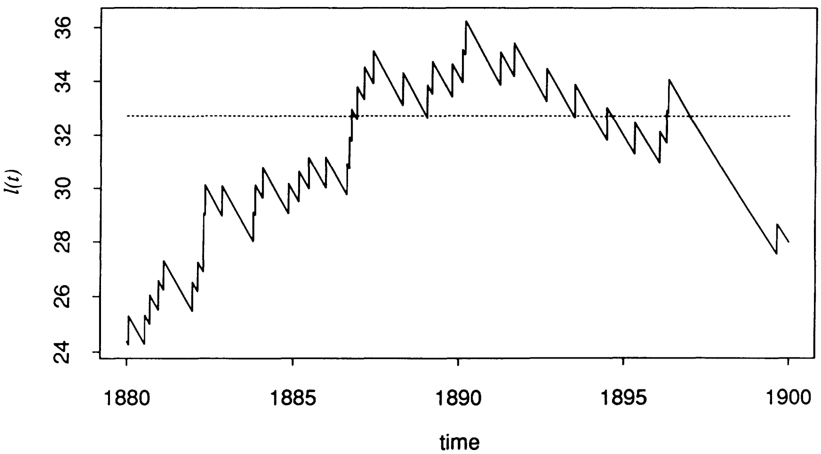

FIG. 6. *95% confidence region for τ using the cp model.*

Confidence regions for τ, computed using the local expansion, are shown for the cp model in Figure 6 and for the llcp model in Figure 5; the confidence region consists of those times for which $l(t)$ is above the dotted line. Since there is evidence of a second change, the peaks around 1930 should be ignored when constructing a confidence region for the change around 1890. The regions around 1890 included in the confidence region have total length 8.08 years using the cp model and 11.31 years using the llcp model.

Approximate 95% joint confidence regions for (τ, δ) are shown in Figure 7, computed using the cp model (solid lines) and llcp model (dotted lines). The llcp model results in a substantially larger confidence region. In particular,

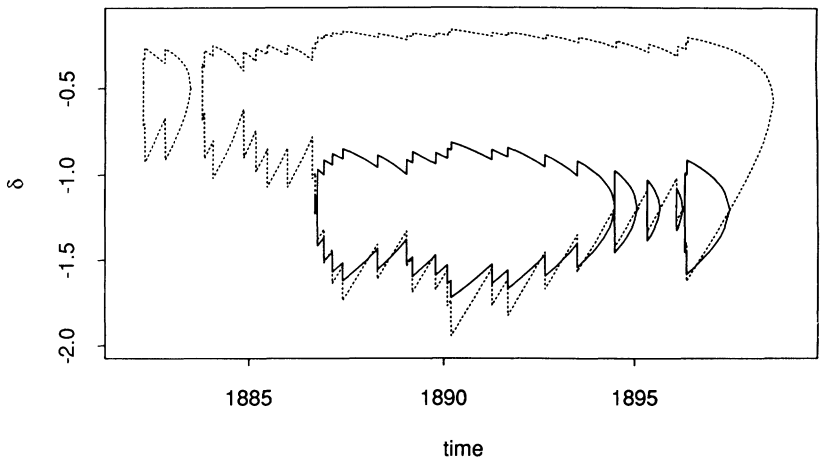

FIG. 7. *95% confidence regions for (τ, δ). The region outlined by the solid lines is computed under the cp model, and the region outlined by the dotted lines is computed under the llcp model.*

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

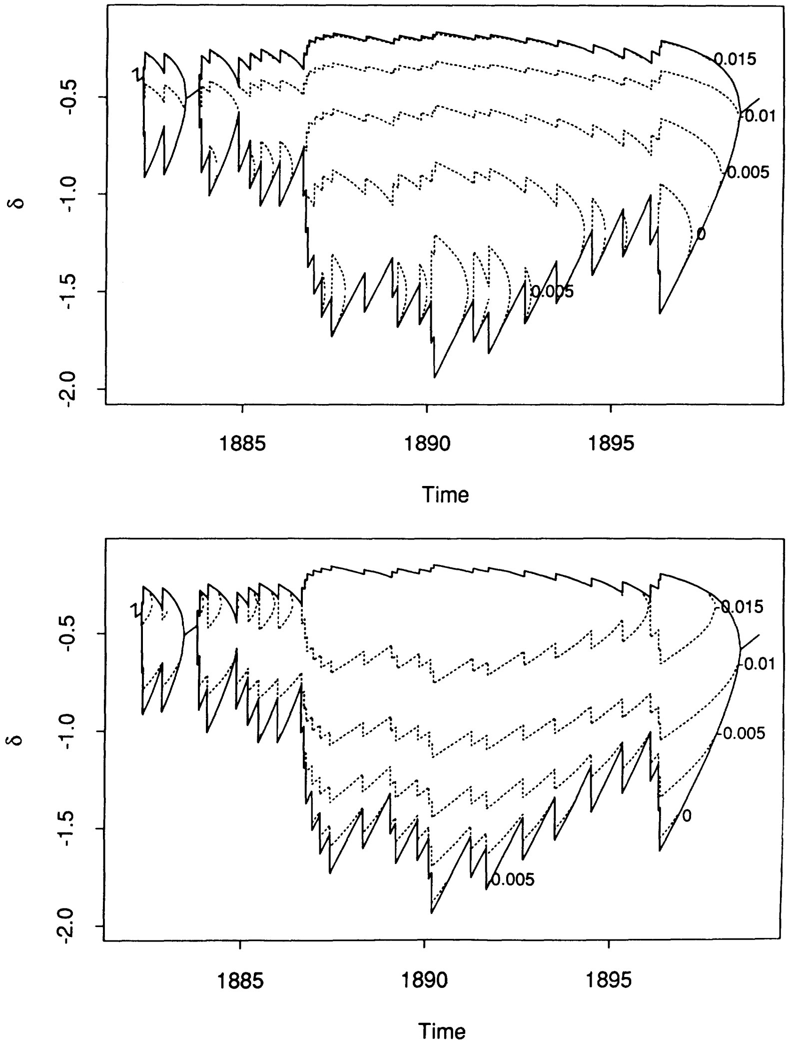

FIG. 8. *90% confidence region for (τ, δ, b). The top panel is a contour plot of the upper bounds for b, as a function of (τ, δ). The bottom panel is a contour plot of the lower bounds.*

larger (closer to 0) values of δ are included in the region, since some of the observed decrease in the accident rate is now modeled by the log-linear term.

