On Exact Inference for Change in a Poisson Sequence

H. Boudjellaba
Department of Mathematics and Computer Science
Laurentian University
Sudbury P3E 2C6

B. MacGibbon
GERAD and Département de mathématiques
Université du Québec à Montréal
C. P. 8888, Succursale Centre-ville
Montréal H3C 3P8

P. Sawyer
Department of Mathematics and Computer Science
Laurentian University
Sudbury P3E 2C6

October, 2000

Les Cahiers du GERAD

G–2000–54

Abstract

Since Hinkley's original work on exact inference for a change in a sequence of random variables, many authors have proposed different methods based either on exact distributions or asymptotic approximations to test for and estimate a change point in such a sequence. Here we concentrate on the Poisson case. We propose estimators, tests and confidence intervals obtained by adapting and modifying Hinkley's and Worsley's work and we do an extensive study of the small sample properties of these proposed methods of inference, also comparing them to methods based on asymptotic approximations. The methods are applied to a neural train data set.

Résumé

Depuis la parution des travaux de Hinkley sur l'inférence exacte pour un point de rupture dans une suite de variables aléatoires, plusieurs auteurs ont proposé différentes méthodes basées soit sur les distributions exactes soit sur des approximations asymptotiques, pour tester et estimer un point de rupture. Nous nous concentrons ici sur le cas des suites de Poisson. Nous proposons des estimateurs, des tests et des intervalles de confiance obtenus en adaptant et en modifiant les résultats de Hinkley et Worsley. À l'aide de simulations intensives, nous étudions leurs propriétés dans le cas d'échantillons de petite taille. Ces méthodes sont aussi comparées aux méthodes basées sur des approximations asymptotiques. Puis, elles sont appliquées à un échantillon de données de transmission de neurones.

Acknowledgments: This work was supported by grants from NSERC of Canada.

Introduction

The change point problem has been extensively discussed in the literature particularly in the case of normal and binomial random variables. In many problems of recent interest in medical research such as in neurophysiological studies of neural spike train data (c.f. Commenges et al (1985, 1986), Brillinger (1992), Joseph and Wolfson (1992) and Bélisle et al (1998)), a model based on a point process such as the Poisson process would be more appropriate (c.f. Kendall and Kendall (1980), Deshayes (1984), Akman and Raftery (1986) and MacNeil and Mao (1994)). The Poisson model is a prototype for a discrete data model and deserves consideration in its own right. Aldous (1989) even argues that Poisson approximations arise in more situations than Gaussian ones. The fundamental work in obtaining asymptotic distributions for a change point in normal and binomial sequences was accomplished by Hinkley (1970) and Hinkley and Hinkley (1970) using random walk results. Worsley (1983) found a simple iterative procedure to calculate the exact null and alternative distributions of the likelihood ratio, and cumulative sum (CUSUM) and related statistics for testing in the binomial problem. These theoretical results were later extended by Worsley (1986) to the general case of exponential family random variables, in particular, the exponential distribution. Gombay and Horvath (1990, 1994, 1996 a, 1996 b) in a series of papers studied asymptotic approximations for maximum likelihood in change point models. Csörgõ and Horvath (1987 a, 1987 b, 1988), Haccou et al (1988), Carlstein (1988), Carlstein and Lele (1994), Ferger (1994, 1995) and many other authors, too numerous to mention, have contributed to the study of non-parametric tests for change point problems and their asymptotic behaviour.

In this paper, we are particularly interested in exact distributions of test statistics and their inversion in order to produce confidence intervals rather than asymptotic approximations to such statistics. An excellent review of confidence intervals for change point problems can be found in Siegmund (1988).

In Section 1, we adapt the results of Hinkley and Hinkley (1970) to the Poisson case. Section 2 gives recursive formulae for the exact distributions of the likelihood ratio test statistic following procedures of Worsley (1983, 1986). Exact confidence intervals for the change point obtained by inverting the likelihood ratio statistic are based on Worsley's method (1986) for exponential family random variables.

We also study a statistic that can be viewed as a modified CUSUM statistic or more usefully as a discrete version of the pontogram introduced by Kendall and Kendall (1980). Confidence intervals based on the exact distribution of the discrete pontogram are also developed here.

In Section 2.2, we consider a suitably curtailed version of the modified CUSUM statistics and by relating it to the pontogram proposed by Akman and Raftery (1986) and Szyszkowicz (1992 a, b) in the study of a change point Poisson process in continuous time, we obtain the asymptotic distribution of the discrete version.

The asymptotic behaviour of suitably normalized pontograms has been studied by Akman and Raftery (1986) and by Szyszkowicz (1992 a, b) and these results will be used to describe the asymptotic behaviour of the test statistics based on the discrete version in Section 3.

The two different methods based on the exact distributions of the likelihood ratio test statistics and the discrete pontogram and on the asymptotic approximations of change point test statistics obtained by Gombay and Horvath (1996 b) are compared in an intensive simulation study in Section 3.

In Section 4, the methods were applied to a real data set of neuronal discharge data originally studied by Commenges et al (1985, 1986) and reanalyzed using different methods by Joseph and Wolfson (1992) and Bélisle et al (1998). Concluding remarks can be found in Section 5.

1 The likelihood function and random walk distributions

Here, we adapt the results of Hinkley and Hinkley (1970) to find the asymptotic distribution of the maximum likelihood estimate of the change point in a Poisson sequence as well as the asymptotic distribution of the likelihood ratio statistic for testing hypotheses about the change point. Some results related to the speed of convergence of the series involved in the computation of these distributions are also discussed. In order to model a change at $\tau < n$ in a sequence of independent Poisson random variables X_1, X_2, ..., X_n, let us assume that

$$P(X_i = k) = \begin{cases} \lambda_0^k \, e^{-\lambda_0}/k! & i = 1, \ldots, \tau \\ \lambda_1^k \, e^{-\lambda_1}/k! & i = \tau+1, \ldots, n \end{cases}$$

where λ_0 and λ_1 are known but τ is unknown. If λ_0 and λ_1 are also unknown, the results stated in this paper are unaffected except for stochastically negligible terms if we use the consistent maximum likelihood estimates for λ_0 and λ_1. Henceforth, we suppose that $\lambda_0 > \lambda_1$.

Maximizing the log likelihood function for a change point at $\tau = t$,

$$L(t) \quad = \quad \log \left(\prod_{i=1}^{t} \frac{\lambda_0^{X_i}}{X_i!} \, e^{-\lambda_0} \prod_{i=t+1}^{n} \frac{\lambda_1^{X_i}}{X_i!} \, e^{-\lambda_1} \right),$$

to obtain the maximum likelihood estimator $\hat{\tau}$ is equivalent to maximizing Y_t where

$$Y_t = \sum_{i=1}^{t} (X_i - \delta), \tag{1}$$

where $\delta = \frac{\lambda_0 - \lambda_1}{\log(\lambda_0/\lambda_1)}$ and $t = 1, \ldots, n-1$.

Adapting Hinkley and Hinkley (1970) to the Poisson problem, let

$$p_{0,0} \quad = \quad \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \sum_{k=0}^{[n\,\delta]} \frac{n^k \, \lambda_0^k \, e^{-n\,\lambda_0}}{k!} \right), \tag{2}$$

$$p_{\ell,m} \quad = \quad \begin{cases} q_{\ell,m} & \text{if } -\ell + \delta\, m \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

where

$$q_{\ell,m} = \begin{cases} \sum_{\ell-\delta\,(m-1) \leq k \leq \ell} \frac{\lambda_0^k}{k!} \, e^{-\lambda_0} \, q_{\ell-k,m-1} & \text{if } \ell \geq 1, \; m \geq 2, \\ \frac{\lambda_0^\ell}{\ell!} \, e^{-\lambda_0} \, p_{0,0} & \text{if } m = 1, \\ e^{-m\,\lambda_0} \, p_{0,0} & \text{if } \ell = 0. \end{cases} \tag{3}$$

Let us also define

$$
\begin{aligned}
p'_{0,0} &= \exp\left(-\sum_{n=1}^{\infty}\frac{1}{n}\sum_{k=[n\,\delta]+1}^{\infty}\frac{n^k\lambda_1^k\,e^{-n\,\lambda_1}}{k!}\right), \qquad (4)\\
p'_{\ell,m} &= \begin{cases} q'_{\ell,m} & \text{if } \ell-\delta\,m\geq 0,\\ 0 & \text{otherwise}\end{cases}
\end{aligned}
$$

where

$$
q'_{\ell,m} = \begin{cases} \sum_{0\leq k\leq \ell-\delta\,(m-1)}\frac{\lambda_1^k}{k!}\,e^{-\lambda_1}\,q'_{\ell-k,m-1} & \text{if } \ell\geq 1,\ m\geq 2,\\ \frac{\lambda_1^\ell}{\ell!}\,e^{-\lambda_1}\,p'_{0,0} & \text{if } m=1,\\ 0 & \text{otherwise.}\end{cases} \qquad (5)
$$

Theorem 1 *If we assume that both τ and $n-\tau$ are large, the distribution of $\hat{\tau}$ is given as follows:*

$$
\begin{aligned}
P(\hat{\tau}=\tau-n)=\pi_{-n} &= \sum_{\ell=0}^{[\delta\,n]} p_{\ell,n}\left\{p'_{00}+\sum_{j=1}^{\infty}\sum_{k=\left[\frac{j+\ell-\delta\,n}{\delta}\right]+1}^{\left[\frac{j}{\delta}\right]} p'_{j,k}\right\} \quad \textit{if } n>0,\\
P(\hat{\tau}=\tau)=\pi_0 &= p_{0,0}\,p'_{0,0},\\
P(\hat{\tau}=\tau+n)=\pi_n &= \sum_{\ell=[n\,\delta]+1}^{\infty} p'_{\ell,n}\left\{p_{00}+\sum_{j=0}^{\infty}\sum_{k=\left[\frac{j}{\delta}\right]+1}^{\left[\frac{j+\ell-\delta\,n}{\delta}\right]} p_{j,k}\right\} \quad \textit{if } n>0.
\end{aligned}
$$

This result, whose proof is relegated to Appendix A, makes it possible to compute the asymptotic bias and variance of $\hat{\tau}$ as well as the percentiles t_α and $t_{1-\alpha}$ for $\alpha = 0.01$, 0.05 and 0.10. Here, t_α is the largest integer such that $P(\hat{\tau}-\tau\leq t_\alpha)\leq\alpha$ when $\alpha = 0.01$, 0.05 and 0.10 while t_α is the largest integer such that $P(\hat{\tau}-\tau<t_\alpha)<\alpha$ when $\alpha = 0.90$, 0.95 and 0.99.

