AN ESTIMATION PROBLEM WITH POISSON PROCESSES

BY

ALAN E. GELFAND

TECHNICAL REPORT NO. 317
APRIL 20, 1982

PREPARED UNDER CONTRACT
NOO014-76-C-0475 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any purpose of the United States Government
Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN ESTIMATION PROBLEM WITH POISSON PROCESSES

By

Alan E. Gelfand

TECHNICAL REPORT NO. 317

April 20, 1982

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
An Estimation Problem with Poisson Processes

By

Alan E. Gelfand

1. Introduction

This paper considers the following problem. Suppose we observe a collection of independent Poisson processes, $X_i(t)$, $i = 1, 2, ..., n$, such that

$$X_i(t) \sim P_0(\lambda_i(t))$$

with

$$\lambda_i(t) = \delta_{i0}(t).$$

Let

$$p(t) = \int_0^t \rho(u)du.$$ 

If all the processes are observed up to a fixed time point $T$, how may we estimate $p(t)$? If we assume $\rho$ integrable on $[0, \infty)$ and without loss of generality set $p(\infty) = 1$ then each $X_i(t)$ converges in distribution to $X_i(\infty)$ where

$$X_i(\infty) \sim P_0(\delta_i).$$

In section 2 we formalize the mathematical setting for this problem. In section 3 we describe two estimation approaches and illustrate these approaches with two choices of $\rho$. Application of this sort of estimation problem arises when there is interest in the conditional distribution of $X(t)$ given $X(\infty)$ that is, in the proportion of
occurrences of a process which may be expected by time \( t \). In section 4 a successful application of the methodology to the active life of a judicial opinion is described.

2. The Set Up

Suppose

\[ X(t) \sim P_0(\lambda(t)) \]

with

\[ \lambda(t) = \delta \rho(t) \]

Suppose further that

(i) \( \rho \) is integrable with

\[ p(t) = \int_0^t \rho(u) \, du \]

and that \( \lim_{t \to \infty} p(t) = p(\infty) = 1 \).

(ii) \( \rho(t) \) strictly decreases in \( t \), \( \rho(0) < \infty \).

(iii) \( \rho'(t), \rho''(t) \) exist, \( t > 0 \).

Two particular examples of interest in the sequel are

\[ \rho_1(t;\alpha) = e^{-\alpha t}, \quad \alpha > 0 \quad \text{with} \]

\[ p_1(t;\alpha) = 1 - e^{-\alpha t} \]

and

\[ \rho_2(t;\alpha) = \frac{\alpha}{(\alpha+t)^2}, \quad \alpha > 0 \quad \text{with} \]

\[ p_2(t;\alpha) = t/(\alpha+t) \].
In general $\rho$ and $p$ tend to look as in Figures 1 and 2.

![Fig 1: A typical $\rho(t)$](image1)

![Fig 2: A typical $p(t)$](image2)

In view of our assumptions

$$X(t) \Rightarrow X(\infty) \quad \text{where} \quad X(\infty) \sim P_0(\delta)$$

with interpretation
\[ X(t) = \# \text{ of occurrences of the process by time } t. \]

\[ X(\infty) = \# \text{ of occurrences of the process at the end of time.} \]

Why is \( p(t) \) of interest? The conditional distribution of \( X(t) \) given \( X(\infty) \) is binomial i.e.

\[ X(t) | X(\infty) \sim B_1(X(\infty), p(t)) \]

and therefore simple computation shows that

\[
E\left( \frac{X(t)}{X(\infty)} \mid X(\infty) > 0 \right) = p(t)
\]

\[
\text{var}\left( \frac{X(t)}{X(\infty)} \mid X(\infty) > 0 \right) = c(\delta) \cdot p(t)(1-p(t))
\]

where

\[ c(\delta) = \left( 1 - e^{-\delta} \right)^{-1} \sum_{y=1}^{\infty} \frac{\delta^y e^{-\delta}}{y!} . \]

Hence \( p(t) \) measures the expected proportion of occurrences of the process by time \( t \). An implication of our assumptions on \( \rho \) is that for any interval of fixed length, the earlier in time it is placed, the greater the expected number of occurrences in the interval.

The significance of \( p(t) \) is more strongly emphasized in the case where we have \( n \) independent processes \( X_i(t) \) such that

\[ X_i(t) \sim P_0(\lambda_i(t)) \quad i = 1, \ldots, n \]

where

\[ \lambda_i(t) = \delta_i \rho(t) . \]

In this case we would like to synthesize information from the \( n \) processes to estimate the common \( p(t) \). In other words, for any
individual process there would be little value in factoring a \( \delta \) from the intensity function. However, for some phenomena that we might sample (see the application in section 4) the behaviour might be such that from sample to sample

i) the absolute intensity (i.e. expected rate of occurrence) varies but

ii) the relative intensity (i.e. ratio of expected rate of occurrence at two different time points) does not.

