

## OPTIMAL STOPPING OF SEQUENTIAL SIZE-DEPENDENT SEARCH

BY ISSA FAKHRE-ZAKERI AND ERIC SLUD<sup>1</sup>

*University of North Carolina, Chapel Hill  
and University of Maryland*

In many areas of application, one searches within finite populations for items of interest, where the probability of sampling an item is proportional to a random size attribute from an i.i.d. superpopulation of attributes which may or may not be observable upon discovery. Here we treat the problem of asymptotically optimal stopping rules for size-dependent searches of this type, as the size of the underlying population grows, where the loss function includes an asymptotically smooth time-dependent cost, a constant cost per item sampled and a cost per undiscovered item which may depend on the size attribute of the undiscovered item. Under some regularity and convexity conditions related to the asymptotic expected loss, we characterize asymptotically optimal rules even when the initial population size and the distribution of size attributes are unknown. We direct especial attention to applications in software reliability, where the items of interest are software faults (“bugs”). In this setting, the size attributes will not be observable when faults are found, and, in addition, our search model allows new bugs to be introduced into the software when faults are detected (“imperfect debugging”). Our results extend those of Dalal and Mallows and Kramer and Starr, and are illustrated in the perfect-debugging case on a previously analyzed dataset of Musa.

**1. Introduction.** There are many problem areas in which scientific investigators search for items within an unknown finite population and have greater chance of detecting items for which some specified size attribute is large. For example, in the search for oil and gas pools, the larger pools are easier to detect [Barouch and Kaufman (1976)]. In wildlife management, it may be easier to catch and tag larger rather than smaller birds or fish, and, at a higher level of aggregation, it is certainly true that larger schools of fish are easier to detect, and that species with many members will be earliest documented [Seber (1992) and references therein]. Similarly, large pottery fragments at an archaeological site are found more often than smaller ones [Kramer and Starr (1987)]; and in software reliability testing, the software faults (“bugs”) which most often cause software failure when run on inputs

---

Received March 1993; revised February 1996.

<sup>1</sup>Research partially supported by Office of Naval Research Contract N00014-92-C0019.

*AMS 1991 subject classifications.*

*Key words and phrases.* Asymptotically optimal stopping rule, exponential order statistic model, imperfect debugging, loss function, mixture model, sequential search, size-dependent successive sampling.

randomly chosen from some ensemble, will be the soonest detected [Miller (1986), Dalal and Mallows (1988), Littlewood (1980, 1981, 1989) and Fakhre-Zakeri and Slud (1995)]. A common feature of these applications is that, after appropriately defining the size attribute, one may model the data-sampling mechanism as “sampling proportional to size” (without replacement). Specifically, if undetected items are assumed initially to have size attributes  $\Lambda_1, \dots, \Lambda_{n_0}$  sampled independently from a superpopulation with distribution  $G$ , then at a time prior to which items with attributes  $\Lambda_{i_1}, \dots, \Lambda_{i_k}$  have been sampled without replacement, the conditional probability that the next item selected has attribute  $\Lambda_j$  (where  $j \notin \{i_1, \dots, i_k\}$ ) can be modeled as

$$\Lambda_j / \left( \sum_{i=1}^{n_0} \Lambda_i - \Lambda_{i_1} - \dots - \Lambda_{i_k} \right).$$

This modeling assumption has been studied from a general perspective by Kramer and Starr (1987), who characterized it qualitatively. Such models have also been treated as variants of “coupon-collector’s problems” by Rosén (1972) and in the literature on “successive sampling” [Gordon (1983), Vardi (1982), Andreatta and Kaufman (1986), Bickel, Nair and Wang (1992) and many other references].

Further common features of these sampling mechanisms are that the total number of items to discover is finite but unknown, that data arrive sequentially and that a natural question concerning the best use of investigative resources is when to stop sampling. In all four examples, one would like at the end to estimate the number of undiscovered items. In the oil-and-gas and software examples, economic interest naturally centers on the time required for the search. A special feature of software testing is that, upon discovery of a bug, its size attribute (frequency of causing failures) will still not be observable. Also Bickel, Nair and Wang (1992) remark that pool size may also be only partially observable in oil-and-gas discovery; in the search for disease, cases due to environmental contamination, in which the “items” are cases and the size attribute is individual exposure to the contaminant, the latter will only be partially observable.

In this paper, we are interested in optimal stopping of such sequential size-biased searches. In specifying appropriate loss functions, there are several costs to be taken into account, all of which may arise in the same application: the overhead of the search itself, costs associated with each item found, opportunity costs due to deferring final decisions based on study results and (especially in the oil-and-gas and software settings) the losses due to items not discovered (e.g., finding software failures after the software has been released for commercial use, as well as customer losses due to recurring failures). In applications, the relative importance of these costs must be weighed carefully in specifying the loss function. Losses for items not found can be more or less serious according to the value of the parameter  $\Lambda_i$ . Thus a general loss function would have the form

$$(1.1) \quad L(t) = L_n(t) = f_n(t) + aN(t) + \sum_{i \leq n_0} r_i(t)h(\Lambda_i),$$

where  $f_n(\cdot)$  is a nonrandom time-dependent loss such that  $f_n(t)/n$  is assumed to converge uniformly on the bounded time horizon  $[0, t_{\max}]$  to a twice-differentiable function  $f(\cdot)$ ,  $N(t)$  denotes the total number of items which have been found up to time  $t$ ,  $n$  is a known order-of-magnitude parameter such that the unknown number  $n_0$  of items satisfies  $n_0/n \equiv \vartheta_n \rightarrow \vartheta$  for an unknown positive parameter  $\vartheta$  as  $n$  becomes large,  $r_i(t)$  is the indicator that the  $i$ th item remains undiscovered as of time on test  $t$  and  $h$  is a nondecreasing positive function expressing the relationship between size attributes and costs of undiscovered items. Since both the second and third terms of  $L_n$  in (1.1) satisfy a law of large numbers with denominator  $n$ , it turns out that, under appropriate conditions on the limiting function  $f$  for  $f_n/n$ , only the order- $n$  behavior of  $f_n$  affects the asymptotics of optimal stopping rules. Note also that we do *not* assume that  $f_n$  itself is either convex or increasing: the only assumptions needed are on the limiting  $f$ , and even it may be highly nonconvex for moderate  $t$ , depending on the behavior of  $h$ . [See conditions (2.3) and (3.2) in the theorems below.] This accommodates applications where time-dependent (but irregular or highly nonconvex) opportunity costs might enter  $f_n$  but scale sublinearly as the parameter  $n$  becomes large. Since it is natural to scale the loss by  $n$ , we define for future reference

$$(1.2) \quad \bar{L}_n(t) \equiv \frac{L_n(t)}{n} = \frac{1}{n}f_n(t) + \frac{\alpha}{n}N(t) + \frac{1}{n} \sum_{i \leq n_0} r_i(t)h(\Lambda_i).$$

Such rescaling, including the assumptions on  $f_n$  and  $n_0$  mentioned above, has the effect of converting the asymptotic-optimality criterion of choosing  $\tau$  to minimize the ratio of  $E(L_n(\tau))/E(L_n(\sigma))$  for competing stopping rules  $\sigma$  into the asymptotically equivalent criterion of minimizing the difference  $E(\bar{L}_n(\tau)) - E(\bar{L}_n(\sigma))$ .

