The Annals of Applied Probability
1992, Vol. 2, No. 3, 752–765

BUYING WITH EXACT CONFIDENCE

BY S. R. DALAL AND C. L. MALLOWS

Bellcore and AT & T Bell Laboratories

We derive some results which may be helpful to buyers of software testing for faults, or to buyers of large lots screening for defectives. Suppose that a fixed but unknown number n of faults or defectives remain before testing. In the testing phase they are observed at random times, $X_1, X_2, \ldots, X_n$, which are order statistics corresponding to n i.i.d. random variables. Since testing is usually an ongoing activity, this distribution is typically known. Under this assumption we derive a stopping criterion that guarantees, for any specified level α and integer m, that for all $n > m$, with probability exactly $1 - \alpha$, when stopping occurs, the software has no more than m faults remaining. We study various properties of this stopping rule, both finite and asymptotic, and show that it is optimal in a certain sense. We modify a conservative stopping rule proposed by Marcus and Blumenthal to make it exact, and we give some numerical comparisons.

1. Introduction. Suppose a fixed but unknown number n of events occur successively at times $X_1, X_2, \ldots$. We assume that unordered X's are independent positive random variables, with a known common distribution F, which we assume is continuous. We can observe these events as they occur, but cannot afford to keep on observing indefinitely. Let $K(t)$ be the number of events that have been observed by time t, that is, the number of indices j such that $X_j \le t$. A stopping rule for this problem will be defined by an increasing sequence $b_1, b_2, \ldots$; we stop at time $\tau = b_J$, where

$$J = \text{the smallest } j \text{ such that } K(b_j) < j. \tag{1}$$

We will show that for any $m \ge 0$ and α in $(0, 1)$, we can choose $b_1, b_2, \ldots$ so that, when we stop, we have confidence (at level exactly $1 - \alpha$) that at most m events have yet to occur. More precisely, we shall achieve:

$$\text{for all } n > m, \quad P(n - K(\tau) > m | n) = \alpha. \tag{2}$$

It will follow that if the number of events is a random variable N with arbitrary distribution supported on $(m + 1, \infty)$, then

$$P(N - K(\tau) > m) = \alpha.$$

Of course, for a general stopping rule b_j may depend on $X_1, X_2, \ldots, X_{j-1}$, but we shall show that constant b's suffice to achieve a certain optimality result.

The above formulation is suggested as a model for the situation facing someone responsible for testing software before release and as an approximation for the problems of (i) deciding when enough proofreading has been done

Received December 1990; revised October 1991.
AMS 1980 *subject classifications*. 60G40, 62G30, 62L15, 62N99.
Key words and phrases. Software testing, optimal stopping, sampling inspection.

and (ii) sequential sampling inspection of finite lots. It is related to some standard models. If we let N be a Poisson random variable with mean λ, then the times $X_1, X_2, \dots$ form a Poisson process with time-dependent rate $\lambda f(t)$. This model is very commonly used in the software field [see Musa, Iannino and Okumoto (1987)], although there it need not be assumed that the integral of f is finite.

Dalal and Mallows (1988) consider the stopping problem from the point of view of the developing organization and derive an optimum rule. The present formulation is more appropriate from the point of view of the purchaser of the software, who may desire to specify m and α and who requires that the software be certified to contain no more than m faults, with confidence α. In Section 8 we examine the economic tradeoff that arises here.

Concerning the proofreading application, several authors [e.g., Chow and Schechner (1985)] have assumed that for each misprint in the text there is a constant probability that it will be found on any one reading, with all misprints acting independently. Thus the number of readings to find a given misprint has a geometric distribution. Our model approximates this by a continuous distribution. Wallenius (1967) discussed sequential sampling of finite lots, each element of which may be defective. If we imagine the items spread out in order along the real axis, the problem is exactly like ours except that time is discrete.

There are several other applications in which a model similar to the one discussed here is appropriate. Marcus and Blumenthal (1974) discuss many of them in our framework and propose a conservative stopping rule. In Section 6 we show that our rule is optimal under broad assumptions. We also show in Section 8 that the Marcus–Blumenthal procedure can be modified to yield an exact confidence statement. We present a numerical comparison of our procedure with this modified procedure.

Our problem is related to that of giving a distribution-free confidence band for an unknown distribution function; see Wald and Wolfowitz (1939). We shall use one of their formulas in our derivation.

2. Preliminaries. Since F is known, without loss of generality we can perform a probability-integral transformation, so we take F to be the standard uniform distribution on $(0, 1)$.

