Simultaneous confidence bands for the rate of a nonhomogeneous Poisson process

Clive R. Loader
AT&T Bell Laboratories, Room 2C-279
600 Mountain Avenue,
Murray Hill, NJ 07974, U.S.A.

Summary

We derive approximate simultaneous confidence bands for the rate of a nonhomogeneous Poisson process observed on a bounded interval. The method is based on a kernel estimate of the rate, which can be approximated by a Gaussian process with mean equal to the true rate. After a square root transformation, the covariance function is nearly independent of the true rate, and standard methods for approximating the distribution of the maxima of Gaussian processes are applied to derive the confidence bands. The method can be adapted to find confidence bands based on kernel density estimates.

Key Words: Confidence bands; density estimate; kernel; Poisson process.

1 Introduction

Let $\{X(t); 0 \le t \le T\}$ be a non-homogeneous Poisson process with rate $\lambda(t)$, which we assume is continuous, differentiable and bounded away from 0 and ∞ on $[0, T]$. Suppose we observe $X(T) = n$ with event times $T_1, \ldots, T_n$. Let $K(x)$ be a kernel function, and $K_h(x) = K(x/h)$. A kernel rate estimate of $\lambda(t)$ with bandwidth h is

$$\hat{\lambda}(t) = \frac{\sum K_h(T_i - t)}{\int_0^T K_h(u-t)du}. \tag{1}$$

We consider approximate confidence bands for $\lambda(t)$ of the form

$$\left[\left\{\sqrt{\hat{\lambda}}(t) - c\sqrt{\sigma}(t,t)\right\}^2, \left\{\sqrt{\hat{\lambda}}(t) + c\sqrt{\sigma}(t,t)\right\}^2\right]. \tag{2}$$

where

$$\sigma(s,t) = \frac{\int_0^T K_h(u-s)K_h(u-t)du}{4\int_0^T K_h(u-s)du \int_0^T K_h(u-t)du}. \tag{3}$$

The simultaneous coverage probability is

$$1 - \gamma = \mathrm{pr}\left\{|\sqrt{\hat{\lambda}}(t) - \sqrt{\lambda}(t)| \le c\sqrt{\sigma}(t,t) \text{ for all } t\right\}.$$

In section 2 we show the dependence of γ on $\lambda(t)$ is small, and derive the approximation

$$\gamma \approx 2\{1 - \Phi(c)\} + \frac{\kappa_0}{\pi}e^{-c^2/2} \tag{4}$$

where

$$\kappa_0 = \int_0^T m(t)^{\frac{1}{2}}dt; \qquad m(t) = \frac{\partial^2}{\partial s \partial t}\frac{\sigma(s,t)}{\{\sigma(s,s)\sigma(t,t)\}^{\frac{1}{2}}}\bigg|_{s=t}.$$

Some manipulation shows

$$m(t) = \frac{\int_0^T K_h(u-t)^2du \int_0^T \frac{\partial}{\partial t}K_h(u-t)^2du - \left\{\int_0^T K_h(u-t)\frac{\partial}{\partial t}K_h(u-t)du\right\}^2}{\left\{\int_0^T K_h(u-t)^2du\right\}^2}.$$

In Section 3 we apply the method to British coal mining accident data, showing how interesting features of the rate function can be distinguished from noise. However, we recommend against using the confidence bands to test the fit of parametric models due to lack of power.

Density estimation is closely related to rate estimation. If $T_1, \ldots, T_n$ are regarded as an independent sample from a density $f(t)$, a kernel density estimate and confidence bands for f are obtained by dividing (1) and (2) by n.

Bickel and Rosenblatt (1973) have studied the distribution of $\sup_x |\hat{f}(x) - f(x)|/f(x)^{\frac{1}{2}}$ and propose confidence bands of a form similar to ours. However, there are some important technical differences in the way limiting Gaussian processes are treated and the probability approximation (4) is usually more accurate. This point will be discussed more at the end of Section 2. Hall and Titterington (1988) construct confidence bands in a different manner which will be discussed in Section 3.

