The Annals of Applied Probability
1992, Vol. 2, No. 1, 199–228

BOUNDARY CROSSING PROBABILITIES FOR LOCALLY POISSON PROCESSES

BY CLIVE R. LOADER

AT&T Bell Laboratories

We derive large deviation approximations to boundary crossing probabilities for a class of point processes which can be approximated locally as Poisson processes. In the special case of empirical processes, we are able to obtain second order correction terms. The methods are applied to Kolmogorov–Smirnov testing, where we are able to obtain accurate approximations to the significance level when the null hypothesis is an exponential family with unknown nuisance parameters.

1. Introduction. Suppose $\{X(t);\ 0 < t < T\}$ is a point process. We are interested in approximating probabilities of the form

$$P\Big\{\inf_{\tau_0 \le t \le \tau_1} (X(t) - c(t)) \le 0\Big\} \tag{1}$$

for constants τ_0 and τ_1, and a boundary function $c(t)$. Exact methods for evaluating (1) are known in only a small number of special cases; for example, when $X(t)$ is a Poisson process or the empirical process of i.i.d. observations. Moreover, these methods involve recursive formulae which can be difficult to evaluate when the number of events grows large; see Shorack and Wellner (1986) for a discussion of several recursive formulae for empirical processes.

When exact methods are unavailable or computationally difficult, a common approach is to approximate $X(t)$ by a Gaussian process with the same mean and covariance structure. General methods for approximating boundary crossing probabilities for Gaussian processes have been widely studied; see for example Durbin (1985). The disadvantage to this approach is that the resulting approximations are not particularly accurate.

In this paper we derive approximations to (1) for a class of processes which can be approximated locally by Poisson process. Our approximations take the form

$$P\Big\{\inf_{\tau_0 \le t \le \tau_1} (X(t) - c(t)) \le 0\Big\} \approx \int_{\tau_0}^{\tau_1} (c'(t) - \mu(t, c(t)))g(t)\,dt, \tag{2}$$

where $\mu(t, c(t))$ approximates the local rate at t when $X(t) = c(t)$ and $g(t)$ is a continuous approximation to $P(X(t) = c(t))$. Generally, (2) must be evaluated by numerical integration. However, this involves less work than exact computations when these are available, and can be applied to problems for which exact methods are not known.

Received June 1990; revised December 1990.

AMS 1980 *subject classifications*. Primary 60G55; secondary 60F10, 60G40.

Key words and phrases. Boundary crossing, empirical process, change point, Kolmogorov–Smirnov test, point process, Poisson process.

In Section 2, we let $X(t)$ be the empirical process of n i.i.d. $\mathscr{U}[0,1]$ observations. We replace the boundary by $na(t)$ and show under regularity conditions that (2) is a first order asymptotic approximation as $n \to \infty$, with error $1 + o(1)$. Second order corrections, with error $1 + o(n^{-1})$, are also derived.

The approach used to derive (2) for empirical processes is to note that the first hitting time, $\tau = \inf\{t: X(t) \leq c(t)\}$, has a discrete distribution, occurring only at times t_j such that $c(t_j) = j$. The approximation is then obtained by summing approximations to the individual probabilities $P(\tau = t_j)$. These approximations are obtained by considering the local behavior of $X(t)$ and $c(t)$ for t close to t_j and noting that the process $\{X(t_j) - X(t_j - t); t > 0\}$ can be well approximated by a Poisson process. This procedure can be applied more generally to processes which can be approximated locally by Poisson process. We explore this idea in Section 3.

Many statistical problems can be expressed in the form (1). The most frequently used example is the significance level of the Kolmogorov–Smirnov test. In Section 4 we study this example and in particular use our results when the null hypothesis consists of an exponential family with unknown parameters. Numerical examples show the results provide reasonable approximations in the case of exponential and normal distributions. Applications of the results to a variety of change point problems may be found in Loader (1992).

2. Boundary crossings for empirical processes.

Suppose $X_1, \dots, X_n$ are i.i.d. $\mathscr{U}[0,1]$ random variables. The empirical process $X(t)$ is defined by

$$X(t) = \sum_{i=1}^{n} I(X_i \leq t).$$

We obtain an approximation to boundary crossing probabilities for $X(t)$ in Theorem 2.1 and obtain second order corrections in Theorem 2.2 below.

THEOREM 2.1. *Suppose $a(t)$ is a continuously differentiable function defined on $[t_0, 1]$ for some $t_0 > 0$. We suppose $a(t_0) = 0$, $a(t) < t$ and $ta'(t) > a(t)$ for all t. Let τ_0 and τ_1 be constants with $t_0 \leq \tau_0 < \tau_1 < 1$. Further, if the function*

$$l(t) = a(t)\log\left(\frac{a(t)}{t}\right) + (1 - a(t))\log\left(\frac{1 - a(t)}{1 - t}\right) \tag{3}$$

attains its minimum over $[\tau_0, \tau_1]$ at either of the endpoints, then its derivative is 0 at these points. Then

$$P\left\{\inf_{\tau_0 \leq t \leq \tau_1} (X(t) - na(t)) \leq 0\right\} = \sqrt{n}\int_{\tau_0}^{\tau_1}\left(a'(t) - \frac{a(t)}{t}\right)\frac{\exp(-nl(t))}{\sqrt{2\pi a(t)(1 - a(t))}}\,dt(1 + o(1)) \tag{4}$$

as $n \to \infty$.

Our approach to proving Theorem 2.1 is to note that the first passage time has a discrete distribution and then use local approximations to the individual probabilities. This is similar to the approach used by Woodroofe (1976) to approximate boundary crossing probabilities arising in sequential analysis. In the interests of clarity, we only sketch the main ideas here; technical details are deferred to Section A of the Appendix. The methods of Aldous [(1989), Chapter D] can also be adapted to derive (4), although proving the stated error term using this approach seems difficult.

If $l(t)$ attains a unique minimum at an interior point s of $[\tau_0, \tau_1]$, then (4) can be approximated by a Laplace approximation, avoiding the need for a numerical integration. However, usually s will still need to be found numerically and the resulting approximation is often less accurate.

Proof of Theorem 2.1. We first suppose $a(\tau_0) > 0$, and define first passage times:

$$\tau = \inf\{t: X(t) < na(t)\},$$

$$\tau' = \inf\{t \geq \tau_0: X(t) < na(t)\}.$$

Then

$$\begin{aligned} P\Big\{\inf_{\tau_0 \leq t \leq \tau_1} (X(t) - na(t)) \leq 0\Big\} & \\ = P\big(X(\tau_0) \leq na(\tau_0)\big) + P(\tau_0 < \tau' \leq \tau_1) & \qquad (5)\\ = P\big(X(\tau_0) \leq na(\tau_0)\big) + P\big(\tau < \tau_0, X(\tau_0) > na(\tau_0), \tau' \leq \tau_1\big) & \\ + P(\tau_0 < \tau \leq \tau_1). & \end{aligned}$$

We concentrate on evaluating the third term on the right of (5). The first two terms represent endpoint corrections and are asymptotically of a smaller order of magnitude; we defer a proof of this to our discussion of second order corrections. Since $X(t)$ is nondecreasing and integer valued, τ can only occur when $na(t)$ is an integer. Let t_j be the solution of $na(t) = j$. Since $a'(t) > 0$, we can choose j_0 and j_1 such that $\tau_0 < t_j \leq \tau_1$ if and only if $j_0 \leq j \leq j_1$. We have

$$\begin{aligned} P(\tau_0 < \tau \leq \tau_1) &= \sum_{j=j_0}^{j_1} P(\tau = t_j) \\ &= \sum_{j=j_0}^{j_1} P\big(\tau = t_j, X(t_j) = j\big) \qquad (6)\\ &= \sum_{j=j_0}^{j_1} P\big(\tau = t_j | X(t_j) = j\big) P\big(X(t_j) = j\big). \end{aligned}$$

The distribution of $X(t_j)$ is binomial, so we can evaluate $P(X(t_j) = j)$. An application of the first order term of Lemma 2.1 shows

$$P(\tau = t_j | X(t_j) = j) = P(X(t) > na(t) \ \forall\, t < t_j | X(t_j) = na(t_j)) \tag{7}$$
$$= 1 - \frac{a(t_j)}{t_j a'(t_j)} + o(1).$$

The approximation (7) can also be derived by a time reversal argument, approximating $X(t_j - t)$ by a Poisson process with rate $na(t_j)/t_j$, replacing the boundary by its tangent at t_j which has slope $na'(t_j)$ and applying Lemma A.2. Further, since $a(\tau_0) > 0$, compactness arguments show that (7) holds uniformly in t_j, and from (6) we get

$$P(\tau_0 < \tau \le \tau_1)$$
$$= \sum_{j=j_0}^{j_1} \left(1 - \frac{a(t_j)}{t_j a'(t_j)}\right) \binom{n}{na(t_j)} t_j^{na(t_j)} (1 - t_j)^{n(1-a(t_j))} (1 + o(1)). \tag{8}$$

Applying Stirling's formula and approximating the sum by an integral leads to the right-hand side of (4). The conditions on the location of the minimum of (3) are necessary to justify passing from the sum to the integral.