The ordered values of a random sample of size n will be written $(X_{1,n}, \dots, X_{n,n})$. Inequalities between vectors are to be interpreted coordinatewise. We write $X \leq_P Y$ to denote that the random vector $X = (X_a, \dots, X_b)$ is stochastically smaller than $(Y_a, \dots, Y_b)$, that is,

$$\text{for all } x = (x_a, \dots, x_b), \qquad P(X \leq x) \geq P(Y \leq x).$$

The binomial coefficient "n choose k" will be written $C(n, k)$. We write B_n for the space of sequences, b's which satisfy

$$0 \leq b_1 \leq b_2 \leq \cdots \leq b_n \leq 1.$$

3. The case $m = 0$. Suppose b is in B_n. If $X_{1,n}, \dots, X_{n,n}$ is an ordered random sample of size n from the uniform distribution on $(0,1)$, we are interested in the probability

$$q_n = q(b_1, \dots, b_n) = P(X_{1,n} < b_1, \dots, X_{n,n} < b_n). \tag{3}$$

In the case $m = 0$ we want $q_n = 1 - \alpha$ for all $n \geq 1$. For $n = 1$ we have $q_1 = b_1$ so $b_1 = 1 - \alpha$. For $n = 2$ we find

$$q_2 = 2b_1 b_2 - b_1^2,$$

so we can take $b_2 = 1 - \alpha/2$. Appendix A, which constitutes a proof of Theorem 1, shows that this construction can be repeated indefinitely, thus defining an increasing sequence $b_1, b_2, \dots$ that satisfies (2) (with $m = 0$) for all $n \geq 1$.

THEOREM 1. *For $m = 0$ and $0 < \alpha < 1$ and for all $k \geq 2$, $b_{k-1} < b_k < 1$.*

For computing these b's it is convenient to work in terms of $c_i = 1 - b_i$. It follows from a result of Wald and Wolfowitz (1939) [their equation (27)] that

$$q_n = \begin{vmatrix} 1 & c_1 & c_1^2 & \cdots & c_1^{n-1} & c_1^n \\ 0 & 1 & 2c_2 & \cdots & (n-1)c_2^{n-1} & nc_2^{n-1} \\ 0 & 0 & 1 & \cdots & C(n-1,2)c_3^{n-3} & C(n,2)c_3^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & nc_n \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{vmatrix}. \tag{4}$$

The (i, j) element in this determinant (for $1 \leq i \leq n$, $1 \leq j \leq n+1$) is

$$C(j-1, i-1)c_i^{j-i}.$$

Expanding the determinant in (4) by its last column, we have

$$q_n + nc_n q_{n-1} + C(n,2)c_{n-1}^2 q_{n-2} + \cdots + nc_2^{n-1} q_1 + c_1^n = 1. \tag{5}$$

Now setting $q_1 = q_2 = \cdots = 1 - \alpha$ we can determine the c's in order.

For $k = 0, 1, \dots, n-1$, the $k+1$-th term on the l.h.s. of (5) is exactly $P(\tau = b_{n+1-k} | n) P(K(\tau) = n - k | n)$. To see this, let us write

$$D(x, n) = \text{largest } j \text{ such that } X_{j,n} < x.$$

Now observe that for $j \geq 0$,

$$\begin{aligned} P(\tau = b_{j+1} | n) &= P(X_{1,n} < b_1, \dots, X_{j,n} < b_j, X_{j+1,n} > b_{j+1}) \\ &= P(X_{1,n} < b_1, \dots, X_{j,n} < b_j, D(b_{j+1}, n) = j). \end{aligned}$$

However,

$$\begin{aligned} P\big(X_{1,n} < b_1, \ldots, X_{j,n} < b_j \mid D(b_{j+1}, n) = j\big) \\ = P\big(X_{1,j} < b_1, \ldots, X_{j,j} < b_j \mid D(b_{j+1}, j) = j\big) \\ = P(X_{1,j} < b_1, \ldots, X_{j,j} < b_j)\big/b_{j+1}^j \end{aligned}$$

so that

$$\begin{aligned} P\big(\tau = b_{j+1} | n\big) &= P(X_{1,j} < b_1, \ldots, X_{j,j} < b_j) P\big(D(b_{j+1}, n) = j\big)\big/b_{j+1}^j \\ &= q_j C(n,j) b_{j+1}^j (1 - b_{j+1})^{n-j} / b_{j+1}^j \qquad (6) \\ &= (1 - \alpha) C(n,j) c_{j+1}^{n-j}. \end{aligned}$$

4. The general case. When $m \geq 1$ and $n = m + k$, $k \geq 1$, the quantity (3) is replaced by

$$(7) \qquad p_{m+k}^{(m)} = p^{(m)}(b_1, \ldots, b_k) = P(X_{1,m+k} < b_1, \ldots, X_{k,m+k} < b_k)$$

and we want $p_{m+k}^{(m)} = 1 - \alpha$ for $k = 1, 2, \ldots$. Notice that $p^{(m)}(b_1, \ldots, b_k)$ concerns a sample of size $m + k$. We can get (7) from (3) by setting $b_{k+1} = \cdots = b_{m+k} = 1$, so we must put $c_{k+1} = \cdots = c_{m+k} = 0$ in (4). Expanding the resulting determinant by its last column, we find

$$(8) \qquad \begin{aligned} p_{m+k}^{(m)} &= 1 - C(m + k, k - 1) q_{k-1} c_k^{m+1} - C(m + k, k - 2) \\ &\quad \times q_{k-2} c_{k-1}^{m+2} - \cdots - (m + k) q_1 c_2^{m+k-1} - c_1^{m+k}, \end{aligned}$$