There is nothing explicitly univariate about our method. When the Poisson process is observed on a compact set in two or more dimensions, the integrals in (1) and (3) are taken over this region. Results of Knowles and Siegmund (1989) and Sun (1992) can be used to derive relations similar to (3) in the multivariate case.

Implementation of the confidence bands requires specification of K and h. We refer to Silverman (1986) for a detailed discussion of these issues; however, there are some additional requirements that we mention here. Firstly, we must assume $m(t)$ exists, which requires K to be absolutely continuous, and $K(x)^2$ and $K'(x)^2$ are both integrable. For computational purposes, it is helpful if K is chosen so that the integrals defining $m(t)$ can be computed in closed form.

Knafl, Sacks and Ylvisaker (1985) and unpublished work by Loader and Sun adjust confidence bands around smooth regression curves for bias by widening the bands, and similar methods could be applied here. The difficulty is that the adjustment depends on the maximum bias, which is unknown and difficult to estimate.

For practical purposes it is sufficient to be slightly conservative in our choice of bandwidth and beware of the effects of bias. An informal procedure to select the bandwidth is to plot the estimate and confidence bands for several values of h, and reduce h until the rate estimate begins appearing noisy.

2 Derivation of the Approximation Formula

Write $\lambda(t) = \alpha f(t)$ where $\int_0^T f(t)dt = 1$. Using our assumptions on $\lambda(t)$, there exists a constant $\beta > 0$ such that

$$|\log\{f(t)\} - \log\{f(s)\}| \leq \beta |t-s| \tag{5}$$

for all s and t. Define

$$\begin{aligned}
\mu(t) = \alpha^{-1} E\{\hat{\lambda}(t)\} &= \frac{\int_0^T K_h(u-t)f(u)du}{\int_0^T K_h(u-t)du} \\
\tilde{\sigma}(s,t) = \alpha^{-1}\mathrm{cov}\{\hat{\lambda}(s), \hat{\lambda}(t)\} &= \frac{\int_0^T K_h(u-s)K_h(u-t)f(u)du}{\int_0^T K_h(u-s)du \int_0^T K_h(u-t)du}.
\end{aligned}$$

For convenience, we assume K is nonnegative and $K(x) = 0$ for $|x| > 1$. By (5),

$$f(t)e^{-\beta h} \leq \mu(t) \leq f(t)e^{\beta h}$$

$$4\sigma(s,t)\{f(s)f(t)\}^{\frac{1}{2}}e^{-\beta h} \leq \tilde{\sigma}(s,t) \leq 4\sigma(s,t)\{f(s)f(t)\}^{\frac{1}{2}}e^{\beta h} \tag{6}$$

Define

$$\begin{aligned}
\tilde{W}(t) &= \alpha^{-\frac{1}{2}}\{X(t) - \int_0^t \lambda(u)du\} \\
&\Rightarrow \mathcal{G}\{0, \int_0^{\min(s,t)} f(u)du\}
\end{aligned}$$

as $\alpha \to \infty$. Here $\mathcal{G}(\,\cdot\,,\,\cdot\,)$ denotes a Gaussian process with the specified mean and covariance functions, and $\Rightarrow$ denotes convergence in distribution. Using (1),

$$\begin{aligned}
\alpha^{-\frac{1}{2}}\{\hat{\lambda}(t) - \alpha\mu(t)\} &= \frac{\int_0^T K_h(t-u)d\tilde{W}(u)}{\int_0^T K_h(t-u)du} \\
&\approx \mathcal{G}\{0, \tilde{\sigma}(s,t)\}
\end{aligned}$$

and by a δ-method type argument,

$$\sqrt{\hat{\lambda}}(t) - \sqrt{\{\alpha\mu(t)\}} \approx \mathcal{G}[0, \frac{\tilde{\sigma}(s,t)}{4\{f(s)f(t)\}^{\frac{1}{2}}}].$$