For selected values of λ_0 and λ_1, Table 1 gives the probability π_0, percentiles t_α and $t_{1-\alpha}$ (α = 0.05 and 0.10) and the asymptotic bias and standard deviation of $\hat{\tau}$. Note that the standard deviation is steadily decreasing when λ_0/λ_1 increases while the bias oscillates around and eventually converges to 0.

Table 1 allows us to calculate the length of the confidence intervals easily. The improvement in the mean, the standard deviation and the confidence intervals as λ_0 and λ_1 diverge should be noted.

For the testing problem, we consider three likelihood ratio tests of significance for τ, namely the two one-sided tests of $H_0\colon\ \tau=\tau_0$ against the alternatives $H_1\colon\ \tau<\tau_0$ and $H'_1\colon\ \tau>\tau_0$ and the two-sided test with alternative $H_2\colon\ \tau\neq\tau_0$. From (1), the likelihood ratio test statistics are respectively

Table 1: Percentile of $(\hat{\tau} - \tau)$, mean of $\tau - \hat{\tau}$ and standard deviation

λ_0	λ_1	$P(\hat{\tau} = \tau)$	$t_{0.01}$	$t_{0.05}$	$t_{0.10}$	$t_{0.90}$	$t_{0.95}$	$t_{0.99}$	$E(\tau - \hat{\tau})$	$\sigma(\hat{\tau})$
1.5	1.0	0.078000	−82	−41	−26	21	36	76	−1.434593	25.186851
2.0	1.0	0.214211	−25	−13	−8	6	10	22	−0.516056	7.382185
3.0	1.0	0.464886	−9	−5	−3	2	3	7	−0.117985	2.326832
4.0	1.0	0.643166	−5	−3	−2	1	1	4	−0.204344	1.211694
6.0	1.0	0.847471	−3	−2	−1	0	1	2	0.005682	0.528833

$$\begin{aligned}
\Lambda_1 &= \max_{t \le \tau_0} (Y_t - Y_{\tau_0}), \\
\Lambda_1' &= \max_{t \ge \tau_0} (Y_t - Y_{\tau_0}), \\
\Lambda_2 &= \max_{t} (Y_t - Y_{\tau_0}) = \max\{\Lambda_1, \Lambda_1'\}.
\end{aligned}$$

We let $p_{\ell,m}(\nu) = \begin{cases} p_{\ell,m} & \text{if } \nu \le 0, \\ q_{\ell,m}(\nu) & \text{if } -\ell + \delta\, m \ge 0, \\ 0 & \text{otherwise} \end{cases}$ and $p'_{\ell,m}(\nu) = \begin{cases} p'_{\ell,m} & \text{if } \nu \le 0, \\ q'_{\ell,m}(\nu) & \text{if } \ell - \delta\, m \ge 0, \\ 0 & \text{otherwise} \end{cases}$
where the terms $q_{\ell,m}(\nu)$ and $q'_{\ell,m}(\nu)$ are defined recursively below. Suppose $\nu \ge 1$. Let

$$\begin{aligned}
p_{0,0}(\nu) &= \sum_{\max\{\delta, -\ell + (m+1)\,\delta\} \le k} \frac{\lambda_1^k}{k!} e^{-\lambda_1} q_{\ell,m}(\nu - 1), \\
p'_{0,0}(\nu) &= \sum_{k \le \delta} \frac{\lambda_0^k}{k!} e^{-\lambda_0} \sum_{m \ge \frac{k+\ell}{\delta} - 1} q'_{\ell,m}(\nu - 1)
\end{aligned}$$

where

$$q_{\ell,m}(\nu) = \begin{cases} \sum_{\ell - \delta\,(m-1) \le k \le \ell} \frac{\lambda_1^k}{k!} e^{-\lambda_1} q_{\ell-k,m-1}(\nu - 1) & \text{if } \ell \ge 1,\ m \ge 2, \\ e^{-m\,\lambda_1} p_{0,0}(\nu - m) & \text{if } \ell = 0,\ m \le \nu, \\ e^{-\nu\,\lambda_1} e^{-(m-\nu)\,\lambda_0} p_{0,0} & \text{if } \ell = 0,\ m > \nu, \\ \frac{\lambda_1^\ell}{\ell!} e^{-\lambda_1} p_{0,0}(\nu - 1) & \text{if } m = 1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$q'_{\ell,m}(\nu) = \begin{cases} \sum_{0 \le k \le \ell - \delta\,(m-1)} \frac{\lambda_0^k}{k!} e^{-\lambda_0} q'_{\ell-k,m-1}(\nu - 1) & \text{if } \ell \ge 1,\ m \ge 2, \\ \frac{\lambda_0^\ell}{\ell!} e^{-\lambda_0} p'_{0,0}(\nu - 1) & \text{if } m = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2 *If we assume that both τ and $n-\tau$ are large, then for any integer ν,*

$$\begin{aligned}
P_1(x;\nu) = P(\Lambda_1 \le x | \tau = \tau_0 - \nu) &= \sum_{J(x)} p_{\ell,m}(\nu), \\
P_1'(x;\nu) = P(\Lambda_1' \le x | \tau = \tau_0 + \nu) &= \sum_{J'(x)} p'_{\ell,m}(\nu), \\
P_2(x;\nu) = P(\Lambda_2 \le x | \tau = \tau_0 + \nu) &= P_1(x;-\nu)\, P_1'(x;\nu)
\end{aligned}$$

where $J(x) = \{(l,m):\ -\ell + \delta\, m \le x\}$ and $J'(x) = \{(l,m):\ l - \delta\, m \le x\}$.

See the proof in Appendix A.

To compute the values p_{jk} and p'_{jk} (and those indexed by ν), we chose a value K for ∞ which could be adjusted if judged necessary (that is, the sums had to be at least 0.999999). We then computed these values for $0 \le j \le \delta\, K$ and $0 \le k \le K$. All our programs are in the C programming language using long double precision.

For selected values of λ_0 and λ_1, Table 2 gives the percentiles l_α ($\alpha = 0.01$ and 0.05) relative to the statistic Λ_2. The critical value l_α is defined to be the smallest x (to three decimal places) such that $P(\Lambda_2 \le x) \ge 1-\alpha$. Since we are dealing with a discrete process, we give the following additional information $Q_\alpha^- = P(\Lambda_2 > l_\alpha - 0.001)$ and $Q_\alpha = P(\Lambda_2 > l_\alpha)$. Naturally, we have $Q_\alpha \le \alpha \le Q_\alpha^-$. We notice more marked differences between Q_α^- and Q_α when λ_0 and λ_1 are further apart.

Table 2: Percentiles of Λ_2, $Q_\alpha^- = P(\Lambda_2 > l_\alpha - 0.001)$ and $Q_\alpha = P(\Lambda_2 > l_\alpha)$

		5 % test			1 % test		
λ_0	λ_1	$l_{0.05}$	$Q_{0.05}^-$	$Q_{0.05}$	$l_{0.01}$	$Q_{0.01}^-$	$Q_{0.01}$
1.5	1.0	8.430000	0.050258	0.049964	12.425000	0.010034	0.009991
2.0	1.0	4.657000	0.050655	0.048850	6.969000	0.010108	0.009965
3.0	1.0	2.564000	0.050498	0.049777	4.103000	0.010479	0.009529
4.0	1.0	1.836000	0.058049	0.045023	2.836000	0.012145	0.009465
6.0	1.0	0.791000	0.085273	0.044789	1.791000	0.023508	0.009777

In Table 3, some results related to the power of Λ_2 are given. More specifically, $Q_\alpha(\nu) = P(\Lambda_2 \le \ell_\alpha | \tau = \tau_0 + \nu)$ where l_α is as above. The values for negative ν's are below the values for the positive ν's. The power is biased because of the asymmetry of the model. We are led to believe that tests based on $\hat{\tau}$ alone are not very powerful and other approaches are necessary.

It is worthwhile to state the following result on the computation of the series (2) and (4).

Proposition 1 *Let $h(t) = -t + \log(1+t)$ for $t > -1$. The terms in the series (2) and (4) are $O\left(e^{-\alpha\, n}/\sqrt{n}\right)$ with $\alpha \sim -\delta\, h(\lambda_0/\delta - 1) = -\delta\, h(\lambda_1/\delta - 1) > 0$.*

See the proof in Appendix A.

Table 3: Power of $Q_{0.05}(\nu)$ and $Q_{0.01}(\nu)$ for Λ_2

		5 % test			1 % test		
λ_0	λ_1	$\nu = 0$	$\nu = \pm 1$	$\nu = \pm 2$	$\nu = 0$	$\nu = \pm 1$	$\nu = \pm 2$
1.5	1.0	0.049964	0.056525	0.064938	0.009991	0.011314	0.013014
			0.054720	0.060338		0.010957	0.012099
2.0	1.0	0.048850	0.084897	0.153662	0.009965	0.017497	0.036326
			0.065882	0.094270		0.013629	0.019715
3.0	1.0	0.049777	0.256744	0.480929	0.009529	0.113283	0.298725
			0.142720	0.464615		0.025121	0.083311
4.0	1.0	0.045023	0.424336	0.729464	0.009465	0.248644	0.581716
			0.425242	0.732222		0.061172	0.448713
6.0	1.0	0.044789	0.854591	0.967043	0.009777	0.719981	0.925181
			0.756038	0.958973		0.401947	0.871678

To understand the significance of this result, we must first observe that $h(t) < 0$ for $t \neq 0$ and that $h(0) = 0$. Equally important is the fact that $\lambda_1 < \delta < \lambda_0$. The plot of $-\delta\, h(\lambda_1/\delta - 1) = -\delta\, h(\lambda_0/\delta - 1)$ in terms of λ_0 (with $\lambda_1 = 1.0$) is given in Figure 1. For instance, when $\lambda_0 = 1.5$ and $\lambda_1 = 1.0$, both series converge like $\sum_{n=1}^{\infty} e^{-0.025\, n}/\sqrt{n}$ while they converge like $\sum_{n=1}^{\infty} e^{-1.073\, n}/\sqrt{n}$ when $\lambda_0 = 6.0$ and $\lambda_1 = 1.0$. In the former case, to obtain an error less than 10^{-6}, we require terms with $n > 500$ while in the later case, we can stop after $n = 12$.