where the q's are still defined by (4). Using (8) and (5) alternately, we can determine in order

$$c_1, q_1, c_2, q_2, c_3, \ldots$$

so that $p_{m+k}^{(m)} = 1 - \alpha$ for $k = 1, 2, \ldots$. We have not seen how to extend the argument in Appendix A to cover this case, but we have strong numerical evidence that for all m and α, $c_1 > c_2 > \cdots$. Specifically, we have computed $c_1, \ldots, c_{10}$ for $m = 1, 2, 4, 8, 16$ and $\alpha = 0.05(0.05)0.95$, finding results that are completely consistent with the asymptotic formulas that we give below.

The same argument that led to (6) now shows that

$$P\big(\tau = b_{j+1} \mid n\big) = q_j C(n,j) c_{j+1}^{n-j}.$$

5. Asymptotics. We can derive the asymptotic forms of these stopping rules, and the asymptotic behavior of $n - K(\tau)$. First consider the case $m = 0$. In (5), multiply both sides by $y^n/n!$ and sum on n from 1 to ∞. This gives

$$(9) \qquad \frac{e^y - \alpha e^{\alpha y}}{1 - \alpha} = e^{yc_1} + y e^{yc_2} + \frac{y^2}{2!} e^{yc_3} + \cdots.$$

The c's can be determined in order by equating coefficients on both sides of (9). Notice that as $\alpha \to 0$, $c_k \to 1$ for all k; but the limit as $\alpha \to 1$ is nontrivial. In this case we find the values

$$(c_1, c_2, c_3, c_4, c_5) = \left(1, \tfrac{1}{2}, \tfrac{5}{12}, \tfrac{35}{96}, \tfrac{22601}{69120}\right),$$

but we have no general formula.

Now multiply (9) by e^{-y} and suppose y is a large integer, n. The r.h.s. of (9) is a Poisson-weighted average of e^{nc_k}, and is dominated by the terms near $k = n$. The l.h.s. is very nearly $1/(1 - \alpha)$. Let us write

$$1 - \alpha = e^{-\lambda}.$$

Then we have the following theorem.

THEOREM 2. *For $m = 0$ and $0 < \alpha < 1$, $c_n = \lambda/n + o(n^{-1})$.*

For a rigorous proof of this theorem see Appendix B. The rest of the results in this section are in analogy with the argument preceding Theorem 2 and are based on formal expansions. We shall not attempt rigorous proofs.

Now consider the terms on the l.h.s. of (5). We have $q_k = 1 - \alpha = e^{-\lambda}$ for all k, so for $k \ll n$ the $(k + 1)$-th term in (5) is approximately

$$e^{-\lambda} \frac{\lambda^k}{k!}.$$

Thus, by the comment immediately following (5), when n is large the stopping time b_J is such that $n + 1 - J$ (i.e., the number of faults not found) has very nearly a Poisson distribution with mean λ.

We can get more detailed results from (5) by assuming

$$c_n \approx \frac{\lambda}{n}\left(1 + \frac{a}{n}\right) + O(n^{-2}).$$

Expanding and collecting terms, we find $a = -\lambda/2$ so that

$$c_n = \frac{\lambda}{n + \lambda/2} + O(n^{-2}). \tag{10}$$

The approximation in (10) is excellent even for small n; for $\alpha = 0.1$ it gives c_n to within 10^{-4} for $n \geq 5$.

Similarly, using (5) and (8) we obtain for the case $m \geq 1$:

$$c_n = \frac{\lambda}{n + (m + \lambda)/2} + O(n^{-2}),$$

$$q_n = e^{-\lambda}\left(1 + \frac{m\lambda}{2n}\right) + O(n^{-2}),$$

where now $\lambda = \lambda_{m,\alpha}$ is defined by

$$1 - \alpha = e^{-\lambda} \sum_{k=0}^{m} \frac{\lambda^k}{k!}. \tag{11}$$

Again, for n large, the stopping time b_J is such that $n + 1 - J$ has very nearly a Poisson distribution with mean λ.

6. An optimality property. We can show that the stopping rule (1) is optimal in the class of all rules that have the exact confidence property. For ease of exposition we consider the case $m = 0$; similar results hold in the general case.

A general stopping rule is defined by an a.e.-monotone sequence of functions $t_1, t_2(x_1), t_3(x_1, x_2), \dots$. It stops at time $\tau = t_J(X_1, \dots, X_{J-1})$, where J is the smallest index such that

$$K\big(t_j(X_1, \dots, X_{j-1})\big) < j.$$

Consider the class Ω of all such rules that satisfy the exact probability requirement, that is,

$$\text{(12)} \qquad \text{for all } n \geq 1, \quad P\big(X_{1,n} < t_1, \dots, X_{n,n} < t_n(X_{1,n}, \dots, X_{n-1,n})\big) = 1 - \alpha.$$

THEOREM 3. *Assume that F is the standard uniform distribution and that the loss incurred by stopping at time t is $L(t, K(t), n) = f(t) + h(n - K(t))$, where f is a monotone increasing convex function and h is an arbitrary function. Then, in the class Ω the stopping rule* (1) [*i.e., with t_j a constant, independent of $(X_1, \dots, X_{j-1})$ minimizes $E(L(\tau, K(\tau), n))$*].