The case $a(\tau_0) = 0$ follows by first choosing $\tau_0' > \tau_0$, applying the preceding argument and letting $\tau_0' \to \tau_0$. □

We note that there is some similarity between the method developed here for Poisson processes and Durbin's (1985) result for Gaussian processes which behave locally like Brownian motion. Define $I(s, X)$ by

$$I(s, X) = \begin{cases} 1, & X(t) > na(t) \ \forall\, t < s, \\ 0, & \text{otherwise,} \end{cases}$$

and write

$$P(\tau = t_j | X(t_j) = j)$$
$$= \lim_{s \to t_j^-} E(I(s, X) | X(t_j) = j)$$
$$= \frac{1}{na'(t_j)} \lim_{s \to t_j^-} \frac{n}{t_j - s} E((a(t_j) - a(s)) I(s, X) | X(t_j) = j) \tag{9}$$
$$= \frac{1}{na'(t_j)} \lim_{s \to t_j^-} \frac{1}{t_j - s} E((X(s) - na(s)) I(s, X) | X(t_j) = j),$$

since if $X(t_j) = j$, $I(s, X) = 1$ and $s > t_{j-1}$, then $X(s) = j = na(t_j)$. By comparison, Durbin (1985) shows the first passage density $p(t)$ for a Gaussian process to a smooth boundary can be factored in the form

$$p(t) = b(t) f(t),$$

where $f(t)$ is the marginal density of $X(t)$ on the boundary and $b(t)$ has a form similar to (9). The extra factor of $na'(t_j)$ in (9) disappears when we approximate the sum over j by an integral over t. The approximation (7) can be obtained by deleting the indicator function $I(s, X)$ from (9) and hence corresponds to Durbin's $p_1(t)$.

THEOREM 2.2. *Suppose the conditions of Theorem* 2.1 *hold and in addition* $a(t)$ *is twice continuously differentiable and* $a(\tau_0) > 0$. *Let* $\mu = a(\tau_0)/\tau_0$, $\lambda = (1 - a(\tau_0))/(1 - \tau_0)$ *and define* μ' *and* λ' *to be the solutions of*

$$\frac{\lambda}{a'(\tau_0)} - \log(\lambda) = \frac{\lambda'}{a'(\tau_0)} - \log(\lambda'),$$

$$\frac{\mu}{a'(t_0)} + \log(\mu) = \frac{\mu'}{a'(\tau_0)} + \log(\mu'),$$

subject to $\mu' > a'(\tau_0) > \lambda'$. *Let* $\tau_0' = a^{-1}([na(\tau_0)] + 0.5)$ *and* $\tau_1' = a^{-1}([na(\tau_1)] + 0.5)$. *Then*

$$P\Big(\inf_{\tau_0 \le t \le \tau_1} (X(t) - na(t)) < 0\Big)$$

$$= \sqrt{n}\left\{\int_{\tau_0'}^{\tau_1'}\left(a'(t) - \frac{a(t)}{t} - \frac{ta(t)a''(t)}{2n(ta'(t) - a(t))^2}\right)\right.$$

$$\times\left(1 - \frac{1 - a(t) + a(t)^2}{12na(t)(1 - a(t))}\right)\frac{\exp(-nl(t))\,dt}{\sqrt{2\pi a(t)(1 - a(t))}}$$

$$+\left(\frac{(\mu/\lambda)^{x_0}}{1 - \mu/\lambda} + \sum_{i=1}^{\infty}\left(\frac{\lambda'}{\mu}\right)^{x_i} h\left(\frac{\mu}{a'(\tau_0)}, \frac{\mu x}{a'(\tau_0)}\right)\right)$$

$$\left.\times\frac{\exp(-nl(\tau_0))}{n\sqrt{2\pi a(\tau_0)(1 - a(\tau_0))}}\right\}\left(1 + o\left(\frac{1}{n}\right)\right),$$

where $h(\cdot, \cdot)$ *is defined by* (39) *and* $x_i = na(\tau_0) - \lfloor na(\tau_0)\rfloor + i$; $i = 0, 1, \dots$. *To a good approximation*,

$$\sum_{i=1}^{\infty}\left(\frac{\lambda'}{\mu'}\right)^{x_i} h\left(\frac{\mu}{a'(\tau_0)}, \frac{\mu x}{a'(\tau_0)}\right) \approx \frac{(a'(\tau_0) - \mu)(\lambda'/\mu')^{x_1}}{(\mu' - a'(\tau_0))(1 - \lambda'/\mu')}. \tag{10}$$

The second order corrections consist of adjustments to the integral in Theorem 2.1 and approximations to the endpoint corrections in (5). The key to adjusting the integral is Lemma 2.1 below, which provides second order corrections to $P(\tau = t_j | X(t_j) = j)$. Similar results are obtained by Woodroofe and Takahashi (1982) for sums of normal random variables.

The endpoint corrections are derived through a local expansion around τ_0. For the accuracy of the $1 + o(n^{-1})$ error term it is only necessary to include the endpoint corrections when $l(\tau)$ is maximized at $\tau = \tau_0$; however, the inclusion of these corrections improves the accuracy of the approximation more generally.

LEMMA 2.1. *Suppose $X(t)$ is the empirical process of n i.i.d. $\mathscr{U}[0,1]$ observations, conditioned on $X(T) = na(T)$. In addition to the conditions of Theorem 2.1, we assume the boundary $a(t)$ is twice differentiable. Then if $\tau_0 < T$,*

$$(11) \qquad \begin{aligned} &P\{\exists\, t < T\colon X(t) < na(t)\} \\ &\quad = \frac{a(T)}{Ta'(T)} + \frac{Ta(T)a''(T)}{2na'(T)(Ta'(T) - a(T))^2} + o(n^{-1}) \end{aligned}$$

as $n \to \infty$. If $a(t)$ is only once differentiable, the first term of (11) *still holds with error* $o(1)$.

PROOF. We will only sketch the main ideas here; technical details are deferred to Section A of the Appendix.

A straightforward argument shows

$$\{Y(t); 0 < t < T\}$$

is a martingale, where

$$Y(t) = \frac{1}{T-t}\left(X(t) - \frac{na(T)}{T}t\right).$$

Let $\tau = \inf\{t\colon X(t) - na(t) \le 0\}$. Choose t such that $na(t) > na(T) - 1$. If $\tau > t$, then $X(t) = na(T)$, so the martingale stopping theorem gives

$$(12) \qquad \begin{aligned} 0 &= E(Y(\tau \wedge t); \tau > t) + E(Y(\tau \wedge t); \tau \le t) \\ &= \frac{na(T)}{T}P(\tau = T) + E\left(\frac{n(Ta(\tau) - \tau a(T))}{T(T-\tau)}; \tau < t\right). \end{aligned}$$

A Taylor series expansion gives $a(\tau) \approx a(T) - (T - \tau)a'(T) + (T-\tau)^2 a''(T)/2$. Substituting this into (12) and rearranging gives

$$P(\tau < T) \approx \frac{a(T)}{Ta'(T)} + \frac{a''(T)}{2a'(T)}E(T - \tau).$$

Approximating $X(T) - X(T-t)$ by a Poisson process with rate $na(T)/T$ and $na(t)$ by $na(T) - n(T-t)a'(T)$ and applying Lemma A.5 gives

$$E(T - \tau) \approx \frac{Ta(T)}{n(Ta'(T) - a(T))^2},$$

which gives (11). □

PROOF OF THEOREM 2.2. We analyze each of the terms in (5) separately. We evaluate the $P(\tau_0 < \tau \le \tau_1)$ term as before. The terms $P(\tau = t_j | X(t_j) = na(t_j))$ are approximated using the second order result given by Lemma 2.1. When approximating the binomial coefficients, the second order correction to Stirling's formula must be used, giving

$$\binom{n}{na(t)} = \frac{1}{\sqrt{2\pi na(t)(1-a(t))}} \times \frac{1}{a(t)^{na(t)}(1-a(t))^{n(1-a(t))}}\left(1 - \frac{1-a(t)+a(t)^2}{12na(t)(1-a(t))} + o\left(\frac{1}{n}\right)\right).$$

Making the approximation of summation by integration accurate to the second order requires some care with the limits. The simple endpoint approximations used in the first order approximation will have error of order $1/n$. If we instead choose limits $\tau_0' = a^{-1}([na(\tau_0)] + 0.5)$ and $\tau_1' = a^{-1}([na(\tau_1)] + 0.5)$, the summation (8) is a midpoint approximation to (4) and hence has the required second order accuracy.

The first endpoint correction is a binomial tail probability,

$$\begin{aligned} P(X(\tau_0) \le na(\tau_0)) &= P(\mathscr{B}(n,\tau_0) \le na(\tau_0)) \\ &= \frac{\exp(-nl(\tau_0))}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}} \\ &\quad \times \frac{(\mu/\lambda)^{(na(\tau_0)-\lfloor na(\tau_0)\rfloor)}}{1+\mu/\lambda}(1+o(1)) \end{aligned}$$

as $n \to \infty$, where $\mu = a(\tau_0)/\tau_0$ and $\lambda = (1 - a(\tau_0))/(1 - \tau_0)$. This approximation uses an expansion of the binomial probabilities in conjunction with Stirling's formula,

$$(13)\quad P(X(\tau_0) = na(\tau_0) + x) = \frac{\exp(-nl(\tau_0))}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}}\left(\frac{\lambda}{\mu}\right)^x (1+o(1))$$

for $x = O(1)$.