In such a case the form for \( \lambda_1(t) \) in (1) is appropriate. In our application we envision considerable variation in the \( \delta_1 \). For some sampled processes, occurrences will be very sparse while for others they will be quite numerous. But again our interest is not in the \( \delta_1 \) but in \( p(t) \).

Suppose then that we observe the \( n \) processes during the interval [0,\( T \)]. How may we estimate \( p(t) \)? (The reader may be curious as to whether differing lengths of observation time, for differing sampled processes can be accommodated. We will return to this question shortly.)
3. **Estimation Approaches**

In attempting to estimate \( p(t; \alpha) \) using data in \([0,T]\) we consider two approaches.

A. Use the actual arrival times of the occurrences

B. Partition \([0,T]\) into intervals and record numbers of arrivals in each interval.

Clearly the latter approach results in some loss of information compared with the former. We shall obtain a measure of this loss and apply it to our \( \rho \)'s of interest. In applications, it may be the case that data is gathered categorically for reasons of convenience or reliability so that only approach B is available. Hence an indication of the loss of information while still of interest may no longer be relevant.

Consider Approach A. From Cox & Hinkley (p.15-16) if \( X(t) \sim P_0(\delta \rho(t)) \) and in the interval \([0,T]\) (ordered) occurrences are observed at \( w_1, w_2, \ldots, w_m \) then the unconditional likelihood becomes

\[
L(\rho, \delta, m, w_1, w_2, \ldots, w_m) = \delta^m \prod_{i=1}^{m} \rho(w_i) e^{-\delta \int_{0}^{T} \rho(u) \, du}.
\]

(2)

The conditional likelihood given \( X(t) = m \) thus becomes
\[ L(\rho, w_1, w_2, w_3, \ldots, w_m | X(T) = m) \]
\[ = m! \delta^m \prod_{i=1}^{m} \rho(w_i) e^{-\delta \int_0^T \rho(u) du} \]
\[ = \delta^m \left[ \int_0^T \rho(u) du \right]^m e^{-\delta \int_0^T \rho(u) du / m} \]

(3)
\[ = m! \prod_{i=1}^{m} \rho(w_i) (p(T))^{-m}, \]
free of \( \delta \).

For \( n \) samples with occurrences time \( w_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i \) (that is \( X_i(T) = m_i \), \( \Sigma m_i = m \)), we similarly obtain

\[ L(\rho, \delta_i, m_i, w_{ij}) \]
\[ = \left[ \prod_{i=1}^{n} \delta_i^{m_i} \right] \prod_{i=1}^{n} \prod_{j=1}^{m_i} \rho(w_{ij}) e^{-(\Sigma \delta_i) p(T)} \]

(4)
and

\[ L(\rho, w_{ij} | X_i(T) = m_i, i = 1, \ldots, n) \]
\[ = \prod_{i=1}^{n} m_i! \prod_{i=1}^{n} \prod_{j=1}^{m_i} \rho(w_{ij}) (p(T))^{-m_i}. \]

(5)
(Note that if sample \( i \) is observed over \([0, T_i]\) the only change to be made in (5) is that \( (p(T))^{-m} \) is replaced by \( \prod_{i=1}^{n} (p(T_i))^{-m_i} \).)

Suppose \( \rho \) is of the form \( \rho(t; \alpha) \) (\( p \) of the form \( p(t; \alpha) \)) so that with estimation of \( \alpha \), \( \rho \) is specified and thus \( p \) is
Then, in (2) the unconditional likelihood equations for \( \alpha \) and \( \delta \) are

\[
\frac{\partial \log L}{\partial \alpha} = \sum \frac{\partial \log \rho(w_i; \alpha)}{\partial \alpha} - \delta \frac{\partial p(T; \alpha)}{\partial \alpha} = 0
\]

\[
\frac{\partial \log L}{\partial \delta} = \frac{m}{\delta} - p(T; \alpha) = 0
\]

whence

\[
(6) \quad \frac{\partial \log L}{\partial \alpha} = \sum \frac{\partial \log \rho(w_i; \alpha)}{\partial \alpha} - m \frac{\partial p(T; \alpha)}{\partial \alpha} / p(T; \alpha) = 0.
\]

In (3) this conditional likelihood equation for \( \alpha \) is

\[
(7) \quad \frac{\partial \log L/m}{\partial \alpha} = \sum \frac{\partial \log \rho(w_i; \alpha)}{\partial \alpha} - m \frac{\partial \log p(T; \alpha)}{\partial \alpha} = 0.
\]

Hence as is well known (e.g. Cox and Lewis, Chap 3) the conditional likelihood equation for \( \alpha \) is the same as the unconditional one.

