The Annals of Statistics
1979, Vol. 7, No. 5, 1040–1051

ASYMPTOTIC OPTIMAL SEQUENTIAL ESTIMATION: THE POISSON CASE[1]

By Y. Vardi

Bell Laboratories
Murray Hill, New Jersey

The problems of estimating sequentially the intensity parameter of a homogeneous Poisson process and of estimating sequentially the mean of a sequence of i.i.d. Poisson rv's, are considered. The procedures suggested are shown to perform well for large values of the parameter and/or for small sampling cost. Having bounded regret, the procedure for estimating the mean of the Poisson sequence is asymptotically Bayes w.r.t. any sequence of a priori densities, which spread mass in a suitably smooth manner.

1. Introduction and results. We consider the problem of estimating the unknown intensity parameter, $\lambda(>0)$, of a homogeneous Poisson process (HPP) with right continuous sample paths $\{X(t);\ t>0\}$, defined on a probability space $(\Omega, \mathfrak{F}, P)$. Having observed the process for the time period $[0, t]$ we estimate λ by $X(t)/t$, so the problem is, simply, that of choosing a stopping time (ST). The loss plus cost incurred by stopping at time $t(>0)$ is assumed to be

$$L_t = A(X(t)/t - \lambda)^2 + t, \tag{1.1}$$

where $A(>0)$ is known. With nomenclature as in Vardi (1978) the following is easily seen. The best fixed stopping time (BFST) is $(A\lambda)^{\frac{1}{2}}$, the 'risk' associated with it is $(2A\lambda)^{\frac{1}{2}}$, and no fixed stopping time (FST) can have bounded regret. Denoting the regret of a family of stopping times (ST's) $\{\nu_A\}_{A>0}$, say, by $\rho_{\nu_A}(A, \lambda) = E_\lambda L_{\nu_A} - 2(A\lambda)^{\frac{1}{2}}$ we measure (following Robbins (1959)) the performance of the family in terms of its regret.

For the ST's $\{\xi_A\}_{A>0}$ that are of interest to us (defined in (2.1) and (2.2)), it is impossible to absorb A into the parameter and thereby reduce the analysis of their regret to a one dimensional analysis (as done, for example, in Starr and Woodroofe (1972)). Here, we must analyze $\rho_{\xi_A}(A, \lambda)$ in the (A, λ)-plane. Facing this situation, natural questions of interest are the behaviour of $\rho_{\xi_A}(A, \lambda)$ as $A\uparrow\infty(\lambda>0$, fixed) or as $\lambda\uparrow\infty(A>0$, fixed). Instead of dealing with these two separate problems, we shall investigate the behaviour of $\rho_{\xi_A}(A, \lambda)$ as $(A\lambda)\to\infty$. Note, however, that since $\rho_{\xi_A}(A, \lambda)$ is not necessarily a function of $(A\lambda)$ alone, attention should be paid to the path through which $(A\lambda)\to\infty$.

In Section 4 we prove the following property of the ST's $\{\xi_A\}_{A>0}$.

Received May 1977; revised July 1978.

[1]This research was partially done at Cornell University and supported by NSF Grant No. MCS 75-22481.

AMS 1970 *subject classifications.* Primary 62L12; secondary 62C10, 62M05.

Key words and phrases. Sequential estimation, homogeneous Poisson process, Poisson distribution, asymptotically Bayes.

THEOREM 1.1. *For every* $\varepsilon > 0$, *as* $(A\lambda) \to \infty$ *on* $(A > \varepsilon) \cap (\lambda > \varepsilon)$ *we have,*

$$\lambda \rho_{\xi_A}(A, \lambda) \leqslant 0(1).$$

COROLLARY 1.1.
(i) $\rho_{\xi_A}(A, \lambda) \leqslant 0(1)$ *as* $A \uparrow \infty$ *for fixed* $\lambda (> 0)$, *and*
(ii) $\rho_{\xi_A}(A, \lambda) \leqslant 0(\lambda^{-1})$ *as* $\lambda \uparrow \infty$ *for fixed* $A (> 0)$.

In Section 5 we extend our discussion to the problem of estimating the mean $\lambda(> 0)$ of a sequence of i.i.d. Poisson rv's. Treating the sample size n as a continuous variable, we see that the BFST and the 'risk' associated with it are the same as in the continuous time Poisson process. We then introduce a family of ST's $\{\eta_A\}_{A>0}$, for which we prove the following (again, the sample mean is taken as an estimator for λ).

THEOREM 1.2. *For every* $\varepsilon > 0$, *as* $(A\lambda) \to \infty$ *on* $(A > \varepsilon) \cap (\lambda > \varepsilon)$ *we have*

$$\lambda[\rho_{\eta_A}(A, \lambda) - d] \leqslant 0(1)$$

where d is any constant satisfying $d > \frac{5}{2}$.

COROLLARY 1.2.
(i) $\rho_{\eta_A}(A, \lambda) \leqslant 0(1)$ *as* $A \uparrow \infty$ *for fixed* $\lambda (> 0)$, *and*
(ii) $\rho_{\eta_A}(A, \lambda) - d \leqslant 0(\lambda^{-1})$ *as* $\lambda \uparrow \infty$ *for fixed* $A (> 0)$.

Asymptotic optimum properties of the procedure $(\eta_A, X(\eta_A)/\eta_A)$, for any fixed $A(> 0)$, follow from Vardi (1978). Specifically, the procedure is asymptotically Bayes w.r.t. any sequence of smooth a priori densities. Details (and examples of such 'smooth sequences') are given there and thus will not be discussed here.