The second endpoint correction can be written

$$\begin{aligned} &P(\tau \le \tau_0, X(\tau_0) \ge na(\tau_0), \tau' \le \tau_1) \\ (14)\quad &= \sum_x P(\tau \le \tau_0, \tau' \le \tau_1 | X(\tau_0) = na(\tau_0) + x) P(X(\tau_0) = na(\tau_0) + x), \end{aligned}$$

where the sum is taken over those values of x for which $na(\tau_0) + x$ is an integer and $x > 0$. Conditional on $X(\tau_0)$, $\{X(t), t < \tau_0\}$ and $\{X(t), t > \tau_0\}$ are independent and

$$\begin{aligned} &P(\tau \le \tau_0, \tau' \le \tau_1 | X(\tau_0) = na(\tau_0) + x) \\ &= P\Big(\inf_{t<\tau_0}(X(t) - na(t)) < 0 | X(\tau_0) = na(\tau_0) + x\Big) \\ &\quad \times P\Big(\inf_{\tau_1 > t > \tau_0}(X(t) - na(t)) < 0 | X(\tau_0) = na(\tau_0) + x\Big), \end{aligned}$$

which can be approximated using local linearizations of $a(t)$ and applying the results of Lemmas A.1 and A.3. This can be justified rigorously using truncation arguments similar to the proof of Lemma 2.1 in Section A of the Appendix. We get

$$P\Big(\inf_{t<\tau_0}(X(t)-na(t))<0|X(\tau_0)=na(\tau_0)+x\Big) \to h\left(\frac{\mu}{a'(\tau_0)},\frac{\mu x}{a'(\tau_0)}\right), \tag{15}$$

$$P\Big(\inf_{\tau_1>t>\tau_0}(X(t)-na(t))<0|X(\tau_0)=na(\tau_0)+x\Big) \to \exp\left(\frac{x}{a'(\tau_0)}(\lambda'-\lambda)\right)=\left(\frac{\lambda'}{\lambda}\right)^x, \tag{16}$$

as $n\to\infty$, where $h(\cdot,\cdot)$ is defined by (39). Using (13) and (14) gives

$$P(\tau\le\tau_0, X(\tau_0)>na(\tau_0),\tau'\le\tau_1) = \frac{e^{-nl(\tau_0)}}{\sqrt{2\pi na(\tau_0)(1-a(\tau_0))}}\sum_{i=1}^{\infty}\left(\frac{\lambda'}{\mu}\right)^{x_i} h\left(\frac{\mu}{a'(\tau_0)},\frac{\mu x}{a'(\tau_0)}\right)(1+o(1)),$$

where $x_i=na(\tau_0)-\lfloor na(\tau_0)\rfloor+i$.

The approximation (10) is obtained by applying Lemma A.4. □

A similar result can be obtained for a Poisson process $Z(t)$ with rate λ. In this case,

$$\begin{aligned} P\Big(\inf_{\tau_0\le t\le\tau_1}(Z(t)-\lambda a(t))\le 0\Big) &= \sqrt{\lambda}\Bigg\{\int_{\tau_0'}^{\tau_1'}\left(a'(t)-\frac{a(t)}{t}-\frac{ta(t)a''(t)}{2\lambda(ta'(t)-a(t))^2}\right) \\ &\quad\times\left(1-\frac{1}{12\lambda a(t)}\right)\frac{\exp(-\lambda m(t))\,dt}{\sqrt{2\pi a(t)}} \\ &\quad+\left(\frac{\mu^{x_0}}{1-\mu}+\sum_{i=1}^{\infty}\left(\frac{\rho}{\mu}\right)^{x_i} h\left(\frac{\mu}{a'(\tau_0)},\frac{\mu x}{a'(\tau_0)}\right)\right)\frac{\exp(-\lambda m(\tau_0))}{\lambda\sqrt{2\pi a(\tau_0)}}\Bigg\} \\ &\quad\times(1+o(\lambda^{-1})) \end{aligned}$$

as $\lambda\to\infty$, where $\mu=a(\tau_0)/\tau_0$, $x_i=\lambda a(\tau_0)-\lfloor\lambda a(\tau_0)\rfloor+i$, $\rho\le 1$ is the smallest positive solution of $\log(\rho)=(\rho-1)/a'(\tau_0)$ and

$$m(t)=a(t)\log\left(\frac{a(t)}{t}\right)+t-a(t).$$

The derivation of this result is very similar to the proof of Theorem 2.2.

3. Locally Poisson processes. The approximation given by Theorem 2.1 is based on a local approximation of the empirical process by a Poisson process and therefore we might expect similar results to hold for other counting processes which behave locally like a Poisson process. To define locally Poisson processes requires magnifying the local behavior of sample paths. For meaningful results, the process must be embedded in a sequence of processes such that the overall rate increases with the magnification.

Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\mathscr{I}[0,1]$ be the class of integer-valued, nondecreasing right-continuous functions f on $[0,1]$ with $f(0) \geq 0$. A counting process on $[0,1]$ is a function $X\colon \Omega \to \mathscr{I}[0,1]$ such that the random variable $X(A) = \int_A dX$ is measurable for all Borel subsets A of $[0,1]$. That is,

$$\{\omega \in \Omega\colon X_\omega(A) = k\} \in \mathscr{F}$$

for all $k \geq 0$ and Borel sets A. Suppose $X(A)$ has a Poisson distribution with mean $\lambda|A|$ for all Borel sets A. Then X is a Poisson process with rate λ. Here, $|A|$ denotes the Lebesgue measure of A.

Let $\{Y_n\}_{n=1}^\infty$ be a sequence of counting processes on Ω. Then Y_n converges weakly to a Poisson process with rate λ if for all finite sequences $A_1, \dots, A_m$ of disjoint Borel subsets of $[0,1]$,

$$P\left(\bigcap_{i=1}^m Y_n(A_i) = k_i\right) \to \prod_{i=1}^m \left(\frac{(\lambda|A_i|)^{k_i}}{k_i!} e^{-\lambda|A_i|}\right) \tag{17}$$

as $n \to \infty$.

Definition 1. Suppose $\{X_n\}_{n=1}^\infty$ is a sequence of counting processes on Ω. For constants $\{c_n\}_{n=1}^\infty$ with $c_n > 0$ and $c_n \to 0$, and fixed T, $0 \leq T < 1$, define

$$Y_n(t) = X_n(T + c_n t) - X_n(T).$$

If, for all $z > 0$, $\{Y_n\}$ converges weakly to a Poisson process with the rate μ^+ on $[0, z]$, then X_n is *locally right-Poisson* at T and X_n has local rate μ^+/c_n. If $T > 0$ and $\{X_n(T) - X_n((T - c_n t)^-)\}$ converges to a Poisson process with rate μ^- on $[0, z]$, then X_n is *locally left-Poisson*. If the pair $\{X_n(T) - X_n(T - c_n t), X_n(T + c_n t) - X_n(T)\}$ converges to independent Poisson processes, then X_n is *locally Poisson* at T.

For our applications, n will represent the total number of events and we take $c_n = 1/n$; we omit the subscript n. A simple example of a locally Poisson process is the sequence of empirical processes $X(t)$ of n i.i.d. $\mathscr{U}[0,1]$ observations, for which the left-hand side of (17) is a multinomial probability and a straightforward expansion shows $\mu^+ = \mu^- = 1$ for all T. Slightly more generally, if we condition on $X(T) = na(T)$, then $\mu^+ = (1 - a(T))/(1 - T)$ and $\mu^- = a(T)/T$.

Establishing results similar to Theorem 2.1 for locally Poisson processes requires an extension of the first order term of Lemma 2.1 and a continuous uniform approximation $g(t_j)$ to $P(X(t_j) = na(t_j))$.

Conditioning on $X(T) = na(T)$, we have

$$P\left\{\inf_{T-z/n \le t < T}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}$$
$$\le P\left\{\inf_{0<t<T}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}$$
$$\le P\left\{\inf_{T-z/n \le t < T}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}$$
$$+ P\left\{\inf_{0<t<T-z/n}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}.$$

Now suppose $X(t)$ is locally left-Poisson at T with rate $\mu(T, a(T)) < a'(T)$. Letting $n \to \infty$ and $z \to \infty$ and using the definition of locally Poisson in conjunction with Lemma A.2, we obtain

$$\frac{\mu(T, a(T))}{a'(T)} + \lim_{z\to\infty}\lim_{n\to\infty} P\left\{\inf_{0<t<T-z/n}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}$$
$$(18) \qquad \ge \lim_{n\to\infty} P\left\{\inf_{0<t<T}(X(t) - na(t)) < 0 | X(T) = na(T)\right\}$$
$$\ge \frac{\mu(T, a(T))}{a'(T)}.$$

We assume the limits in (18) exist; otherwise, similar relations hold with lim sup's and lim inf's. The extension of the first order term of Lemma 2.1 to locally Poisson processes now requires showing the last term of (18) 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

$$(19) \qquad \liminf_{n\to\infty} \frac{1}{n} E(X(t) | X(T) = na(T)) > a(t)$$

for all T and $t < T$, which generalizes the condition $ta'(t) - a(t) > 0$ in Theorem 2.1 and ensures boundary crossing is a rare event. Subtracting $a(T)$ from each side of (19) and letting $t \to T$ shows a necessary condition is $\mu(t, a(t)) < a'(t)$; unfortunately, this is not a sufficient condition.

THEOREM 3.1. *Suppose there exists a large deviation approximation* $g(t) = n^{\gamma}u(t)e^{-nv(t)}$ *such that*

$$P\big(X(t_j) = na(t_j)\big) = g(t_j)(1 + o(1))$$

uniformly in t_j; $u(t)$ *and* $v(t)$ *are continuous and if* $v(t)$ *attains its minimum over* $[\tau_0, \tau_1]$ *at one of the endpoints, its derivative exists and is* 0 *at these points. Suppose also that* $X(\cdot)$, *conditioned on* $X(t) = na(t)$, *is locally Poisson* *with left rates* $\mu(t, a(t))$ *and the remainder term of* (18) *is* 0. *Then*

$$P\left\{\inf_{\tau_0 \le t \le \tau_1} (X(t) - na(t)) \le 0\right\} = \int_{\tau_0}^{\tau_1} n(a'(t) - \mu(t, a(t)))g(t)\,dt(1 + o(1)) \tag{20}$$

as $n \to \infty$.