Sequential finite-population search problems of this sort have been treated in Kramer and Starr (1987, 1990) and Dalal and Mallows (1988). The loss function in Kramer and Starr (1990) had the form (1.1) (except for a random constant unaffected by the number of previously discovered items) with  $\alpha = 0$ . In Dalal and Mallows (1988), the loss function for stopping software testing as of time  $t$  was (1.1) with  $h(\cdot) \equiv b$  constant, but the number  $n_0$  was taken to be random (Poisson-distributed) and  $G$  was given a parametric Gamma form. In the settings just described, Kramer and Starr (1990) (with  $n_0$  assumed known) and Dalal and Mallows (1988) found optimal stopping rules for known  $G$ . For related results of Dalal and Mallows on asymptotically optimal rules, see Remark 2.2 following Theorem 2.1.

In the next sections, we investigate asymptotically optimal rules based on the loss function (1.1)—modified for software reliability applications as in Theorem 3.1—and show under some regularity conditions that the “myopic” stopping rule (i.e., the rule with smallest current expected loss) is asymptotically optimal as  $n \rightarrow \infty$ , in the sense described below. This extends the work of Dalal and Mallows (1988) and Kramer and Starr (1990) in three directions:

1. to the case of unknown distribution  $G$  of parameters  $\Lambda_i$ ;

2. to unobservable mixture-parameter values: this is needed to make the model relevant to software reliability;
3. to a model incorporating “offspring” bugs (due to imperfect debugging) in software testing.

Extension (3) is undertaken only in Section 3, where the software reliability setting is elaborated.

In Section 2 we study the asymptotically optimal stopping rules for the size-dependent search model without replacement, obtaining results which apply also in the examples mentioned above where size attributes are unobserved. In Section 3 we extend the discussion to the case of software reliability with imperfect debugging. Section 4 contains an illustrative example using real software reliability data. All proofs are deferred to Section 5.

**2. Asymptotic optimality in size-dependent search without replacement.** We begin by introducing some notation. Index the undiscovered items of interest by  $i = 1, \dots, n_0$ ,  $n_0 \equiv n\vartheta_n$ , where  $n$  denotes a known parameter indicating the order of magnitude or scale of the search, and  $\vartheta_n \rightarrow \vartheta$ , an unknown positive parameter. Let  $(\Lambda_i, T_i)$  be i.i.d. for  $i = 1, \dots, n_0$ , with the random size parameter  $\Lambda_i$  for item  $i$  distributed according to d.f.  $G$  on  $(0, \infty)$  and with  $T_i$  conditionally exponential  $(\Lambda_i)$  distributed given  $\Lambda_i$ , and define  $r_i(t) \equiv I_{[T_i \geq t]}$  for  $t \geq 0$ . The unconditional distribution function  $Q$  and density  $q$  of  $T_i$  are respectively given by

$$Q(t) \equiv \int_0^\infty (1 - e^{-\lambda t})G(d\lambda),$$

$$q(t) \equiv \int_0^\infty \lambda e^{-\lambda t}G(d\lambda).$$

The observable process  $N(t) \equiv \sum_i I_{[T_i \leq t]}$  counts the total number of discoveries through time  $t$ . Let  $\mathcal{F}_t \equiv \mathcal{F}_t^n$  be the  $\sigma$ -field generated by all r.v.'s  $\Lambda_i$ , as well as all  $T_i$ 's which are less than or equal to  $t$ . The infinitesimal operator associated with the cost function  $L$  of (1.1) is defined by

$$\alpha(t) \equiv \alpha_n(t) = \lim_{x \rightarrow 0+} x^{-1}E[L(t+x) - L(t) | \mathcal{F}_t],$$

when the limit exists. For positive increasing functions  $U(\lambda)$ , let

$$(2.1) \quad \psi_U(t) \equiv E(r_i(t)U(\Lambda_i)) = \int e^{-\lambda t}U(\lambda)G(d\lambda).$$

In what follows,  $\psi_{\text{id}}$  and  $\psi_h$  respectively denote the functions  $\psi_U$  with  $U(x)$  replaced by the identity function  $U(x) = x$  or the function  $U(x) = h(x)$  appearing in (1.1). In order that these functions be twice continuously differentiable at 0, we assume that

$$(2.2) \quad \int \lambda^2(1 + h(\lambda))G(d\lambda) < \infty.$$

Our method follows Kramer and Starr (1990) in investigating the monotonicity of the infinitesimal cost operator in order to characterize the optimal rule. In the first part of this section, until stated otherwise, the distribution  $G$  is assumed known.

It is known that if the cost function is such that  $\alpha(t)$  exists and is monotonically nondecreasing with probability 1, then the stopping time  $\tau = \inf\{t \geq 0: \alpha_n(t) \geq 0\}$  is optimal in the sense that  $E[L_n(\tau)] \leq E[L_n(\tau')]$  for all  $\{\mathcal{F}_t\}$  stopping times  $\tau'$ . For details, see Ross (1971) and Shapiro and Wardrop (1980). Here the model for the counting process of observed software failures for large  $n$  is of a triangular-array type such that  $\alpha_n(t)$  tends a.s. by the law of large numbers to the nonrandom limit

$$\nu(t) \equiv \lim_{n \rightarrow \infty} \frac{d}{dt} E[\bar{L}_n(t)]$$

when  $\bar{L}_n$  is given by (1.2). Although  $\alpha_n(t)$  is not monotone in  $t$ , the function  $\nu(t)$  will be, under restrictions on  $G$  and  $L_n$ . [Indeed, it will be shown in Theorem 5.1 that the first part of condition (2.3) below is equivalent to the monotonicity of  $\nu(t)$ .]

**THEOREM 2.1.** *Assume that  $L_n$  is given by (1.1), with  $f_n(\cdot)/n$  converging uniformly on  $[0, t_{\max}]$  as  $n \rightarrow \infty$  to  $f$  which is twice boundedly differentiable, where  $t_{\max} < \infty$  does not vary with  $n$ , and assume also*

$$(2.3) \quad f''(t) + \vartheta \int \lambda^2 (h(\lambda) - a) e^{-\lambda t} dG(\lambda) > 0, \quad t \geq 0,$$

$$f'(0) - \vartheta \int \lambda (h(\lambda) - a) dG(\lambda) < 0$$

for the true parameters  $(\vartheta, G)$  governing the discovery-count process  $N(t)$ . If (2.2) holds and the stopping time  $t_n$  is defined by

$$(2.4) \quad t_n \equiv \min \left\{ t_{\max}, \inf \left\{ t > 0: \frac{nf'(t)Q(t)}{\int e^{-\lambda t} \lambda (h(\lambda) - a) G(d\lambda)} \geq N(t) \right\} \right\},$$

then  $t_n$  is an asymptotically optimal stopping time in the sense that, for any alternative  $\mathcal{F}_t$  stopping times  $\tau_n$  bounded by  $t_{\max}$ ,

$$\liminf_{n \rightarrow \infty} EL_n(\tau_n) / EL_n(t_n) \geq 1$$

or

$$\liminf_{n \rightarrow \infty} E(\bar{L}_n(\tau_n) - \bar{L}_n(t_n)) \geq 0.$$

Moreover, as  $n \rightarrow \infty$  the times  $t_n$  tend almost surely to

$$t_* \equiv \inf \left\{ t \in [0, t_{\max}]: f'(t) - \vartheta \int \lambda (h(\lambda) - a) e^{-\lambda t} G(d\lambda) \geq 0 \right\}.$$

**REMARK 2.1.** The asymptotic property established for  $t_n$  with respect to  $L_n$  is asymptotic optimality in the sense of Bickel and Yahav (1969). Asymp-

otic pointwise optimality [Bickel and Yahav (1967, 1968)] could also be shown, but would require some modification of the hypotheses and conclusions of Lemma 2.3 below.