We may rewrite (7) as

\[
(8) \quad \sum \frac{\partial \log (\rho(w_i; \alpha)/p(T; \alpha))}{\partial \alpha} = 0.
\]

In (8) the conditional nature of the likelihood equation is most apparent i.e. \( \rho(u; \alpha)/p(T; \alpha) \) is the density function for an occurrence given it occurred by time \( T \). Thus if \( \hat{\alpha} \) is a solution (8) and the usual regularity conditions are satisfied, \( \hat{\alpha} \) will be asymptotically normal with mean the true \( \alpha \) and asymptotic variance \((I(\alpha))^{-1}\) where

\[
I(\alpha) = \sum \left( \frac{\partial \log \rho(w_i; \alpha)}{\partial \alpha} \right)^2 / p(T; \alpha) \sum \frac{\partial \log p(T; \alpha)}{\partial \alpha} / p(T; \alpha)
\]
\[
I(\alpha) = E \left[ \frac{\partial}{\partial \alpha} \log(p(W;\alpha)/p(T;\alpha)) \right]^2
\]

(9) \[
= E \left[ \frac{\partial}{\partial \alpha} \log(p(W;\alpha)) \right]^2 - \left( \frac{\partial}{\partial \alpha} \log(p(T;\alpha)) \right)^2.
\]

Computationally a more convenient form may be

\[
I(\alpha) \equiv -E \left[ \frac{\partial^2}{\partial \alpha^2} \log(p(W;\alpha)/p(T;\alpha)) \right]
\]

(10) \[
= -E \frac{\partial^2}{\partial \alpha^2} \log(p(W;\alpha)) + \frac{\partial^2}{\partial \alpha^2} \log(p(T;\alpha))
\]

For the case of \( n \) samples, \( \alpha \) is a solution to

\[
\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\partial}{\partial \alpha} \log(p(W_{ij};\alpha)) = m \frac{\partial}{\partial \alpha} \log(p(T;\alpha))
\]

(11)

from which we see that the separation of the occurrence times by samples is superfluous. The total of \( m \) occurrence times may be considered as a single sample. This may be noted directly by considering the occurrence times of the process \( Y(t) = \sum_{i=1}^{n} X_i(t) \) in \([0,T]\).

Returning to (7), consider \( \rho_1(t;\alpha) = ae^{-\alpha t} \). We obtain

\[
\sum \left( \frac{1}{\alpha - w_i} \right) - \frac{m T e^{-\alpha T}}{1 - e^{-\alpha T}} = 0
\]

or

\[
\sum w_i = \frac{m}{\alpha} - \frac{m T e^{-\alpha T}}{1 - e^{-\alpha T}}
\]

and
(12) \[ I(\alpha) = \frac{1}{\alpha^2} - \frac{T^2 e^{-\alpha T}}{(1-e^{-\alpha T})^2}. \]

(Again see Cox and Lewis, p.47).

For \( \rho_2(t;\alpha) = \alpha/(\alpha+t)^2 \) we obtain

\[
\sum \left[ \frac{1}{\alpha} - \frac{2}{\alpha+w_i} \right] + \frac{m}{\alpha+T} = 0
\]

or

\[
\sum \left[ \frac{1}{\alpha+w_i} \right] \frac{m(2\alpha+T)}{2\alpha(\alpha+T)}
\]

and

(13) \[ I(\alpha) = \frac{T^2}{3\alpha^2(\alpha+T)^2}. \]

We now consider approach B. Suppose we define \( k \) intervals by \( 0 < t_1 < t_2 < \ldots < t_k = T \). Then for one sample

\[ X(t_1), X(t_2) - X(t_1), \ldots, X(t_k-1) - X(t_{k-2})|X(T) = m \]

\[ \sim \text{Multinomial } (m, q_1, q_2, \ldots, q_k) \]

where

\[ q_j = \frac{p(t_j) - p(t_{j-1})}{p(T)}, \quad j = 1, \ldots, k \]

\[ (t_0 \equiv 0) \]

For \( n \) samples let

\[ Z_j = \sum_{i=1}^{n} (X_i(t_j) - X_i(t_{j-1})) \]
Then \( Z = (Z_1, Z_2, \ldots, Z_{k-1}) \) is complete and sufficient given
\[ X_i(T_i) = m_i \] and in fact is multinomially distributed as well i.e.
\[ (\sum m_i = m) \]
\[ Z \sim \text{Multinomial } (m, q_1, \ldots, q_k) \]

Two natural estimation approaches arise at this point: the method of minimum chi square and the method of maximum likelihood (method of moments estimation might be considered as well). The former leads to consideration of the function
\[
\sum_{j=1}^{k} \frac{(Z_j - mq_j)^2}{mq_j}
\]
while the latter leads to the (conditional) likelihood function
\[
L(\rho, Z_1, \ldots, Z_k/m)
\]
\[
= m! \prod_{j=1}^{k} q_j^{Z_j/Z_j!} \]
\[ \{Z_k \equiv m - \sum_{i} Z_i\} \]

Again presuming \( \rho \) is of the form \( \rho(t; \alpha) \), the resulting likelihood equation for \( \alpha \) is
\[
\frac{\partial \log L/m}{\partial \alpha} = \sum \frac{\partial \log q_j(\alpha)}{\partial \alpha} = 0
\]
or
\[
(14) \sum Z_j \frac{\partial \log (p(t_1; \alpha) - p(t_{j-1}; \alpha))}{\partial \alpha} - \frac{m \partial \log p(T; \alpha)}{\partial \alpha} = 0
\]