The proof of Theorem 3.1 essentially follows that of Theorem 2.1 and is omitted.

Often, it is difficult to verify directly that a process is locally Poisson. However, for processes arising naturally in statistics, we may be prepared to assume without rigorous proof that a process can be approximated locally as a Poisson process. In this case, it will still be necessary to find $\mu(t, a(t))$. There are several possible representations; one which may be useful is

$$\lim_{s \to t^-} \frac{1}{t - s} E(X(t) - X(s) | X(t) = na(t)) \approx n\mu(t, a(t)). \tag{21}$$

Extension of the second order correction to locally Poisson processes appears difficult; moreover, our applications involve two boundary problems for which our numerical examples suggest the effect of the second boundary is often larger than the error of the approximation (20).

4. Application to Kolmogorov–Smirnov testing. The Kolmogorov–Smirnov test is designed to test the null hypothesis that independent observations come from a prespecified distribution. The significance level is defined as a boundary crossing probability for the empirical process for which we can apply the methods developed in Section 2.

Suppose $X_1, \dots, X_n$ are i.i.d. from a continuous distribution F. The Kolmogorov–Smirnov statistic for testing $\mathscr{H}_0$: $F = F_0$ versus $\mathscr{H}_1$: $F \ne F_0$ is

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

where $\hat{F}(x)$ is the empirical distribution function. The distribution of D_n under $\mathscr{H}_0$ does not depend on F_0, so without loss of generality we assume $F_0(x) = x$ for $0 < x < 1$.

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

$$\alpha = P_0\left(\sup_{0 \le t \le 1} |X(t) - nt| \ge n\eta\right). \tag{22}$$

Letting $a(t) = t - \eta$ and applying (4) gives

$$\alpha = 2\eta \int_\eta^1 \frac{\exp(-nl(t))}{t\sqrt{2\pi n(t - \eta)(1 - t + \eta)}}\,dt(1 + o(1)) \tag{23}$$

TABLE 1
Approximations to significance level of the Kolmogorov–Smirnov test

n	Method	c = 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

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 (23) arises from the upper boundary which is treated by time reversal. Since $a''(t) = 0$, the second order term in lemma 2.1 disappears. To obtain second order corrections, we use the second term of Stirling's formula, replace the lower limit by $\eta + 1/2n$ and add $(1 - \eta)^n$, which is the probability of hitting the boundary at η.

Several other methods for approximating and evaluating (22) are available. These include combinatorial recursions and Gaussian approximations; see Durbin (1973a) for details. Smirnov's approximation is

$$(24) \qquad P_0\left(\sup_{0 \le x \le 1} \left(\hat{F}(x) - x\right) \ge \frac{c}{\sqrt{n}}\right) \approx e^{-2c(c + 1/3\sqrt{n})},$$

which we multiply by 2 to obtain a two-sided approximation; see also Section 3.6 of Durbin (1973a). The approximation given by Siegmund (1982) can be obtained by applying a Laplace approximation to (23).

The approximations are compared in Table 1. Here, $c = \eta\sqrt{n}$. The large deviation approximation performs very well, with relative error 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 for $n = 50$ and $n = 100$. 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 [obtained by letting $n \to \infty$ in (24)] performs poorly. Smirnov's approximation (24) performs slightly better than the large deviation

TABLE 2
Approximations to the significance level of the likelihood ratio test for a Poisson process change point with 10% truncation

n	Method	c 2.0	2.5	3.0	3.5	4.0
	Gaussian	0.4639	0.1890	0.0578	0.01332	0.00234
50	Exact	0.4399	0.1771	0.0563	0.01430	0.00230
	Large Dev'n	0.4484	0.1836	0.0576	0.01356	0.00218
	2nd Order	0.4313	0.1782	0.0569	0.01460	0.00234
100	Exact	0.4431	0.1782	0.0539	0.01258	0.00225
	Large Dev'n	0.4501	0.1820	0.0553	0.01279	0.00228
	2nd Order	0.4333	0.1770	0.0539	0.01261	0.00227

approximation for small c and slightly worse for large c; however, (24) does not require numerical integration.

As an example where the second order terms play a greater role, we consider the likelihood ratio test for a Poisson process change point. Suppose a Poisson process $Z(t)$ is observed on $[0, 1]$, with rate λ_0 for $0 < t < \tau$ and λ_1 for $\tau < t < 1$; λ_0, λ_1 and τ are all unknown. The log-likelihood ratio statistic for testing $\mathscr{H}_0$: $\lambda_0 = \lambda_1$ versus $\mathscr{H}_1$: $\lambda_0 \neq \lambda_1$ is $\sup_{0<t<1} nl(t)$, where $l(t)$ is defined by (3) with $a(t) = Z(t)/n$ and $n = Z(1)$; see Loader (1992) for more details. We reject $\mathscr{H}_0$ if $\sup_{\tau_0 \le t \le \tau_1} l(t) \ge \eta^2/2$ for some $\eta > 0$ and $0 < \tau_0 < \tau_1 < 1$. Defining $p_t < t < q_t$ to be the solutions for $a(t)$ obtained by setting (3) equal to $\eta^2/2$, the significance level is then

$$\alpha = 1 - P_0\big(np_t \le Z(t) \le nq_t \;\forall\, \tau_0 \le t \le \tau_1 | Z(1) = n\big).$$

We condition on $Z(1)$ to remove dependency on the common value of λ_0 and λ_1. Note $p_t = t - \eta\{t(1-t)\}^{1/2} + o(\eta)$ and $q_t = t + \eta\{t(1-t)\}^{1/2} + o(\eta)$ and hence these boundaries are similar to those proposed by Anderson and Darling (1955) for testing goodness of fit.

We compare the approximations in Table 2 for $n = 50$ and $n = 100$, $\tau_0 = 1 - \tau_1 = 0.1$ and various values of $c = \eta\sqrt{\eta}$. The Gaussian approximation used is

$$P\Big(\sup_{\tau_0 \le t \le \tau_1} \big(|W_0(t)| - c\sqrt{t(1-t)}\big) > 0\Big)$$
$$\approx \phi(c)\left((c - c^{-1})\log\left(\frac{\tau_1(1-\tau_0)}{\tau_0(1-\tau_1)}\right) + 4c^{-1}\right),$$

where $W_0(t)$ is a Brownian bridge; this may be derived using the methods of Jennen (1985). The exact calculation again uses recursive methods described in Durbin (1973a). The Gaussian approximation performs much better in this case than in many problems arising in sequential analysis [see Siegmund

(1985) for some examples]. The large deviation approximation is slightly better and the second order approximation is generally excellent, especially for the small significance levels of interest for testing purposes.

4.1. *The unknown parameter case.* Let $\mathscr{F} = \{F(x,\theta), \theta \in \Theta\}$ be a one-parameter exponential family, with densities

$$f(x,\theta) = \exp(\theta x - \psi(\theta)) f(x) \tag{25}$$

for some nonnegative function $f(x)$ and $\Theta = \{\theta\colon \int e^{\theta x} f(x)\,dx < \infty\}$. The null hypothesis is now $\mathscr{H}_0$: $F \in \mathscr{F}$ and the Kolmogorov–Smirnov statistic is

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

where $\hat{\theta}$ is the maximum likelihood estimate of θ. We assume $\hat{\theta}$ exists and is the solution of the likelihood equation

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

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 reject $\mathscr{H}_0$ if $D_n \geq \eta$ for some $\eta > 0$. The significance level of the test is

$$\alpha = P_0\left(D_n \geq \eta \,\Big|\, \sum_{i=1}^{n} X_i = ny\right), \tag{26}$$

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

Durbin (1973b) shows that under $\mathscr{H}_0$ the process

$$Y_n(t) = \sqrt{n}\left(\hat{F}(t) - F(t,\hat{\theta})\right)$$

converges weakly to a Gaussian process $Y(t)$ with mean 0 and covariance

$$\sigma(s,t) = F(s,\theta)(1 - F(t,\theta)) - \frac{g(s,\theta)g(t,\theta)}{I(\theta)}, \qquad s \leq t, \tag{27}$$

where $g(t,\theta) = (\partial/\partial\theta)F(t,\theta)$ and $I(\theta) = -E((\partial^2/\partial\theta^2)\log(f(X_1,\theta)))$. Although Durbin works with the unconditional distribution of $Y_n(t)$, the result also holds conditionally if y is fixed and $\theta = \hat{\theta}$ in (27). Standard methods for approximating boundary crossing probabilities for Gaussian processes, such as those given by Durbin (1985), can be used to approximate $P(\sup_t Y(t) \geq c)$.

To apply our large deviation methods, we let $X(t) = n\hat{F}(t)$ and define boundaries $p_t = F(t,\hat{\theta}) - \eta$ and $q_t = F(t,\hat{\theta}) + \eta$. To achieve suitable approximations to the local rate and distribution of $X(t)$, we embed $\mathscr{F}$ in two-parameter exponential families $\mathscr{F}_t$ with densities

$$f_t(x,\theta,\delta) = \exp(\theta x + \delta I(x > t) - \psi_t(\theta,\delta)) f(x), \tag{28}$$

where

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

The reason for choosing this exponential family is that $X(t)$ is the sufficient statistic for δ.