REMARK 2.2. The rule (2.4) of the preceding theorem, in the case where  $h(\cdot) \equiv b > a$  is constant, coincides with the asymptotic stopping rule of Dalal and Mallows (1988).

REMARK 2.3. The role of the second part of condition (2.3) is to ensure that the optimal stopping time  $t_*$  is strictly positive. Then  $\vartheta$  is consistently estimated by  $N(t)/(nQ(t))$  for  $t \geq \varepsilon > 0$ .

REMARK 2.4. A remarkable feature of the asymptotically optimal rule of Theorem 2.1 is that it does not depend on the size-attribute values  $\Lambda_i$ . Thus these rules can be adapted to the case of unknown  $G$ .

REMARK 2.5. In Theorem 2.1, we have assumed that  $f_n/n$  has a finite nonzero limit. In some applications, it may turn out that the limit together with its derivatives is close to 0, that is, the fixed time-dependent costs may be of much smaller order than the random costs. In that case, it is easy to see that the stopping times  $t_n$  approach  $t_{\max}$ . At the other extreme,  $f_n/n$  together with its derivatives may be uniformly large for large  $n$ , and therefore by (2.3) nearly convex:  $t_n$  then approximates the first time at which  $f'_n \geq 0$ .

Since one cannot usually claim to know  $G$  (or equivalently  $Q$ ) in advance, our next result allows the optimal stopping times  $t_n$  to be approximated by times constructed from estimated functions  $G$ .

THEOREM 2.2. Assume all hypotheses of Theorem 2.1, except that now  $G$  is not assumed known [and (2.2) need not hold], and, for each  $t > 0$ ,  $\lambda^2(1 + h(\lambda))e^{-\lambda t}$  is a bounded continuous function of  $\lambda > 0$ . Then  $G$  is identifiable from data  $\{N(u): 0 \leq u \leq \varepsilon\}$  for any  $\varepsilon > 0$ , and if  $\hat{G}_t \equiv \hat{G}_{n,t}$  for  $t \geq \varepsilon$  is any system of a.s. uniformly consistent estimators of  $G$  based on data  $(N(u): 0 \leq u \leq t)$ , the stopping times

$$\xi_n \equiv \min \left\{ t_{\max}, \inf \left\{ t \geq \varepsilon: \frac{nf'(t) \int (1 - e^{-\lambda t}) \hat{G}_t(d\lambda)}{\int e^{-\lambda t} \lambda (h(\lambda) - a) \hat{G}_t(d\lambda)} \geq N(t) \right\} \right\}$$

share the same optimality properties as  $t_n$  in Theorem 2.1 whenever  $t_* \geq \varepsilon > 0$ , where  $t_*$  is as defined in Theorem 2.1.

The following lemma, which is used to prove Theorem 2.1, may be of some independent interest. Suppose that, for each positive integer  $n$ , a (vector-valued) stochastic process  $(X_n(t): 0 \leq t \leq 1)$  is observed, where  $X_n(\cdot) = X_n(\cdot, \omega)$  is a  $D[0, 1]$ -valued random process on a probability space  $(\Omega_n, \mathcal{F}^n, P_n)$ . Suppose also that a nonrandom cost function  $C_n(\cdot, \cdot)$  is given, with  $c_n(t) \equiv C_n(t, X_n(t))$  representing the cost of stopping a random experiment generating data  $X_n$  at time  $t \in (0, 1]$ .

LEMMA 2.3. Assume that  $E(c_n(t)) \equiv E(C_n(t, X_n(t)))$  is a differentiable function of  $t$  for each  $n$ , and that, as  $\varepsilon \rightarrow 0+$ ,

$$(2.5) \quad \limsup_{n \rightarrow \infty} E \sup_{u, v: |u-v| \leq \varepsilon} |c_n(u) - c_n(v)| \rightarrow 0,$$

$$(2.6) \quad \limsup_{n \rightarrow \infty} \sup_{0 \leq t \leq 1-\varepsilon} \left| \frac{1}{\varepsilon} E(c_n(t+\varepsilon) - c_n(t)) - \frac{d}{dt} E(c_n(t)) \right| \rightarrow 0,$$

$$(2.7) \quad \limsup_{n \rightarrow \infty} E \sup_{0 \leq t \leq 1-\varepsilon} \left| \frac{1}{\varepsilon} [c_n(t+\varepsilon) - c_n(t)] - c'(t) \right| \rightarrow 0,$$

where  $c'(t)$  is a strictly increasing and continuous function of  $t$ . Then the nonrandom times

$$t_n = \min \left\{ 1, \inf \left\{ t \in (0, 1] : \frac{d}{dt} EC_n(t, X_n(t)) \geq 0 \right\} \right\}$$

converge to  $t_* = \min\{1, \inf\{t \in [0, 1] : c'(t) \geq 0\}\}$  and satisfy

$$(2.8) \quad \liminf_{n \rightarrow \infty} E[C_n(\tau_n, X_n(\tau_n)) - C_n(t_n, X_n(t_n))] \geq 0$$

for any sequence  $\{\tau_n\}$  of  $\sigma(X_n(v): v \leq t)$  measurable random times on  $\Omega_n$ .

There are several methods available for nonparametric estimation of  $(G, \vartheta)$ , which Fakhre-Zakeri and Slud (1995) found to be identifiable even if the size attributes  $\Lambda_i$  are not observable. (If they were observable, then  $G$  would be estimated by an empirical d.f.) One approach, justified by identifiability plus an argument of Blum and Susarla (1977), is to discretize the  $\lambda$ -axis finely and treat  $G$  as a superposition of finitely many point masses, and then to solve

$$\frac{N(u)}{n} = \vartheta \int (1 - e^{-\lambda u}) G(d\lambda)$$

for a succession of values  $u \leq t$  as a system of linear equations in the jumps of  $\vartheta G(\cdot)$ . Such an estimation method can be refined by estimating  $\vartheta$  and the jumps of  $G$  by a minimum-chi-square method. See Slud (1983) for details and asymptotic justification. Note that the nonparametric maximum likelihood estimator [Lindsay (1983)] for the mixture d.f.  $G$  could not be used directly because  $n_0$  is unknown.

If the unknown d.f.  $G$  were assumed to belong to a suitably regular finite-parameter family  $\{G(\cdot, \mu): \mu \in \mathbb{R}^k\}$ , then the parametric asymptotic theory of van Pul (1992) justifies simultaneous MLEs  $\hat{\vartheta}_t, \hat{\mu}_t$  for  $\vartheta$  and  $\mu$  based on data  $(N(u): 0 \leq u \leq t), t \geq \varepsilon > 0$ , in terms of which the estimators  $\hat{G}_{n,t} \equiv G(\cdot, \hat{\mu}_t)$  will satisfy the requirements of Theorem 2.2

**3. Software reliability application and extensions.** We have mentioned in the Introduction that size-dependent search arises naturally in the statistical analysis of failure data obtained through software reliability testing [Dalal and Mallows (1988) and Fakhre-Zakeri and Slud (1995)]. In this

setting, the scale factor  $n$  in the model of Section 2 might measure the size or scope of applicability of the software system, for example, through the number of modules, total lines of code or complexity of logic. The search in software testing has the features (i) that the size attributes (or hazard for occurrence of failures) due to individual faults or “bugs” will not be observable and (ii) that detection (and immediate fixup and removal) of bugs from a software system may also result in the introduction of new faults (“imperfect debugging”). To model the latter phenomenon, Fakhre-Zakeri and Slud (1995) introduced the following extension and modification of the model of Section 2.