Naturally, when λ_0 tends to 1.0, $-\delta\, h(\lambda_0/\delta - 1)$ tends to 0, namely the series does not converge.

2 Exact distributions of statistics for a change point

The small sample behaviour of statistics plays an important role in applications. In this section, exact distributions for likelihood ratio and a cumulative sum statistic and its modification are derived. We first introduce these statistics, give their exact null distributions conditional on a fixed number of occurrences using the same procedures as those of Worsley (1983) in the binomial case. Then we find the exact alternative distributions. In Section 2.1, following Worsley (1986), confidence regions about the change point can be constructed based on the "inverted" likelihood ratio statistic. As suggested by Bickel and Doksum (1977), this method of test "inversion" can be used for other statistics and here we use the CUSUM statistic and its modification.

We wish to test $H_0 : \lambda_i = \lambda,\ i = 1, \ \ldots,\ n$ against $H_1 : \begin{cases} \lambda_i &= \lambda_0 & i = 1, \ \ldots, \tau \\ \lambda_i &= \lambda_1 & i = \tau + 1, \ \ldots, n \end{cases}$.

Let $S_k = \sum_{i=1}^{k} X_i$, $S = \sum_{i=1}^{n} X_i$, $S'_k = \sum_{i=k+1}^{n} X_i = S - S_k$ and $\sigma^2 = S/n$.

When λ_0 and λ_1 are unknown and are replaced by S_k/k and $S'_k/(n-k)$, minus the log likelihood ratio for fixed k is

$$L_k \;=\; S_k \log(S_k/k) + S'_k \log(S'_k/(n-k)) - S \log(S/n).$$

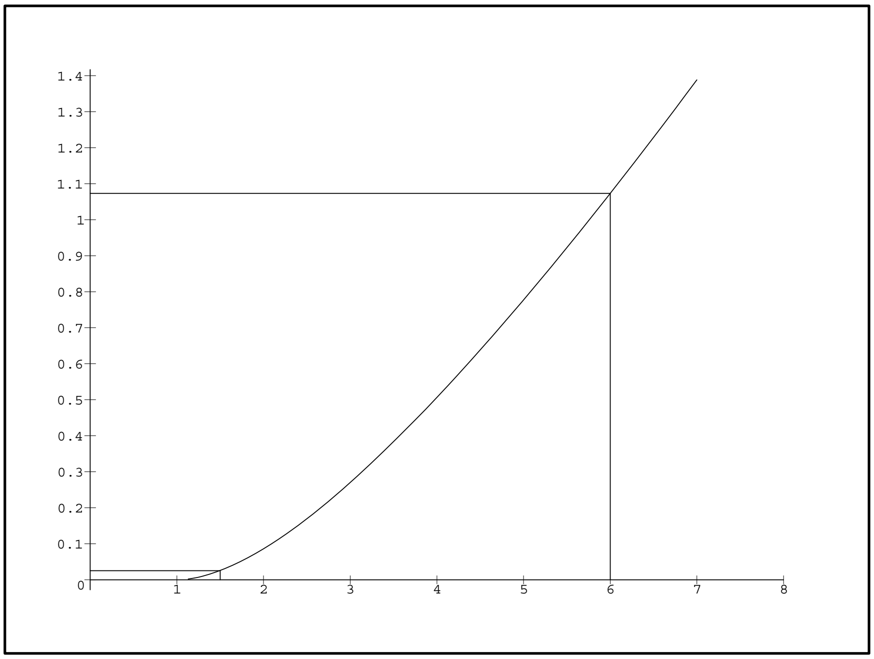

Figure 1: Speed of convergence (the 2 axes are on different scales)

The maximum likelihood estimate of τ is the value of k that maximizes L_k. It is the same value that maximizes expression (1) of Section 1 when λ_0 and λ_1 are replaced by their estimates. The corresponding statistic for testing H_0 against H_1 is $L = \max_k L_k$.

Following Worsley (1983), the cumulative sum (CUSUM) statistic used to test whether a change occurs is given by the maximum of the absolute values of $Q_k = \frac{S_k - \frac{k}{n} S}{\sqrt{n\,\sigma^2}}$. The corresponding statistic for testing H_0 against H_1 is $Q = \max_k |Q_k|$.

In order to estimate the position of the change, we propose to study a modified CUSUM statistic: for $k < n$,

$$Z_k = \frac{S_k - \frac{k}{n} S}{\sqrt{S \, \frac{k}{n} \, (1 - \frac{k}{n})}}$$

while $Z_n = 0$. This is a discrete version of the pontogram introduced by Kendall and Kendall (1980) and henceforth it shall be called a discrete pontogram. Now, $\hat{\tau}$ is the smallest value of k that maximizes $|Z_k|$ and $Z = \max_k |Z_k|$ is the corresponding statistic for testing H_0 against H_1.

Here, continuing with our study of the exact distribution of Z_k and L_k, we use an iterative procedure for calculating $P(Z \leq t)$, and $P(L \leq t)$ under H_0 conditional on a fixed number of occurrences S as suggested by Worsley (1986). We will start with Z.

Let us define, for $k = 1, \ldots, n$,

$$F_{k,n}(t; S_k, S) = P(|Z_i| \le t\colon\ i \le k | S_k, S). \tag{6}$$

We wish to compute $P(Z \le t|S) = F_{n,n}(t; S, S)$. The next result gives the exact distribution of $(Z \le t|S)$ under the null hypothesis H_0.

Theorem 3

Under the hypothesis H_0, $F_{1,n}(t; S_1, S) = \begin{cases} 1 & \text{if } |Z_1| \le t \\ 0 & \text{if } |Z_1| > t \end{cases}$ and, for $k \ge 1$,

$$F_{k+1,n}(t; S_{k+1}, S) = \begin{cases} \displaystyle\sum_{m=0}^{S_{k+1}} F_{k,n}(t; m, S) \binom{S_{k+1}}{m} \left(\frac{k}{k+1}\right)^m \left(\frac{1}{k+1}\right)^{S_{k+1}-m} & \text{if } |Z_{k+1}| \le t \\ 0 & \text{if } |Z_{k+1}| > t \end{cases}$$

for $k = 1, \ldots, n-1$.

Proof: Following Worsley (1986), from (6),

$$F_{k+1,n}(t; S_{k+1}, S) = \sum_{m=0}^{S_{k+1}} P(|Z_i| \le t\colon\ i \le k+1 | S_k = m,\ S_{k+1}, S)\, P(S_k = m | S_{k+1},\ S).$$

If we know S_{k+1}, the knowledge of S brings no new information on S_k since $S = S_{k+1} + \sum_{i>k+1} X_i$ and the sum $\sum_{i>k+1} X_i$ is independent of S_k and S_{k+1}. It is a simple exercise to show that $P(S_k = m|S_{k+1}) = \binom{S_{k+1}}{m} (k/(k+1))^m\, (1/(k+1))^{S_{k+1}-m}$. Moreover,

$$P(|Z_i| \le t\colon\ i \le k+1 | S_k = m,\ S_{k+1}, S) = P(|Z_i| \le t\colon\ i \le k | S_k = m, S)$$

if $|Z_{k+1}| \le t$ and 0 otherwise. We use the fact that $S - S_{k+1}$, $S_{k+1} - S_k$ and S_k are independent to obtain

$$\begin{aligned}
P(|Z_i| \le t, i \le k | S_k = a, S_{k+1} = b, S = c) &= \frac{P(|Z_i| \le t, i \le k, S_k = a, S_{k+1} = b, S = c)}{P(S_k = a, S_{k+1} = b, S = c)} \\
&= \frac{P(|Z_i| \le t, i \le k, S_k = a)\, P(S_{k+1} - S_k = b - a)\, P(S - S_{k+1} = c - b)}{P(S_k = a)\, P(S_{k+1} - S_k = b - a)\, P(S - S_{k+1} = c - b)} \\
&= \frac{P(|Z_i| \le t, i \le k, S_k = a)}{P(S_k = a)} = \frac{P(|Z_i| \le t, i \le k, S_k = a)\, P(S - S_k = c - a)}{P(S_k = a)\, P(S - S_k = c - a)} \\
&= \frac{P(|Z_i| \le t, i \le k, S_k = a, S = c)}{P(S_k = a, S = c)} = P(|Z_i| \le t, i \le k | S_k = a, S = c).
\end{aligned}$$

∎

The computation of the values $F_{k,n}$ is quite demanding numerically. Refer to Remark 1 in Appendix A where this concern is addressed.

To compute the null distribution of L and of Z it suffices to realize that the proofs of Theorem 3 remain essentially the same if we replace the statistic Z_k by another statistic which depends only on S_k and S.

We now give the distribution of $(Z \leq t|S)$ under the alternative hypothesis $H_0^{(\tau)}$ that the change point occurs at τ.