THEOREM 4.1. *Assume for each t the maximum likelihood estimates $\hat{\theta}_1$ and $\hat{\delta}_1$, dependent on t, exist and are the solutions of*

$$y = \psi'(\theta) + \frac{(1 - e^{\delta})(\partial/\partial\theta)F(t,\theta)}{F(t,\theta) + e^{\delta}(1 - F(t,\theta))}, \tag{29}$$

$$1 - p_t = \frac{e^{\delta}(1 - F(t,\theta))}{F(t,\theta) + e^{\delta}(1 - F(t,\theta))}. \tag{30}$$

Then

$$\begin{aligned} &P\left(\inf_t (X(t) - np_t) \le 0 \,\Big|\, \sum_{i=1}^{n} X_i = ny\right) \\ &\quad \approx \sqrt{n}\int_{t_0}^{\infty}\big(f(t,\hat{\theta}) - \mu(t,p_t)\big)\frac{\psi''(\hat{\theta})^{1/2}}{\big(2\pi|\psi_t''(\hat{\theta}_1,\hat{\delta}_1)|\big)^{1/2}}\exp(-nl(t))\,dt, \end{aligned} \tag{31}$$

where t_0 is the solution of $F(t,\hat{\theta}) = \eta$, $|\psi_t''|$ denotes the determinant of the second derivative matrix of ψ_t and

$$\mu(t,p_t) = f_t\big(t^-,\hat{\theta}_1,\hat{\delta}_1\big) = \frac{p_t e^{\hat{\theta}_1 t} f(t,0)}{\int_{-\infty}^{t} e^{\hat{\theta}_1 x} f(x,0)\,dx},$$

$$\begin{aligned} l(t) = \big(\hat{\theta}_1 - \hat{\theta}\big)y + \hat{\delta}_1(1 - p_t) + \psi(\hat{\theta}) - \psi\big(\hat{\theta}_1\big) \\ - \log\Big(F\big(t,\hat{\theta}_1\big) + e^{\hat{\delta}_1}\big(1 - F\big(t,\hat{\theta}_1\big)\big)\Big). \end{aligned}$$

A similar approximation to the probability of crossing q_t is found by time reversal.

PROOF. This is just an application of (20). The large deviation approximation to the conditional distribution of $X(t_j)$ is derived in Lemma B.1 in Section B of the Appendix. We give a heuristic derivation of $\mu(t, p_t)$ here; a rigorous proof that $X(t)$ is locally Poisson is outlined in Lemma B.2.

By the strong law of large numbers,

$$\frac{X(t) - X(s)}{n} \to P_{\theta,\delta}(s < X_1 < t) \quad \text{a.e.}$$

This suggests

$$\frac{1}{n}E\left(X(t) - X(s)|X(t), \sum_{i=1}^{n} X_i\right) \to P_{\theta,\delta}(s < X_1 < t) \quad \text{a.e.}$$

Now suppose $\theta = \hat{\theta}_1$ and $\delta = \hat{\delta}_1$. By the strong law of large numbers, $X(t)/n \to p_t$ and $\sum_{i=1}^{n} X_i/n \to y$ a.e., suggesting

$$\frac{1}{n}E\left(X(t) - X(s)|X(t) = np_t, \sum_{i=1}^{n} X_i = ny\right) \to P_{\hat{\theta}_1,\hat{\delta}_1}(s < X_1 < t).$$

The representation (21) conditioned on $\sum_{i=1}^{n} X_i$, then gives

$$\mu(t, p_t) = f_t\left(t^-, \hat{\theta}_1, \hat{\delta}_1\right) = \frac{p_t e^{\hat{\theta}_1 t} f(t,0)}{\int_{-\infty}^{t} e^{\hat{\theta}_1 x} f(x,0)\,dx}.$$

To justify the local approximation, there is a two-sided generalization of (19),

$$q_u > \left(\limsup_{n\to\infty} \ge \liminf_{n\to\infty}\right)\frac{1}{n}E(X(u)|X(t) = np_t) > p_u$$

for all t and $u \neq t$, and similarly when conditioning on $X(t) = nq_t$. For the examples below, this condition is not satisfied for all values of t; however, the exceptions generally occur only for extreme values of t when the probability of being on the boundary is very small and hence this does not cause any difficulty. □

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, (29) 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 (31).

4.2. *Examples.* We use the exponential and normal distributions as examples to compare Gaussian and large deviation approximations to the significance level of the Kolmogorov–Smirnov test.

The exponential density is

$$f(x,\theta) = \exp(\theta x + \log(-\theta))I(x \ge 0),$$

where $\theta < 0$. Letting $y = \sum_{i=1}^{n} X_i/n$, we have

$$\hat{\theta} = -\frac{1}{y},$$

$$F(t,\hat{\theta}) = 1 - \exp\left(-\frac{t}{y}\right).$$

The covariance function (27) 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 the approximation $p_1(t)$ given by Durbin (1985), which involves a one-dimensional numerical integral.

The embedded exponential family (28) 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

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

In this case (29) and (30) can be written as

$$y = -\frac{1}{\hat{\theta}_1} - \frac{(1 - e^{\hat{\delta}_1})te^{\hat{\theta}_1 t}}{1 - (1 - e^{\hat{\delta}_1})\, e^{\hat{\theta}_1 t}}, \tag{32}$$

$$1 - p_t = \frac{e^{\hat{\delta}_1}e^{\hat{\theta}_1 t}}{1 - (1 - e^{\hat{\delta}_1})\, e^{\hat{\theta}_1 t}}, \tag{33}$$

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

$$\frac{y}{t} = -\frac{1}{\hat{\theta}_1 t} - \frac{e^{\hat{\theta}_1 t} - 1 + p_t}{1 - e^{\hat{\theta}_1 t}}. \tag{34}$$

There is no closed form solution for $\hat{\theta}_1$ so (34) must be solved numerically. The second derivative matrix ψ_t'' is found by differentiating (32) and (33).

When evaluating the integral (31) over the lower boundary, the lower limit is $t_0 = -y\log(1 - \eta)$. Also, if $X(t) = np_t$, then $\Sigma_{i=1}^n X_i \geq nt(1 - p_t)$ and hence we require $1 - p_t \leq y/t$. Since $y/t \to 0$ and $1 - p_t \to \eta$ as $t \to \infty$, we have a finite upper limit of integration defined by the solution of $t(1 - p_t) = y$. 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 (31). Since θ is a scale parameter, the distribution of D_n is independent of $\Sigma_{i=1}^n X_i$ so we can fix y to be an arbitrary positive number. The probability of crossing the upper boundary $q_t = F(t,\hat{\theta}) + \eta$ is evaluated by time reversal. The limits

TABLE 3
Approximations to significance level of the Kolmogorov–Smirnov test for the exponential distribution

		c				
n	**Method**	**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

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 (26) exactly. We use simulations of size 10,000 to check the performance of the Gaussian and large deviation approximations. The results are shown in Table 3.

The large deviation 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.

Durbin (1985) reports exact calculations for the asymptotic Gaussian processes for the one-sided case. When the true asymptotic boundary crossing probability p is 0.05, Durbin obtains the first order approximation $p_1 = 0.0507$. This will correspond to a value of c close to 1.0. Table 3 shows the error arising from the approximation of $X(t)$ by a Gaussian process is much larger than the difference between $p(t)$ by $p_1(t)$. Using more accurate and computational approximations to the Gaussian process boundary crossing probability will not improve the performance.

To evaluate (31) for the normal distribution with unknown mean and variance, we parameterize the density as

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

where $y = x^2$, $\alpha = -1/(2\sigma^2) < 0$ and $\beta = \mu/\sigma^2$. This leads to a pair of

TABLE 4
Approximations to significance level of the Kolmogorov–Smirnov test for the normal distribution

		c			
n	**Method**	**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

equations similar to (29) and (30) to define the maximum likelihood estimates of α and β under the embedded change point family.

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

The Gaussian approximation is fairly straightforward. When $\mu = 0$ and $\sigma^2 = 1$, the covariance function is

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

for $s < t$, and Durbin's approximation $p_1(t)$ can be evaluated by numerical integration.

The approximations are compared with simulations of size 10,000 in Table 4. The large deviation approximation tends to overestimate the true probability for small c but performs well for the larger values of c which are of interest for significance testing. The Gaussian approximation substantially overestimates the true probability.

APPENDIX

A. Derivation of boundary crossing results. In this section we present some results used in the proof of Theorems 2.1 and 2.2 and complete the proof of Lemma 2.1. We begin with a series of lemmas concerning level crossing probabilities for the process $aZ(t) + \gamma t$, where $Z(t)$ is a Poisson process with constant rate λ and a and γ are constants. We then complete the proof of Lemma 2.1.

Most of the results are not new; see Pyke (1959) for related work. The approximation given in Lemma A.4 is used in Siegmund (1988). However, our approach appears to be simpler than that taken by other authors and is included here for completeness.

LEMMA A.1. *Suppose $a < 0$ and $\gamma > 0$. Then*

$$P\Big\{\sup_{t>0}(aZ(t) + \gamma t) \geq c\Big\} = \exp\left(\frac{c}{\gamma}(\lambda' - \lambda)\right), \tag{35}$$

where

$$\frac{a\lambda}{\gamma} + \log(\lambda) = \frac{a\lambda'}{\gamma} + \log(\lambda') \tag{36}$$

and $\lambda' \leq \gamma/|a|$.

PROOF. Let $\mathscr{M} = \sup_{t>0}(aZ(t) + \gamma t)$ and $\tau_c = \inf\{t\colon aZ(t) + \gamma t \geq c\}$. Since $a < 0$ implies all jumps of $aZ(t) + \gamma t$ are downward, applying the strong Markov property to the process restarted at time τ_c gives

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

This lack of memory property characterizes the exponential distribution. Therefore,

$$P(\mathscr{M} > c) = e^{-\beta c} \tag{37}$$

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

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

Letting $\delta \to 0$ gives

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

which reduces to (36) if we let $\lambda' = \lambda - \beta\gamma$ and (37) gives (35). □

LEMMA A.2. *Suppose $a > 0$ and $\gamma < 0$. Let $\mathscr{T}_+ = \inf\{t\colon aZ(t) + \gamma t > 0\}$. Then*:

(i) $P(\mathscr{T}_+ < \infty) = \lambda a/|\gamma|$.
(ii) *Conditionally on $\mathscr{T}_+ < \infty$, $aZ(\mathscr{T}_+) + \gamma\mathscr{T}_+$ is distributed uniformly on* $[0, a]$.