Let the unknown parameter  $p \in [0, 1)$  denote the probability of introducing a new bug (a “descendant” of the one just removed) at the time of a software failure, and assume that such a new bug has size attribute  $\Lambda'$  independently sampled from the same d.f.  $G$  as those initially embedded in the software. [Although it might seem natural to choose a different distribution than  $G$  for new bugs, there is no obvious choice. The issue of estimating parameters related to imperfect debugging is difficult enough, cf. Slud (1994), that we consider here only the model of Fakhre-Zakeri and Slud (1995).] To simplify the analysis, we assume that whenever a bug causes failure, it is removed and is replaced by either zero or one new bugs. [Models with a Poisson random number of new bugs were given by Sumita and Shantikumar (1986) and van Pul (1991) and would be tractable and lead to a development similar to the one below.] Unremoved bugs, old and new, are assumed to cause failure and produce descendants independently.

In this setting, where the large parameter  $n$  both measures the extent of the software and is proportional to the initial number of bugs, it seems natural to allow the time of testing to grow proportionately to some function  $m(n)$  increasing to  $\infty$  with  $n$  and to regard the instantaneous hazard of failure due to any single bug to be inversely proportional to  $m(n)$ . Similar rescaling of time may also be appropriate in other applications of size-dependent search.

The conditional intensities of failure given the bug severities  $\Lambda_i$  are specified here to be constant between failures because the operation of the software in continuous time is idealized as a continuously restarted repetition of the same program with randomly chosen inputs [Littlewood (1989)]. Especially in the usual case where the running time for the program with each input set is much smaller than the total CPU time to be expended on software testing, the failure intensity remains essentially constant between successive observed failures in the gross scale of cumulative CPU time. One consequence of the constancy of hazards is that when hazards are scaled inversely with the large factor  $m(n)$  and also the observation horizon scales proportionately with  $m(n)$ , the counting processes  $N(t)$  of observed failures up to time  $t$  are identical in law to processes  $N(s)$  for  $s \equiv t/m(n)$  within which neither the time horizon nor the intensities vary with  $n$  for each  $i$ . Thus, although time scaling and hazard scaling are appropriate in the software application, we do not treat them explicitly, since all stopping times would scale up by the same factor  $m(n)$ .

For the remainder of this section, the following model is in force. Associated with the initial bugs, labeled  $i = 1, \dots, n_0$ ,  $n_0 = n\vartheta_n$ , where  $\vartheta_n \rightarrow \vartheta$  as  $n \rightarrow \infty$ , we assume that there is an i.i.d. sample  $\Lambda_1, \Lambda_2, \dots, \Lambda_{n_0}$  of real random variates from an unknown distribution function  $G$ . Assume further that the waiting times  $T_i$  until system failure due to the  $i$ th bug are (conditionally given  $\Lambda_1, \Lambda_2, \dots, \Lambda_{n_0}$ ) independent exponential random variables with hazard rates  $\Lambda_i$ . At time  $T_i$ , the bug labeled  $i$  is assumed to be instantaneously removed but with probability  $p$  is replaced by a new bug whose rate  $\Lambda_i$  of causing failure is an independently sampled variable with d.f.  $G$ . The observable data consist only of the failure times which occur up to time  $t_{\max}$ , where  $t_{\max} > 0$  is known.

As in Section 2, let  $r_i(t)$  denote the indicator that bug  $i$  (or its descendant) is still present in the system as of time  $t$ . When  $r_i(t) = 1$ , let  $\Lambda_i(t)$  denote the severity attribute of the descendant of bug  $i$  in the software just before time  $t$ . Note that  $\Lambda_i(t) = \Lambda_i$  (the initial severity) only if bug  $i$  has caused no failures up to  $t$ ; otherwise  $\Lambda_i(\cdot)$  is left-continuous with jumps at each occurrence time for failure due to bug  $i$  (or its descendant), with distinct values i.i.d. with d.f.  $G$ . In the loss functions  $L_n(t)$  and  $\bar{L}_n(t)$  of (1.1) and (1.2), the term  $h(\Lambda_i)$  should now be understood to be replaced by  $h(\Lambda_i(t))$ , which measures losses linked to the frequency of occurrence of bugs in normal use after commercial release of the software. In the following theorem, we replace the previous notation  $\psi_U(t)$  in (2.2) by

$$(3.1) \quad \psi_U(t) = E\{r_1(t)U(\Lambda_1(t))\}$$

noting that, in the case  $p = 0$ , the new definition simplifies to the old one since then  $r_1(t) = I_{[T_1 \geq t]}$  and  $\Lambda_1(t) \equiv \Lambda_1$ .

**THEOREM 3.1.** *Assume that  $L_n$  is given by (1.1), with  $h(\Lambda_i)$  replaced by  $h(\Lambda_i(t))$  and with  $f_n(\cdot)/n$  converging uniformly in  $t \in [0, t_{\max}]$  to twice boundedly differentiable  $f$  as  $n \rightarrow \infty$ . Assume also*

$$(3.2) \quad \begin{aligned} f''(t) + \vartheta(\alpha\psi'_{id} + \psi''_h) &> 0, & t \geq 0, \\ f'(0) + \vartheta(\alpha\psi'_{id} + \psi'_h) &< 0 \end{aligned}$$

for the parameters  $(\vartheta, G)$  governing the discovery-count process  $N(t)$ .

(i) If (2.2) holds and the stopping time  $t_n$  is defined (for known  $p, G$ ) by

$$(3.3) \quad t_n \equiv \min \left\{ t_{\max}, \inf \left\{ t > 0: \frac{-nf'(t)Q(t)}{\alpha\psi'_{id} + \psi'_h} \geq N(t) \right\} \right\},$$

then  $t_n$  is an asymptotically optimal stopping time in the sense that, for any alternative  $\mathcal{F}_t$  stopping times  $\tau_n$  bounded by  $t_{\max}$ ,

$$\liminf_{n \rightarrow \infty} EL_n(\tau_n)/EL_n(s_n) \geq 1$$

or

$$\liminf_{n \rightarrow \infty} E(\bar{L}_n(\tau_n) - \bar{L}_n(s_n)) \geq 0.$$

Moreover, as  $n \rightarrow \infty$  the times  $t_n$  tend almost surely to

$$t_* \equiv \inf\{t \in [0, t_{\max}]: f'(t) + \vartheta(a\psi_{\text{id}} + \psi'_h) \geq 0\}.$$

(ii) Suppose that  $(p, G)$  is not assumed known and that (2.2) need not hold, but that, for data  $\{N(u): 0 \leq u \leq t\}$ , a consistent estimator  $\hat{p}$  of  $p$  and consistent estimators  $\hat{G}_\lambda \equiv \hat{G}_{n,\lambda}$  of  $G$ , uniformly on compact sets of  $\lambda$ , are available. Assume that, for  $t > 0$ ,  $\lambda^2(1 + h(\lambda))e^{-\lambda t}$  is bounded and continuous in  $\lambda > 0$ . Then the stopping times

$$\xi_n \equiv \min\left\{t_{\max}, \inf\left\{t \geq \varepsilon: \frac{nf'(t) \int (1 - e^{-\lambda t}) \hat{G}(d\lambda)}{-a\hat{\psi}_{\text{id}} - \hat{\psi}'_h} \geq N(t)\right\}\right\}$$

share the same optimality properties as  $t_n$  whenever  $t_* \geq \varepsilon > 0$ , where the estimators  $\hat{\psi}_{\text{id}}$  and  $\hat{\psi}'_h$  are defined by substituting  $\hat{p}$  for  $p$  and  $\hat{G}$  for  $G$  in the integral equation (5.8).