Theorem 4

Under the hypothesis $H_0^{(\tau)}$,

$$\begin{aligned} P(Z \leq t|S) &= \sum_{m=0}^{S} F_{\tau,n}(t;m,S)\, F_{n-\tau,n}(t;S-m,S) \\ &\qquad \cdot \binom{S}{m} \left(\frac{\tau\,\lambda_0}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^m \left(\frac{(n-\tau)\,\lambda_1}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^{S-m}. \end{aligned}$$

Proof: Note first that under $H_0^{(\tau)}$,

$$\begin{aligned} P(S_\tau = m|S) &= P(S_\tau = m|S_\tau + S'_\tau = S) \\ &= \binom{S}{m} \left(\frac{\tau\,\lambda_0}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^m \left(\frac{(n-\tau)\,\lambda_1}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^{S-m}. \end{aligned}$$

$$\begin{aligned} P(Z \leq t|S) &= \sum_{m=0}^{S} P(|Z_i| \leq t; i \leq n|S_\tau = m, S)\, P(S_\tau = m|S) \\ &= \sum_{m=0}^{S} P(|Z_i| \leq t; i \leq \tau|S_\tau = m, S)\, P(|Z_i| \leq t; \tau+1 \leq i \leq n|S_\tau = m, S) \\ &\qquad \cdot \binom{S}{m} \left(\frac{\tau\,\lambda_0}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^m \left(\frac{(n-\tau)\,\lambda_1}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^{S-m} \\ &= \sum_{m=0}^{S} F_{\tau,n}(t;m,S)\, F_{n-\tau,n}(t;S-m,S) \\ &\qquad \cdot \binom{S}{m} \left(\frac{\tau\,\lambda_0}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^m \left(\frac{(n-\tau)\,\lambda_1}{\tau\,\lambda_0+(n-\tau)\,\lambda_1}\right)^{S-m}. \end{aligned}$$

■

2.1 Confidence regions for the change point by inversion of tests

As indicated by Bickel and Doksum (1977), there is a duality between tests and confidence intervals and tests can be "inverted" in order to form confidence intervals. We follow this procedure for the likelihood ratio test and the statistic Z based on the discrete pontogram as Worsley (1983) did in the binomial case. We will do the computations for $Z = \max_k |Z_k|$ but the same reasoning applies to $L = \max_k |L_k|$. We now derive the confidence region C_α which contains the values k that cannot be rejected as the change point using the statistic Z and an appropriate α level test.

Under the hypothesis that the change point occurs at k, using Theorem 3, we have

$$\begin{aligned} P(Z \leq t|S_k, S) &= P\left(\max_{1\leq i\leq k} |Z_i| \leq t|S_k, S\right) P\left(\max_{k+1\leq i\leq n} |Z_i| \leq t|S - S_k, S\right) \\ &= F_{k,n}(t; S_k, S)\, F_{n-k,n}(t; S - S_k, S). \end{aligned}$$

The confidence region C_α can therefore be given as

$$C_\alpha = \{k\colon\ F_{k,n}(Z; S_k, S)\, F_{n-k,n}(Z; S - S_k, S) \leq 1 - \alpha\}.$$

If C_α consists of disjoint clusters of consecutive integers, this may indicate the existence of several change points.

2.2 The asymptotic behaviour of the discrete pontogram

Akman and Raftery (1986) used a continuous version of the pontogram, originally introduced by Kendall and Kendall (1980), to detect the change point. They showed that under certain conditions, the estimator for τ given by their pontogram is consistent. We will show that, under the same conditions, our version of the estimator has the same properties. The results are stated more precisely in Theorem 5.

Let $\theta = \tau/n$ and assume that we know a, b such that $0 < a \leq \theta \leq b < 1$. For $c \leq s \leq d$, let

$$R(s; c, d) = \sqrt{s'\,(1 - s')}\,\left[\frac{\Pi(s) - \Pi(c)}{s'} - \frac{\Pi(d) - \Pi(s)}{1 - s'}\right] \tag{7}$$

where $0 \leq c < d < 1$, $s' = (s - c)/(d - c)$, $\Pi(s) = N(s\,n)$ and $N(t)$ is the number of observations occurring before time t. We are interested in the statistic $\sup_{a\leq t\leq b} R(t; 0, 1)$ and the estimator for θ is the smallest value of t for which this maximum occurs. The corresponding estimator for $\hat{\tau}$ is then $n\,\hat{\theta}$.

Let us consider (7) with $c = d = 0$ and $s = k/n$, k a positive integer. We find

$$R(k/n; 0, 1) = \sqrt{(k/n)\,(1 - (k/n))}\,\left[\frac{S_k}{k/n} - \frac{S - S_k}{1 - k/n}\right] = Z_k.$$

In other words, our statistic is a "discrete" version of the statistic used by Akman and Raftery (1986) in their study of a continuous time Poisson process.

Like Akman and Raftery (1986), we could use instead the statistic

$$\tilde{Z} = \max_{a\,n\leq k\leq b\,n} |Z_k|$$

and define an estimator for τ as the smallest value of k where this maximum occurs.

Theorem 1 of Akman and Raftery (1986) can then be adapted to our situation. In order to show that this is valid, we need to take care of two differences: we use a discrete variable and we use absolute values. For the first difference : on the interval $[\frac{k-1}{n}, \frac{k}{n})$, $R(t;0,1) = \frac{S_{k-1} - t\,S}{\sqrt{t\,(1-t)}}$ is decreasing so its maximum is $R((k-1)/n;0,1) = Z_{k-1}$. Therefore,

$$\max_{a \le t \le b} R(t;0,1) = \max_{a\,n \le k \le b\,n} Z_k.$$

Consequently, the estimators corresponding to these statistics have to be the same.

The matter of the absolute value has a resolution which requires a bit more work but is not too difficult. It suffices to show that whenever $a\,n \le k \le b\,n$, $P(Z_k < 0) = P(S_k - \frac{k}{n}\,S < 0) = O(\alpha_n)$ where $\lim_{n\to\infty}(n\,\alpha_n) = 0$ (it is not enough to show that $\lim_{n\to\infty} P(S_k - \frac{k}{n}\,S < 0) = 0$ for a specific value of k). The statement of the result then becomes:

Theorem 5 *Let $\hat{\tau}$ be the smallest value of k where the maximum of $\tilde{Z} = \max_{a\,n \le k \le b\,n} |Z_k|$ occurs. Suppose that $n \to \infty$ and $\tau \to \infty$ while a, b and $\theta = \tau/n$ remains constant $(0 < a \le \theta \le b < 1)$. Then*

1. *The estimators $\hat{\theta} = \hat{\tau}/n$, $\hat{\lambda}_0 = S_{\hat{\tau}}/\hat{\tau}$ and $\hat{\lambda}_1 = (S - S_{\hat{\tau}})/(n - \hat{\tau})$ are consistent.*
2. $$\frac{\lambda_1\,\hat{\tau} - S_{\hat{\tau}}}{\sqrt{S_{\hat{\tau}}}} \to_d N(0,1),$$
$$\frac{\lambda_1\,(n-\hat{\tau}) - (S - S_{\hat{\tau}})}{\sqrt{S - S_{\hat{\tau}}}} \to_d N(0,1).$$

3 Comparison of the methods via simulations

To supplement these analytical results and compare the different methods, we conducted a series of Monte Carlo experiments. There are several interesting issues that can be addressed in such experiments.

Using the results of Section 2, we can compute the critical values for the statistics L and Z under the null hypothesis that there is no change point. We define the critical value ℓ_α as the largest (correct to three decimal places) x such that $P(L \le x|S) \le 1 - \alpha$ (similarly for the statistics Z).

One goal of this section is to compare the "exact" critical values for $L = \max_k |L_k|$ with the estimates given by Gombay and Horvath (1996 b) in Tables I to VI of their paper for different values of n and for $\lambda = 1$, 5 and 10 (in their paper, L is actually denoted by Z!).

Table 4 gives the results of the simulations (each of them repeated 5000 times). The values r^* are computed using Theorem 2.1 of Gombay and Horvath (1996 b) and correspond to the "Asymptotic critical values" of Table I to VI of that paper. The values $r = r(h(n), l(n))$ correspond to their asymptotic approximation given by Theorem 3.1 together with formula (3.8) in their paper.

We produced Poisson simulations for $\lambda = 1$, 5 and 10 respectively. The numbers in parenthesis under r^*, r and l_α correspond to the relative frequency the observed maximum was below these values in our simulations (for $\lambda = 1.0$, $\lambda = 5.0$ and $\lambda = 10.0$ respectively).

Table 4: Critical values for the maximum likelihood estimator

Sample size	$1-\alpha$	r^*	r	l_α $\lambda=1$	l_α $\lambda=5$	l_α $\lambda=10$
10	0.90	3.057 (0.982,0.974,0.987)	2.364 (0.890,0.895,0.900)	2.494 (0.890)	2.351 (0.880)	2.413 (0.907)
	0.95	3.615 (0.997,0.997,0.998)	2.663 (0.922,0.941,0.945)	2.670 (0.922)	2.665 (0.941)	2.703 (0.945)
	0.99	4.877 (1.000,1.000,1.000)	3.247 (0.993,0.990,0.993)	3.196 (0.982)	3.245 (0.978)	3.223 (0.990)
20	0.90	3.113 (0.983,0.968,0.975)	2.523 (0.875,0.897,0.977)	2.549 (0.875)	2.599 (0.899)	2.585 (0.894)
	0.95	3.599 (0.996,0.997,0.996)	2.814 (0.945,0.943,0.947)	2.855 (0.951)	2.844 (0.951)	2.833 (0.949)
	0.99	4.700 (1.000,1.000,1.000)	3.383 (0.990,0.993,0.990)	3.392 (0.989)	3.292 (0.991)	3.356 (0.989)
100	0.90	3.226 (0.973,0.969,0.967)	2.786 (0.875,0.888,0.886)	2.857 (0.894)	2.838 (0.904)	2.809 (0.892)
	0.95	3.637 (0.990,0.992,0.991)	3.056 (0.950,0.938,0.940)	3.084 (0.953)	3.130 (0.949)	3.079 (0.945)
	0.99	4.570 (1.000,1.000,1.000)	3.589 (0.990,0.992,0.989)	3.548 (0.989)	3.581 (0.991)	3.590 (0.989)

The work of Szyszkowicz (1992 a, 1992 b) ensures that we can make the same comparisons using $Z = \max_k |Z_k|$. The results appear in Table 5.

Table 4 clearly demonstrates that the critical values r^* computed using Theorem 2.1 of Gombay and Horvath (1996) are too conservative and tend to overestimate the "true" critical value. The asymptotic approximation given by Gombay and Horvath (1996) is an improvement since the significance levels with this statistic are extremely close to the nominal values. Also our test statistic using the exact inference outlined here was extremely close in value to r for $\lambda = 1$, 5 and 10. Table 5 illustrates the same phenomenon when the pontogram is used instead of the maximum likelihood estimator is used; that is, the asymptotic approximation r should be used rather than r^* and that the results using r or exact inference methods are extremely close.

A second goal is to compare the confidence intervals obtained by the method of test inversion outlined in Section2. We are also particularly interested in the following questions.

1. Is the performance of the tests or confidence intervals sensitive to the position of the change point (whether it is at either end of the sequence or near the middle) ?
2. What are the properties of the estimators introduced for small samples ?