2. The procedure. In analogy with the derivation of the BFST, $(A\lambda)^{\frac{1}{2}}$, we define a family of ST's as follows. For each $A > 0$ let

$$\xi_A = \text{first time } t \geqslant r(A) \text{ for which } X(t) \leqslant A^{-1}t^3. \tag{2.1}$$

Here $r(A)$ is a positive real valued function whose choice will be discussed below. Having observed the process until time ξ_A, we estimate λ by $X(\xi_A)/\xi_A$.

To understand the choice of $r(A)$, assume temporarily that $r(A) = r$ for some fixed positive r. Then, with the notation $E[Y; A] = \int_A Y dP$ (for $A \in \mathfrak{F}$ and Y a rv defined on $(\Omega, \mathfrak{F}, P)$), we have

$$AE(X(\xi_A)/\xi_A - \lambda)^2 \geqslant AE[(X(\xi_A)/\xi_A - \lambda)^2; \xi_A = r]$$
$$= Ar^{-2}\Sigma_{j=0}^{[A^{-1}r^3]}(j - \lambda r)^2 e^{-\lambda r}\frac{(\lambda r)^j}{j!} = \lambda^2 A e^{-\lambda r}$$

for all values of A larger than r^3, where we used $[x]$ to denote the integer part of x. Now, if the procedure is to have regret bounded in A and λ, the minimal sampling time, r, must be an increasing function of A. In what follows, we assume (unless otherwise stated) that $r(A)$ satisfies the following

$$
\begin{array}{ll}
 & \text{(i)} \quad a(\varepsilon) > 0 \ \forall \varepsilon > 0, \text{ where } a(\varepsilon) \equiv \inf_{A \geqslant \varepsilon} r(A). \\
 & \text{(ii)} \quad r(A) \text{ is bounded on bounded intervals.} \\
(2.2) & \text{(iii)} \quad r(A) \text{ is an eventually increasing function of } A. \\
 & \text{(iv)} \quad r(A) \leqslant A^{\frac{1}{3}} \text{ for all large values of } A. \\
 & \text{(v)} \quad r(A)/\log A \to \infty \text{ as } A \uparrow \infty.
\end{array}
$$

REMARK. The conditions above give a great flexibility in the choice of $r(A)$. As a practical rule one might choose $r(A) = A^p$ for some fixed $p \in (0, \frac{1}{3}]$. Guidelines for the choice of p can then be determined by a Monte Carlo study, when consideration of the 'relevant' parameter space (the parameter space, in practice, is always bounded) is taken into account.

3. Assumptions, notation and preliminaries.

(a) In order to exclude from our analysis the case where $(A\lambda) \to \infty$ but $\min(A, \lambda) \to 0$, we assume that $A \geqslant \varepsilon$ and $\lambda \geqslant \varepsilon$, where $\varepsilon > 0$ is arbitrary.

(b) The product $A\lambda$ appears very often in the computations and, for convenience, will be denoted ψ.

(c) A word about our use of $o(\cdot)$ and $O(\cdot)$ with functions of two variables is in order. For such a function $f(A, \lambda)$ and any real number k we write

$$f(A, \lambda) \leqslant O(\psi^{-k}) \text{ iff } \limsup\, (A\lambda)^k f(A, \lambda) < \infty,$$

$$f(A, \lambda) \leqslant o(\psi^{-k}) \text{ iff } \limsup\, (A\lambda)^k f(A, \lambda) \leqslant 0$$

$$f(A, \lambda) = O(\psi^{-k}) \text{ iff } |f(A, \lambda)| \leqslant O(\psi^{-k}),$$

$$f(A, \lambda) = O(\psi^{-k}) \text{ iff } |f(A, \lambda)| \leqslant o(\psi^{-k}).$$

The limsup above is taken for $(A\lambda) \to \infty$ on $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$. Also if $f(A, \lambda) \leqslant o(\psi^{-k})$ for all $k > 0$ we write $f(A, \lambda) \leqslant o(\psi^{-\infty})$. Similar notation are used, when needed, for functions of only one variable.

(d) If $f(A, \lambda)$, $g(A, \lambda)$ are two functions satisfying $f(A, \lambda) - g(A, \lambda) = o(\psi^{-\infty})$ we shall write $f(A, \lambda) \sim g(A, \lambda)$.

(e) Since the process $\{X(t);\ t \geqslant 0\}$ is continuous from the right, nondecreasing, and has unit jumps, the following holds:

(i) On $\xi_A > r(A)$, $X(\xi_A) = A^{-1}\xi_A^3$.

(ii) ξ_A assumes its values in the set $\{r(A)\} \cup \{[r(A), \infty) \cap \{r_1, r_2, \cdots\}\}$, where $r_i = (Ai)^{\frac{1}{3}}$, $i = 1, 2, \cdots$.

(iii) On $\xi_A > r(A)$, $[\xi_A > t] \Rightarrow [X(\xi_A) > A^{-1}t^3]$.

(iv) If $0 < s \leqslant t$ then, $[s \leqslant \xi_A \leqslant t] \Rightarrow [X(s) \leqslant A^{-1}t^3]$.

(f) For $t \geqslant 0$ we denote the σ-field $\sigma[X(s);\ 0 \leqslant s \leqslant t]$ by $\mathcal{F}_t$. If T is a ST, the σ-field $\mathcal{F}_T$ is the collection of all sets prior to T (A set $A \in \mathcal{F}$ is said to be prior to T iff $A \cap [T \leqslant t] \in \mathcal{F}_t$ for all $t \geqslant 0$).