REMARK 3.1. Here the situation is not so satisfactory as in Section 2. When  $p > 0$  is unknown, Fakhre-Zakeri and Slud (1995) have shown that  $(G, \vartheta, p)$  is not identifiable from the function  $(EN(t): 0 \leq t \leq t_{\max})$ . However, identifiability can be restored by restricting  $G$  to lie in finite-dimensional families  $\{G(\cdot, \mu)\}$ , and the theory of simultaneous likelihood-based estimation of  $(\mu, \vartheta, p)$  is treated in Slud (1994).

**4. Illustrative example.** We illustrate graphically the stopping rules given by Theorem 2.2, using the system 1 dataset of Musa (1980). The data consist of 136 intervals between observed failures, in CPU time during the system testing phase of a software development project concerning real-time command and control, followed by a final interval of observation during which no further system-test failures occurred. Musa (1980) argued that these data fit a nonhomogeneous Poisson process model well, and elsewhere [Musa and Okumoto (1985)] argued for a particular form of such a model, called the “logarithmic Poisson model.” Fakhre-Zakeri and Slud (1995) showed that this model is a limiting case, as the number  $n_0$  of initial bugs in the software goes to  $\infty$  and the parameter  $\alpha$  goes to 0 as  $k/n_0$  for a constant  $k$ , of the Littlewood (1980, 1981) model in which  $G$  is the Gamma distribution function with parameters  $\alpha, \beta$ . Due to the excellent fit to the data of this model in Figure 1, there is no need to assume imperfect debugging, and we do not. The logarithmic Poisson model used in this reanalysis has the parameterization  $EN(t) = k \ln(1 + t/\beta)$ . The joint maximum likelihood estimators for  $(k, \beta)$  can be found readily, with  $\hat{k}$  given as a function of  $\hat{\beta}$ , and the latter found as the numerical solution of a nonlinear equation.

We consider the problem of optimally stopping the system phase of testing for the Musa system 1 data, with respect to the loss function (1.1) with  $h(\cdot) = b$  constant and with  $f_n(t) = f_0(t + e^{d(t-\eta)})$ , where  $f_0, d$  and  $\eta$  are constants. This is the same choice of loss function used by Dalal and Mallows (1988), Sections 4 and 5, for their illustrations. According to Theorem 2.2, one

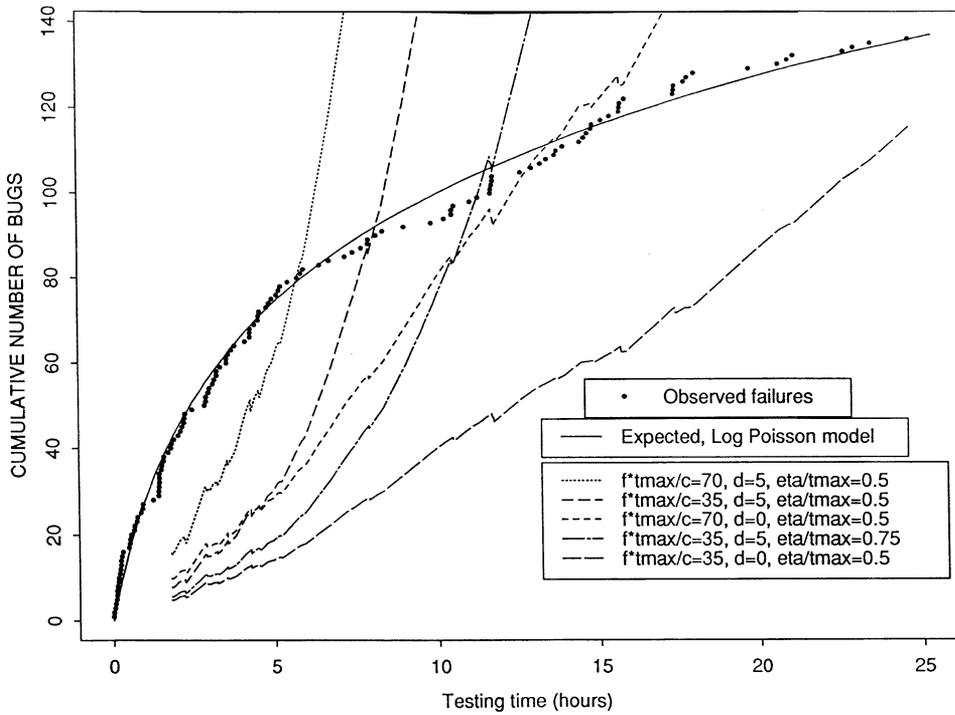


FIG. 1. System 1 data and stopping boundaries.

should optimally stop testing at the first time  $t$ , after some minimum  $\varepsilon$ , at which

$$(4.1) \quad N(t) \leq (f_0/(b-a))(\hat{\beta}_t + t)\ln(1 + t/\hat{\beta}_t),$$

where  $\hat{\beta}_t$  is the MLE of  $\beta$  based on data  $(N(s), 0 \leq s \leq t)$ . The formula (4.1) arises from taking the limit of estimators  $\hat{G}_t$  in a model with  $n_0$  initial bugs and  $G$  the Gamma( $\alpha, \beta$ ) d.f., where  $n_0, n$  and  $\alpha^{-1}$  increase to  $\infty$  in such a way that  $n_0/n \rightarrow \vartheta$ ,  $n_0\alpha \rightarrow k$ . [See Fakhre-Zakeri and Slud (1995), Example 6.2 and Appendix A.3.] Substitution of a Gamma( $\alpha, \beta$ ) distributional form for  $G$  in Theorem 2.2 gives a stopping rule not depending on  $\alpha$  or  $n_0\alpha$  (in the limit of small  $\alpha$ ), so that only the estimator  $\hat{\beta}_t$  enters into the (asymptotic form of the) stopping times  $\xi_n$  in Theorem 2.2.

Figure 1 displays the Musa (1980) system 1 data [the plotted points  $(t, N(t))$ ], the fitted logarithmic-Poisson model [the plotted solid line  $(t, EN(t))$ ] and several choices of the optimal stopping boundary [the right-hand side of (4.1) plotted vs.  $t$  as dashed curves] for various choices of the scaled parameters  $f_0 t_{\max}/c$  (where  $c = b - a$ ),  $d$  and  $\eta/t_{\max}$ . The parameter choices and appearance of the graphs in Figure 1 are similar to those of Dalal and Mallows' (1988) Figure 4. The minimum possible stopping time  $\varepsilon$  was chosen

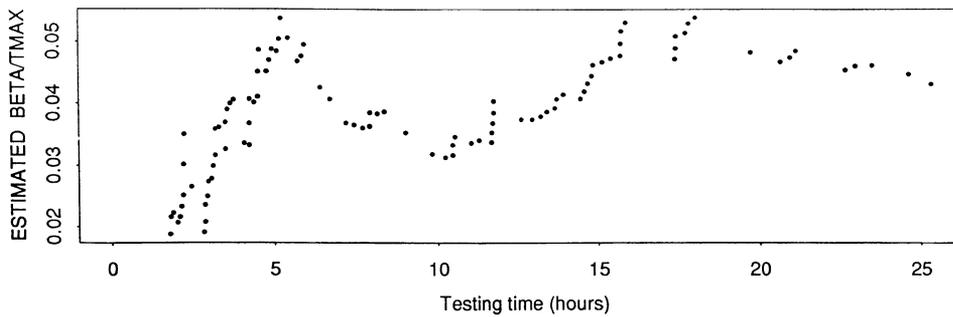


FIG. 2. Estimated beta values as a function of time, system 1.

as the 40th observed failure time, so that the MLEs  $\hat{\beta}_t$  were calculated only for  $t = N^{-1}(40), \dots, N^{-1}(136)$ . Several irregularities in the plotted stopping boundaries can be understood clearly as being due to fluctuations between estimated values  $\hat{\beta}_t$  at  $t$  ranges from the 40th through 136th observed failure times, and these MLEs are plotted as a function of testing time  $t$  in Figure 2. The considerable variability of  $\hat{\beta}_t$  over the first half of the testing period  $[0, t_{\max}]$  strongly suggests that very early stopping of software reliability testing would be unwise.