We remark that this result is not as powerful as it might seem, since the exactness requirement severely restricts the class of stopping rules. However, from the practical point of view such requirements are quite meaningful and in Section 8, we define and exhibit another class of rules different from (1) belonging to Ω.

PROOF OF THEOREM 3. First consider the case $h = 0$. t_1 is a constant $(= 1 - \alpha)$, determined by the exactness condition at $n = 1$. For $n \geq 1$ we determine t_{n+1} to satisfy (13) and to minimize the expected loss at n:

$$E\big(L(\tau, K(\tau), n)\big) = E\big(f(\tau)|n\big) = \sum_0^n P\big(K(\tau) = k|n\big)E\big(f(\tau)|K(\tau) = k, n\big).$$

Stopping can occur at any one of $t_1, \dots, t_{n+1}$; the only term involving t_{n+1} is

$$\text{(13)} \qquad F_n = \int_0^{t_1} dx_1 \int_{X_1}^{t_2(x_1)} dx_2 \cdots \int_{x_{n-1}}^{t_n(x_1, \dots, x_{n-1})} n!\, f(t_{n+1})\, dx_n.$$

The exactness condition (12) for $n + 1$ requires

$$(n+1)! \int_0^{t_1} dx_1 \cdots \int_{x_{n-1}}^{t_n(x_1, \dots, x_{n-1})} dx_n \int_{x_n}^{t_{n+1}(x_1, \dots, x_n)} dx_{n+1} = 1 - \alpha.$$

Subtracting the $n-1$ version of this equation, we have

$$n!\int_0^{t_1} dx_1 \cdots \int_{x_{n-1}}^{t_n(x_1,\dots,x_{n-1})}\left[(n+1)\left(t_{n+1}(x_1,\dots,x_n)-x_n\right) -1\right] dx_n = 0. \tag{14}$$

We need to minimize (13) subject to the condition (14). Using a Lagrange multiplier, we need to minimize

$$n!\int_0^{t_1} dx_1 \cdots \int_{x_{n-1}}^{t_n(x_1,\dots,x_{n-1})}\left[f\left(t_{n+1}(x_1,\dots,x_n)\right) +\lambda\left((n+1)\left(t_{n+1}(x_1,\dots,x_n)-x_n\right)-1\right)\right] dx_n.$$

This is accomplished by minimizing the term inside the square brackets for each fixed $x_1,\dots,x_n$. This gives the condition $f'(t_{n+1})+\lambda(n+1)=0$, and consequently, by the monotonicity of f, t_{n+1} does not depend on $x_1,\dots,x_n$. Because of the convexity of f this stationary point is a minimum.

Now consider the case where h is an arbitrary function. The term involving t_{n+1} will involve only $h(0)$, which is a constant. Thus, the same optimality result holds.

It is interesting to notice that the stopping rule (1) is asymptotically optimal in a quite different sense also; see Section 8.

7. A Bayesian approach. Suppose that $m=0$ and that the total number of faults is a random variable, N. Given (only) that observation stops at $\tau=b_j$, a Bayesian with a prior distribution $\pi(n)$ on N will make inferences from the posterior distribution

$$p(n|j) = \frac{\pi(n)C(n,j-1)(1-b_j)^{n+1-j}}{\sum_{m=j-1}^{\infty}\pi(m)C(m,j-1)(1-b_j)^{m+1-j}}, \qquad n \ge j-1.$$

Thus his posterior probability that $N=j-1$ will be

$$\frac{\pi(j-1)}{\sum_{m=j-1}\pi(m)C(m,j-1)c_j^{m+1-j}}. \tag{15}$$

If j is large and $\pi(n)$ is locally uniform, this will be approximately

$$\left(1+jc_j+C(j+1,2)c_j^2+C(j+2,3)c_j^3+\cdots\right)^{-1},$$

which [using (10)] is nearly $e^{-\lambda}=1-\alpha$. Thus there is no great discrepancy between the Bayesian and confidence approaches. However, there is no proper prior distribution $\pi(n)$ that makes them agree exactly. To see this, set (15) equal to $1-\alpha$ and multiply both sides by the denominator. Summing the resulting series of equations and using (5), we find that $\sum\pi(j)\alpha^j=0$, which is impossible. It is conceivable that some improper prior will work, or that for some proper prior the approaches can be made to agree exactly for all $j\ge 2$.