(g) In the sequel, we shall use the following application of Markov's inequality: if X is a nonnegative rv and $a > 0$ is a real number, then for all $\theta > 0$ we have

$$P\{X \geqslant a\} \leqslant e^{-\theta a} E e^{\theta X}, \tag{3.1}$$
$$P\{X \leqslant a\} \leqslant e^{\theta a} E e^{-\theta X}.$$

(h) The subscript λ (as in P_λ (an event) or E_λ (a rv)) will mean "under the assumption that λ is the true value of the parameter". However, if there is no ambiguity the subscript will be omitted. We shall often use the above (mainly e(i) and g) without repeatedly citing them.

4. Analysis of the regret (large values of ψ).

LEMMA 4.1.
(i) *For every* $A > 0$ *and* $\lambda > 0$, $P_\lambda(\xi_A < \infty) = 1$.
(ii) *For* $0 < \alpha < 1$, *as* $\psi \to \infty$ *on* $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, *we have*

$$P_\lambda\big(\xi_A < \alpha\psi^{\frac{1}{2}}\big) = o(\psi^{-\infty}).$$

PROOF. Temporarily denoting $r = r(A)$, we observe that for all $\theta > 0$ we have

$$P\big(\xi_A \leqslant \alpha\psi^{\frac{1}{2}}\big) \leqslant P\big(X(r) \leqslant A^{-1}r^3\big) + \Sigma P\big(X(r_i) \leqslant A^{-1}r_i^3\big)$$
$$\leqslant \exp\{\theta A^{-1}r^3 + \lambda r(e^{-\theta} - 1)\} + \Sigma \exp\{\theta A^{-1}r_i^3 + \lambda r_i(e^{-\theta} - 1)\}$$

where the summations above (and below) are on the set $\{i;\, r \leqslant r_i \leqslant \alpha\psi^{\frac{1}{2}}\}$. It is not difficult to verify that, for all sufficiently large values of ψ and for θ satisfying $\theta < 2(1 - \alpha^2)$, $\max_x[\theta A^{-1}x^3 + \lambda x(e^{-\theta} - 1)]$ subject to $r \leqslant x \leqslant \alpha\psi^{\frac{1}{2}}$ is attained at $x = r$. Choosing $\theta = 1 - \alpha^2$, we get

$$P\big(r \leqslant \xi_A \leqslant \alpha\psi^{\frac{1}{2}}\big) \leqslant (1 + \Sigma 1)\exp\{(1 - \alpha^2)A^{-1}r^3 + \lambda r(e^{\alpha^2-1} - 1)\} \tag{4.1}$$
$$\leqslant \big(1 + A^{-1}\alpha^3\psi^{\frac{3}{2}}\big)\exp\{(1 - \alpha^2)A^{-1}r^3 + \lambda r(e^{\alpha^2-1} - 1)\}.$$

Now, from (2.2) (iv) we see that the last expression is $o(\psi^{-\infty})$ provided $\lambda r(A)(e^{\alpha^2-1} - 1) + k\log A + k\log\lambda$ approaches $-\infty$ as $(A\lambda) \to \infty$ on $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, for every positive k. However, this is easily seen to be the case upon observing that $\lambda r(A) \leqslant \frac{1}{2}\lambda a(\varepsilon) + \frac{1}{2}\varepsilon r(A)$ on $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$ and then using (2.2) (v). From this the proof of (ii) follows. The proof of (i) is an immediate consequence of the strong law of large numbers (SLLN), thus the lemma is proved.

LEMMA 4.2. *On* $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, *as* $\psi \to \infty$ *we get*
(i) $\lambda^k E_\lambda[(\xi_A - \psi^{\frac{1}{2}})/\psi^{\frac{1}{4}}]^{2k} = 0(1)$ $k = 1, 2, \cdots,$
(ii) $E_\lambda(\xi_A/\psi^{\frac{1}{2}})^k \leqslant 0(1)$ $k = 1, 2, \cdots$.

PROOF. For (i) we write

$$(4.2)\quad E\left[\frac{\xi_A-\psi^{\frac12}}{\psi^{\frac14}}\right]^{2k}=\int_0^\infty P\left[\left[\frac{\xi_A-\psi^{\frac12}}{\psi^{\frac14}}\right]^{2k}>y\right]dy$$

$$=2k\int_0^\infty u^{2k-1}P\left(\xi_A>\psi^{\frac12}+u\psi^{\frac14}\right)du$$

$$+2k\int_0^\infty u^{2k-1}P\left(\xi_A<\psi^{\frac12}-u\psi^{\frac14}\right)du=I+II,\text{ say.}$$

To bound I, we observe that for all $\theta>0$

$$(4.3)\quad P\left(\xi_A>\psi^{\frac12}+u\psi^{\frac14}\right)\leqslant P\left[X\left(\psi^{\frac12}+u\psi^{\frac14}\right)>A^{-1}\left(\psi^{\frac12}+u\psi^{\frac14}\right)^3\right]$$

$$\leqslant\exp\left\{-\theta A^{-1}\left(\psi^{\frac12}+u\psi^{\frac14}\right)^3+\lambda\left(\psi^{\frac12}+u\psi^{\frac14}\right)(e^\theta-1)\right\}$$

$$=\exp\left\{(e^\theta-1-\theta)\lambda\psi^{\frac12}-3\theta u\lambda\psi^{\frac14}\right.$$

$$\left.-3\theta u^2\lambda-\lambda\theta u^3\psi^{-\frac14}+\lambda\psi^{\frac14}u(e^\theta-1)\right\}.$$