PROOF. The first part of this result can be derived in many ways; the simplest is to apply Corollary 8.39 from Siegmund (1985). The second part is most easily obtained using the Wiener–Hopf factorization.

Note that $\mathscr{T}_+ < \infty$ is equivalent to $\sup_{j>0} aj + \gamma T_j > 0$, where T_j is the jth event time. If $S_j = aj + \gamma T_j$, then $\{S_j, j \geq 0\}$ is a random walk and S_1 has density

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

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

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

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,

$$G_-(\theta) = \frac{\lambda}{\lambda - i\theta\gamma}. \tag{38}$$

Applying the Wiener–Hopf factorization [cf. Siegmund (1985), Theorem 8.41] gives

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

and therefore

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

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

LEMMA A.3. *Suppose $a > 0$ and $a\lambda + \gamma < 0$. Then*

$$P\left\{\sup_{t>0} (aZ(t) + \gamma t) > c\right\} = h\left(\frac{\lambda a}{|\gamma|}, \frac{\lambda c}{|\gamma|}\right),$$

where

$$1 - h\left(\frac{\lambda a}{|\gamma|}, \frac{\lambda c}{|\gamma|}\right) = \left(1 - \frac{\lambda a}{|\gamma|}\right) \sum_{j=0}^{\infty} \left(\frac{\lambda a}{|\gamma|}\right)^j P(aU_j < c) \tag{39}$$

$$= \left(1 - \frac{\lambda a}{|\gamma|}\right) \sum_{k=0}^{[c/a]} \frac{(-1)^k}{k!} \left(\frac{\lambda}{|\gamma|}(c - ak)\right)^k e^{\lambda(c-ak)/|\gamma|}, \tag{40}$$

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

PROOF. Let J be the number of finite ascending ladder times. Then J has a geometric distribution, and by Lemma A.2,

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

Also by Lemma A.2, $P(\mathscr{M} < c | J = j) = P(aU_j < c)$, from which (39) follows. Writing $c/a = m + y$ where m is an integer and $0 \le y < 1$ and substituting

$$P(aU_j < c) = \frac{1}{j!} \sum_{k=0}^{m \wedge j} \binom{j}{k} (-1)^k (m + y - k)^j \tag{41}$$

gives (40) after some rearrangement. □

Both forms of Lemma A.3 can be numerically unstable and are especially unsuitable for use in Theorem 2.2 since this requires evaluation of $h(\cdot, \cdot)$ many times. The following lemma gives a useful and fairly accurate approximation.

LEMMA A.4. *Suppose $a > 0$ and $\gamma < 0$. Then*

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

where λ' is defined by (36) and $\lambda' > |\gamma|/a$.

PROOF. Let $\tau_c = \inf\{j: S_j > c\}$, where S_j is as defined in the proof of Lemma A.2. Then by Wald's likelihood ratio identity [cf. Siegmund (1985), page 13],

$$\begin{aligned} P_\lambda(\tau_c < \infty) &= \sum_{n=1}^{\infty} \int_{\{\tau_c = n\}} \frac{dP_\lambda}{dP_{\lambda'}} \, dP_{\lambda'} \\ &= \sum_{n=1}^{\infty} \int_{\{\tau_c = n\}} \left(\frac{\lambda}{\lambda'} \right)^n e^{-(\lambda - \lambda')(S_n - an)/\gamma} \, dP_{\lambda'} \end{aligned}$$

$$= e^{-c(\lambda - \lambda')/\gamma} E_{\lambda'} \left\{ \exp\left(-(\lambda - \lambda')(S_{\mathscr{T}_c} - c)/\gamma \right) \right\} \tag{42}$$

$$\sim e^{-c(\lambda - \lambda')/\gamma} \frac{1 - E_{\gamma'}\left\{ \exp\left(-(\lambda - \lambda') S_{\tau_+}/\gamma \right) \right\}}{(\lambda - \lambda') E_{\lambda'}(S_{\tau_+})/\gamma} \quad \text{as } c \to \infty. \tag{43}$$

We have used (8.46) of Siegmund (1985) to obtain (43) from (42). Using Wald's equation,

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

and

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

by (38), so

$$E_{\lambda'}(S_{\tau_+}) = \frac{a\lambda' + \gamma}{\lambda' - \lambda}. \tag{44}$$

Similarly,

$$E_{\lambda'} \left\{ \exp\left(-\frac{1}{\gamma} (\lambda - \lambda') S_{\tau_+} \right) \right\} = P_\lambda(\tau_+ < \infty) = \frac{\lambda a}{|\gamma|} \tag{45}$$

by Lemma A.2. Substituting (44) and (45) into (43) completes the proof. □

LEMMA A.5. *Let $Z(t)$ be a Poisson process with rate $\lambda < \gamma$ and $\tau = \sup\{t: Z(t) \geq \gamma t\}$. Then*

$$E(\tau) = \frac{\lambda}{(\gamma - \lambda)^2}. \tag{46}$$

PROOF. We assume $\gamma = 1$; the general case follows by a straightforward time transformation. Since $Z(t)$ is nondecreasing, we must have $Z(\tau) = \tau$ and therefore τ is an integer-valued random variable. We have

$$\begin{aligned} P(\tau = j) &= P(\tau = j, Z(\tau) = j) \\ &= P(\tau = j | Z(j) = j) P(Z(j) = j) \\ &= (1 - \lambda) \frac{(\lambda j)^j}{j!} e^{-\lambda j}, \end{aligned}$$

using Lemma A.2. Summing over j, we get

$$\begin{aligned} E(\tau) &= (1 - \lambda) \sum_{j=1}^{\infty} \frac{(\lambda j)^j}{(j-1)!} e^{-\lambda j} \\ &= (1 - \lambda) \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} (-1)^k \frac{(\lambda j)^{j+k}}{(j-1)! k!} \\ &= (1 - \lambda) \sum_{l=1}^{\infty} \lambda^l \sum_{k=0}^{l-1} (-1)^k \frac{(l-k)^l}{(l-k-1)! k!} \qquad (47) \\ &= (1 - \lambda) \sum_{l=1}^{\infty} \lambda^l \frac{l(l+1)}{2}. \qquad (48) \end{aligned}$$

Here, we have used a combinatorial identity given on page 60 of Feller (1968) to evaluate the inner sum of (47). Evaluating (48) gives (46). □

PROOF OF LEMMA 2.1 (Continued). For the remainder of this section, all probabilities and expectations are evaluated conditionally on $X(T) = na(T)$. Using the conditions $a(t) < t$ and $ta'(t) > a(t)$ for all t, we can find a constant m such that

$$\frac{ta(T)}{T} > a(T) - m(T - t) \geq a(t) \qquad \forall\, t < T$$

and since $a(t)$ is twice differentiable, there exists v such that

$$a(t) \geq a(T) - (T - t)a'(T) + \frac{v}{2}(T - t)^2 \qquad \forall\, t < T.$$

Moreover, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$a(t) \geq a(T) - (T - t)a'(T) + \frac{a''(T) - \varepsilon}{2}(T - t)^2 \qquad \forall\, T - \delta < t < T.$$

Define $\tau = \inf\{t: X(t) \le na(t)\}$. Fix $k > 0$ and choose n large enough so that $k/n < \delta$. Choose t such that $T - k/n < t < T$ with $na(t) > na(T) - 1$. Following the derivation of (12), we get

$$P(\tau < T) \ge \frac{a(T)}{Ta'(T)} + \frac{v}{2a'(T)} E\left(T - \tau; \tau \le T - \frac{k}{n}\right) + \frac{a''(T) - \varepsilon}{2a'(T)} E\left(T - \tau; T - \frac{k}{n} < \tau < T\right). \tag{49}$$

We suppose $v < 0$; otherwise, the term involving v can be omitted. We then want an upper bound for $E(T - \tau; \tau \le T - k/n)$. Define $\tau' = \inf\{t: X(t) \le n(a(T) - m(T - t))\}$ so $\tau' < \tau$ and hence

$$E\left(T - \tau; \tau \le T - \frac{k}{n}\right) \le E\left(T - \tau'; \tau' \le T - \frac{k}{n}\right).$$

Let $j_1 = \lceil mk \rceil$ and $t_j = T - j/(nm)$. Then

$$\begin{aligned} E\left(T - \tau'; \tau' \le T - \frac{k}{n}\right) &= \sum_{j=j_1}^{na(T)} E\left(\frac{j}{nm}; \tau' = t_j\right) \\ &\le \sum_{j=j_1}^{na(T)} \frac{j}{nm} P\big(X(t_j) = na(T) - j\big). \end{aligned} \tag{50}$$

Split the sum into $j \le \sqrt{n}$ and $j > \sqrt{n}$. For $j < \sqrt{n}$,

$$\begin{aligned} P\big(X(t_j) = na(T) - j\big) &= \binom{na(T)}{j}\left(1 - \frac{t_j}{T}\right)^j \left(\frac{t_j}{T}\right)^{na(T)-j} \\ &\le \frac{(na(T))^j}{j!}\left(\frac{j}{nmT}\right)^j \left(1 - \frac{j}{nmT}\right)^{na(T)} \left(1 - \frac{j}{nmT}\right)^{-j} \\ &\le \left(\frac{ea(T)}{mT}\right)^j \exp\left(-\frac{ja(T)}{mT}\right)\left(1 - \frac{1}{\sqrt{n}\, mT}\right)^{-\sqrt{n}}. \end{aligned}$$

Summing (50) over $j < \sqrt{n}$ gives

$$\begin{aligned} &\sum_{j=j_1}^{\sqrt{n}} \frac{j}{nm} P\big(X(t_j) = na(T) - j\big) \\ &\quad \le \left(1 - \frac{1}{\sqrt{n}\, mT}\right)^{-\sqrt{n}} \sum_{j=j_1}^{\infty} \frac{j}{nm}\left(\frac{a(T)e^{1-a(T)/(mT)}}{mT}\right)^j \\ &\quad \sim \frac{1}{n} u(j_1) \end{aligned}$$

as $n \to \infty$, where $u(j_1) \to 0$ as $j_1 \to \infty$. To establish convergence of the sum, note that the relation $\log(x) < x - 1$ with $x = a(T)/mT \ne 1$ shows $a(T)e^{1-a(T)/(mT)}/mT < 1$.