8. Related work. Marcus and Blumenthal (1974) propose a stopping rule of a different kind; assuming the failure times $X_1, X_2, \ldots, X_N$ are i.i.d. exponential, they consider the waiting times between successive failures, $G_j = G_{j,N} = X_j - X_{j-1}$, $j = 1, 2, \ldots, N+1$ (where $X_{0,N} = 0$ and $X_{N+1,N} = \infty$), and stop as soon as some G_j exceeds a chosen constant g. They show that, for this stopping rule, the probability that no more than m items remain is

$$\prod_{m+1}^{n} (1 - e^{-jg}). \tag{16}$$

When this probability is small, the distribution of the number of remaining items is approximately geometric, with mean $1/(e^g - 1)$. Since (16) depends on the unknown n they derive a lower bound on the probability (16).

It is easy to modify this procedure to yield a procedure with the exact confidence property. Consider a sequence g_j and stop as soon as $G_j > g_j$ for some j. Now determine the sequence g_i to make

$$pr\big(G_j \le g_j, j = 1, \ldots, n - m|n\big) = 1 - \alpha.$$

Using the fact that the G's are independent exponential random variables with scales $1/n, 1/(n-1), \ldots, 1/(m+1)$, this reduces to solving (iteratively) the system of equations

$$\prod_{j=m+1}^{\infty} (1 - e^{-jg_{n-j+1}}) = 1 - \alpha. \tag{17}$$

Of course as $n \to \infty$, $g_n \to g$, where $\prod_{j=m+1}^{b}(1 - e^{-jg}) = 1 - \alpha$. We can now compute the expected stopping time $E(\tau)$ and the expected number of unobserved events $E(n - K(\tau))$ and compare them with those corresponding to the procedure (1). For this comparison we take $\alpha = 0.05$, $m = 0$ and F to be the exponential distribution with scale 1. Table 1 gives the comparisons. As is qualitatively suggested by our optimality result, both $E(n - K(\tau))$ and $E(\tau)$ are larger for the stopping rule (17) than for (1). The first of these differences

TABLE 1
Comparison of stopping rules (1) *and* (17) *for* $m = 0$, $\alpha = 0.05$

	Stopping rule (1)		**Stopping rule (17)**	
n	$E(\tau)$	$E(n - K(\tau))$	$E(\tau)$	$E(n - K(\tau))$
1	3.654	0.05	3.842	0.05
2	4.060	0.0525	4.339	0.0525
4	4.574	0.0518	4.922	0.0525
8	5.165	0.0515	5.559	0.0525
16	5.802	0.0514	6.22	0.0525
32	6.466	0.0513	6.897	0.0525
64	7.144	0.0513	7.583	0.0525
∞		0.0513		0.0525

is stable at around 0.0012, while the difference in $E(\tau)$ increases slowly and for $n = 64$ it is about 0.44.

Starr (1974) considered independent exponentially distributed capture times for each of n prey, with a payoff equal to the number of prey caught less a linear time cost, ct. He derived a stopping rule that asymptotically agrees with ours, with

$$k = \frac{c}{\mu}(e^{\mu t_k} - 1)$$

corresponding to (after transforming to the uniform distribution)

$$c_k = \frac{\lambda}{k + \lambda}, \qquad \lambda = \frac{c}{\mu}.$$

Starr showed that this rule is almost asymptotically optimal, in the sense that as $n \to \infty$ its loss differs from that of the best (known, fixed n) rule by less than $1/2 + n^{-\gamma}$, for all $\gamma > 0$.

Vardi (1980) considered the same setup with a loss of the form

$$(\hat{n} - n)^2 + ct + bK(t),$$

where $\hat{n}$ is the estimate $K(\tau)/(1 - e^{-\mu t})$ of n. The same stopping rule, with $\lambda = c/(1 - b)$, is obtained and is almost asymptotically optimal in a similar sense. Vardi also proved the limiting Poisson distribution property of $n - K(\tau)$ in this case.

In Dalal and Mallows (1988, 1992) we generalized the Starr–Vardi approach to allow for a known arbitrary distribution of lifetimes and general testing costs. We assumed that the total cost (to the supplier) when testing terminates at time τ is

$$\text{Cost} = f(\tau) + aK(\tau) + b(n - K(\tau)),$$

where $f(t)$ is the cost of testing through time t, a is the cost per fault found during testing and b is the cost per fault not found before release. Typically b is much larger than a. We assumed that n, the total number of faults, is not known. In the special case $f(t) = tf$, with exponentially distributed fault lifetimes with known intensity μ, the optimal stopping rule is asymptotically

$$\text{(18)} \qquad \text{stop when } K(t) \le \frac{f}{c\mu}(e^{\mu}t - 1),$$

where $c = b - a$. The number of faults not found is asymptotically Poisson with mean $f/c\mu$. Remarkably, this rule has asymptotically the same form as the one we have derived here when the objective is exact confidence; we have only to replace the multiplier $f/c\mu$ by $\lambda_{m;\alpha}$ determined from (11). Under the same economic assumptions as led to (18), the total cost to the supplier of providing the (m, α) certification is approximately

$$an + c\lambda + \frac{f}{\mu}\log\frac{n}{\lambda}.$$

If some prior estimate of n could be made, this would provide a basis for negotiating a price for the (m, α) guarantee.