**5. Proofs.** We now prove the main theorems of the previous sections, as well as some sufficient conditions for the first lines of each of the hypotheses (2.3) and (3.2). Since Theorems 2.1 and 2.2 are just the special cases of Theorem 3.1 in which  $p = 0$ , we prove only the latter theorem. In order to understand the proofs more simply when  $p = 0$ , that is, in the usual case where there are no replacements of discovered items, simply read  $\Lambda_i$  for  $\Lambda_i(t)$  in what follows.

**PROOF OF LEMMA 2.3.** First, since  $c'$  is strictly increasing,  $t_*$  is well defined, and, by the continuity of  $c'$ , either  $c'(1) < 0$  or  $c'(t_*) = 0$ . Without loss of generality, assume that  $EC_n(0, X_n(0)) = 0$ , since otherwise we could replace  $c_n(t)$  by  $C_n(t, X_n(t)) - C_n(0, X_n(0))$ , leaving (2.5)–(2.7) still valid and the definition of  $t_n$  and assertion of (2.8) unchanged. It follows immediately from (2.5) that

$$(5.1) \quad \limsup_{n \rightarrow \infty} E \sup_{0 \leq t \leq 1} |c_n(t)| < \infty$$

and, as  $\varepsilon \rightarrow 0+$ ,

$$(5.2) \quad \limsup_{n \rightarrow \infty} E \sup_{0 \leq t \leq 1 - \varepsilon} \left| \frac{1}{\varepsilon} \int_t^{t+\varepsilon} c_n(u) du - c_n(t) \right| \rightarrow 0.$$

Now, with  $c(t) \equiv \int_0^t c'(s) ds$ , use the calculation

$$\begin{aligned} & \int_s^t \left[ \frac{c_n(u + \varepsilon) - c_n(u)}{\varepsilon} - c'(u) \right] du \\ &= \varepsilon^{-1} \left\{ \int_t^{t+\varepsilon} c_n(u) du - \int_s^{s+\varepsilon} c_n(u) du \right\} - c(t) + c(s) \end{aligned}$$

to conclude from (2.7), (5.1), (5.2) and the triangle inequality that

$$(5.3) \quad E \sup_{0 \leq s, t \leq 1} |c_n(t) - c_n(s) - c(t) + c(s)| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

By (2.6), (2.7) and the triangle inequality,

$$(5.4) \quad \limsup_{n \rightarrow \infty} \sup_{0 \leq t < 1} \left| \frac{d}{dt} E c_n(t) - c'(t) \right| = 0.$$

Therefore, by the definition of  $t_n$  and the continuity of  $c'$ , either  $c'(1) < 0$  and

$$\left. \frac{d}{dt} E(c_n(t)) \right|_{t=1} < 0$$

and  $t_n = 1$  for all large  $n$ , or, as  $n \rightarrow \infty$ ,

$$\left. \frac{d}{dt} E c_n(t) \right|_{t=t_n} \rightarrow 0$$

and

$$c'(t_n) \rightarrow 0.$$

In either case, the strict increase and continuity of  $c'$  imply  $t_n \rightarrow t_*$ . Apply (2.5) to conclude that  $E|C_n(t_n, X_n(t_n)) - C_n(t_*, X_n(t_*))|$  converges to 0. Now, since  $t_*$  is the unique strict minimizer of  $c$  on  $[0, 1]$  apply (5.3) with  $s$  replaced by the arbitrary stopping times  $\tau_n$  and  $t$  replaced by  $t_n$  to obtain

$$\liminf_{n \rightarrow \infty} E[c_n(\tau_n) - c_n(t_n)] = \liminf_{n \rightarrow \infty} E[c(\tau_n) - c(t_*)] \geq 0,$$

which completes the proof of (2.8).  $\square$

PROOF OF THEOREM 3.1(i). Lemma 2.3 applies directly, with

$$X_n(t) \equiv \left[ \begin{array}{c} N(t)/n \\ n^{-1} \sum_{i \leq n_0} r_i(t) h(\Lambda_i(t)) \end{array} \right],$$

where  $\Lambda_i(t)$  can be replaced by  $\Lambda_i$  in the setting of Section 2, and

$$C_n(t, x) = n^{-1} f_n(t) + ax_1 + x_2,$$

so that  $C_n(t, X_n(t)) \equiv \bar{L}_n(t)$  as in (1.2). We proceed to check the hypotheses (2.5)–(2.7) of Lemma 2.3 for  $X_n$  and  $C_n$  just defined (where we ignore the difference between  $[0, t_{\max}]$  and  $[0, 1]$ ). Observe that  $X_n(t)$  satisfies a uniform law of large numbers as  $n \rightarrow \infty$  [Ranga Rao (1962), Theorem 6.5, on the separable Banach space  $D[0, 1]$  with Skorohod's  $J_1$ -topology metrized by the

$d_0$ -norm; cf. Billingsley (1968), pages 112–115], with the limit

$$(5.5) \quad \vartheta \left[ \int_0^t \psi_{\text{id}}(u) \, du, \psi_h(t) \right]^\top.$$

Moreover, the limiting vector function (5.5) is twice continuously differentiable in  $t$  on all of  $[0, t_{\max}]$ , as long as (2.2) holds. [If all that is desired is two continuous derivatives on  $[\varepsilon, t_{\max}]$  for  $\varepsilon > 0$  as in Theorem 3.1(ii), then (2.2) is not necessary.] Apart from the nonrandom terms involving  $f_n/n$ , the expectations in each of (2.5)–(2.7) depend on  $n$  only through  $n_0/n \equiv \vartheta_n \rightarrow \vartheta$ , and the hypotheses (2.5)–(2.7) of Lemma 2.3 are easily verified. The only hypothesis left to check is the strict monotonicity of the function  $c'$ , that is, that

$$\lim_{n \rightarrow \infty} \frac{d}{dt} E[\bar{L}_n(t)]$$

is increasing, and Theorem 5.1 below establishes this under the hypothesis (3.2). Thus the lemma applies, yielding the conclusion that the stopping times

$$s_n \equiv \min \left\{ t_{\max}, \inf \left\{ t > 0: \frac{1}{n} f'_n(t) + \vartheta_n (\alpha \psi_{\text{id}}(t) + \psi'_h(t)) \geq 0 \right\} \right\}$$

are asymptotically optimal and converge to the nonrandom time  $t_*$  defined as the smaller of  $t_{\max}$  and the smallest  $t \in [0, t_{\max}]$ , where

$$f'(t) + \vartheta (\alpha \psi_{\text{id}}(t) + \psi'_h(t)) \geq 0.$$

The second part of (3.2) implies that  $t_* > 0$ . Then, since  $N(t)/n\vartheta$  is uniformly convergent (a.s. and in the mean) to  $Q(t)$  as  $n \rightarrow \infty$ , the proof of the lemma implies that  $t_n$  given in the theorem converges in probability to  $t_*$ , hence that  $t_n$  is also asymptotically optimal.  $\square$

PROOF OF THEOREM 3.1(ii). The proof builds upon the proof of Theorem 3.1(i). As remarked following (2.2), the twice continuous differentiability of  $\psi_{\text{id}}$  and  $\psi_h$  on  $[\varepsilon, t_{\max}]$  for  $\varepsilon > 0$  does not require any integrability assumption such as (2.2). Therefore, since the time interval of interest is now  $[\varepsilon, t_{\max}]$ , we apply the lemma on that interval in place of  $[0, 1]$ , and we dispense with (2.2).