Choosing $\theta=\lambda^{-\frac12}\psi^{-\frac14}$, we see that $\theta\leqslant\varepsilon^{-\frac12}\psi^{-\frac14}\to0$ as $\psi\to\infty$, so we get

$$(4.4)\quad\begin{aligned}&\text{(i)}\quad(e^\theta-1-\theta)\lambda\psi^{\frac12}=(e^\theta-1-\theta)\theta^{-2}=\tfrac12+o(1)\text{ as }\psi\to\infty.\\&\text{(ii)}\quad-3u\theta\lambda\psi^{\frac14}=-3u\lambda^{\frac12}.\\&\text{(iii)}\quad u\lambda\psi^{\frac14}(e^\theta-1)\leqslant u\lambda\psi^{\frac14}2\theta=2u\lambda^{\frac12},\text{ for all large }\psi.\end{aligned}$$

Applying (4.4) to the appropriate terms on the right hand side of (4.3), we see that the right hand side of (4.3) does not exceed $\exp\{\frac12+o(1)-\lambda^{\frac12}u\}$, as $\psi\uparrow\infty$. Hence,

$$(4.5)\quad I\leqslant 2ke^{\frac12+o(1)}\int_0^\infty u^{2k-1}e^{-\lambda^{\frac12}u}du=\frac{(2k)!\exp\left(\frac12+o(1)\right)}{\lambda^k}.$$

For II of (4.2), let $0<\delta<1$ and write

$$(4.6)\quad II\leqslant 2k\int_0^{\delta\psi^{\frac14}}u^{2k-1}P\left(\xi_A<\psi^{\frac12}-u\psi^{\frac14}\right)du$$

$$+2k\int_{\delta\psi^{\frac14}}^{\psi^{\frac14}}u^{2k-1}P\left(\xi_A<\psi^{\frac12}-u\psi^{\frac14}\right)du=III+IV,$$

say.

From Lemma 4.1(ii) we have

$$(4.7)\quad IV<2k\psi^{k/2}P\left(\xi_A<(1-\delta)\psi^{\frac12}\right)=o(\psi^{-\infty}).$$

For III of (4.6), let γ be such that $\delta<\gamma<1$, then, again by Lemma 4.1(ii), we have

$$(4.8)\quad III\sim 2k\int_0^{\delta\psi^{\frac14}}u^{2k-1}P\left[(1-\gamma)\psi^{\frac12}<\xi_A<\psi^{\frac12}-u\psi^{\frac14}\right]du.$$

Let $r=(1-\gamma)\psi^{\frac12}$ and $s=\psi^{\frac12}-u\psi^{\frac14}$. Then the event in question implies that $Y(t)=\lambda^{-1}t^{-1}X(t)\leqslant\psi^{-1}s^2$ for some $t\in[r,s]$. The right hand side of (4.8) is thus bounded above by

(4.9) $$2k + \int_1^{\delta\psi^{\frac{1}{4}}} u^{2k-1} P\left[\max_{t\in[r,s]}(1 - Y(t)) \geq 1 - \psi^{-1}s^2\right] du$$
$$\leq 2k + \int_1^{\delta\psi^{\frac{1}{4}}} u^{2k-1} P\left[\max_{t\in[r,s]}(1 - Y(t))^{2m} \geq \left(u\psi^{-\frac{1}{4}}\right)^{2m}\right] du,$$

where we used the fact that for $u \in [0, \delta\psi^{\frac{1}{4}}]$, $1 - \psi^{-1}s^2 > u\psi^{-\frac{1}{4}}$. Applying the maximal inequality to the backward submartingale $(1 - Y(t))^{2m}$, the right hand side of (4.9) can be bounded above by

(4.10) $$2k + \psi^{m/2} \int_1^{\delta\psi^{\frac{1}{4}}} u^{-2(m-k)-1} E(1 - Y(r))^{2m} du$$
$$= 2k + \psi^{m/2}(\lambda r)^{-2m} 0((\lambda r)^m) \int_1^{\delta\psi^{\frac{1}{4}}} u^{-2(m-k)-1} du$$
$$= 2k + 0(\psi^{-m/2}),$$

for all $m > k$. Combining (4.8) through (4.10) we get

(4.11) $$III \leq 2k + o(\psi^{-\infty}) = 0(1).$$

Expressions (4.2), (4.5), (4.6), (4.7) and (4.11) combine to prove (i). For (ii) we write

$$E\left(\xi_A/\psi^{\frac{1}{2}}\right)^k = \int_0^\infty P\left[\left(\xi_A/\psi^{\frac{1}{2}}\right)^k > y\right] dy$$
$$\leq 2^k + k\int_2^\infty u^{k-1} P\left[X\left(u\psi^{\frac{1}{2}}\right) > A^{-1}u^3\psi^{\frac{3}{2}}\right] du$$
$$\leq 2^k + k\int_2^\infty u^{k-1} \exp\left\{\lambda\psi^{\frac{1}{2}}\left[-u^3 + u(e-1)\right]\right\} du = 0(1),$$

which completes the proof.

The following lemma is similar to related results contained in Y. S. Chow, H. Robbins and H. Teicher (1965). Nevertheless, since they deal only with sequences of rv's and since the proof for the homogeneous Poisson process is short, we present the proof here.