In Dalal and Mallows (1990) we considered the case of an unknown lifetime distribution, and provided some convenient graphical devices for this case. Recently we have extended our present work to the case of exponentially distributed variables with unknown scale and are studying various stopping rules and their properties.

APPENDIX A

We prove that for $m = 0$ and $0 < \alpha < 1$ the recursive construction of the sequence "b_j" is feasible, that is, that for all $k \geq 2$ we find $b_{k-1} < b_k < 1$. This property is rather delicate, since if we start with an arbitrary monotone sequence $b_1, \dots, b_n$ such that for some n, $q(b_1, \dots, b_n) = 1 - \alpha$, then in general we cannot find b_{n+1} in $(b_n, 1)$ such that $q(b_1, \dots, b_{n+1}) = 1 - \alpha$. For example, if $m = 0$ and $(b_1, b_2) = (0.1, 0.9)$, then $q(b_1, b_2) = 0.17$, but $q(b_1, b_2, b_3) = 0.51b_3 - 0.242$, which goes from 0.217 at $b_3 = 0.9$ to 0.268 at $b_3 = 1$. Thus no feasible b_3 gives the value 0.17.

The following argument is due to Henry J. Landau, to whom we are extremely grateful. We write β for $1 - \alpha$. The condition (3) can be written

$$1 = \frac{n!}{\beta} \int_0^{b_1} dx_1 \int_{x_1}^{b_2} dx_2 \cdots \int_{x_{n-1}}^{b_n} dx_n = \int_0^{b_n} dx_n \, w_n(x_n),$$

where the functions $w_1, w_2, \dots$ and the limits $b_1, b_2, \dots$ are determined sequentially (as functions of β) as follows:

$$w_1(x) = \frac{1}{\beta},$$

$$\int_0^{b_1} w_1(u)\, du = 1,$$

$$w_2(x) = \int_0^{\min(x, b_1)} 2w_1(u)\, du,$$

$$\int_0^{b_2} w_2(u)\, du = 1$$

and so on;

$$w_n(x) = \int_0^{\min(x, b_{n-1})} n w_{n-1}(u)\, du, \tag{A.1}$$

$$\int_0^{b_n} w_n(u)\, du = 1. \tag{A.2}$$

The function $w_n(x)$ is monotone, piecewise polynomial, of degree $n - k$ in (b_{k-1}, b_k) and equals n for $b_{n-1} \leq x$. We define $w_n(x) = 0$ for $x \leq 0$. Defining

w_n by (A.1), we need to show that

$$\int_0^{b_n} w_{n+1}(u)\,du < 1 < \int_0^1 w_{n+1}(u)\,du$$

so that b_{n+1} can be defined to satisfy (A.2).

The right-hand inequality is easy; this integral is

$$(1/\beta)P(X_{1,n+1} < b_1, \dots, X_{n,n+1} < b_n)$$
$$> (1/\beta)P(X_{1,n} < b_1, \dots, X_{n,n} < b_n) = 1$$

since $(X_{1,n+1}, \dots, X_{n,n+1}) <_P (X_{1,n}, \dots, X_{n,n})$. Now by definition of w_{n+1},

$$\int_0^{b_n} w_{n+1}(u)\,du = \int_0^{b_n} n(b_n - u)w_n(u)\,du$$
$$= f_n(\beta), \quad \text{say}.$$

We shall show that if $0 < \beta < \gamma \le 1$, then $f_n(\beta) < f_n(\gamma)$, and that $f_n(1) = 1$. It will follow that $f_n(\beta) < 1$ for all $\beta < 1$. We need to make the dependence of $w_n(x)$ and b_n on β explicit, so from here on we write these as $w_n^{(\beta)}(x)$ and $b_n^{(\beta)}$, respectively.

First, at $\beta = 1$ we have, for all k, $b_k^{(1)} = 1$, $w_k^{(1)}(x) = kx^{k-1}$, so $f_k(1) = 1$ as claimed. Now over the interval $(0, b_n^{(\beta)})$, $w_n^{(\beta)}$ is a density function (i.e., positive and integrating to 1); in fact it is the density of $X_{n,n}|(X_{1,n} < b_1^{(\beta)}, \dots, X_{n,n} < b_n^{(\beta)})$. In stochastic language, we propose to show the following theorem.

THEOREM A.1.

$$n(b_n^{(\beta)} - X_{n,n})|(X_{1,n} < b_1^{(\beta)}, \dots, X_{n,n} < b_n^{(\beta)}) \quad \textit{is stochastically increasing in } \beta. \tag{A.3}$$

Since $f_n^{(\beta)}$ is the expectation of this random variable, this will prove what we want. We have not seen how to translate the following argument into stochastic language.