An approximate 90% confidence region for (τ, δ, b) is shown in Figure 8. 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 8 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).

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

APPENDIX

Conditional distribution of $X(\tau)$. In this Appendix we derive a large deviation approximation to the conditional distribution of $X(\tau)$ given $X(T)$ and $\sum_{i=1}^{n} T_i$. The approach we take is based on saddle point approximations to the distribution of sums of random variables. These methods have received considerable attention recently and we only give the important details for our application. A useful reference for further reading is Field and Ronchetti (1990); Sections 5.4 and 5.5 are particularly relevant to our application. The main advantage of the saddle point approach over the central limit theorem approach is greater accuracy, especially for estimating small tail probabilities.

Conditionally on $X(T) = n$, $T_1, \dots, T_n$ are distributed as the order statistics of a sample of size n from a distribution

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

where $\psi(\delta, b)$ is defined by (12). We can write

$$P_{\tau,\delta}^{(n)}(X(\tau) = np | S_n = ny) = \frac{P_{\tau,\delta,b}^{(n)}(n - X(\tau) = n(1-p), S_n = ny)}{P_{\tau,\delta,b}^{(n)}(S_n = ny)};$$

the quantities on the right-hand side being densities in S_n. Let b_δ be the mle of b when δ is known and $\psi_\delta(\delta, b)$ the second derivative of ψ with respect to b. Using standard saddle point approximations gives

$$P_{\tau,\delta,b}^{(n)}(X(\tau) = np, S_n = ny)$$
$$\approx \frac{1}{2\pi n \left|\psi''(\hat{\delta}, \hat{b})\right|^{1/2}} \left(\frac{\exp[by + \delta(1-p) - \psi(\delta, b)]}{\exp\left[\hat{b}y + \hat{\delta}(1-p) - n\psi(\hat{\delta}, \hat{b})\right]} \right)^n$$

and

$$P_{\tau,\delta,b}^{(n)}(S_n = ny)$$
$$\approx \frac{1}{\left(2\pi n \psi_\delta''(\delta, b_\delta)\right)^{1/2}} \exp((b - b_\delta)y - n(\psi(\delta, b) - \psi(\delta, b_\delta))),$$

where b_δ is defined by (23) and $\psi_\delta''(\delta, b)$ denotes the second derivative of ψ with δ fixed. This leads to

$$(24) \qquad P_{\tau,\delta}^{(n)}(X(\tau) - np | S_n = ny) \approx \frac{\sqrt{\psi_\delta''(\delta, b_\delta)}}{\sqrt{2\pi n \left|\psi''(\hat{\delta}, \hat{b})\right|}} \exp(-nl_\delta(\tau)),$$

where

$$l_\delta(\tau) = \left(\hat{b} - b_\delta\right)y + (\hat{\delta} - \delta)(1-p) - \left(\psi(\hat{\delta}, \hat{b}) - \psi(\delta, b_\delta)\right).$$

Note that $nl_\delta(\tau)$ is the log-likelihood ratio statistic for testing the specified value of δ against the llcp alternative when τ is known.

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms

Acknowledgments. This work is based on parts of the author's Ph.D. dissertation. I thank my advisor, Professor D. Siegmund, for many helpful suggestions.

REFERENCES

AKMAN, V. E. and RAFTERY, A. E. (1986a). Asymptotic inference for a change-point Poisson process. *Ann. Statist.* **14** 1583–1590.

AKMAN, V. E. and RAFTERY, A. E. (1986b). Bayes factors for non-homogeneous Poisson processes with vague prior information. *J. Roy. Statist. Soc. Ser. B* **48** 322–329.

CHERNOFF, H. and RUBIN, H. (1956). The estimation of the location of a discontinuity in density. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 19–37. Univ. California Press, Berkeley.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.

FIELD, C. and RONCHETTI, E. (1990). *Small Sample Asymptotics*. IMS, Hayward, Calif.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 71–83.

JARRETT, R. G. (1979). A note on the intervals between coal-mining disasters. *Biometrika* **66** 191–193.

KIM, H. J. and SIEGMUND, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76** 409–423.

LOADER, C. R. (1992). Boundary crossing probabilities for locally Poisson processes. *Ann. Appl. Probab.* **2** 199–228.

MAGUIRE, E. A., PEARSON, E. S. and WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **39** 168–180.

PYKE, R. (1959). The supremum and infimum of the Poisson process. *Ann. Math. Statist.* **30** 568–576.

RAFTERY, A. E. and AKMAN, V. E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika* **73** 85–89.

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56** 31–48.

WORSLEY, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika* **73** 91–104.

YAO, Y. C. (1986). Maximum likelihood estimation in hazard rate models with a change-point. *Comm. Statist. Theory Methods* **15** 2455–2466.

AT&T BELL LABORATORIES
ROOM 2C-279
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974-2070

This content downloaded from 128.104.46.206 on Wed, 24 Jan 2018 22:40:09 UTC
All use subject to http://about.jstor.org/terms