LEMMA 4.3. *Let* $U(t) = X(t) - \lambda t$, $V(t) = (X(t) - \lambda t)^2 - \lambda t$ *and* $W(t) = (X(t) - \lambda t)^3 - 3\lambda t(X(t) - \lambda t) - \lambda t$. *Then,*

(i) $U(t)$, $V(t)$, $W(t)$ *are all martingales* w.r.t. $\mathcal{F}_t$, $t \geq 0$ *and* $E_\lambda U(t) = E_\lambda V(t) = E_\lambda W(t) = 0$.

(ii) $\{U(0), U(\xi_A)\}$, $\{V(0), V(\xi_A)\}$, $\{W(0), W(\xi_A)\}$ *are all martingales* w.r.t. $\{\mathcal{F}_0, \mathcal{F}_{\xi_A}\}$.

PROOF. The proof of (i) is a straightforward computation and will be omitted. To prove (ii) we have to verify the conditions of the optional sampling theorem. We demonstrate it for $W(\xi_A)$; the proof for $V(\xi_A)$ and $U(\xi_A)$ is analogous and will be omitted. Thus, we need only show

(a) $E|W(\xi_A)| < \infty$,

(b) $\liminf_{t\to\infty} \int_{[\xi_A > t]} |W(t)| dP = 0$. From the Schwarz inequality, we have

$$\int_{[\xi_A > t]} |W(t)| dP \leq \left(EW(t)^2 P[\xi_A > t]\right)^{\frac{1}{2}}.$$

Now, $E(W(t))^2 = O(t^k)$ for some positive k and

$$P[\xi_A > t] < P[X(t) > A^{-1}t^3] \leqslant \exp\{-A^{-1}t^3 + \lambda t(e-1)\} = o(t^{-\infty});$$

thus (b) follows. The proof of (a) is standard and will be omitted.

LEMMA 4.4. *On* $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, *as* $\psi \to \infty$ *we have*

$$E_\lambda \xi_A^i \leqslant \psi^{i/2} + o(\psi^{-\infty}), \qquad \textit{for} \quad i = 1, 2, 3.$$

PROOF. From Lemma 4.3 we have

$$\lambda E\xi_A = EX(\xi_A) = E[X(r(A)); \xi_A = r(A)] + A^{-1}E[\xi_A^3; \xi_A > r(A)]$$
$$\sim A^{-1}E\xi_A^3$$

(by Lemma 4.1(ii)). Hence, $\psi E_\lambda \xi_A \sim E_\lambda \xi_A^3$. Also the Jensen inequality implies that

$$(E\xi_A)^2 \leqslant E\xi_A^2 \leqslant (E\xi_A^3)^{\frac{2}{3}} \sim (\psi E\xi_A)^{\frac{2}{3}}.$$

Thus, the result follows.

LEMMA 4.5. *As* $\psi \to \infty$, *on* $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, *we have,*

$$A\psi^{-\frac{3}{2}}|E_\lambda(\lambda\xi_A - X(\xi_A))^3| = 0(1).$$

PROOF. From Lemma 4.3(ii) we have

$$(4.12) \qquad E(\lambda\xi_A - X(\xi_A))^3 = -3\lambda E\xi_A(X(\xi_A) - \lambda\xi_A) - \lambda E\xi_A$$

and from Lemma 4.4

$$(4.13) \qquad A\psi^{-\frac{3}{2}}\lambda E\xi_A \leqslant \psi^{-\frac{1}{2}}\left(\psi^{\frac{1}{2}} + o(\psi^{-\infty})\right) = 1 + o(\psi^{-\infty}).$$

Also, from Lemma 4.3(ii), the Schwarz inequality and Lemma 4.2(i), we have

$$(4.14) \quad \begin{aligned} A\psi^{-\frac{3}{2}}3\lambda|E\xi_A(X(\xi_A) - \lambda\xi_A)| &= 3\psi^{-\frac{1}{2}}|E\left(\xi_A - \psi^{\frac{1}{2}}\right)(X(\xi_A) - \lambda\xi_A)| \\ &\leqslant 3\psi^{-\frac{1}{2}}\left[E\left(\xi_A - \psi^{\frac{1}{2}}\right)^2 E(X(\xi_A) - \lambda\xi_A)^2\right]^{\frac{1}{2}} \\ &\leqslant 3\psi^{-\frac{1}{2}}\left[\lambda^{-1}\psi^{\frac{1}{2}}0(1)\lambda\left(\psi^{\frac{1}{2}} + o(\psi^{-\infty})\right)\right]^{\frac{1}{2}} \\ &= 0(1). \end{aligned}$$

Applying the triangle inequality to the absolute value of (4.12) and using (4.13) and (4.14) the result follows.

