OPTIMAL SAMPLING SCHEMES FOR ESTIMATING SYSTEM RELIABILITY BY TESTING COMPONENTS--
I: FIXED SAMPLE SIZES

by

Donald A. Berry

Technical Report No. 173

April 1972

University of Minnesota
Minneapolis, Minnesota

1. Introduction.

The reliability of a system of r components connected in series is $p = p_1 p_2 \ldots p_r$, where $0 < p_i < 1$ for all i and p_i is the (unknown) probability that a component of type i functions properly. For many systems the most efficient estimate (perhaps the only estimate) of p can be obtained by testing the components individually. The usual assumption is that the components are each tested a fixed number of times giving point estimates or interval estimates of p (see Myrhe and Saunders (1971)for an example of the latter, and for further references). Hwang and Buehler (1971) explore an inverse sampling scheme, where each component is tested until it yields a fixed number of failures.

Such an assumption seems to be made separate from considerations of sampling costs (in terms of time and resources). The approach here will consider sampling costs explicitly. In particular, it will be assumed that the cost of testing n_i components of type i observing s_i successes, components which function properly, $i = 1, \ldots, r$, is

$$(1.1) \qquad c(J) = \sum_{i=1}^{r} c_i n_i \ ,$$

where J denotes the accumulated data $(s_1, n_1; \ldots ; s_r, n_r)$ and each $c_i > 0$. The number c_i can be thought of as the purchase price of a type i component which will be destroyed in testing--whether the component functions or not. Applications can be envisaged where the cost of sampling is proportional to the number of successes (or failures)--these will not be addressed here. The problem is to determine the $(n_1, \ldots, n_r)$ which corresponds to smallest total cost among sampling schemes which have the same expected information value in estimating p.

The parameters $p_1, \ldots, p_r$ are not known precisely but are themselves random variables. The Bernoulli trials associated with a sequence of successes and failures associated with component type i are therefore not independent, but are independent conditional on the unknown quantity p_i, so that the trials are exchangeable--see, for example, Feller(1966), Section VII 4 . At any stage of sampling the information about p is given by the accumulated data J, which can always be regarded as a probability distribution on the parameters $p_1, \ldots, p_r$. The J corresponding to no data is $J_0 = (0, 0; \ldots ; 0, 0)$, the initial distribution. Throughout this paper the parameters $p_1, \ldots, p_r$ are assumed to be initially, and therefore also henceforth, statistically independent. If the initial distribution were such that this assumption is violated then more information would be present in any subsequent J than otherwise, leading to a more efficient estimate of p.

The best sampling scheme will depend on J_0 and the goodness of any scheme varies greatly with J_0. One cannot expect, therefore, to find a scheme that is good (or even reasonable) for all possible J_0. Several different kinds of distributions of the p_i will be examined here.

Throughout this paper quadratic loss is assumed; specifically, the loss associated with an estimate $\hat{p}$ is

$$L(p, \hat{p}) = (p-\hat{p})^2. \tag{1.2}$$

At any stage of sampling the estimate of p which minimizes the Bayes risk is the expected value of p, the (unconditional) probability that the system functions properly:

$$\hat{p} = E(p|J) = E(p_1|J) \ldots E(p_r|J), \tag{1.3}$$

in view of the independence of the p_i. The Bayes risk corresponding to $\hat{p} = E(p|J)$ is the variance of p:

$$(1.4) \qquad \mathrm{var}(p|J) = E(p_1^2|J) \ldots E(p_r^2|J) - E^2(p_1|J) \ldots E^2(p_r|J),$$

again in view of independence.

The problem is to determine an <u>optimal</u> sampling scheme, call it $(n_1^0, \ldots, n_r^0)$, where the n_i^0 are values of n_i which minimize expected Bayes risk plus cost:

$$(1.5) \qquad B(n_1, \ldots, n_r) = E_0 \, \mathrm{var}(p|J) + \sum_{i=1}^{r} c_i n_i,$$

where expectation E_0 refers to the initial distribution of the p_i and averages over the possible values of s_i for fixed n_i. A good sampling scheme will be driven by the following consideration: for each i, n_i should be relatively large if c_i is relatively small, but on the other hand it should be relatively small if, according to J_0, much is known about p_i. The problem is then properly regarded as a search for an appropriate balance of sampling costs and information. It is clear, for example, that an optimal scheme would sample from all r component types (provided the sampling costs are sufficiently small) since there is limited information about p available in any subset of the components (unless some of the p_i are known, a trivial case that has been excluded). To be specific, suppose that J is such that $n_1 = 0$ while $\min\{n_2, \ldots, n_r\} \to \infty$, then $\mathrm{var}(p|J) \to p_2^2 \ldots p_r^2 \, \mathrm{var}(p_1|J_0) > 0$.