To deal with these issues, 1000 replications of a Poisson process were generated first for $\lambda_0/\lambda_1 < 3$ for each of $n = 20$, 50, 100, 200 and when the change occurs at the beginning of the

Table 5: Critical values for the pontogram

Sample size	$1-\alpha$	r^*	r	l_α $\lambda=1$	l_α $\lambda=5$	l_α $\lambda=10$
10	0.90	3.057 (0.964,0.980,0.989)	2.364 (0.893,0.904,0.905)	2.371 (0.893)	2.357 (0.873)	2.333 (0.884)
	0.95	3.615 (0.995,0.995,0.998)	2.663 (0.954,0.949,0.951)	2.581 (0.935)	2.598 (0.938)	2.666 (0.951)
	0.99	4.877 (0.999,1.000,1.000)	3.247 (0.993,0.986,0.991)	3.162 (0.964)	3.299 (0.986)	3.250 (0.991)
20	0.90	3.113 (0.974,0.975,0.978)	2.523 (0.905,0.902,0.884)	2.581 (0.905)	2.520 (0.894)	2.592 (0.891)
	0.95	3.599 (0.990,0.994,0.996)	2.814 (0.939,0.955,0.943)	3.077 (0.951)	2.800 (0.949)	2.828 (0.945)
	0.99	4.700 (0.999,1.000,1.000)	3.383 (0.990,0.990,0.990)	3.726 (0.992)	3.360 (0.989)	3.366 (0.989)
100	0.90	3.226 (0.958,0.971,0.965)	2.786 (0.863,0.895,0.890)	2.931 (0.893)	2.824 (0.903)	2.822 (0.895)
	0.95	3.637 (0.981,0.992,0.989)	3.056 (0.940,0.945,0.941)	3.135 (0.950)	3.132 (0.953)	3.086 (0.945)
	0.99	4.570 (0.997,1.000,1.000)	3.589 (0.980,0.988,0.986)	4.020 (0.987)	3.651 (0.992)	3.614 (0.987)

sequence, in the middle or at the end regardless of the sample size. Neither of the methods seems adequate to test for and estimate a change point for samples of size 20 or 50. Nevertheless, for large samples ($n = 100$, 200), the performance of the tests is fairly good but the variance and the bias of $\hat{\tau}$ remains quite large even when λ_0/λ_1 was as large as 6.

Next, we considered when $\lambda_0/\lambda_1 \geq 3$. To evaluate the small sample properties of the estimators and the power of the tests, we did extensive simulations. For $n = 10$, 20, 100 and 200, we generated 1000 realizations of each of the different change point Poisson processes with $\lambda_0/\lambda_1 = 3$ and $\lambda_0/\lambda_1 = 6$. The results for λ_0/λ_1 are summarized in Table 6. Here a is the average length of the confidence regions and p is the proportion of times τ belongs to the confidence region, E is the empirical value of $E(\hat{\tau}-\tau)$, σ is the empirical value of $\mathrm{sd}(\hat{\tau})$ and Power is the power of the test when $\alpha = 0.05$ (the proportion of times we reject H_0, the hypothesis that there is no change point).

Furthermore, I corresponds to the situation where τ is near the beginning of the Poisson sequence (i.e. $\tau = 3$ when $n = 10$ or $n = 20$ and $\tau = 5$ when $n = 100$ or $n = 200$), II corresponds to the situation where τ is in the middle (i.e. $\tau = n/2$) and III corresponds to the situation where τ is towards the end (i.e. $\tau = n - 3$ when $n = 10$ or $n = 20$ and $\tau = n - 5$ when $n = 100$ or $n = 200$).

It is clear that when τ/n is small ($\tau/n < 1/2$), that although L works quite well, Z gives better results in most cases (in terms of the standard deviation and the length of the confidence intervals). When τ/n is close to 1, which means that the change point occurs near the end of the sequence, L gives the best results in terms of the variance and bias of $\hat{\tau}$ and the power of the tests. When $\tau \approx n/2$ both methods perform well although, when n is small, the length of the confidence intervals for Z are slightly larger than when $\tau/n < 1/2$. For large samples, L and Z perform equally well in all cases.

The bias is generally negative and Z usually gives the smallest bias. However, for $\lambda_0 = 3.0$, $\lambda_1 = 1.0$, the bias is generally positive at the beginning of the sequence. The estimator of the change point given by L is always greater than or equal to that given by Z.

The standard deviation of Z is lower than that of L when the change is at the beginning of the sequence and this trend is reversed when the change is at the end of the sequence. Both methods perform well if the change is in the middle of the sequence. For $\lambda_0/\lambda_1 = 6$, the power of both tests is good. However when $\lambda_0/\lambda_1 = 3$, the power of both tests seems inadequate.

Table 6: Coverage probabilities, lengths of confidence intervals, bias and standard deviation for 1000 replications of a change point with $\lambda_0 = 6.0$, $\lambda_1 = 1.0$, $\alpha = 0.05$.

		$n = 10$		$n = 20$		$n = 100$		$n = 200$	
		L	Z	L	Z	L	Z	L	Z
I	p	0.82	0.93	0.86	0.91	0.90	0.93	0.91	0.94
	a	1.91	1.89	2.09	1.84	1.70	1.87	2.10	1.94
	E	0.10	-0.17	0.14	-0.11	0.15	-0.21	0.23	-0.20
	σ	0.78	0.56	1.17	0.97	2.39	0.74	5.64	0.67
	Power	0.96	0.97	0.98	0.99	1.00	1.00	1.00	1.00
II	p	0.84	0.93	0.90	0.93	0.93	0.94	0.94	0.95
	a	1.68	2.13	1.54	2.06	1.34	1.73	1.29	1.67
	E	-0.07	-0.36	-0.07	-0.31	-0.01	-0.25	0.03	-0.23
	σ	0.72	0.89	0.66	0.91	0.54	0.74	0.53	0.68
	Power	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00
III	p	0.82	0.95	0.86	0.94	0.89	0.92	0.92	0.95
	a	1.90	3.06	2.39	4.73	1.83	3.50	1.77	3.62
	E	-0.10	-0.42	-0.24	-0.85	-0.20	-0.71	-0.11	-1.03
	σ	0.67	1.09	1.26	2.35	1.60	2.85	0.82	8.43
	Power	0.95	0.93	0.97	0.94	1.00	1.00	1.00	1.00

We also performed the simulations for the confidence intervals by test inversion based on the asymptotic approximations based on r proposed by Gombay and Horvath (1996 b) for the test statistic L and by Szyszkowicz (1992 a, 1992 b) for Z. In both cases, although the power of the tests were similar to their exact counterparts, the average length of the intervals was much shorter than the exact ones and the coverage was much poorer for the asymptotic confidence intervals obtained this way.

4 An illustrative example

Stochastic modeling and various statistical methods for the analysis of nerve cell spike train data has attracted considerable attention (see, e.g. Brillinger (1992) and Wegman and Habib (1992). Using parametric neural process theory, Commenges and Seal (1985) analyzed the results of an experiment designed to study the time-to-reaction following the introduction of a stimulus to a neuron located in the posterior parietal cortex of a monkey. These reaction times are of primary interest and give insight into the function of important structures in the central nervous system. The data were collected using micro-electrodes that detect and record electrical activity in the neuron where this activity is observed as a sequence of discrete electrical discharges, called "spike trains" and occur even in non-stimulated states. The counts of spike train in intervals of equal length may be modeled as a Poisson process.

The usual response to a stimulus often results in a sudden change in the parameter of this process, although the time of increase in electrical activity lags behind the time that the stimulus is applied. In the experiment analyzed by Commenges et al (1985), a stimulus was applied after 500 ms to the same neuron 35 times, resulting in 35 data sequences. The electrical discharge of the neuron was observed for a time before and after the stimulus. The objects of interest are the time lags or reaction times. These results are reanalyzed in Joseph and Wolfson (1992) by multi-path change point methods and also in Bélisle et al (1998) using a hierarchical Bayesian model. With the use of Bayes factors they showed that the model of independence fit the data better than models with correlation.

We have thus assumed the independence model here and we use several of these data sequences (as shown in Figure 2) here to illustrate our methods. Each row represents an experiment; the numbers represent the times in microseconds (ms) between each discharge. We used intervals of 50 ms and our results agree with those published by Joseph and Wolfson; their result of 200 ms corresponds to our $\hat{\tau} \approx 14$. In Table 2, $P_L = P(L \leq L_{\text{observed}})$, a_L is the length of the confidence region for the given value of α and $(\hat{\lambda}_0/\hat{\lambda}_1)_L$ is the ratio of the estimated values of the parameters (N/A means that the estimated value of $\hat{\lambda}_1$ is zero). The same information is also given for the statistic Z. The results concerning $P = P(\max \leq \text{observed maximum})$ and the length of the confidences intervals are similar for both methods.

It should be noted that in the example, $\hat{\tau}_L \leq \hat{\tau}_Z$ and $(\hat{\lambda}_0/\hat{\lambda}_1)_L \geq (\hat{\lambda}_0/\hat{\lambda}_1)_Z$. This may seem contrary to the simulations, but it can be explained by the fact that in the data example, $\lambda_1 > \lambda_0$ whereas, throughout the article we assumed $\lambda_0 > \lambda_1$. We obtain results consistent with the theory if we work with the reversed sequence.

5 Conclusion

As argued in the introduction, the Poisson model as a prototype for discrete data deserves to be studied in its own right. We have adapted the results of Hinkley and Hinkley (1970) to the study of a change in a Poisson model and obtained asymptotic approximations to the distribution of the change point. When $\lambda_0/\lambda_1 \geq 6$, the bias and standard deviation are small and the power of the test is quite good. This leads us to consider the development of the exact distribution of two statistics, the log-likelihood ratio L and a discrete pontogram Z, using techniques of Worsley (1983). Confidence regions were also obtained by inversion of these tests.