For $na(T) > j > \sqrt{n}$, we use standard bounds on factorials to obtain

$$P(X(t_j) = na(T) - j)$$
$$\le e^{1/4}\sqrt{\frac{na(T)}{2\pi j(na(T) - j)}}\left(\frac{a(T)}{mT}\right)^j\left(\frac{1 - j/(nmT)}{1 - j/(na(T))}\right)^{na(T)-j}$$
$$\le \sqrt{na(T)}\exp\left(-nw\left(\frac{j}{n}\right)\right),$$

where

$$w(x) = x\log\left(\frac{mT}{a(T)}\right) + (a(T) - x)\log\left(\frac{1 - x/a(T)}{1 - x/(mT)}\right)$$
$$= \sup_p(x\log(p) + (a(T) - x)\log(1 - p))$$
$$-\left(x\log\left(\frac{x}{mT}\right) + (a(T) - x)\log\left(1 - \frac{x}{mT}\right)\right).$$

Since $w(0) = 0$, $w(x) > 0$ for $x > 0$ and $w'(0) > 0$, $w(j/n)$ is minimized over $j \ge \sqrt{n}$ at $j = \sqrt{n}$, at least for n sufficiently large. Hence

$$\sum_{j=\sqrt{n}}^{na(T)} P(X(t_j) = na(T) - j) \le \exp\left(-nw\left(\frac{1}{\sqrt{n}}\right)\right)\{1 + (na(T))^{3/2}\}$$
$$= o(n^{-1})$$

as $n \to \infty$ since the first factor converges exponentially to 0. Substituting into (49) gives

$$P(\tau < T) \ge \frac{a(T)}{Ta'(T)} = \frac{a''(T) - \varepsilon}{2a'(T)}E\left(T - \tau; T - \frac{k}{n} < \tau < T\right)$$
(51)
$$+ \frac{v}{2na'(T)}u(mk) + o(n^{-1}).$$

A similar limiting argument as $n \to \infty$ and an application of Lemma A.5 show

$$\lim_{k\to\infty}\lim_{n\to\infty} nE\left(T - \tau; T - \frac{k}{n} < \tau < T\right) = \frac{Ta(T)}{(Ta'(T) - a(T))^2}.$$

Substituting into (51) and letting $\varepsilon \to 0$ gives

$$P(\tau < T) \ge \frac{a(T)}{Ta'(T)} + \frac{Ta(T)a''(T)}{2na'(T)(Ta'(T) - a(T))^2} + o(n^{-1}).$$

A similar upper bound completes the proof. □

B. Conditional distributions. In this section we derive the approximation to the conditional distribution of $X(t)$ used in Theorem 4.1 and sketch a proof that $X(t)$ is indeed locally Poisson. Our approach will be based on large deviation approximations to the distribution of the sum of independent random variables from an exponential family. Suppose $X_1, \dots, X_n$ are i.i.d. random variables in $\mathscr{R}^d$, from an exponential family $\mathscr{F} = \{F_\theta, \theta \in \Theta\}$ with densities

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

where $\Theta = \{\theta: \int \exp(\langle \theta, x \rangle) f(x)\, dx < \infty\}$ and $\psi(\theta)$ is the normalizing function. Denote by $f_\theta^{(n)}$ the n-fold convolution of f_θ. Assume for each x there exists a solution $\hat{\theta}$ of the likelihood equations

$$x_i = \psi_i(\hat{\theta}); \qquad i = 1, \dots, d,$$

where $x = (x_1, \dots, x_d)$. The large deviation approximation to $f_\theta^{(n)}$ is

$$f_\theta^{(n)}(nx) \sim \frac{1}{(2\pi n)^{d/2} |\psi''(\hat{\theta})|^{1/2}} \exp(-n l_\theta(x)), \tag{52}$$

where $n l_\theta(x)$ is the log-likelihood ratio statistic for testing the null hypothesis of known θ against the general alternative,

$$l_\theta(x) = \langle \hat{\theta} - \theta, x \rangle - \big(\psi(\hat{\theta}) - \psi(\theta)\big).$$

For our purposes the important feature of (52) is that it provides an asymptotic approximation as $n \to \infty$ with x fixed; see Borovkov and Rogozin (1965). By contrast, a central limit approximation will only hold for values of x in a set which shrinks to the point $E_\theta(X_1)$ as $n \to \infty$. We can view (52) as applying the central limit theorem to $f_{\hat{\theta}}^{(n)}$ and applying a likelihood ratio argument to obtain $f_\theta^{(n)}$.

LEMMA B.1. *In the setting of Theorem* 4.1,

$$P\left(X(t) = np_t \Big| \sum_{i=1}^n X_i = ny \right) = \frac{|\psi''(\hat{\theta})|^{1/2}}{\left(2\pi n \left|\psi''\left(\hat{\theta}_1, \hat{\delta}_1\right)\right|\right)^{1/2}} e^{-nl(t)}(1 + o(1))$$

as $n \to \infty$ *with* t, p_t *and* y *fixed*.

PROOF. The proof involves writing the conditional probability as the ratio of densities,

$$P\left(X(t) = np_t \Big| \sum_{i=1}^n X_i = ny \right) = \frac{P(n - X(t) = n(1 - p_t), \sum_{i=1}^n X_i = ny)}{P(\sum_{i=1}^n X_i = ny)},$$

writing $n - X(t) = \sum_{i=1}^n I(X_i > t)$ and applying (52) to both the numerator and denominator. The exponential family (25) is used for the denominator and the embedded exponential family (28) is used for the numerator. □

LEMMA B.2. *Let* $X(t)$ *be as in Theorem* 4.1. *Condition on* $X(t) = np_t$ *and* $\sum_{i=1}^n X_i = ny$. *Then* $X(t)$ *is locally Poisson at* t, *with left rate*

$$\mu(t, p_t) = f_t\left(t^-, \hat{\theta}_1, \hat{\delta}_1\right) = \frac{p_t e^{\hat{\theta}_1 t} f(t, 0)}{\int_{-\infty}^t e^{\hat{\theta}_1 x} f(x, 0)\, dx}.$$

PROOF. To simplify the notation, we only show for each $z > 0$, $X(t) - X(t - z/n)$ converges in law to a Poisson random variable with mean $\mu(t, p_t)z$. Establishing that (17) holds more generally is similar. In addition, some uniformity conditions are omitted.

Let $S_n = \Sigma_{i=1}^n X_i$. Denote by $P^{(m)}$ the measure P conditioned on $X(t) = m$ and $m = np_t$. Then

$$P^{(m)}\left(X(t) - X\left(t - \frac{z}{n}\right) = k | S_n = ny\right)$$

$$(53) \qquad = \frac{P^{(m)}(X(t) - X(t - z/n) = k, S_n = ny)}{P^{(m)}(S_n = ny)}$$

$$\sim \frac{P^{(m)}(X(t) - X(t - z/n) = k) P^{(m-k)}\left(\tilde{S}_{n-k} = ny - kt\right)}{P^{(m)}(S_n = ny)}$$

as $n \to \infty$, where $\tilde{S}_{n-k}$ denotes the sum of $n - k$ random variables $\tilde{X}_1, \dots, \tilde{X}_{n-k}$ from a density

$$\tilde{f}(x) = \frac{f(x, \theta)}{F(t - z/n, \theta) + 1 - F(t, \theta)} \left(1 - I(t - \frac{z}{n}, t]\right).$$

We treat each of the terms in (53) separately. Under $P^{(m)}$, $X(t) - X(t - z/n)$ has a $\mathscr{B}\{m, (F(t) - F(t - z/n))/F(t)\}$ distribution and the standard Poisson limit theorem gives

$$(54) \quad P^{(m)}\left(X(t) - X\left(t - \frac{z}{n}\right) = k\right) \to \frac{1}{k!}\left(\frac{p_t z f(t, \theta)}{F(t, \theta)}\right)^k \exp\left(-\frac{p_t z f(t, \theta)}{F(t, \theta)}\right)$$

as $n \to \infty$.