According to the way in which the problem has been formulated it is clear that its solution is identical with the solution for a system of components connected in parallel (where system reliability is

$1 - (1-p_1)(1-p_2) \ldots (1-p_r))$; namely, the one in which the p_i are replaced with $1 - p_i$.

The case where each of the p_i are initially uniform on $(0, 1)$ and $r = 2$ is considered in Section 2. The density for p_i is proportional to $p_i^{m_i}$ and r is arbitrary in Section 3, thereby generalizing Section 2 in several ways. Some readers may want to omit Section 2 and others may want to omit Section 3. Distributions specified for p rather than for the p_i are considered briefly in Section 4.

Part II of this paper will consider the problem sequentially, wherein the component type sampled at any stage depends on the history of components sampled and the results obtained. The approach of both Parts I and II follows modern Bayesian decision theory as espoused, for example, in Degroot (1970) or in Raiffa and Schlaifer (1961).

2. $r = 2$, p_1 and p_2 Uniformly Distributed.

When there are two components and according to J_0 the parameters p_1 and p_2 both have uniform densities in $(0, 1)$, the initial density of p is

$$f(p) = -\ln p,\ p \in (0, 1). \tag{2.1}$$

The initial ("no data") estimate of p is $E(p|J_0) = E_0 p = 1/4$, with Bayes risk $\operatorname{var}(p|J_0) = \operatorname{var}_0 p = 7/144$. Obviously, if the sampling cost associated with either component is greater than $7/144$ then the optimal sample size for that component is zero. (Stronger statements are possible--analysis not in order here reveals that sampling is optimal if and only if c_1 or c_2 is less than or equal to $1/144$.)

A word is necessary about the assumption that the p_i are initially uniform on $(0, 1)$. Such an initial distribution has been proposed by some to represent "complete ignorance." That there cannot be a distribution which represents complete ignorance is easily seen by the following well-known argument. If one is completely ignorant about p_1 and p_2 then he is also completely ignorant about $p = p_1 p_2$ as well. But these three parameters, regarded as random variables, cannot be subject to the same distribution (save the one-point distributions concentrated at 0 or 1, neither of which can reasonably qualify). To accomplish a solution some assumption must be made and the one made above, while restrictive, is appealing on several grounds. It represents a certain amount of "openmindness" about the p_i (but not p); see Edwards et al. (1963) for a discussion of "openminded" distributions. But more importantly, it means that the joint density of p_1 and p_2 at any stage of sampling, which is specified by $J = (s_1, n_1; s_2, n_2)$, can be written as the product of two beta densities; in particular, as proportional to

$$p_1^{s_1}(1-p_1)^{n_1-s_1} p_2^{s_2}(1-p_2)^{n_2-s_2} . \tag{2.2}$$

The latter reason by itself does not dictate the uniform distribution, for the same is true (with modified exponents in (2.2)) if the p_i initially have arbitrary beta densities (see Degroot(1970), p. 160 for a discussion of the conjugate nature of the beta family in Bernoulli sampling).

Though redundant, it will be convenient in the case of two components to have the additional notation

$$k = c_1/c_2. \tag{2.3}$$

For reasons of symmetry, it is clear that $n_1^0 = n_2^0$ if $k = 1$. If the costs are unequal then it seems reasonable to expect that $n_1^0 > n_2^0$ or $n_2^0 > n_1^0$ according as $k < 1$ or $k > 1$. That this is the case will be verified. Furthermore, it will be seen that n_1^0 and n_2^0 are approximately (i.e., asymptotically) related as follows:

$$\frac{n_2^0}{n_1^0} \doteq \sqrt{k} \;,$$

and approximations for the n_i^0 will be obtained.

In view of the well-known relation:

$$\text{(2.4)} \qquad E_0 \operatorname{var}(p|J) = \operatorname{var}_0 p - \operatorname{var}_0 E(p|J) = \frac{7}{144} - \operatorname{var}_0 E(p|J),$$

where $J = (s_1, n_1; s_2, n_2)$, $\operatorname{var}_0 E(p|J)$ can be regarded as the expected worth of the information about p provided by n_1 observations on component type 1 and n_2 of observations on component type 2. The quantity $\operatorname{var}_0 E(p|J)$ is an increasing function of n_1 for all n_2 and of n_2 for all n_1. For, since

$$\text{(2.5)} \qquad \Pr(s_i|J_0, n_i) = \frac{1}{n_i+1} \;, \; s_i = 0,1,\ldots, n_i; \; i = 1, 2$$