$\hat{\tau}_L$	$\hat{\tau}_Z$	P_L	P_Z	a_L	a_Z	a_L	a_Z	$(\hat{\lambda}_0/\hat{\lambda}_1)_L$	$(\hat{\lambda}_0/\hat{\lambda}_1)_Z$
				$\alpha = 0.1$		$\alpha = 0.05$			
14	14	0.997326	0.985535	4	5	4	7	19.600000	19.600000
12	12	0.874254	0.926091	15	9	19	18	6.000000	6.000000
12	12	0.827751	0.903313	17	10	17	17	4.800000	4.800000
13	16	0.996450	0.995879	4	4	6	5	10.400000	8.000000
14	14	0.986319	0.982103	6	4	9	7	10.500000	10.500000
10	10	0.544582	0.700601	15	15	15	15	3.000000	3.000000
14	14	0.995150	0.962508	3	8	4	11	N/A	N/A
11	11	0.999808	0.985067	1	5	2	7	N/A	N/A
14	14	0.989189	0.990109	3	2	3	3	17.500000	17.500000
13	13	0.999559	0.994925	1	3	2	4	N/A	N/A
12	12	0.997093	0.988395	2	4	4	6	N/A	N/A
7	15	0.986425	0.990827	8	4	11	8	N/A	7.500000
14	14	0.965531	0.969927	7	4	11	10	7.000000	7.000000
15	15	0.976200	0.981443	7	4	10	9	15.000000	15.000000
14	14	0.993930	0.984586	6	6	8	9	16.800000	16.800000
10	16	0.975283	0.988781	9	4	11	8	N/A	12.000000
14	14	0.944814	0.977645	7	4	10	6	8.750000	8.750000
11	11	0.959571	0.960015	6	5	8	7	7.700000	7.700000
14	14	0.999404	0.996505	3	4	3	5	22.400000	22.400000
14	14	0.663965	0.820047	18	14	18	18	3.500000	3.500000
12	12	0.999596	0.994924	3	4	4	6	21.600000	21.600000
13	13	0.999840	0.997673	1	5	2	6	N/A	N/A
13	15	0.968903	0.981861	10	5	11	8	13.000000	9.375000
14	14	0.995656	0.993595	2	3	3	4	11.200000	11.200000
13	13	0.895697	0.915428	13	8	14	14	4.333333	4.333333
5	14	0.821167	0.782002	19	19	19	19	N/A	4.666667
9	9	0.985449	0.895888	4	11	6	15	N/A	N/A
13	13	0.952921	0.953307	7	8	12	10	7.583333	7.583333
16	16	0.865098	0.965427	15	3	18	7	8.000000	8.000000
10	10	0.525184	0.619728	15	15	15	15	4.000000	4.000000
7	13	0.752381	0.791800	17	18	18	18	N/A	7.800000
14	14	0.965453	0.977923	6	5	12	7	6.125000	6.125000
14	14	0.994069	0.994897	4	2	5	5	5.833333	5.833333
7	12	0.986792	0.982572	9	9	11	10	10.230769	4.500000
14	14	0.928444	0.948838	7	6	12	11	4.480000	4.480000

Figure 2: Results for the electrical discharge of a neuron

Both the discrete pontogram and the maximum likelihood method were successful in finding point and interval estimates of a change point in a Poisson sequence as long as the ratio λ_1/λ_0 is sufficiently large. This has been illustrated both by extensive simulations and on an interesting data set. We recommend that both methods be used as long as $\lambda_0/\lambda_1 \geq 3$ because, as indicated above, there is often concordance or almost concordance between the two methods. Care must be taken, however for $\lambda_0/\lambda_1 \approx 3$, because the power of the tests for detecting a change is not very good.

We also compared our exact results to the asymptotic approximations obtained by Gombay and Horvarth (1996 b) for L and by Szyszkowicz (1992 a, 1992 b) for Z. However, in our simulation study, the asymptotic tests had similar properties to the exact ones, but the confidence intervals were inferior in terms of coverage.

A Proofs of the results of Section 1

We use here the notation of Section 1.

A.1 Proof of Theorem 1

Henceforth, we will assume that $\delta = \frac{\lambda_0-\lambda_1}{\log(\lambda_0/\lambda_1)}$ is an irrational number (in particular, for integers l and m, the only way $-\ell + \delta\, m = 0$ can occur is when $\ell = m = 0$).

Let us define, as in Hinkley and Hinkley (1986) for the binomial case,

$$\begin{array}{rclcll} V_i &=& Y_{\tau-i} - Y_{\tau-i+1} &=& -X_{\tau-i+1} + \delta & (i = 1, \ldots, \tau-1), \\ V_i' &=& Y_{\tau+i} - Y_{\tau+i-1} &=& X_{\tau+i} - \delta & (i = 1, \ldots, n-\tau-1) \end{array}$$

and the random walks essential to the definition of the maximum likelihood estimator,

$$\begin{array}{rcl} W &=& (0, Y_{\tau-1} - Y_\tau, Y_{\tau-2} - Y_\tau, \ldots, Y_1 - Y_\tau) = (0, V_1, V_1 + V_2, \ldots, \sum_{j=1}^{\tau-1} V_j), \\ W' &=& (0, Y_{\tau+1} - Y_\tau, Y_{\tau+2} - Y_\tau, \ldots, Y_{n-1} - Y_\tau) = (0, V_1', V_1' + V_2', \ldots, \sum_{j=1}^{n-\tau-1} V_j'). \end{array}$$

The inference about τ will depend on the maxima of these random walks. In what follows, the fact that δ is an irrational number precludes any other possibility.

- $\hat{\tau} = \tau - k$, $k > 0$ if and only if $\max W > \max W' \geq 0$ and the maximum of the random walk W occurs at k;
- $\hat{\tau} = \tau + k$, $k > 0$ if and only if $\max W' > \max W \geq 0$ and the maximum of the random walk W' occurs at k;
- $\hat{\tau} = \tau$ if and only if $\max W = 0 = \max W'$.

We can now define the terms necessary in order to specify the distribution of $\hat{\tau} - \tau$.

Let $U = \max_{n\geq 1} \left\{ \sum_{j=1}^n V_j \right\}$ where $P(V_i = -k+\delta) = \dfrac{\lambda_0^k e^{-\lambda_0}}{k!}$. We also define $M = \max(0, U)$ (i.e. M is the maximum of the random walk W).

We also define as in Hinkley and Hinkley (1970), $p_{\ell,m} = P(M = -\ell + \delta\, m)$ and $q_{\ell,m} = P(U = -\ell + \delta\, m)$.

Let us compute the probabilities $q_{\ell,m}$:

If $\ell \geq 1$ and $m \geq 2$: $\{U = -\ell + \delta\, m\} = \cup_{k \in U_{\ell,m}} \left(\{V_1 = -k + \delta\} \cup \{\tilde{U} = -(l-k) + \delta\,(m-1)\} \right)$ where $\tilde{U}$ corresponds to the maximum of V_2, $V_2 + V_3$, $V_2 + V_3 + V_4, \ldots$ and $U_{\ell,m} = \{k\colon\ 0 \leq k \leq \ell,\ -k + \delta \leq -\ell + \delta\, m\}$. Therefore, $q_{\ell,m}$ is given by (3).

We suppose here that U and $\tilde{U}$ have the same distribution and therefore that τ and $n - \tau$ are large. We will continue to use that hypothesis.

Now for $\ell \geq 0$ and $m = 1$, $q_{\ell,1} = P(V_1 = -\ell + \delta)\, P(\tilde{M} = 0)$ where $\tilde{M} = \max\{0, \tilde{U}\}$, so $q_{\ell,1}$ is given by (3). Analogous reasoning works for $q_{0,m}$ ($m \geq 1$).

We now compute the probability that $M = 0$. If we refer to Theorem 1, page 395 of Feller (1966),

$$p_{0,0} = \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n}\, P\left(\sum_{j=1}^{n} V_j > 0 \right) \right)$$

(corresponds to $1 - \tau(1)$ in Feller's notation).

We also have

$$P\left(\sum_{j=1}^{n} V_j > 0 \right) = P\left(\sum_{j=1}^{n} (-X_j + \delta) > 0 \right) = P\left(\sum_{j=1}^{n} X_j < n\,\delta \right) = \sum_{k=0}^{[n\,\delta]} \frac{n^k\, \lambda_0^k\, e^{-n\,\lambda_0}}{k!}.$$

Let $U' = \max_{n \geq 1} \left\{ \sum_{j=1}^{n} V'_j \right\}$ where $P(V'_i = k - \delta) = \dfrac{\lambda_1^k e^{-\lambda_1}}{k!}$. We also define $M' = \max(0, U')$ (i.e. M' is the maximum of the random walk W').

We also define, $p'_{\ell,m} = P(M' = l - \delta\, m)$ and $q'_{\ell,m} = P(U' = l - \delta\, m)$.

Without repeating the computations that gave us the $q_{\ell,m}$'s, it is clear that the corresponding formulae for the probabilities $q'_{\ell,m}$ are given by (5).

We now compute the probability that $M' = 0$. As before, we have

$$p'_{0,0} = \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n}\, P\left(\sum_{j=1}^{n} V'_j > 0 \right) \right)$$

while

$$P\left(\sum_{j=1}^{n} V'_j > 0 \right) = P\left(\sum_{j=1}^{n} (X_j - \delta) > 0 \right) = P\left(\sum_{j=1}^{n} X_j > n\,\delta \right) = \sum_{k=[n\,\delta]+1}^{\infty} \frac{n^k \lambda_1^k\, e^{-n\,\lambda_1}}{k!}.$$

It is easy to see that for every ℓ and m, $q_{\ell m} = a_{\ell m}\, \frac{\lambda_0^\ell}{\ell!}\, e^{-m\,\lambda_0}\, p_{0,0}$ and $q'_{\ell m} = a'_{\ell m}\, \frac{\lambda_1^\ell}{\ell!}\, e^{-m\,\lambda_1}\, p'_{0,0}$ where the coefficients $a_{\ell m}$ and $a'_{\ell m}$ are nonnegative integers.