We are now ready to give the

PROOF OF THEOREM 1.1. From Lemma 4.4 we have $\lambda(E_\lambda\xi_A - \psi^{\frac{1}{2}}) \leqslant \varepsilon^{-1}\psi o(\psi^{-\infty})$. Thus, we need only show that on $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$ as $\psi \to \infty$, we have

$$(4.15) \qquad \lambda\left\{AE_\lambda(X(\xi_A)/\xi_A - \lambda)^2 - \psi^{\frac{1}{2}}\right\} \leqslant 0(1).$$

Write

$$(4.16)\qquad \lambda\left[AE(X(\xi_A)/\xi_A - \lambda)^2 - \psi^{\frac{1}{2}}\right] = E\left[\left(\frac{\psi}{\xi_A^2} - 1\right)(X(\xi_A) - \lambda\xi_A)^2\right] + E\left[X(\xi_A) - \lambda\xi_A\right]^2 - \lambda\psi^{\frac{1}{2}};$$

then from Lemmas 4.3(ii) and 4.4 we have

$$E\left[X(\xi_A) - \lambda\xi_A\right]^2 - \lambda\psi^{\frac{1}{2}} = \lambda\left(E\xi_A - \psi^{\frac{1}{2}}\right) \leqslant \varepsilon^{-1}\psi o(\psi^{-\infty}) = o(\psi^{-\infty}).$$

It remains to show that

$$(4.17)\qquad E\left[\left(\frac{\psi}{\xi_A^2} - 1\right)(X(\xi_A) - \lambda\xi_A)^2\right] \leqslant 0(1).$$

From Lemma 4.1(ii) we have, for $0 < \alpha < 1$,

$$(4.18)\qquad E\left[\left(\frac{\psi}{\xi_A^2} - 1\right)(X(\xi_A) - \lambda\xi_A)^2\right]$$

$$\sim E\left[\left(\frac{\psi}{\xi_A^2} - 1\right)(A^{-1}\xi_A^3 - \lambda\xi_A)^2;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$= A^{-2}E\left[(\psi - \xi_A^2)^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$= A^{-2}E\left[(\psi - AX(\xi_A)/\xi_A)^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$= A\psi^{-\frac{3}{2}}E\left[\frac{\psi^{\frac{3}{2}}}{\xi_A^3}(\lambda\xi_A - X(\xi_A))^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right].$$

Using the mean value theorem (MVT) on U^{-3} we write

$$(4.19)\qquad \frac{\psi^{\frac{3}{2}}}{\xi_A^3} = 1 - 3\psi^{\frac{3}{2}}d^{-4}\left(\xi_A - \psi^{\frac{1}{2}}\right),$$

where d lies between ξ_A and $\psi^{\frac{1}{2}}$. Applying (4.19) to the right hand side of (4.18) we see that it can be written as

$$(4.20)\qquad A\psi^{-\frac{3}{2}}E\left[(\lambda\xi_A - X(\xi_A))^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right] - 3AE\left[d^{-4}\left(\xi_A - \psi^{\frac{1}{2}}\right)(\lambda\xi_A - X(\xi_A))^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right] = I + II, \text{ say.}$$

From Lemma 4.5 we have

$$(4.21)\qquad |I| = 0(1).$$

For the computation of II, observe that on $\xi_A > \alpha\psi^{\frac{1}{2}}$ we have $d > \alpha\psi^{\frac{1}{2}}$ and also sign $(\psi^{\frac{1}{2}} - \xi_A) =$ sign $(\lambda\xi_A - X(\xi_A))$. Therefore, we get

$$(4.22) \qquad II \leqslant -\frac{3A}{\alpha^4\psi^2} E\left[\left(\xi_A - \psi^{\frac{1}{2}}\right)(\lambda\xi_A - X(\xi_A))^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$= -\frac{3\lambda^2}{\alpha^4\psi^4} E\left[\left(\xi_A - \psi^{\frac{1}{2}}\right)(\psi\xi_A - \xi_A^3)^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$= \frac{3\lambda^2}{\alpha^4} E\left[\left[\frac{\xi_A - \psi^{\frac{1}{2}}}{\psi^{\frac{1}{4}}}\right]^4 \left(\frac{\xi_A}{\psi^{\frac{1}{2}}}\right)^3 \left(\frac{\xi_A}{\psi^{\frac{1}{2}}} + 1\right)^3;\ \xi_A > \alpha\psi^{\frac{1}{2}}\right]$$

$$\leqslant \frac{3\lambda^2}{\alpha^4}\left[E\left[\frac{\xi_A - \psi^{\frac{1}{2}}}{\psi^{\frac{1}{4}}}\right]^{16}\right]^{\frac{1}{4}} \left[E\left(\frac{\xi_A}{\psi^{\frac{1}{2}}}\right)^{12}\right]^{\frac{1}{4}}$$

$$\times \left[E\left(\frac{\xi_A}{\psi^{\frac{1}{2}}} + 1\right)^{12}\right]^{\frac{1}{4}} \left[P\left(\xi_A > \alpha\psi^{\frac{1}{2}}\right)\right]^{\frac{1}{4}}$$

$$= \frac{3\lambda^2}{\alpha^4}\left[\lambda^{-8} 0(1)\right]^{\frac{1}{4}} 0(1) = 0(1).$$

Here we used the Hölder inequality and, for the last step, Lemma 4.2. Relations (4.20)–(4.22) combine to prove (4.17) and thus the proof of the theorem is completed.

5. The case of i.i.d. Poisson rv's. Let $Y_1, Y_2, \cdots$ be a sequence of i.i.d. rv's having Poisson distribution with mean $\lambda > 0$. For the problem of estimating λ we choose the estimator $\Sigma_i^n Y_i/n$ (whenever n observations are available) and we assume the loss plus cost function (1.1). The BFST is either $[\psi^{\frac{1}{2}}]$ or $[\psi^{\frac{1}{2}}] + 1$, where $[x]$ denotes the integer part of $[x]$. Nevertheless, we shall take the BFST to be $\psi^{\frac{1}{2}}$ so that the resulting 'risk' is $2\psi^{\frac{1}{2}}$. Note that the 'risk' associated with either $[\psi^{\frac{1}{2}}]$ or $[\psi^{\frac{1}{2}}] + 1$ is always $\geqslant 2\psi^{\frac{1}{2}}$, which makes our bounds for the regret somewhat conservative. It follows that the regret function $\rho_{\eta_A}(A, \lambda)$, for a ST η_A, is the same as defined in Section 1. The model we consider here is a discrete analogue of the continuous case described in the previous sections and thus, naturally, we will use some of the results obtained earlier.