Again if \( \delta_G \) (G for grouped) is a solution to (14), under the usual regularity conditions \( \delta_G \) will be asymptotically normal with mean
the true $\alpha$ and asymptotic variance $(I_G(\alpha))^{-1}$ where

$$I_G(\alpha) = -\frac{1}{m} E \left[ \sum_j z_j \frac{\partial^2 \log q_j(\alpha)}{\partial \alpha^2} \right]$$

(15)

$$= -\sum_j q_j(\alpha) \frac{\partial^2 \log q_j(\alpha)}{\partial \alpha^2}$$

or in terms of the $p_j(\alpha)$

$$p(T;\alpha))^{-1} \sum_j (p(t_j;\alpha) - p(t_{j-1};\alpha))^{-1} \left[ \frac{\partial (p(t_j;\alpha) - p(t_{j-1};\alpha))}{\partial \alpha} \right]^2 \left[ \frac{\partial \log p(T;\alpha)}{\partial \alpha} \right]^2$$

(16)

The similarity between expressions (7) and (14) and between (9) and (16) is not unexpected. In fact, if as $k \rightarrow \infty$, $\sup_{1 \leq j \leq k} (t_j - t_{j-1}) \rightarrow 0$ it is straightforward to show that the "categorical" expressions converge to the corresponding "continuous" ones.

In (14), consider $\rho_1(t;\alpha) = e^{\alpha t}$. Since

$$q_j(\alpha) = \frac{e^{-\alpha t_j} - e^{-\alpha t_{j-1}}}{1 - e^{-\alpha T}}$$

we obtain

$$-\sum_j z_j \frac{t_{j-1} e^{-\alpha t_j} - t_j e^{-\alpha t_{j-1}}}{e^{-\alpha t_j} - e^{-\alpha t_{j-1}}} - \frac{mT e^{-\alpha T}}{1 - e^{-\alpha T}} = 0$$

(17)

and
\[ I_G(\alpha) = \sum \frac{e^{-\alpha(t_{j-1}+t_j)}}{(1-e^{-\alpha T})(e^{-\alpha t_{j-1}}-e^{-\alpha t_j})} - \frac{T^2 e^{-\alpha T}}{(1-e^{-\alpha T})^2}. \]

Expressions (17) and (18) can be simplified considerably if we use equally spaced intervals. Let \( \bar{t} = T/k \) whence \( t_j = j \bar{t} \). Then (17) becomes

\[ \sum jz_j = m(1-e^{-\alpha \bar{t}})^{-1} - m k e^{-\alpha T} (1-e^{-\alpha T})^{-1} \]

and (18) becomes

\[ t^2 e^{-\alpha \bar{t}} (1-e^{-\alpha \bar{t}})^{-2} - T^2 e^{-\alpha T} (1-e^{-\alpha T})^{-2}. \]

In (14) consider \( \rho_2(t; \alpha) = \alpha/(\alpha+t)^2 \). Now

\[ q_j(\alpha) = \frac{\alpha (\alpha+T)}{T} \frac{(t_j-t_{j-1})}{(\alpha+t_j)(\alpha+t_{j-1})} \]

and we obtain

\[ \sum \frac{z_j(t_j^2-t_{j-1}^2-\alpha^2)}{\alpha(\alpha+t_j)(\alpha+t_{j-1})} + \frac{m}{\alpha+T} = 0 \]

and

\[ I_G(\alpha) = \frac{\alpha+T}{T} \sum \left[ \frac{t_{j-1}}{\alpha+t_{j-1}} - \frac{t_j}{\alpha+t_j} \right]^{-1} \left[ \frac{t_j}{(\alpha+t_j)^2} - \frac{t_{j-1}}{(\alpha+t_{j-1})^2} \right] - \frac{1}{(\alpha+T)^2}. \]

Again, equally spaced intervals simplify matter Expression (21) becomes

\[ \sum \frac{(2\alpha+2j-1)\bar{t}}{(\alpha+j\bar{t})(\alpha+(j-1)\bar{t})} = \frac{m(2\alpha+T)}{\alpha(\alpha+T)} \]

and (22) becomes

\[ \frac{(\alpha+t)}{T} \frac{\bar{t}}{\alpha} \sum \frac{(\alpha^2-j(j-1)\bar{t}^2)^2}{(\alpha+j\bar{t})^3(\alpha+(j-1)\bar{t})^3} - \frac{1}{(\alpha+T)^2}. \]
We now attempt comparison of the estimation approaches. Again \( \hat{\theta} \) is the MLE from (7) while \( \hat{\theta}_G \) is the MLE from (14).

We employ ideas similar to those of Lindley. Given \( \hat{\theta}_G \) how may we "correct" to \( \hat{\theta} \). We obtain an approximation for \( \Delta \equiv \hat{\theta}_G - \hat{\theta} \). Lindley considers the reverse situation; from the functional form for the ungrouped MLE he obtains a correction to the grouped MLE.