We now compute $\pi_n = P(\hat{\tau} = \tau + n)$, for $n = 0, \pm 1, \pm 2, \ldots$

Let us compute π_{-n} $(n > 0)$:

$\hat{\tau} = \tau - n$ means that $M = -\ell + \delta\, n$ and $M' = j - \delta\, k$ with $\overbrace{-\ell + \delta\, n}^{>0} > \overbrace{j - \delta\, k}^{\geq 0}$ (it is possible that $j = k = 0$). Therefore,

$$\pi_{-n} = \sum_{\ell=0}^{[\delta\, n]} p_{\ell,n} \left\{ p'_{0,0} + \sum_{j=1}^{\infty} \sum_{k=\left[\frac{j+\ell-\delta\, n}{\delta}\right]+1}^{\left[\frac{j}{\delta}\right]} p'_{j,k} \right\}$$

Let us now compute π_n $(n > 0)$:

$\hat{\tau} = \tau + n$ means that $M' = l - \delta\, n$ and $M = -j + \delta\, k$ with $\overbrace{l - \delta\, n}^{>0} > \overbrace{-j + \delta\, k}^{\geq 0}$ (it is possible that $j = k = 0$). Therefore,

$$\pi_n = \sum_{\ell=[n\,\delta]+1}^{\infty} p'_{\ell,n} \left\{ p_{0,0} + \sum_{j=0}^{\infty} \sum_{k=\left[\frac{j}{\delta}\right]+1}^{\left[\frac{j+\ell-\delta\, n}{\delta}\right]} p_{j,k} \right\}$$

Finally,

$$\pi_0 = p_{0,0}\, p'_{0,0}.$$

A.2 Proof of Theorem 2

Now, clearly, under the hypothesis $\tau = \tau_0$, $\Lambda_1 = M$, $\Lambda'_1 = M'$ and $\Lambda'_2 = \max\{M, M'\}$. In order to determine critical values for the likelihood ratio test and to discuss the power, we now consider the alternative hypothesis $\tau = \tau_0 - \nu$ where $\nu > 0$ is an integer.

$$\begin{aligned}
\Lambda_1 &= \max_{t \leq \tau+\nu} (Y_t - Y_{\tau+\nu}) \\
&= \max\{0, Y_{\tau+\nu-1} - Y_{\tau+\nu}, Y_{\tau+\nu-2} - Y_{\tau+\nu}, \ldots, Y_{\tau-1} - Y_{\tau+\nu}, Y_{\tau-2} - Y_{\tau+\nu}, \ldots, Y_1 - Y_{\tau+\nu}\} \\
&= \max\{0, Z_1 = -V'_\nu, Z_2 = -V'_\nu - V'_{\nu-1}, \ldots, Z_\nu = -\sum_{j=1}^{\nu} V'_j, Z_{\nu+1} = -\sum_{j=1}^{\nu} V'_j + V_1, \ldots, \\
&\qquad Z_{\nu+\tau-1} = -\sum_{j=1}^{\nu} V'_j + \sum_{j=1}^{\tau-1} V_j\}.
\end{aligned}$$

Let $\mathcal{U} = \max\{Z_i\}$ and $\mathcal{M} = \max\{0, \mathcal{U}\}$. We define $p_{\ell,m}(\nu) = P(\mathcal{M} = -\ell + \delta\, m | \tau = \tau_0 - \nu)$ and $q_{\ell,m}(\nu) = P(\mathcal{U} = -\ell + \delta\, m | \tau = \tau_0 - \nu)$. Note that $p_{\ell m}(0) = p_{\ell m}$ and $q_{\ell m}(0) = q_{\ell m}$ given above. We are still assuming that τ and n are large.

If $\ell \geq 1$ and $m \geq 2$: $\{\mathcal{U} = -\ell + \delta\, m\} = \cup_{k \in U_{\ell,m}} \Big(\{-V'_\nu = -k + \delta\} \cup \{\tilde{\mathcal{U}} = -(l-k) + \delta\,(m-1)\} \Big)$ where $\tilde{\mathcal{U}}$ corresponds to the maximum of $Z_2 + (-V'_\nu),\ Z_3 + (-V'_\nu),\ \ldots$ under the hypothesis $\tau = \tau_0 - (\nu - 1)$ and $U_{\ell,m} = \{k\colon\ 0 \leq k \leq \ell,\ \ -k + \delta \leq -\ell + \delta\, m\}$.

Therefore if $\ell \geq 1$ and $m \geq 2$,

$$q_{\ell,m}(\nu) = \sum_{\ell - \delta\,(m-1) \leq k \leq \ell} \frac{\lambda_1^k}{k!}\, e^{-\lambda_1}\, q_{\ell-k,m-1}(\nu - 1)$$

If $\ell = 0$ and $m \geq 1$:

$$\begin{aligned} & q_{0,m}(\nu) \\ & = \begin{cases} \left[\prod_{k=1}^m P(-V'_{\nu+1-k} = \delta)\right] P(\hat{\mathcal{M}} = 0) & = e^{-m\,\lambda_1}\, p_{0,0}(\nu - m) \quad \text{if } m \leq \nu, \\ \left[\prod_{k=1}^m P(-V'_{\nu+1-k} = \delta)\right] \left[\prod_{k=1}^{m-\nu} P(V_k = \delta)\right] P(\hat{M} = 0) & = e^{-\nu\,\lambda_1}\, e^{-(m-\nu)\,\lambda_0}\, p_{0,0} \quad \text{if } m > \nu \end{cases} \end{aligned}$$

If $\ell \geq 0$ and $m = 1$:

$$q_{\ell,1}(\nu) \;=\; P(V'_\nu = -\ell + \delta)\, P(\tilde{\mathcal{M}} = 0) = \frac{\lambda_1^\ell}{\ell!}\, e^{-\lambda_1}\, p_{0,0}(\nu - 1).$$

Clearly, for all other values of ℓ and m, $q_{\ell,m}(\nu) = 0$.

Now $p_{0,0}(\nu) = P(\mathcal{M} = 0 | \tau = \tau_0 - \nu)$; $\{\mathcal{M} = 0\} = \cup_{-k+\delta \leq 0} \{-V'_\nu = -k + \delta$ and $(Z_i + V'_\nu) \leq k - \delta,\ i \geq 2\}$. Hence,

$$p_{0,0}(\nu) \;=\; \sum_{k \geq \delta} \frac{\lambda_1^k}{k!}\, e^{-\lambda_1} \sum_{-\ell + m\,\delta \leq k - \delta} q_{\ell,m}(\nu - 1) = \sum_{\max\{\delta, -\ell + (m+1)\,\delta\} \leq k} \frac{\lambda_1^k}{k!}\, e^{-\lambda_1}\, q_{\ell,m}(\nu - 1).$$

Note that under the hypothesis $\tau = \tau_0 - \nu$, $\nu < 0$, Λ_1 has the same distribution as M.

Therefore,

$$P_1(x;\nu) = P(\Lambda_1 \leq x | \tau = \tau_0 - \nu) \;=\; \begin{cases} \sum_{J(x)} p_{\ell,m}(\nu) & \text{if } \nu \geq 1, \\ P_1(x) & \text{if } \nu \leq 0. \end{cases}$$

where

$$\begin{aligned} J(x) &= \{(l,m)\colon\ -\ell + \delta\, m \leq x\}, \\ J'(x) &= \{(l,m)\colon\ l - \delta\, m \leq x\}. \end{aligned}$$

We now discuss the hypothesis $\tau = \tau_0 + \nu$ where $\nu > 0$ is an integer.

$$\begin{aligned}
\Lambda_1' &= \max_{t\geq \tau-\nu}(Y_t - Y_{\tau-\nu})\\
&= \max\{0, Y_{\tau-\nu+1}-Y_{\tau-\nu}, Y_{\tau-\nu+2}-Y_{\tau-\nu},\ldots, Y_\tau - Y_{\tau-\nu}, Y_{\tau+1}-Y_{\tau-\nu}, Y_{\tau+2}-Y_{\tau-\nu},\\
&\qquad \ldots, Y_{n-\tau-1}-Y_{\tau-\nu}\}\\
&= \max\{0, Z_1' = -V_\nu, Z_2' = -V_\nu - V_{\nu-1},\ldots, Z_\nu' = -\sum_{j=1}^{\nu} V_j, Z_{\nu+1}' = -\sum_{j=1}^{\nu} V_j + V_1',\ldots,\\
&\qquad\qquad Z_{\nu+n-\tau-1}' = -\sum_{j=1}^{\nu} V_j + \sum_{j=1}^{n-\tau-1} V_j'\}.
\end{aligned}$$

Let $\mathcal{U}' = \max\{Z_i'\}$ and $\mathcal{M}' = \max\{0,\mathcal{U}'\}$. We also define $p_{\ell,m}'(\nu) = P(\mathcal{M}' = l - \delta\, m|\tau = \tau_0+\nu)$ and $q_{\ell,m}'(\nu) = P(\mathcal{U}' = l-\delta\, m|\tau=\tau_0+\nu)$. Note that $p_{\ell m}'(0) = p_{\ell m}'$ and $q_{\ell m}'(0) = q_{\ell m}'$ of Section 1.

In this case, the computations would be essentially the same as above so we will not repeat them.

If $\ell \geq 1$ and $m \geq 2$:

$$q_{\ell,m}'(\nu) = \sum_{0\leq k\leq \ell-\delta\,(m-1)} \frac{\lambda_0^k}{k!}\, e^{-\lambda_0}\, q_{\ell-k,m-1}'(\nu-1)$$

If $\ell = 0$ and $m \geq 1$:

$$q_{0,m}'(\nu) = \begin{cases} e^{-\lambda_0}\, p_{0,0}'(\nu-1) & \text{if } m=1,\\ 0 & \text{if } m>1\end{cases}$$

If $\ell \geq 0$ and $m = 1$:

$$q_{\ell,1}'(\nu) = \frac{\lambda_0^\ell}{\ell!}\, e^{-\lambda_0}\, p_{0,0}'(\nu-1).$$

Clearly, for all other values of ℓ and m, $q_{\ell,m}'(\nu) = 0$.

As for $p_{0,0}'(\nu) = P(\mathcal{M}' = 0|\tau=\tau_0+\nu)$,

$$p_{0,0}'(\nu) = \sum_{k\leq\delta} \frac{\lambda_0^k}{k!}\, e^{-\lambda_0} \sum_{\ell-m\,\delta\leq -k+\delta} q_{\ell,m}'(\nu-1) = \sum_{k\leq\delta}\frac{\lambda_0^k}{k!}\, e^{-\lambda_0}\sum_{m\geq\frac{k+\ell}{\delta}-1} q_{\ell,m}'(\nu-1)$$

Note that under the hypothesis $\tau = \tau_0+\nu$, $\nu<0$, Λ_1' has the same distribution as M'.