If h is fixed as $\alpha \to \infty$, these are genuine limit theorems. However, we prefer to take $h = o(\alpha^{-1/3})$ with $\alpha h \to \infty$, so the bias $\surd\{\alpha\mu(t)\} - \surd\lambda(t) = O(h\surd\alpha)$ is negligible compared to the standard deviation $\surd\tilde{\sigma}(t,t) = O(h^{-1/2})$. Also, (6) can be used to approximate $\tilde{\sigma}(s,t)$. Then

$$\surd\hat{\lambda}(t) - \surd\lambda(t) \approx \mathcal{G}\{0, \sigma(s,t)\}.$$

Therefore,

$$\begin{aligned} \gamma &\approx 1 - \mathrm{pr}\left\{|Z_h(t)| \leq c\surd\sigma(t,t) \text{ for all } t\right\} \\ &\leq 2\{1 - \Phi(c)\} + \frac{\kappa_0}{\pi} e^{-c^2/2} \end{aligned} \tag{7}$$

where $\{Z_h(t), 0 < t < T\}$ is a Gaussian process with mean $E\{Z_h(t)\} = 0$ and covariance $E\{Z_h(s)Z_h(t)\} = \sigma(s,t)$. If K is absolutely continuous, the covariance function $\sigma(s,t)$ and sample paths of $Z_h(t)$ are also absolutely continuous. The distribution of the maxima of Gaussian processes of this type has been studied by several authors; (7) applies (3.7) of Davies (1977). For a fixed covariance function, (7) provides an asymptotic approximation to the Gaussian process tail probability as $c \to \infty$, and numerical examples in Knowles and Siegmund (1989) show the approximation is good for the small tail probabilities of interest.

Following Bickel and Rosenblatt (1973) a more rigorous limit theorem may be derived by letting $c \to \infty$ at rate $\surd\log(1/h)$. If some effects from the boundaries at 0 and T are ignored, the process $Z_h(t)$ is stationary and results from extreme value theory can be used to show

$$1 - \lim_{h\to 0} \mathrm{pr}\left\{|Z_h(t)| < (a_h x + b_h)\surd\sigma(t,t) \text{ for all } t\right\} = 1 - \exp(-2e^{-x}) \tag{8}$$

where

$$\begin{aligned} a_h &= \{2\log(T/h)\}^{-\frac{1}{2}} \\ b_h &= \{2\log(T/h)\}^{\frac{1}{2}} + \frac{\log\{m/(2\pi)\}}{\{2\log(T/h)\}^{\frac{1}{2}}} \\ m^2 &= \frac{\int_{-\infty}^{\infty} K'(x)^2 dx}{\int_{-\infty}^{\infty} K(x)^2 dx}. \end{aligned}$$

Unfortunately, the rate of convergence of (8) is $1/|\log(h)|$. For values of h and x of practical interest, (8) is often larger than the upper bound given by (7).

3 Example: Coal Mining Accidents

Jarrett (1979) lists 191 accidents in British coal mines resulting in 10 or more deaths between 15 March, 1851 and 22 March, 1962. This is an extended and corrected version of an earlier dataset in Maguire, Pearson and Wynn (1952). Several parametric models have been fitted to this dataset; for example Cox and Lewis (1966); Akman and Raftery (1986) and Loader (1992).