Let $\{X(t);\ t \geqslant 0\}$ be a HPP with rate $\lambda > 0$; then, restricting the process to integer times, the resulting sequence $\{X(n);\ n = 0, 1, \cdots\}$ has the same probability law as the sequence $\{\Sigma_{j=0}^n Y_j;\ n = 0, 1, \cdots\}$ with $Y_0 \equiv 0$. We therefore assume in what follows that the sequence $\{\Sigma_{j=0}^n Y_j;\ n = 0, 1, \cdots\}$ emerged from the HPP $\{X(t);\ t \geqslant 0\}$ restricted to the time points $0, 1, 2, \cdots$. Naturally we then write $X(n) = \Sigma_0^n Y_j$ and we continue working within the same framework laid out in the previous sections. In particular, the assumptions of Section 3 apply to this section as well.

We define the following family of ST's. For each $A > 0$ let

$$\eta_A = \text{first integer } n \geqslant r(A) \text{ for which } X(n) \leqslant A^{-1}n^3, \tag{5.1}$$

where $r(A)$ is given by (2.2). It is clear from the SLLN that for every $\lambda > 0$ and $A > 0$, $P_\lambda(\eta_A < \infty) = 1$.

Before proving Theorem 1.2 we need two lemmas.

LEMMA 5.1. *As in Lemma* 4.3, *let* $U(t) = X(t) - \lambda t$, $V(t) = (X(t) - \lambda t)^2 - \lambda t$. *Then* $\{U(\xi_A), U(\eta_A)\}$, $\{V(\xi_A), V(\eta_A)\}$ *are martingales* w.r.t. $\{\mathcal{F}_{\xi_A}, \mathcal{F}_{\eta_A}\}$.

The proof is a standard application of the optional sampling theorem and will be omitted.

In the following lemma we consider the rv

$$Z(\xi_A) = E_\lambda(\eta_A | \mathcal{F}_{\xi_A}) - \xi_A.$$

LEMMA 5.2. *On* $(A \geqslant \varepsilon) \cap (\lambda \geqslant \varepsilon)$, *as* $\psi \to \infty$ *we have*

$$E_\lambda[Z(\xi_A)]^i \leqslant b^i + o(\psi^{-\infty})$$

where b is an arbitrary real number satisfying $b > \frac{3}{2}$.

PROOF. To simplify the notation we shall use the following abbreviations: $\xi \equiv \xi_A$, $\eta \equiv \eta_A$, $Z = Z(\xi_A)$. From Lemma 5.1, $E[X(\eta) - \lambda\eta | \mathcal{F}_\xi] = X(\xi) - \lambda\xi$, therefore

$$\begin{aligned}\lambda E(\eta | \mathcal{F}_\xi) &= E(X(\eta)|\mathcal{F}_\xi) - X(\xi) + \lambda\xi \\ &> E\left(A^{-1}(\eta - 1)^3 | \mathcal{F}_\xi\right) - A^{-1}\xi^3 + \lambda\xi \\ &> A^{-1}\left(E(\eta - 1|\mathcal{F}_\xi)\right)^3 - A^{-1}\xi^3 + \lambda\xi.\end{aligned}$$

Hence,

$$\psi E(\eta - 1|\mathcal{F}_\xi) > \left(E(\eta - 1|\mathcal{F}_\xi)\right)^3 - \xi^3 + \psi\xi - \psi,$$

from which we conclude, after some elementary algebra, that

$$0 > (Z - 1)\left((\xi + Z - 1)^2 + (\xi + Z - 1)\xi + \xi^2\right) - \psi Z. \tag{5.2}$$

Now, on $Z \geqslant s \geqslant 1$ (5.2) implies

$$\begin{aligned}\psi\frac{s}{s-1} \geqslant \psi\frac{Z}{Z-1} &> (\xi + Z - 1)^2 + (\xi + Z - 1)\xi + \xi^2 \\ &> (\xi + s - 1)^2 + (\xi + s - 1)\xi + \xi^2 > 3\xi^2 + 3(s-1)\xi.\end{aligned}$$

Hence, for $s > 1$, we have

$$P(Z > s) \leqslant P\left(3\xi^2 + 3(s-1)\xi < \frac{s}{s-1}\psi\right). \tag{5.3}$$

Using this and the fact that $s/(s-1)$ is a decreasing function of s we have

$$EZ^i \leqslant b^i + i\int_b^\infty s^{i-1}P\{Z > s\}ds$$

$$\leqslant b^i + i\int_b^\infty s^{i-1}P\left\{3\xi^2 + 3(s-1)\xi < \frac{s}{s-1}\psi\right\}ds$$