In our applications the grouped MLE is all that is available. Lindley also considers an i.i.d. sample while our \( W_j \) are clearly not so.

We also seek to estimate the loss of efficiency due to grouping. Since \( \hat{\theta} \) is most efficient we calculate the (asymptotic) relative efficiency of \( \hat{\theta}_G \) to \( \hat{\theta} \) by \( I(\alpha)/I_G(\alpha) \). In fact the proportion of information lost is \( (I(\alpha) - I_G(\alpha))/I(\alpha) \).

To approximate \( \Delta \) we assume equally spaced intervals. Using Taylor's theorem

\[
p(j+b\alpha) - p((j-1)+b\alpha) \approx b \cdot \rho((j-1)+b\alpha) + \frac{b^2}{24} \cdot \rho''((j-1)+b\alpha)
\]

whence

\[
\frac{\partial \log(p(j+b\alpha) - p((j-1)+b\alpha))}{\partial \alpha} \approx \frac{\partial \log(\rho((j-1)+b\alpha))}{\partial \alpha} + \frac{b^2}{24} \cdot \frac{\partial}{\partial \alpha} \left\{ \frac{\rho''((j-1)+b\alpha)}{\rho((j-1)+b\alpha)} \right\}
\]

Hence using (14) and suppressing the arguments of \( \rho \) and \( p \) we have

\[
\sum z_j \frac{\partial \log \rho}{\partial \alpha} \bigg|_{\hat{\theta}_G} + \frac{b^2}{24} \sum z_j \frac{\partial}{\partial \alpha} \left( \frac{\rho''}{\rho} \right) \bigg|_{\hat{\theta}_G} \approx 0.
\]
In (7), presume that if \( w_j \in ((j-1)t, jt) \) then
\[
\rho(w_j; \alpha) \approx \rho((j-\frac{1}{2})t; \alpha)
\]
so that
\[
\sum z_j \frac{\partial \log \rho}{\partial \alpha} \bigg|_{\delta} - m \frac{\partial \log p}{\partial \alpha} \bigg|_{\delta} \approx 0.
\]

To a first approximation, then
\[
\Delta = -\frac{t^2}{24} \sum z_j \frac{\partial}{\partial \alpha} \left\{ \frac{\rho''}{\rho} \right\} \bigg|_{\delta_G} - \frac{m}{\delta^2} \frac{\partial \log p}{\partial \alpha} \bigg|_{\delta_G}
\]
\[
(25)
\]
Iteration can be employed to better the approximation. For
\[\rho_1(t; \alpha) = ae^{-at}\]
(25) becomes
\[
\frac{t^2}{12} \frac{\delta_G}{\delta_G^2} \left[ \frac{1}{\delta_G^2} - \frac{t^2 e^{-\delta_G T}}{\left(1-e^{-\delta_G T}\right)^2} \right]^{-1}
\]
For \[\rho_2(t; \alpha) = \alpha/\left(\alpha t\right)^2\]
, (25) becomes
\[
\frac{t^2}{2} \sum z_j b_j^{-3} \left[ \frac{m ((\hat{\alpha}_G + T)^2 + \hat{\alpha}_G^2)}{\delta_G^2 (\hat{\alpha}_G + T)^2} - 2 \sum z_j b_j^{-2} \right]^{-1}
\]
where \[b_j = \delta_G + (j-\frac{1}{2})t\].

15
For the relative efficiency of $\theta_G$ to $\theta$, from (9) and (16) we obtain

$$\text{Eff}(\theta_G|\theta) = \frac{\mathbb{E} \left[ \frac{\partial \log p(w;\alpha)}{\partial \alpha} \right]^2 - \left( \frac{\partial \log p(T;\alpha)}{\partial \alpha} \right)^2}{(p(T;\alpha))^{-1} \sum (p(t_j;\alpha) - p(t_{j-1};\alpha))^{-1} \left[ \frac{\partial}{\partial \alpha} (p(t_j;\alpha) - p(t_{j-1};\alpha)) \right]^2 \left( \frac{\partial \log p(T;\alpha)}{\partial \alpha} \right)^2}$$

Let us consider our illustrative $\rho_1, \rho_2$ using equally spaced intervals. For $\rho_1$, we have

$$(26) \quad \text{Eff}(\theta_G|\theta) = \frac{\frac{1}{\alpha} - \frac{T^2 e^{-\alpha T}}{(1-e^{-\alpha T})^2}}{\frac{-t^2 e^{-\alpha t}}{(1-e^{-\alpha t})^2} - \frac{T^2 e^{-\alpha T}}{(1-e^{-\alpha T})^2}}.$$

For fixed $T$, the following limits may be easily discerned in (26).