LEMMA A.1. *Suppose* $0 < \beta < \gamma \le 1$ *and* $0 < \lambda < b_n^{(\gamma)} - b_n^{(\beta)}$. *Then* $w_n^{(\gamma)}(x + \lambda)$ *and* $w_n^{(\beta)}(x)$ *intersect only once in their increasing sections.*

PROOF. This is easily verified when $n = 1$. Suppose we know it for $m - 1$. First, compare $w_m^{(\gamma)}(x + b_{m-1}^{(\gamma)} - b_{m-1}^{(\beta)})$ with $w_m^{(\beta)}(x)$, in the range $-(b_{m-1}^{(\gamma)} - b_{m-1}^{(\beta)}) \le x \le b_{m-1}^{(\beta)}$. That is, we shift $w_m^{(\gamma)}$ to the left until the right-hand endpoints of their increasing sections coincide. Then

$$w_m^{(\gamma)}(x + b_{m-1}^{(\gamma)} - b_m^{(\beta)}) - w_m^{(\beta)}(x) \begin{cases} > 0, & -b_{m-1}^{(\gamma)} - (b_{m-1}^{(\beta)}) < x < 0, \\ = 0, & \text{at } x = b_{m-1}^{(\beta)}. \end{cases}$$

However, the derivative of this function is

$$w_{m-1}^{(\gamma)}(x + b_{m-1}^{(\gamma)} - b_{m-1}^{(\beta)}) - w_{m-1}^{(\beta)}(x),$$

which by the inductive hypothesis has one change of sign. So the function is everywhere nonnegative. Therefore the integral of the increasing portion of $w_m^{(\gamma)}$ exceeds the integral of the increasing portion of $w_m^{(\beta)}$, and so [by the definition (A.2) of $b_m^{(\gamma)}$ and $b_m^{(\beta)}$],

$$b_m^{(\gamma)} - b_{m-1}^{(\gamma)} < b_m^{(\beta)} - b_{m-1}^{(\beta)},$$

that is,

$$b_m^{(\gamma)} - b_m^{(\beta)} < b_{m-1}^{(\gamma)} - b_{m-1}^{(\beta)}. \tag{A.4}$$

Now with $0 \le \lambda \le b_m^{(\gamma)} - b_m^{(\beta)}$, consider

$$w_m^{(\gamma)}(x + \lambda) - w_m^{(\beta)}(x)$$

over the range $-\lambda \le x \le b_{n-1}^{(\beta)}$, which covers the increasing part of $w_m^{(\beta)}$. As before, the difference is positive in $-\lambda < x < 0$ and negative at $b_{m-1}^{(\beta)}$, while the derivative is

$$w_{m-1}^{(\gamma)}(x + \lambda) - w_{m-1}^{(\beta)}(x).$$

By (A.4) and the inductive hypothesis, this derivative changes sign only once, hence the difference itself has a single zero. This proves the lemma. □

Taking $\lambda = b_n^{(\gamma)} - b_n^{(\beta)}$ Lemma A.1 shows that indeed (A.3) holds.

APPENDIX B

For proving Theorem 2 we shall need several results which we state as lemmas. First let us introduce some notation. Let Z_i's be a sequence of i.i.d. Poisson r.v.'s with mean 1. Let

$$p_{i,n} = P\{\bar{Z}_n = i/n\} = e^{-n} n^i / i! \quad \text{and} \quad P_{i,n} = P\{\bar{Z}_n \le i/n\}.$$

LEMMA B.1. *For any* $k > n - 2$, $P_{k,n} > 1 - p_{k+1,n}(k+2)/(k+2-n)$.

PROOF.

$$\begin{aligned} P_{k,n} &= 1 - \sum_{k+1}^{\infty} e^{-n} n^i / i! \\ &= 1 - p_{k+1,n}\left\{1 + \frac{n}{(k+2)} + \frac{n^2}{(k+2)(k+3)} + \cdots\right\} \\ &> 1 - p_{k+1,n}\left\{1 + \frac{n}{(k+2)} + \frac{n^2}{(k+2)^2} + \cdots\right\} \\ &= 1 - p_{k+1,n} 1/(1 - n/(k+2)). \end{aligned}$$

□

LEMMA B.2. *For any $k < n$, $P_{k,n} < p_{k,n} n/(n-k)$.*

PROOF. By using the well-known relationship between the Poisson c.d.f. and incomplete gamma functions, we have

$$P_{k,n} = \int_n^\infty e^{-t} t^k \, dt/k! = p_{k,n} \int_n^\infty f(t) \, dt,$$

where $f(t) = e^{-(t-n)}(t/n)^k$. Now we will replace $f(t)$ by $g(t) = e^{-D(t-n)}$, where D is selected so that $f(n) = g(n)$ and $f'(n) = g'(n)$. This gives us $D = 1 - k/n$. Now using the fact that $e^x \geq 1 + x$, it is easily seen that $f(t) < g(t)$. Thus the lemma. □

Finally, using Stirling's formula, and $e^{-x} \geq 1 - x$, we get the following lemma.

LEMMA B.3. *For any large k and n, $k < n$,*

$$P_{k,n} \leq \exp(-n\{k/n \ln k/n - 1 - k/n\})/\sqrt{2\pi k}\,(1 - k/n)\{1 + o(1)\}.$$

The term in curly brackets in the numerator is positive for $k < n$.