As before, Lemma 2.3 yields asymptotically optimal stopping times  $t_n$  converging to  $t_*$  (assumed  $\geq \varepsilon$ ), as well as asymptotic optimality of any other sequence of stopping times  $t_n$  converging in probability to  $t_*$ . Substituting  $\hat{G}$  for  $G$  and  $\hat{p}$  for  $p$  in the definition of (estimators  $\hat{\psi}_U$  and  $\hat{\psi}'_U$  in) the stopping times  $t_n$  of part (i) yields new stopping times  $\xi_n$  which are easily seen to differ asymptotically negligibly in probability from  $t_n$ . (A key point here is the uniform consistency on  $[\varepsilon, t_{\max}]$  of the estimators  $\hat{\psi}_U$  and  $\hat{\psi}'_U$ , which follows from the assumptions on  $\hat{p}$ ,  $\hat{G}$  and  $h$ .) Thus the stopping times  $\xi_n$  of part (ii) of the theorem are asymptotically optimal and converge to  $t_*$  as asserted.  $\square$

This completes the proof of Theorem 3.1. Important special cases will be treated in detail after Theorem 5.1, and Corollaries 5.2 and 5.3 below provide conditions under which the first line of (3.2) is satisfied. In particular, when  $p = 0$  and  $f$  is convex and the minimum cost  $h(0)$  of an undiscovered bug is larger than the cost  $a$  of finding it [i.e.,  $h(0) \geq a$ ], then the first line of (3.2) holds and need not be separately assumed.

**THEOREM 5.1.** *Assume first that  $f(t) \equiv f_n(t)/n$  and that (2.2) holds. Then*

$$\nu(t) \equiv \frac{d}{dt} E[\bar{L}_n(t)]$$

*is strictly increasing if and only if*

$$(5.6) \quad f''(t) + \vartheta_n(a\psi'_{id}(t) + \psi''_h(t)) > 0 \quad \text{for all } t \geq 0,$$

*where  $\vartheta_n \equiv n_0/n$ . If, instead,  $f(t)$  exists as the uniform limit of  $f_n(t)/n$ , then (5.6) implies that  $(d/dt)E[\bar{L}_n(t)]$  is asymptotically increasing in the sense that, for  $t > s$ ,*

$$\liminf_{n \rightarrow \infty} \left( \frac{d}{dt} E[\bar{L}_n(t)] - \frac{d}{dt} E[\bar{L}_n(s)] \right) > 0.$$

**PROOF.** First, note that if  $\Lambda_i(u)$  again denotes the failure-rate parameter for the descendant of bug  $i$  still present in the software as of time  $u$ , for  $i = 1, 2, \dots, n_0$ , then the compensator of the counting process  $N(t)$  with respect to the filtration  $\sigma(N(u), \Lambda_i(u), r_i(u); i = 1, \dots, n_0, 0 \leq u \leq t)$  is

$$\int_0^t \sum_{i=1}^{n_0} r_i(u) \Lambda_i(u) du,$$

which implies that

$$(5.7) \quad x^{-1} E[N(t+x) - N(t)] \rightarrow n_0 E\{r_1(t) \Lambda_1(t)\} \quad \text{as } x \rightarrow 0+.$$

Thus, to compute the function

$$\nu(t) = \frac{d}{dt} E[\bar{L}_n(t)],$$

it suffices to study (the derivatives of)  $\psi_{id}(t)$  and  $\psi_h(t)$ . Properties of the latter functions arise from the following application of the renewal argument. For given nondecreasing function  $U$  which is integrable with respect to  $dG$ , by conditioning on the time  $T_1$  of occurrence of the first software failure due to bug 1,

$$(5.8) \quad \begin{aligned} \psi_U(t) &= \int_0^\infty e^{-\lambda t} U(\lambda) dG(\lambda) + p \int_0^\infty \int_0^t \lambda e^{-\lambda u} \psi_U(t-u) du dG(\lambda) \\ &= \int_0^\infty e^{-\lambda t} U(\lambda) dG(\lambda) + p \int_0^t q(t-u) \psi_U(u) du, \end{aligned}$$

where we recall that  $q(t) = \int_0^\infty \lambda e^{-\lambda t} dG(\lambda)$  is the density of  $T_i$ . Thus  $\psi_U$  is twice continuously differentiable if  $\int \lambda^2 U(\lambda) dG(\lambda) < \infty$ . Since  $f$  is twice differentiable, by (1.2) and (5.7),

$$\nu'(t) = f''(t) + \vartheta_n(a\psi'_{id}(t) + \psi''_h(t)).$$

The theorem is proved.  $\square$

Next, we provide a sufficient condition for (5.6) to hold, when  $h(\cdot) \equiv b > 0$  is constant. First, let  $\gamma^*\psi_U(t) = \int_0^t \gamma(x)\psi_U(t-x) dx$  for an arbitrary differentiable function  $\gamma$ . Then

$$(5.9) \quad \begin{aligned} (\gamma^*\psi_U)'(t) &= \gamma(t)\psi_U(0) + \gamma^*\psi'_U(t), \\ (\gamma^*\psi_U)''(t) &= \gamma'(t)\psi_U(0) + \gamma(t)\psi''_U(0) + \gamma^*\psi''_U(t). \end{aligned}$$

From (5.8), with  $\gamma(t) \equiv q(t) = \int_0^\infty \lambda e^{-\lambda t} dG(\lambda)$ , we have

$$(5.10) \quad \begin{aligned} \psi_{id}(t) &= \int_0^\infty \lambda e^{-\lambda t} dG(\lambda) + pq^*\psi_{id}(t), \\ \psi_h(t) &= \int_0^\infty h(\lambda)e^{-\lambda t} dG(\lambda) + pq^*\psi_h(t). \end{aligned}$$

Write  $\mathcal{Z}(t) = a\psi'_{id}(t) + \psi''_h(t)$ . Then, in view of (5.9) and (5.10),

$$\begin{aligned} \mathcal{Z}(t) &= \int \lambda^2 e^{-\lambda t} (h(\lambda) - a) dG(\lambda) \\ &\quad + p[aq(t)\psi_{id}(0) + q'(t)\psi_h(0) + q(t)\psi'_h(0) + q^*\mathcal{Z}(t)]. \end{aligned}$$

**COROLLARY 5.2.** *Assuming (2.2) and  $h(\cdot) \equiv b \geq a/(1-p) > 0$ , a sufficient condition for (5.6) is that, for all  $t \geq 0$ ,*

$$(5.11) \quad \begin{aligned} f''(t) &\geq \vartheta_n(b(1-p) - a) \left\{ p \int \lambda dG(\lambda) \right. \\ &\quad \left. - \inf(\text{supp}(G)) \right\} \sum_{m=0}^{\infty} p^m q^{*(m+1)}(t). \end{aligned}$$

**PROOF.** The renewal-type equation for  $\mathcal{Z}$  preceding the corollary is  $\mathcal{Z}(t) = \nu(t) + pq^*\mathcal{Z}(t)$ , where

$$\begin{aligned} \nu(t) &\equiv \int \lambda^2 e^{-\lambda t} (h(\lambda) - a) dG(\lambda) \\ &\quad + p[aq(t)\psi_{id}(0) + q'(t)\psi_h(0) + q(t)\psi'_h(0)]. \end{aligned}$$