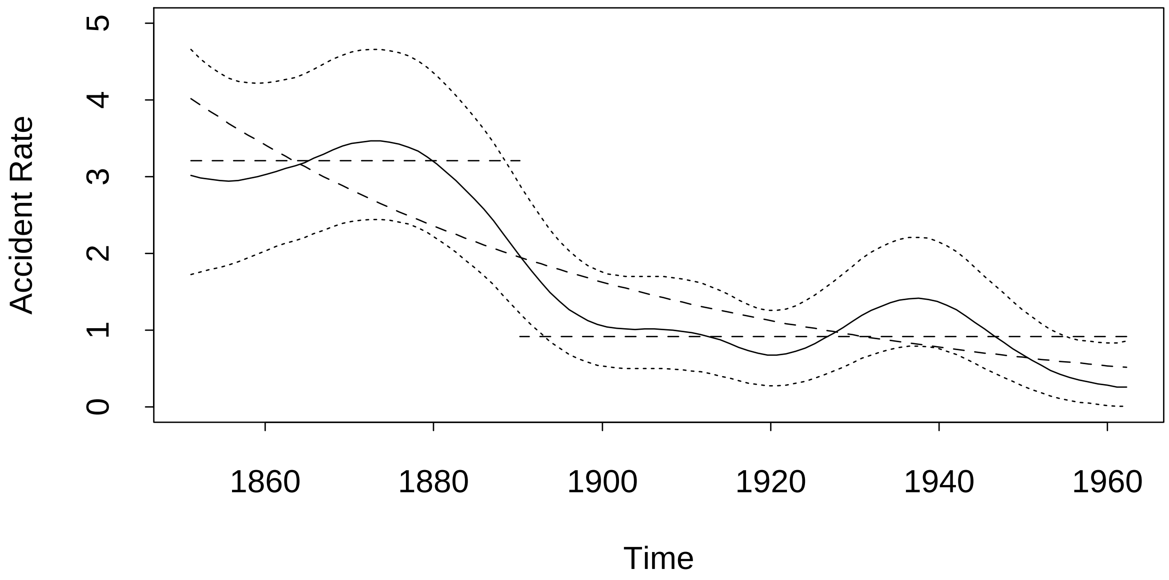

Figure 1: Kernel rate estimate in Accidents per year and 90% confidence bands for British coal mining data. Also shown are two parametric fits: a change point model and a log-linear rate function.

We display a kernel estimate of the rate together with 90% confidence limits in Figure

1. We used the bisquare kernel,

$$K(x) = \begin{cases} (1-x^2)^2 & |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

with bandwidth $h = 15$; this corresponds to $\kappa_0 = 12.31$ and $c = 2.73$. Also displayed are the change point model fitted by Akman and Raftery (1986) and the model $\log\{\lambda(t)\} = a + bt$, fitted by maximum likelihood.

The rate estimate and confidence bands indicate a significant drop in the accident rate between 1880 and 1900; possibly a small rate increase between 1930 and 1940, and another significant drop after 1940. The change point model picks up the big decrease between 1880 and 1900, but fails to find the later decrease. The change point model does not satisfy our smoothness assumptions and should not be rejected on the basis of our confidence bands.

The confidence bands cover the log-linear fit, although this model has been rejected in parametric tests with more specific alternatives; see for example Raftery and Akman (1986) and Loader (1992). While the confidence bands are useful for assessing the significance of specific features of the rate estimate, they are less useful in constructing formal tests of parametric models due to large variability and resulting lack of power.

We conclude with brief remarks about the method of Hall and Titterington (1988). If we divide $[0, T]$ into nonoverlapping windows of width $2h$, then independent rate estimates can be found at the midpoint of each window and simultaneous bands for these point estimates are easily found. If a bound on the derivative $\lambda'(t)$ is known, then these bands are easily interpolated to find continuous confidence bands. However, this approach is inadequate unless it is appropriate to choose h very small. For the coal mining data set, with h slightly less than 15 we can choose four nonoverlapping windows, which is insufficient to capture all the structure of $\lambda(t)$. Choosing smaller windows will substantially increase the variability.

REFERENCES

Akman, V. E. and Raftery, A. E. (1986). Asymptotic inference for a change-point Poisson process. *Ann. Statist.* **14**, 1583-90.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071-95.

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

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247-54.

HALL, P. and TITTERINGTON, D. M. (1988). On confidence bands in nonparametric density estimation and regression, *J. Mult. Anal.* **27**, 228-254.

JARRETT, R. G. (1979). A note on the intervals between coal-mining disasters. *Biometrika* **74**, 71-83.

KNAFL, G.; SACKS, J. and YLVISAKER, D. (1985). Confidence bands for Regression functions. *J. Amer. Statist. Assn.* **80**, 683-91.

KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's geometric approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57**, 205-20.

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

MAGUIRE, E. A.; PEARSON, E. S. and WYNN, A. H. A. (1952). The time intervals between industrial accidents. *Biometrika* **38**, 168-80.

RAFTERY, A. E. and AKMAN, V. E. (1986). Bayesian Analysis of a Poisson process with a change-point. *Biometrika* **73**, 85-9.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

SUN, J. (1992). Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab*, to appear.