i) $\overline{t}$ i.e. $k$ fixed, $\alpha \to 0$, Eff $\to k^2/(k^2-1)$

ii) $\overline{t}$ i.e. $k$ fixed, $\alpha \to \infty$, Eff $\to \infty$

iii) $\alpha$ fixed, $\overline{t} \to 0$ i.e. $k \to \infty$, Eff $\to 1$

iv) $\alpha$ fixed, $\overline{t} \to T$ i.e. $k \to 1$, Eff $\to \infty$

Letting $s = \alpha T$, with $k \overline{t} = T$ (26) becomes

$$(27) \quad \frac{k^2 \frac{g(s/k)}{g(s) - s^2}}{S^2(g(s) - k^2 g(s/k))}$$

where

$$g(s) = (e^s-1)(1-e^{-s}) = \sum_{j=1}^{\infty} \frac{2^j s^j}{(2j)!}.$$
Table 1 evaluates (27) over a range of $s$ and $k$ values. For $\rho_2$

\begin{equation}
\text{Eff}(\hat{\alpha}_G | \alpha) = \frac{T^2}{3\alpha^2(\alpha+T)^2} \cdot \frac{\bar{t}(\alpha+t)}{\alpha T} \frac{(\alpha^2 - j(j-1)t^2)}{(\alpha+jt)^3(\alpha+(j-1)t)^3} \frac{1}{(\alpha+T)^2} \Sigma (\alpha^2 - j(j-1)t^2)
\end{equation}

Letting $s = \alpha/T$ with $kt = T$ (28) becomes

\begin{equation}
\left\{ 3sk(s+1) \Sigma \frac{(s^2k - j(j-1))^2}{(sk+j)^3(sk+(j-1))^3} - 3s^2 \right\}^{-1}
\end{equation}

Table 2 evaluates (29) over a range of $s$ and $k$ values. The limiting behavior noted above for (26) also occurs for (28) with "$\alpha = 1/\alpha".\)

Regardless of what parametric form was adopted for $\rho$ and of what estimation approach was used the following offers a simple way of assessing fit. For any two time points $t_1, t_2, t_1 < t_2$, $X(t_1) | X(t_2) \sim B_1(X(t_2), p(t_1)/p(t_2))$. Hence (if $X(t_2) > 0$) $X(t_1)/X(t_2)$ is the UMVU of $p(t_1)/p(t_2)$. For any estimate $\hat{p}(t) \equiv p(t, \theta)$ we can compare $p(t_1; \theta)/(p(t_2; \theta)$ with the UMVU ratio.

17
4. **An Application**

One attractive application of the above methodology is to the matter of describing the active lifetime of a court case. More specifically, associated with court opinions (particularly higher court opinions such as those of the U.S. appellate or supreme courts) is the issue of precedential value. How much effect will an opinion have on later opinions? One way to study this point is to record citations of this opinion in later cases. It is thus of interest to examine the pattern of subsequent citation of a case. It seems reasonable to assume that subsequent citations arrive according to a Poisson process. Moreover, although the intensity function will vary greatly from case to case the expected proportion of citations to a given time need not. (For the data to be presented, some preliminary empirical work using fixed points \( t_1, t_2 \) and comparing binomials has supported this premise.) That is, sparsely cited cases will receive about the same relative proportion of total cites by a fixed time as will heavily cited cases. Our assumptions regarding \( p(t) \) made in section 2 seem appropriate in this context since an opinion is a bit like wine. At first it improves with age (i.e. is more often cited) but then it deteriorates (i.e. is less often cited as newer opinion diminish its value). Estimation approach B was forced in this situation. Data collection recorded the date an opinion was handed down along with the dates of subsequent citation. It was decided more convenient and reliable to categorize subsequent citations into intervals after decision date than to convert to the exact number of days after the decision date.
The court cases studied in this application are all from the U.S. Court of Appeals. They are drawn from a larger study (see Schuchman and Gelfand) of the Federal appellate courts. Two independent samples were analyzed. One consists of 100 cases from the Fifth circuit court of appeals while the other consists of 100 cases from the remaining ten circuits. In each sampled case the date of decision was recorded (all dates are between 1960 and 1962). By use of Shepard's Index we may find all later cases citing any sampled case. By examining these citing opinions, we may obtain the total number of citations for any sampled case and a categorical time from date of decision to each such citation. Subsequent citation of each case in each sample was considered through December 1976 thus achieving a minimum of 13 full years of observation with numerous cases reaching to nearly 16 full years.

For convenience we set \( T = 12 \) and consider \( \bar{t} = 2, k = 6 \).

The resulting \( Z \) vectors are given in Table 3. Note that the fifth circuit opinions are more frequently cited than their other circuit counterparts. Furthermore, we find an inexplicable, but common, bump in both samples occurring at \( Z_4 \), i.e. for the period between 6 and 8 years after date of decision. A test of homogeneity between the samples yields \( \chi^2 = 1.290 \) with descriptive level = .93.

Table 4 presents the estimation of \( \alpha \) using both \( p_1(t;\alpha) \) and \( p_2(t;\alpha) \). We see that
(i) Both functions $p_1$ and $p_2$ fit the data well

(ii) Estimation by MLE and MME$^2$ yield very similar estimates of $\alpha$

(iii) The estimates of $\alpha$ for the two samples are fairly close
     (as expected from the test of homogeneity)

(iv) For the 5th circuit data using $p_1$ and $\delta_G \approx 0.09, \Delta = 0.0027$
     while for $p_2$ and $\delta_G \approx 18, \Delta = 1.038.$

(v) From Tables 1 and 2 the relative efficiency for $p_1$ under the given
     grouping is 1.03, for $p_2,$ 1.04.