(5.4)

$$\leqslant b^i + i\int_b^\infty s^{i-1}P\left\{\xi^2 + (s-1)\xi < \frac{b}{3(b-1)}\psi\right\}ds$$

$$= b^i + i\int_{b-1}^\infty (s+1)^{i-1}P\left\{\xi < -\frac{s}{2} + \left[\left(\frac{s}{2}\right)^2 + \frac{b}{3(b-1)}\psi\right]^{\frac{1}{2}}\right\}ds$$

$$= b^i + i\int_{(b-1)/2\psi^{\frac{1}{2}}}^\infty (2\psi^{\frac{1}{2}}u + 1)^{i-1}P\left\{\xi < \left(-u + \left[u^2 + \frac{b}{3(b-1)}\right]^{\frac{1}{2}}\right)\psi^{\frac{1}{2}}\right\}du.$$

Now, $(-u + [u^2 + (b/3(b-1))]^{\frac{1}{2}})\psi^{\frac{1}{2}}$ is a decreasing function of u which attains the value $r(A)$ at

$$u = \psi^{\frac{1}{2}}/4r(A) - \tfrac{1}{2} < \psi^{\frac{1}{2}}/a(\varepsilon).$$

Therefore, the rightmost member of (5.4) does not exceed

$$(5.5)\quad b^i + i2^{i-1}\psi^{(i-1)/2}\int_{(b-1)/2\psi^{\frac{1}{2}}}^{\psi^{\frac{1}{2}}/a(\varepsilon)}(u + 2^{-1}\psi^{-\frac{1}{2}})^{i-1}P\left\{\xi < -\frac{b-1}{2} + \left[\left(\frac{b-1}{2}\right)^2 + \frac{b}{3(b-1)}\psi\right]^{\frac{1}{2}}\right\}du$$

$$< b^i + i2^{i-1}\psi^{(i-1)/2}(\psi^{\frac{1}{2}}/a(\varepsilon) + e^{-1}\psi^{-\frac{1}{2}})^{i-1}P\left\{\xi < -\frac{b-1}{2} + \left[\left(\frac{b-1}{2}\right)^2 + \frac{b}{3(b-1)}\psi\right]^{\frac{1}{2}}\right\}.$$

Now, b was chosen to satisfy $0 < b/3(b-1) < 1$, so by Lemma 4.1 (ii) the second term on the right hand side of (5.5) is $o(\psi^{-\infty})$. (5.4) and (5.5) combine to give $EZ^i \leqslant b^i + o(\psi^{-\infty})$ as desired.

REMARK. Using Lemma 5.2 (with $i = 1$) and Lemma 4.4 we get $E_\lambda\eta_A \leqslant \psi^{\frac{1}{2}} + b + \theta(\psi^{-\infty})$ ($b > \frac{3}{2}$ arbitrary). One can use, however, Theorem 2 of Starr and Woodroofe (1968) to get a slightly better approximation. Specifically,

$$(5.6)\quad E_\lambda\eta_A \leqslant 1 + \left\{\psi + (r(A)-1)^2P\left[X(\eta_A) \leqslant A^{-1}\eta^3;\ \forall n \geqslant r(A)\right]\right\}^{\frac{1}{2}}$$

$$= 1 + \psi^{\frac{1}{2}} + o(\psi^{-\infty}).$$

PROOF OF THEOREM 1.2. Let $M_t = A(t^{-1}X(t) - \lambda)^2$. Then, with the abbreviations $\xi = \xi_A$, $\eta = \eta_A$ and $Z = Z(\xi_A)$, we have

$$(5.7) \qquad \begin{aligned} E\{M_\eta|\mathcal{F}_\xi\} &\leqslant AE\{\xi^{-2}(X(\eta) - \lambda\eta)^2|\mathcal{F}_\xi\} \\ &= AE\{\xi^{-2}[(X(\eta) - \lambda\eta)^2 - \lambda\eta]|\mathcal{F}_\xi\} + \psi\{\xi^{-2}\eta|\mathcal{F}_\xi\} \\ &= M_\xi + \psi\xi^{-2}Z. \end{aligned}$$

$$\begin{aligned} E\{M_\eta\} &\leqslant E\{M_\xi\} + E\{\psi\xi^{-2}Z;\ \xi > \alpha\psi^{\frac{1}{2}}\} + o(\psi^{-\infty}), \\ &\leqslant E\{M_\xi\} + \alpha^{-2}E(Z) + o(\psi^{-\infty}), \end{aligned}$$

where $\alpha \in (0, 1)$ is arbitrary. Lemma 5.2 (with $i = 1$), (5.6) and (5.7) combine to complete the proof.

Acknowledgment. I would like to express my sincerest thanks to Professor Jack Kiefer for his encouragement, guidance and many helpful discussions. Thanks are also due to the referee for his excellent suggestions which helped simplify some of the proofs.

REFERENCES

[1] CHOW, Y. S., ROBBINS, H. and TEICHER, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

[2] DOOB, J. L. (1953). *Stochastic Process*. 376. Wiley, New York.

[3] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics*. (Harald Cramér, ed). Almquist and Vikell, Uppsalla, Sweden. 235–245.

[4] STARR, N. and WOODROOFE, M. B. (1968). Remarks on a stopping time. *Proc. Nat. Acad. Sci. U.S.A.* **61** 1215–1218.

[5] STARR, N. and WOODROOFE, M. B. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. U.S.A.* **63** 285–288.

[6] STARR, N. and WOODROOFE, M. B. (1972). Further remarks on sequential estimation: the exponential case. *Ann. Math. Statist.* **43** 1147–1154.

[7] VARDI, Y. (1979). Asymptotic optimality of certain sequential estimators. *Ann. Statist.* **7** 1034–1039.

BELL LABORATORIES
600 MOUNTAIN AVE.
MURRAY HILL, NEW JERSEY 07974