COMMENT. Chernoff's large deviations theorem says that, for any $k < n$,

$$p_{k,n} \leq \exp(-n\{k/n \ln k/n + 1 - k/n\}).$$

Clearly, for any $k \to \infty$, and $k/n < 1 - \varepsilon$, $\varepsilon > 0$, Lemma B.3 indicates that the bound in Lemma B.2 is better than Chernoff's bounds.

The following lemma gives an upper bound on $\lim e^{nc_n}$.

LEMMA B.4. $\lim_{n\to\infty} e^{nc_n} \leq 1/(1-\alpha)$.

PROOF. Let $j = n + n^{3/4}$. Then upon rearranging (9) we get

$$\begin{aligned} \{1 - \alpha e^{-(1-\alpha)n}\}/(1-\alpha) &= \sum_{i=1}^{\infty} (e^{nc_i} - e^{nc_{i+1}}) P_{i-1,n} \\ &> e^{nc_j} P_{j-1,n} \\ &> e^{nc_j}\left(1 - (1 + n^{1/4} + n^{-3/4}) p_{j,n}\right) \end{aligned}$$

by Lemma B.1.

Now the l.h.s. $\to 1/(1-\alpha)$. Also $p_{j,n} \approx e^{-n^{1/2}/2}$, so for any $\varepsilon > 0$ we can choose n so large that

$$e^{nc_j} \leq 1/(1-\alpha) + \varepsilon.$$

Now since $n/j \to 1$, we get Lemma B.4. □

Finally, to conclude the proof of Theorem 2 we shall need the following lower bound on $\lim e^{nc_n}$.

LEMMA B.5. $\lim_n e^{nc_n} \geq 1/(1-\alpha)$.

PROOF. Let $1 < j < k < n$, where $j = n^{1-\varepsilon}$ and $k = n(1-\delta)$ for fixed positive ε and δ. Then we have by equation (9):

$$\{1 - \alpha e^{-(1-\alpha)n}\}/(1-\alpha) = \sum_{i=1}^{\infty} e^{nc_i} e^{-n} n^i / i!$$
$$\leq e^{nc_1} P_j + e^{nc_j} P_k + e^{nc_k}, \quad \text{(by monotonicity of } c\text{'s)}.$$

By Lemma B.2 and using the fact that $\lim e^{nc_n} \leq 1/(1-\alpha)$, it follows that the first two terms on the right-hand side tend to 0. Finally, since the left-hand side tends $1/(1-\alpha)$, we have for any ε', for n sufficiently large,

$$e^{nc_k} \geq 1/(1-\alpha) - \varepsilon'.$$

Finally, since $n/k = 1/(1-\delta)$, we have $\lim e^{nc_n} \geq 1/(1-\alpha)$, thus concluding the proof of Theorem 2. □

Acknowledgments. We wish to record our thanks to T. Hastie and D. Lambert for suggesting that in the proof of Theorem 3, f need not be taken to be linear, and to Henry J. Landau for supplying the proof in Appendix A.

REFERENCES

CHAPMAN, D. G. (1958). A comparative study of several goodness-of-fit tests. *Ann. Math. Statist.* **29** 655–674.

CHOW, C.-W. and SCHECHNER, Z. (1985). On stopping rules in proofreading. *J. Appl. Probab.* **22** 971–977.

DALAL, S. R. and MALLOWS, C. L. (1988). When should one stop testing software? *J. Amer. Statist. Assoc.* **83**, 872–879.

DALAL, S. R. and MALLOWS, C. L. (1990). Some graphical aids for deciding when to stop testing software. *IEEE J. Selected Areas in Communications* **8** 169–175.

DALAL, S. R. and MALLOWS, C. L. (1992). When to stop testing software—some exact and asymptotic results. *Bayesian Analysis in Statistics and Econometrics. Lecture Notes in Statist.* **75** 267–276. Springer, New York.

MARCUS, R. and BLUMENTHAL, S. (1974). A sequential screening procedure. *Technometrics* **16** 229–234.

MUSA, J. D., IANNINO, A. and OKUMOTO, K. (1987). *Software Reliability Measurement, Prediction, Application*. McGraw-Hill, New York.

STARR, N. (1974). Optimal and adaptive stopping based on capture times. *J. Appl. Probab.* **11** 294–301.

VARDI, Y. (1980). On a stopping time of Starr and its use in estimating the number of transmission sources. *J. Appl. Probab.* **17** 235–242.

WALD, A. and WOLFOWITZ, J. (1939). Confidence limits for continuous distribution functions. *Ann. Math. Statist.* **10** 105–118.

WALLENIUS, K. T. (1967). Sampling for confidence. *J. Amer. Statist. Assoc.* **62** 540–547.

BELLCORE
ROOM 2P-392
445 SOUTH STREET
MORRISTOWN, NEW JERSEY 07960-1910

AT&T BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974