In Table 5 we compare the fits of $p_1$ and $p_2$ with the "best" (i.e. UMVU) estimate as described at the end of the previous section. We look at $t_i$ relative to $t_{i+1}, t_i = \bar{t}.$ Again the fits look quite good i.e. the maximum relative error for the exponential model is .03, for the bilinear model .04.

How can it be that two such different functions $p_1$ and $p_2$
fit the data almost identically? Table 6 clarifies the matter.
It also provides representative values of the function over time using $\alpha = 0.09$ in $p_1$ and $\alpha = 18$ in $p_2$. We see that during the time interval $t = 2$ through 12 years $p_1(t)/p_2(t)$ is virtually constant. Since the assessment of fit involves $p(t)$ in a relative way, over the 12 years of observation $p_1$ and $p_2$
will be virtually indistinguishable. However in an absolute sense $p_1$ is preferred to $p_2.$ This follows by asking how many years
we expect to wait for a specified proportion of subsequent citations
e.g. what is $t$ such that $p_1(t) = .9,$ such that $p_2(t) = .9$?

20
For $p_1$ the answer is 25.6 years, for $p_2$, 162 years. Additionally $p_1(33.3) = .95$ and $p_1(51.1) = .99$. These proportions are reasonable both empirically and intuitively. Thus an exponential intensity function seems an effective descriptor for subsequent citation of an opinion.
### Table 1: Relative Efficiency for \( p_1(t;\alpha) = 1 - e^{-\alpha t} \)

<table>
<thead>
<tr>
<th>k</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>1.333</td>
<td>1.125</td>
<td>1.067</td>
<td>1.029</td>
<td>1.010</td>
<td>1.003</td>
<td>1</td>
</tr>
<tr>
<td>.5</td>
<td>1.337</td>
<td>1.127</td>
<td>1.068</td>
<td>1.029</td>
<td>1.010</td>
<td>1.003</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1.350</td>
<td>1.131</td>
<td>1.070</td>
<td>1.030</td>
<td>1.011</td>
<td>1.003</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1.893</td>
<td>1.323</td>
<td>1.170</td>
<td>1.072</td>
<td>1.025</td>
<td>1.006</td>
<td>1.001</td>
</tr>
<tr>
<td>10</td>
<td>5.990</td>
<td>2.361</td>
<td>1.647</td>
<td>1.255</td>
<td>1.087</td>
<td>1.021</td>
<td>1.003</td>
</tr>
<tr>
<td>50</td>
<td>&gt;10^8</td>
<td>&gt;10^8</td>
<td>&gt;10^8</td>
<td>59.879</td>
<td>5.857</td>
<td>1.642</td>
<td>1.086</td>
</tr>
</tbody>
</table>

### Table 2: Relative Efficiency for \( p_2(t;\alpha) = t/(\alpha + t) \)

<table>
<thead>
<tr>
<th>k</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01</td>
<td>34.337</td>
<td>11.896</td>
<td>7.412</td>
<td>4.428</td>
<td>2.480</td>
<td>1.445</td>
</tr>
<tr>
<td>.05</td>
<td>7.683</td>
<td>3.034</td>
<td>2.119</td>
<td>1.529</td>
<td>1.178</td>
<td>1.035</td>
</tr>
<tr>
<td>.1</td>
<td>4.364</td>
<td>1.951</td>
<td>1.491</td>
<td>1.210</td>
<td>1.061</td>
<td>1.010</td>
</tr>
<tr>
<td>.5</td>
<td>1.778</td>
<td>1.170</td>
<td>1.075</td>
<td>1.027</td>
<td>1.007</td>
<td>1.001</td>
</tr>
<tr>
<td>1</td>
<td>1.500</td>
<td>1.102</td>
<td>1.044</td>
<td>1.016</td>
<td>1.004</td>
<td>1.001</td>
</tr>
<tr>
<td>5</td>
<td>1.344</td>
<td>1.067</td>
<td>1.303</td>
<td>1.010</td>
<td>1.003</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1.336</td>
<td>1.067</td>
<td>1.029</td>
<td>1.010</td>
<td>1.003</td>
<td>1</td>
</tr>
<tr>
<td>Sample</td>
<td>$Z_1$</td>
<td>$Z_2$</td>
<td>$Z_3$</td>
<td>$Z_4$</td>
<td>$Z_5$</td>
<td>$Z_6$</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Fifth Circuit</td>
<td>225</td>
<td>191</td>
<td>145</td>
<td>152</td>
<td>112</td>
<td>85</td>
</tr>
<tr>
<td>Other Circuits</td>
<td>182</td>
<td>144</td>
<td>114</td>
<td>123</td>
<td>84</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 3: The Fifth Circuit and Other Circuit Data
<table>
<thead>
<tr>
<th>$\hat{\alpha}(\text{MLE})$</th>
<th>$\theta_G(\text{MLE}^2)$</th>
<th>$\hat{\Sigma}_G(\text{MLE}^2)$</th>
<th>$\hat{\Sigma}_G(\text{D.L.})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0893</td>
<td>.0888</td>
<td>4.679 (.32)</td>
<td>4.480 (.35)</td>
</tr>
<tr>
<td>.0831</td>
<td>.0827</td>
<td>4.681 (.22)</td>
<td>4.386 (.36)</td>
</tr>
</tbody>
</table>