Now substitute  $q(t) = \int \lambda e^{-\lambda t} dG(\lambda)$ ,  $h \equiv b$ , and use (5.9) and (5.10) at  $t = 0$  to conclude that

$$\nu(t) = (b(1-p) - a) \left( \int \lambda^2 e^{-\lambda t} dG(\lambda) - pq(t) \int \lambda dG(\lambda) \right).$$

Since  $\int \lambda^2 e^{-\lambda t} dG(\lambda) / \int \lambda e^{-\lambda t} dG(\lambda)$  is a decreasing function of  $t$  and has limit

equal to  $\inf(\text{supp}(G)) = \inf\{y > 0: G(y) > 0\}$ , we have shown

$$\nu(t) \geq q(t)(b(1-p) - a) \left( \inf(\text{supp}(G)) - p \int \lambda dG(\lambda) \right).$$

Now, by repeated substitution in the renewal-type equation  $\mathcal{Z}(t) = \nu(t) + pq^* \mathcal{Z}(t)$ ,

$$\mathcal{Z}(t) = \nu(t) + pq^* \nu(t) + \cdots + p^m q^{*(m)} * \nu(t) + p^{m+1} q^{*(m+1)} * \mathcal{Z}(t).$$

It follows by letting  $m \rightarrow \infty$  that

$$\begin{aligned} \mathcal{Z}(t) &= \sum_{m \geq 0} p^m q^{*(m)} * \nu(t) \\ &\geq (b(1-p) - a) \left\{ \inf(\text{supp}(G)) - p \int \lambda dG(\lambda) \right\} \sum_{m=0}^{\infty} p^m q^{*(m+1)}(t). \end{aligned}$$

Therefore, by (5.11),  $f''(t) + \vartheta_n \mathcal{Z}(t) \geq 0$  as asserted in (5.6).  $\square$

Another special case arises if  $p = 0$ , when condition (5.6) becomes

$$(5.12) \quad f''(t) + \vartheta_n \int \lambda^2 (h(\lambda) - a) e^{-\lambda t} dG(\lambda) \geq 0 \quad \text{for } t \geq 0.$$

Since  $h$  is assumed to be a nondecreasing function, it follows that

$$\frac{\int \lambda^2 e^{-\lambda t} h(\lambda) dG(\lambda)}{\int \lambda^2 e^{-\lambda t} dG(\lambda)}$$

is a decreasing function of  $t$ , and the limit of this ratio as  $t$  becomes large is easily seen to be  $h(\inf(\text{supp}(G))) \geq h(0)$ . Thus we have the following result.

**COROLLARY 5.3.** *When  $p = 0$  and  $h$  is nondecreasing, a sufficient condition for (5.6) is*

$$(5.13) \quad \inf_{t \geq 0} \left\{ f''(t) / \int \lambda^2 e^{-\lambda t} dG(\lambda) \right\} + \vartheta_n (h(0) - a) > 0.$$

## REFERENCES

- ANDREATTA, G. and KAUFMAN, G. (1986). Estimation of finite population properties when sampling is without replacement and proportional to magnitude. *J. Amer. Statist. Assoc.* **81** 657–666.
- BAROUCH, E. and KAUFMAN, G. (1976). Probabilistic modelling of oil and gas discovery. In *Energy: Mathematics and Models* (Fred Roberts, ed.) 133–152. SIAM, Philadelphia.
- BICKEL, P., NAIR, V. and WANG, P. (1992). Nonparametric inference under biased sampling from a finite population. *Ann. Statist.* **20** 853–878.
- BICKEL, P. and YAHAV, J. (1967). Asymptotically pointwise optimal procedures in sequential analysis. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 401–413. Univ. California Press, Berkeley.
- BICKEL, P. and YAHAV, J. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.* **39** 442–456.
- BICKEL, P. and YAHAV, J. (1969). An asymptotically pointwise optimal rule in sequential estimation with quadratic loss. *Ann. Math. Statist.* **40** 417–426.

- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- BLUM, J. and SUSARLA, V. (1977). Estimation of a mixing distribution function. *Ann. Probab.* **5** 200–209.
- DALAL, S. R. and MALLOWS, C. (1988). When should one stop testing software? *J. Amer. Statist. Assoc.* **83** 872–879.
- FAKHRE-ZAKERI, I. and SLUD, E. (1995). Mixture models for reliability of software with imperfect debugging: identifiability of parameters. *IEEE Trans. Reliability* **44** 104–113.
- GORDON, L. (1983). Successive sampling in large finite populations. *Ann. Statist.* **11** 702–706.
- KRAMER, M. and STARR, N. (1987). Discovery rates for a search in which sampling is proportional to size. *Adv. in Appl. Math.* **8** 222–236.
- KRAMER, M. and STARR, N. (1990). Optimal stopping in a size dependent search. *Sequential Anal.* **9** 59–80.
- LINDSAY, B. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11** 86–94.
- LITTLEWOOD, B. (1980). Theories of software reliability: how good are they and how can they be improved? *IEEE Trans. Software Engrg.* **6** 489–500.
- LITTLEWOOD, B. (1981). Stochastic reliability growth: a model for fault-removal in computer-programs and hardware-designs. *IEEE Trans. Reliability* **30** 313–320.
- LITTLEWOOD, B. (1989). Predicting software reliability. *Philos. Trans. Roy. Soc. London Ser. A* **327** 513–527.
- MILLER, D. (1986). Exponential order statistics of software reliability growth. *IEEE Trans. Software Engrg.* **12** 12–24.
- MUSA, J. (1980). *Software Reliability Data*. Data and Analysis Center for Software, Rome, NY.
- MUSA, J. and OKUMOTO, K. (1985). A logarithmic Poisson execution time model for software reliability measurement. *Proceedings of the Seventh International Conference on Software Engineering* 343–348. IEEE, New York.
- RANGA RAO, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.
- ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I, II. *Ann. Math. Statist.* **43** 373–397, 748–776.
- ROSS, S. (1971). Infinitesimal look-ahead stopping-rules. *Ann. Math. Statist.* **42** 297–303.
- SEBER, G. (1992). A review of estimating animal abundance. II. *Internat. Statist. Rev.* **60** 129–166.
- SHAPIRO, C. and WARDROP, R. (1980). Dynkin's identity applied to Bayes sequential estimation of a Poisson process rate. *Ann. Statist.* **8** 171–182.
- SLUD, E. (1983). An empirical Bayes goodness of fit test. Technical Report 598, School of Operations Research and Industrial Engineering, Cornell Univ.
- SLUD, E. (1994). Testing for imperfect debugging in software reliability. Unpublished manuscript.
- SUMITA, U. and SHANTIKUMAR, J. (1986). A software reliability model with multiple-error introduction and removal. *IEEE Trans. Reliability* **35** 459–462.
- VAN PUL, M. (1991). Modelling imperfect repair and software growth. *Proceedings of the Interface '91 Conference, Seattle*.
- VAN PUL, M. (1992). Asymptotic properties of statistical models in software reliability. *Scand. J. Statist.* **19** 235–254.
- VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *Ann. Statist.* **10** 772–785.

DEPARTMENT OF STATISTICS  
 UNIVERSITY OF NORTH CAROLINA  
 AT CHAPEL HILL  
 CB #3260, PHILLIPS HALL  
 CHAPEL HILL, NORTH CAROLINA 27599-3260  
 E-MAIL: issa@stat.unc.edu

DEPARTMENT OF MATHEMATICS  
 UNIVERSITY OF MARYLAND  
 COLLEGE PARK, MARYLAND 20742  
 E-MAIL: evs@math.umd.edu