Therefore,

$$P_1'(x;\nu) = P(\Lambda_1'\leq x|\tau=\tau_0+\nu) = \begin{cases} \sum_{J'(x)} p_{\ell,m}'(\nu) & \text{if } \nu\geq 1,\\ P_1'(x) & \text{if } \nu\leq 0.\end{cases}$$

Finally,

$$P_2(x;\nu) = P(\Lambda_2 \le x | \tau = \tau_0 + \nu) \quad = \quad \begin{cases} P_1(x;-\nu)\, P_1'(x) & \text{if } \nu \le 0, \\ P_1'(x;\nu)\, P_1(x) & \text{if } \nu \ge 1. \end{cases}$$

A.3 Proof of Proposition 1

We want to estimate the size of the following expressions in terms of n. If $n\,\delta \ge 1$ then

$$\sum_{k=0}^{[n\,\delta]} \frac{n^k\,\lambda_0^k\,e^{-n\,\lambda_0}}{k!} \quad = \quad \frac{\int_{n\,\lambda_0}^{\infty} e^{-x}\,x^{[n\,\delta]}\,dx}{\int_0^{\infty} e^{-x}\,x^{[n\,\delta]}\,dx} \le \frac{\int_{n\,\lambda_0}^{\infty} e^{-x}\,x^{n\,\delta}\,dx}{[n\,\delta]!}, \tag{8}$$

$$\sum_{k=[n\,\delta]+1}^{\infty} \frac{n^k\lambda_1^k\,e^{-n\,\lambda_1}}{k!} \quad = \quad \frac{\int_0^{n\,\lambda_1} e^{-x}\,x^{[n\,\delta]}\,dx}{\int_0^{\infty} e^{-x}\,x^{[n\,\delta]}\,dx} \le \frac{\int_0^{n\,\lambda_1} e^{-x}\,x^{n\,\delta}\,dx}{[n\,\delta]!}. \tag{9}$$

Let $h(t) = -t + \log(1+t)$ for $t > -1$. Note that $h(t)$ is increasing for $-1 < t < 0$, that $h(0) = 0$ and that $h(t)$ is decreasing for $t > 0$ (in particular, $h(t) < 0$ for $t \ne 0$). It is also useful to note that $\lambda_1 < \delta < \lambda_0$.

If we write $e^{-x}\,x^{n\,\delta} = e^{-x+n\,\delta\,\log x}$ and $x = n\,\delta\,(1+t)$ then

$$\int_{n\,\lambda_0}^{\infty} e^{-x}\,x^{n\,\delta}\,dx \quad = \quad (n\,\delta)^{n\,\delta+1} \int_{\lambda_0/\delta-1}^{\infty} e^{N\,h(t)}\,dt \le (n\,\delta)^{n\,\delta+1}\,e^{(N-1)\,h(\lambda_0/\delta-1)} \int_{-1}^{\infty} e^{-h(t)}\,dt$$

which gives (8) when combined with Stirling's formula. Equation (9) is proven in a similar manner.

Remark 1 *The recurrence relations of Theorem 3 and Theorem 4 imply some numerical complications. For a sample of size n whose sum is S, we need to reserve a space of size $n\,(S+1)$ real variables to avoid having to compute the same value of $F_{k,n}(t;m,S)$ more than once. In fact, it is better to start by computing all the $F_{k,n}(t;m,S)$ in the right order and not to use recursion in the computer program (recursion is rather "costly" in computer programs). The sum*

$$\sum_{m=0}^{S_{k+1}} F_{k,n}(t;m,S) \binom{S_{k+1}}{m} \left(\frac{k}{k+1}\right)^m \left(\frac{1}{k+1}\right)^{S_{k+1}-m} \tag{10}$$

must also be computed carefully. Indeed the terms $\binom{S_{k+1}}{m}$, $\left(\frac{k}{k+1}\right)^m$ and $\left(\frac{1}{k+1}\right)^{S_{k+1}-m}$ require many multiplications to be computed. In addition, computing $\binom{S_{k+1}}{m}$ directly can cause an overflow when m is large while $\left(\frac{1}{k+1}\right)^{S_{k+1}-m}$ can cause an underflow (the number becomes indistinguishable form 0) when S_{k+1} is large and m is small. Therefore computing each term in (10) and adding them up is inefficient and collects roundoff errors. For a discussion of these basic principles see for instance Cheney and Kincaid (1994).

The following algorithm circumvents these pitfalls to some extent. The basic idea is to use the "nested form": for example, to compute the polynomial $a_0\,x^n + \cdots + a_{n-1}\,x + a_n$ (which requires

$n\,(n+1)/2$ multiplications and n additions) it is more economical to compute $a_n + x\,(a_{n-1} + x\,(a_{n-2} + x\,(\cdots + x\,(a_2 + x\,(a_1 + x\,a_0)))))$ (which requires n multiplications and n additions).

sum $\leftarrow$ *0*
factor $\leftarrow k+1$
for m *from 0 to* S_{k+1} *do*
  sum $\leftarrow (sum + F_{k,n}(t; m, S) * factor)/(k+1)$
  factor $\leftarrow factor * k * (S_{k+1} - m)/((m+1) * (k+1))$
end for loop
$F_{k+1,n}(t; S_{k+1}, S) \leftarrow sum$

References

[Akman, V. E. and Raftery A. E. (1986)] *Asymptotic inference for a change point Poisson process*, The Annals of Statistics, 14, 1583–1590.

[Aldous, D. (1989)] *Probability approximations via the Poisson clumping heuristic*, Springer-Verlag, New-York.

[Bélisle P., Joseph, L., MacGibbon, B., Wolfson D. and du Berger, R. (1998)] *Change point analysis of neural spike train data*, Biometrics, 54, 113–123.

[Bickel, P. J. and Doksum (1977)] *Mathematical Statistics*, Holden-Day, Oakland.

[Brillinger, D. R. (1992)] *Nerve cell spike train data analysis: a progression of techniques*, Journal of the American Statistical Association, 87, 260–271.

[Carlstein, E. (1988)] *Nonparametric change point estimation*, Annals of Statistics, 16, 188–197.

[Carlstein, E. and Lele S. (1994)] *Nonparametric change point estimation for data from an ergodic sequence*, Theory of Probability and its Applications, 38, 726–733.

[Cheney, W. and Kincaid D. (1994)] *Numerical mathematics and computing*, Third Edition, Brooks/Cole, 1994.

[Commenges, D. and Seal, J. (1985)] *The analysis of neuronal discharge sequences: change point estimation and comparison of variances*, Statistics in Medicine, 5, 91–104.

[Commenges, D. Seal and J. and Pinatel, P. (1986)] *Inference about a change point in experimental neurophysiology*, Mathematical Biosciences, 80, 81–108.

[Csörgõ M. and Horvath L. (1987 a)] *Asymptotic distributions of pontograms*, Mathematical Proceedings of the Cambridge Philosophical Society 101, 131–139.

[Csörgõ M. and Horvath L. (1987 b)] *Nonparametric tests for the change-point*, Journal of Statistical Planning and Inference, 17, 1–9.

[Csörgõ M. and Horvath L. (1988)] *Nonparametric methods for the change-point*. In: P. R. Krishnaiah and C. R. Rao, Eds., Handbook of Statistics, Vol. 7, Elsevier, Amsterdam, 403–425.

[Deshayes, J. (1984)] *Ruptures de modèles pour des processus de Poisson*, Annales Scientifiques de Clermont-Ferrand, 11, 78, 1–7.

[Feller W. (1966)] *An introduction to probability theory and its applications, Volume II*, John Wiley & Sons, Inc.

[Ferger, D. (1994)] *Asymptotic distribution theory of change point estimators and confidence intervals based on bootstrap approximation*, Mathematical Methods of Statistics, 3, 362–378.

[Ferger, D. (1995)] *Nonparametric tests for nonstandard change point problems*, Annals of Statistics, 23, 1848–1861.

[Gombay, E. and Horvath L. (1990)] *Asymptotic distributions of maximum likelihood tests for change in the mean*, Biometrika, 77, 411–414.

[Gombay, E. and Horvath L. (1994)] *An application of the maximum likelihood test to the change-point problem*, Stochastic Processes and their Applications, 50, 161–171.

[Gombay, E. and Horvath L. (1996 a)] *Approximations for the time of change and the power function in change-point models*, Journal of Statistical Planning and Inference, 52, 43–66.

[Gombay, E. and Horvath L. (1996 b)] *On the rate of approximations for maximum likelihood tests in change-point models*, Journal of Multivariate Analysis, 56, 120–152.

[Haccou, P., Meelis, E. and Van De Geer (1988)] *The likelihood ratio test for the change point problem for exponentially distributed random variables*, Stochastic Processes and their Applications, 27, 121–139.

[Hinkley, D. V. (1970)] *Inference about the change point in a sequence of random variables*, Biometrika, 57, 1, 1–17.

[Hinkley, D. V. and Hinkley E. A. (1970)] *Inference about the change point in a sequence of binomial variables*, Biometrika, 57, 3, 477–488.

[Joseph L. and Wolfson D. (1992)] *Estimation in multipath change-point problems*, Communications in Statistics, 21, 897–913.

[Kendall D. G. and Kendall W. S. (1980)] *Alignments in two-dimensional random sets of points*, Advances in Applied Probability, 12, 380–424.

[MacNeil, I. B. and Mao, Y. (1994)] *Change-point-analysis for mortality and morbidity rate*, Journal of Statistical Applied Science, 1, 359–377.

[Siegmund, D. (1988)] *Confidence sets in change-points problems*, International Statistical Review, 56, 1, 31–48.

[Szyszkowicz, B. (1992 a)] *Asymptotic distributions of weighted pontograms under contiguous alternatives*, Mathematical Proceedings of the Cambridge Philosophical Society 112, 431–447.

[Szyszkowicz, B. (1992 b)] *Asymptotic distributions of weighted compound Poisson bridges*, Mathematical Proceedings of the Cambridge Philosophical Society 112, 613–629.

[Wegman, E. J. and Habib M. K. (1992)] *Stochastic methods for neural systems*, Journal of Statistical Planning and Inference, 33, 5-26.

[Worsley K. J. (1983)] *The power of likelihood ratio and cumulative sum tests for a change in a binomial probability*, Biometrika, 70, 455–464.

[Worsley (1986)] *Confidence regions and tests for a change point in a sequence of exponential family of random variables*, Biometrika, 73,l 91–104.