Table 4: Estimation of $\hat{\alpha}$.
<table>
<thead>
<tr>
<th>Fifth Circuit</th>
<th>( t_1/t_2 )</th>
<th>( t_2/t_3 )</th>
<th>( t_3/t_4 )</th>
<th>( t_4/t_5 )</th>
<th>( t_5/t_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Best&quot; (UMVU)</td>
<td>1.849</td>
<td>1.349</td>
<td>1.271</td>
<td>1.157</td>
<td>1.103</td>
</tr>
<tr>
<td>( \hat{\alpha} (MM^2) )</td>
<td>1.837</td>
<td>1.382</td>
<td>1.231</td>
<td>1.157</td>
<td>1.114</td>
</tr>
<tr>
<td>( \hat{\alpha} (MLE) )</td>
<td>1.836</td>
<td>1.381</td>
<td>1.231</td>
<td>1.157</td>
<td>1.113</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>( \hat{\alpha} (MM^2) )</td>
<td>1.815</td>
<td>1.373</td>
<td>1.229</td>
<td>1.160</td>
</tr>
<tr>
<td>( \hat{\alpha} (MLE) )</td>
<td>1.814</td>
<td>1.372</td>
<td>1.229</td>
<td>1.160</td>
<td>1.119</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>( \hat{\alpha} (MM^2) )</td>
<td>1.791</td>
<td>1.350</td>
<td>1.280</td>
<td>1.149</td>
</tr>
<tr>
<td>( \hat{\alpha} (MLE) )</td>
<td>1.849</td>
<td>1.389</td>
<td>1.237</td>
<td>1.163</td>
<td>1.118</td>
</tr>
<tr>
<td>( p_1 )</td>
<td>( \hat{\alpha} (MM^2) )</td>
<td>1.847</td>
<td>1.388</td>
<td>1.237</td>
<td>1.162</td>
</tr>
<tr>
<td>( \hat{\alpha} (MLE) )</td>
<td>1.825</td>
<td>1.380</td>
<td>1.234</td>
<td>1.164</td>
<td>1.122</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>( \hat{\alpha} (MM^2) )</td>
<td>1.824</td>
<td>1.379</td>
<td>1.234</td>
<td>1.163</td>
</tr>
</tbody>
</table>

Table 5: Examining the Exponential & Bilinear Fits.
<table>
<thead>
<tr>
<th>$t$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1(t; .09)$</td>
<td>.165</td>
<td>.302</td>
<td>.417</td>
<td>.513</td>
<td>.593</td>
<td>.660</td>
<td>.741</td>
<td>.835</td>
<td>.895</td>
<td>.933</td>
</tr>
<tr>
<td>= $1 - e^{-0.09t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_2(t; 18)$</td>
<td>.1</td>
<td>.182</td>
<td>.25</td>
<td>.308</td>
<td>.357</td>
<td>.40</td>
<td>.455</td>
<td>.526</td>
<td>.581</td>
<td>.625</td>
</tr>
<tr>
<td>= $\frac{t}{18+t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_1/p_2$</td>
<td>.165</td>
<td>1.66</td>
<td>1.67</td>
<td>1.67</td>
<td>1.66</td>
<td>1.65</td>
<td>1.63</td>
<td>1.58</td>
<td>1.54</td>
<td>1.49</td>
</tr>
</tbody>
</table>

Table 6: Comparing $p_1$ and $p_2$. 

26
References


AN ESTIMATION PROBLEM WITH POISSON PROCESSES

ALAN E. GELFAND

Department of Statistics
Stanford University
Stanford, CA 94305

Office Of Naval Research
Statistics & Probability Program Code 411SP
Arlington, VA 22217

Poisson Processes, Intensity Function, Maximum Likelihood Estimation, "Correction" Factor, Relative Efficiency.

PLEASE SEE REVERSE SIDE.
#317

AN ESTIMATION PROBLEM WITH POISSON PROCESSES

This paper considers an estimation problem involving \( n \) independent Poisson processes such that the \( i \)-th process has intensity function \( \lambda_i(t) = \delta \rho(t; \alpha) \). It is of interest to estimate \( \rho(t; \alpha) \). Two estimation procedures are developed, one using the exact arrival times of observations, the second using categorical arrival times of observations. Two specific instances of \( \rho(t) \), an exponential and a bilinear form are investigated further. An example applying the methodology to the active life of a judicial opinion is described.