The second term in the numerator of (53) can be written

$$(55) \quad P^{(m-k)}\left(\tilde{S}_{n-k} = ny - kt\right) = \frac{P\left(\tilde{S}_{n-k} = ny - kt, \tilde{X}(t) = m - k\right)}{P\left(\tilde{X}(t) = m - k\right)}.$$

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 (55) by

$$P\left(\tilde{S}_{n-k} = ny - kt, \tilde{X}(t) = m - k\right)$$

$$\sim \frac{1}{2\pi n |\tilde{\psi}_t''(\tilde{\theta}, \tilde{\delta})|^{1/2}} \exp\left(-(n-k)\left((\tilde{\theta} - \theta)\frac{ny - kt}{n - k} + \tilde{\delta}\frac{n - m}{n - k}\right.\right.$$

$$\left.\left. -\tilde{\psi}(\tilde{\theta}, \tilde{\delta}) + \tilde{\psi}(\theta, 0)\right)\right)$$

as $n \to \infty$, where $\tilde{b}$ and $\tilde{\delta}$ are the maximum likelihood estimates of b and δ

and

$$\tilde{\psi}(\theta,\delta) = \psi(\theta) + \log\left(F\left(t - \frac{z}{n}, \theta\right) + e^{\delta}(1 - F(t,\theta))\right)$$
$$= \psi(\theta,\delta) - \frac{zf(t,\theta)}{n\left(F(t,\theta) + e^{\delta}(1 - F(t,\theta))\right)} + o(n^{-1}).$$

This gives

$$\begin{aligned} &P\left(\tilde{S}_{n-k} = ny - kt, \tilde{X}(t) = m - k\right) \\ &\quad \sim \frac{1}{2\pi n|\tilde{\psi}_t''(\tilde{\theta},\tilde{\delta})|^{1/2}} \exp\Big(-(\tilde{\theta} - \theta)(ny - kt) - \tilde{\delta}(n - m) \\ &\qquad\qquad + (n - k)\left(\psi(\tilde{\theta},\tilde{\delta}) - \psi(\theta,0)\right)\Big) \\ &\qquad \times \exp\left(-\frac{zf(t,\tilde{\theta})}{F(t,\tilde{\theta}) + e^{\tilde{\delta}}\left(1 - F(t,\tilde{\theta})\right)} + zf(t,\theta)\right). \end{aligned} \tag{56}$$

When $\delta = 0$, $\tilde{X}(t)$ has a binomial distribution with success probability

$$\frac{F(t - z/n,\theta)}{F(t - z/n,\theta) + 1 - F(t,\theta)}$$

which gives

$$\begin{aligned} &P\left(\tilde{X}(t) = m - k\right) \\ &\quad = \binom{n-k}{m-k}\left(\frac{F(t - z/n,\theta)}{F(t - z/n,\theta) + 1 - F(t,\theta)}\right)^{m-k} \\ &\qquad \times \left(\frac{1 - F(t,\theta)}{F(t - z/n,\theta) + 1 - F(t,\theta)}\right)^{n-m} \\ &\quad \sim \binom{n-k}{m-k} F(t,\theta)^{m-k}(1 - F(t,\theta))^{n-m} \exp\left(zf(t,\theta)\left(1 - \frac{p_t}{F(t,\theta)}\right)\right) \end{aligned} \tag{57}$$

as $n \to \infty$.

We can write the denominator of (53) as

$$P^{(m)}(S_n = ny) = \frac{P(S_n = ny, X(t) = m)}{P(X(t) = m)}$$

and using the large deviation approximation,

$$\begin{aligned} &P(S_n = ny, X(t) = m) \\ &\quad \sim \frac{1}{2\pi n|\psi_t''(\hat{\theta}_1,\hat{\delta}_1)|^{1/2}} \exp\Big(-n\Big((\hat{\theta}_1 - \theta)y + \hat{\delta}_1(1 - p_t) \\ &\qquad\qquad - \left(\psi_t(\hat{\theta}_1,\hat{\delta}_1) - \psi(\theta,0)\right)\Big)\Big) \end{aligned} \tag{58}$$

as $n \to \infty$. Also,

$$P(X(t) = m) = \binom{n}{m} F(t,\theta)^m (1 - F(t,\theta))^{n-m}. \tag{59}$$

Substituting (54), (56), (57), (58) and (59) into (53) gives

$$
\begin{aligned}
&P^{(m)}\left(X(t)-X\left(t-\frac{z}{n}\right)=k \mid S_n=ny\right)\\
&\sim \frac{1}{k!}\left(\frac{p_t z f(t,\theta)}{F(t,\theta)}\right)^k \exp\left(-\frac{p_t z f(t,\theta)}{F(t,\theta)}\right)\\
&\quad\times \frac{n!(m-k)!}{m!(n-k)!}F(t,\theta)^k \exp\left(-zf(t,\theta)\left(1-\frac{p_t}{F(t,\theta)}\right)\right)\\
&\quad\times \exp\left(\left(\hat\theta_1-\theta\right)ny-(\tilde\theta-\theta)(ny-kt)+n(1-p)\left(\hat\delta_1-\tilde\delta\right)\right)\\
&\quad\times \exp\left((n-k)\left(\psi(\tilde\theta,\tilde\delta)-\psi(\theta,0)\right)-n\left(\psi\left(\hat\theta_1,\hat\delta_1\right)-\psi(\theta,0)\right)\right)\\
&\quad\times \exp\left(-\frac{zf(t,\tilde\theta)}{F(t,\tilde\theta)+e^{\tilde\delta}\left(1-F(t,\tilde\theta)\right)}+zf(t,\theta)\right)
\end{aligned}
\tag{60}
$$

$$
\begin{aligned}
&\sim \frac{1}{k!}\left(\frac{p_t z f(t,\theta)}{F(t,\theta)}\right)^k\left(\frac{n}{m}\right)^k F(t,\theta)^k \exp(k(\psi(\theta,0)-\theta t))\\
&\quad\times \exp\left(k\left(\tilde\theta t-\psi(\tilde\theta,\tilde\delta)\right)-\frac{zf(t,\tilde\theta)}{F(t,\tilde\theta)+e^{\tilde\delta}\left(1-F(t,\tilde\theta)\right)}\right)\\
&\quad\times \exp\left(\left(\hat\theta_1-\tilde\theta\right)ny+n(1-p_t)\left(\hat\delta_1-\tilde\delta\right)+n\left(\psi(\tilde\theta,\tilde\delta)-\psi\left(\hat\theta_1,\hat\delta\right)\right)\right).
\end{aligned}
\tag{61}
$$

The first line of (60) arises from (54), the second line comes from (57) and (59) and the remaining lines come from (56) and (58). It is easy to show $\tilde\theta-\hat\theta_1=O(n^{-1})$ and $\tilde\delta-\hat\delta_1=O(n^{-1})$ as $n\to\infty$. This implies

$$
\begin{aligned}
\psi(\tilde\theta,\tilde\delta)&=\psi\left(\hat\theta_1,\hat\delta\right)+\left(\tilde\theta-\hat\theta_1\right)\psi_1\left(\hat\theta_1,\hat\delta\right)+\left(\tilde\delta-\hat\delta_1\right)\psi_2\left(\hat\theta_1,\hat\delta\right)+o\left(\frac{1}{n}\right)\\
&=\psi\left(\hat\theta_1,\hat\delta\right)+\left(\tilde\theta-\hat\theta_1\right)y+\left(\tilde\delta-\hat\delta_1\right)(1-p_t)+o\left(\frac{1}{n}\right)
\end{aligned}
$$

and hence the third line of (61) converges to 1. Writing $f(t,\theta)=\exp(\theta t-\psi(\theta))f(t)$ and canceling terms in the first line of (61), we get

$$
\begin{aligned}
&P^{(m)}\left(X(t)-X\left(t-\frac{z}{n}\right)=k \mid S_n=ny\right)\\
&\sim \frac{1}{k!}\left(z\exp\left(\tilde\theta t-\psi(\tilde\theta,\tilde\delta)\right)f(t)\right)^k \exp\left(-\frac{zf(t,\tilde\theta)}{F(t,\tilde\theta)+e^{\tilde\delta}\left(1-F(t,\tilde\theta)\right)}\right)\\
&\to \frac{1}{k!}\left(zf_t\left(t^-,\hat\theta_1,\hat\delta_1\right)\right)^k \exp\left(-zf_t\left(t^-,\hat\theta_1,\hat\delta_1\right)\right)
\end{aligned}
$$

as $n\to\infty$, which completes the proof. □

Acknowledgments. This work forms part of the author's Ph.D. dissertation, written at Stanford University. I thank my advisor, Professor D. Siegmund, for his encouragement and many suggestions, and the referees and Editor for helpful comments.

REFERENCES

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

ANDERSON, T. W. and DARLING, D. A. (1955). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

BOROVKOV, A. A. and ROGOZIN, R. A. (1965). On the multidimensional central limit theorem. *Theory Probab. Appl.* **10** 55–62.

DURBIN, J. (1973a). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.

DURBIN, J. (1973b). Weak convergence of the sample distribution function when the parameters are estimated. *Ann. Statist.* **1** 279–290.

DURBIN, J. (1985). The first passage density of a Gaussian process to a general boundary. *J. Appl. Probab.* **22** 99–122.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

JENNEN, C. (1985). Second-order approximations to the density, mean and variance of Brownian first-exit times. *Ann. Probab.* **13** 126–144.

LOADER, C. R. (1992). A log-linear model for a Poisson process change point. *Ann. Statist.* To appear.

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

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

SIEGMUND, D. O. (1982). Large deviations for boundary crossing probabilities. *Ann. Probab.* **10** 581–588.

SIEGMUND, D. O. (1985). *Sequential Analysis. Tests and Confidence Intervals*. Springer, New York.

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

WOODROOFE, M. (1976). Frequentist properties of Bayesian sequential tests. *Biometrika* **63** 101–110.

WOODROOFE, M. and TAKAHASHI, H. (1982). Asymptotic expansions for the error probabilities of some repeated significance tests. *Ann. Statist.* **10** 895–908.

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