(cf. (3.6)), it follows that

$$\text{(2.6)} \qquad \operatorname{var}_0 E(p|J) = \operatorname{var}_0 \left(\frac{s_1+1}{n_1+2} \; \frac{s_2+1}{n_2+2} \right)$$

$$= E_0 \left(\frac{s_1+1}{n_1+2} \right)^2 E_0 \left(\frac{s_2+1}{n_2+2} \right)^2 - E_0^2 \left(\frac{s_1+1}{n_1+2} \right) E_0^2 \left(\frac{s_2+1}{n_2+2} \right)$$

$$= \frac{1}{n_1+1} \sum_{s_1=0}^{n_1} \left(\frac{s_1+1}{n_1+2} \right)^2 \frac{1}{n_2+1} \sum_{s_2=0}^{n_2} \left(\frac{s_2+1}{n_2+2} \right)^2 - \left(\frac{1}{n_1+1} \sum_{s_1=0}^{n_1} \frac{s_1+1}{n_1+2} \right)^2 \cdot \left(\frac{1}{n_2+1} \sum_{s_2=0}^{n_1} \frac{s_2+1}{n_2+2} \right)^2$$

$$= \frac{2n_1+3}{6(n_1+2)} \; \frac{2n_2+3}{6(n_2+2)} - \frac{1}{16} \;.$$

The first difference (and the partial derivative as well) of the last quantity with respect to both n_1 and n_2 is positive. Also, both second differences are negative, indicating that the increment in expected worth of information when n_i is increased by 1 is greater for smaller values of n_i.

Rewriting (1.5) for this case in view of (2.4) and (2.6),

$$(2.7)\qquad \begin{aligned} B(n_1, n_2) &= E_0\mathrm{var}(p|J) + c_1n_1 + c_2n_2 \\ &= \frac{7}{144} - \frac{2n_1+3}{6(n_1+2)}\,\frac{2n_2+3}{6(n_2+2)} + \frac{1}{16} + c_1n_1 + c_2n_2 \\ &= \frac{1}{9} - \frac{1}{36}\,\frac{2n_1+3}{n_1+2}\,\frac{2n_2+3}{n_2+2} + c_1n_1 + c_2n_2 , \end{aligned}$$

which is to be minimized. Regard n_1 and n_2 as nonnegative real variables rather than just integers. Taking the partial derivatives of $B(n_1, n_2)$ and equating them to zero yields the pair of equations:

$$(2.8a)\qquad -\frac{1}{36}\,\frac{2n_2^0 + 3}{(n_1^0 + 2)^2(n_2^0 + 2)} + c_1 = 0,$$

$$(2.8b)\qquad -\frac{1}{36}\,\frac{2n_1^0 + 3}{(n_2^0 + 2)^2(n_1^0 + 2)} + c_2 = 0.$$

It is clear from (2.8a) that for fixed n_2^0, n_1^0 is the order of $1/\sqrt{c_1}$, and from (2.8b) that for fixed n_1^0, n_2^0 is the order of $1/\sqrt{c_2}$; in particular, $n_i^0 \to \infty$ as $c_i \to 0$. In view of equations (2.8) the values n_2^0 and n_1^0 can be obtained by solving the cubic equation:

$$(2.9)\qquad (n_2^0)^3 + 2(2\sqrt{k} + 1)(n_2^0)^2 + \left(4(k+\sqrt{k}) - \frac{k}{18c_1}\right)n_2^0 - \frac{k}{12c_1} = 0,$$

for which there is exactly one positive root (provided c_1 is sufficiently small--less than $1/72$ suffices), and in view of (2.8b),

$$(2.10)\qquad n_1^0 = \frac{3 - 72c_2(n_2^0 + 2)^2}{36c_2(n_2^0 + 2)^2 - 2} .$$

However, the solution of (2.9) is not trivially arrived at and is real rather than integer in any case. Furthermore, the generalization of (2.9) for the case considered in Section 3 is impossible to solve explicitly.

A simple though very accurate approximation for n_1^0 and n_2^0 can be obtained from a single iteration in equations (2.8). Writing (2.8a) and (2.8b) as

$$(2.11a)\qquad 1 - 2(n_2^0 + 2) + 36(n_2^0 + 2)(n_1^0 + 2)^2 c_1 = 0,$$

$$(2.11b)\qquad 1 - 2(n_1^0 + 2) + 36(n_1^0 + 2)(n_2^0 + 2)^2 c_2 = 0,$$

and ignoring the constant 1 in each, yields, respectively,

$$(2.12)\qquad n_i^0 + 2 \doteq \frac{1}{\sqrt{18c_i}}\ , \ i = 1, 2,$$

where the approximate equalities approach equalities as $c_1, c_2 \to 0$. Using this approximation of $n_2^0 + 2$ in (2.11a) and of $n_1^0 + 2$ in (2.11b) yields the better--particularly for small n_i^0--approximations:

$$(2.13a)\qquad n_1^0 \doteq \frac{2 - \sqrt{18c_2}}{36c_1} - 2\ ,$$

$$(2.13b)\qquad n_2^0 \doteq \frac{2 - \sqrt{18c_1}}{36c_2} - 2.$$

The following calculations illustrate the accuracy of approximations (2.13). Table 2.1 gives these numbers for four examples--in each $k = c_1/c_2$ is 4. For each of these examples the function $B(n_1, n_2)$ is given in Table 2.2 for the nine pairs (n_1, n_2) closest to the values given in Table 2.1. In each case n_1^0 and n_2^0 are seen to be the values obtained by rounding to the nearest integer in approximations (2.13). This is typical in view of the "smoothness" of B (especially for small c_1 and c_2) but, of course, cannot be guaranteed.

TABLE 2.1

c_1	c_2	n_1^0(app.)	n_2^0(app.)
4×10^{-3}	10^{-3}	1.60	4.94
4×10^{-4}	10^{-4}	9.66	21.06
4×10^{-5}	10^{-5}	35.14	72.03
4×10^{-6}	10^{-6}	115.73	233.20

TABLE 2.2: $B(n_1, n_2)$

$c_1 = 4 \times 10^{-3}$, $c_2 = 10^{-3}$			
n_1 \ n_2	4	5	6
1	.0342346	.0341323	.0343056
2	.0339907	.0338333	.0339653
3	.0354444	.0352540	.0353611

$c_1 = 4 \times 10^{-4}$, $c_2 = 4 \times 10^{-4}$			
n_1 \ n_2	20	21	22
9	.0130610	.0130562	.0130601
10	.0130497	.0130444	.0130480
11	.0131016	.0130961	.0130993

$c_1 = 4 \times 10^{-5}$, $c_2 = 10^{-5}$			
n_1 \ n_2	71	72	73
34	.00436367	.00436353	.00436366
35	.00435225	.00435211	.00436223
36	.00436301	.00435386	.00435298

$c_1 = 4 \times 10^{-6}$, $c_2 = 10^{-6}$			
n_1 \ n_2	232	233	234
115	.001403236	.001403230	.001403233
116	.001403221	.001403215	.001403217
117	.001403268	.001403267	.001403269

To find the expected worth of sampling the entries in Table 2.2 can be compared with $B(0, 0) = 7/144 \doteq .0486111$. Evidently the expected worth of sampling tends to $7/144$ as the costs tend to zero; that is, $B(n_1^0, n_2^0)$ tends to zero. In fact, using the asymptotic result (2.12) in (2.7), as c_1 and $c_2 \to 0$ keeping $c_1/c_2 = k$,

$$(2.14) \qquad B(n_1^0, n_2^0) \doteq \frac{1}{9} - \frac{(2-3\sqrt{2c_1})(2-3\sqrt{2c_2})}{36} + \frac{\sqrt{c_1} + \sqrt{c_2}}{3\sqrt{2}}$$

$$\doteq \frac{\sqrt{2}}{3}(\sqrt{c_1} + \sqrt{c_2}) = \frac{\sqrt{2c_1}}{3}\left(1 + \frac{1}{\sqrt{k}}\right),$$

where terms which tend to zero at the same rate as c_1 have been ignored.

It can be noticed from Table 2.1 that the ratio of n_2^0 to n_1^0 is nearly constant at 2. That the limit of n_2^0/n_1^0 is 2 as the costs go to zero is easily seen from (2.12), or by dividing (2.8a) by (2.8b). The latter approach yields

$$(2.15) \qquad \frac{c_1}{c_2} = \frac{(2n_2^0 + 3)(n_2^0 + 3)}{(2n_1^0 + 3)(n_1^0 + 2)} \quad .$$

As c_1 and $c_2 \to 0$ keeping $c_1/c_2 = k$, n_1^0 and $n_2^0 \to \infty$ and the right side of (2.15) tends to $(n_2^0/n_1^0)^2$. Therefore,

$$(2.16) \qquad \frac{n_2^0}{n_1^0} \to \frac{c_1}{c_2} = \sqrt{k} \quad \text{as} \quad c_1, c_2 \to 0.$$

An easy corollary of this fact is that (2.16) holds as well if p_1 and p_2 initially have arbitrary distributions in the beta family. While, practically speaking, most initial distributions can reasonably be approximated by beta distributions for the purposes of this problem, certain special forms of initial information cannot. For example, if p_1 is initially 0 or 1 with probability 1/2 each, then $n_1^0 = 0$ for large values of c_1 and $n_1^0 = 1$ for small values of c_1 (the cutoff point of c_1, where both $n_1^0 = 0$ and $n_1^0 = 1$, depends on c_2 and the initial information about p_2), since one observation on component type 1 delivers complete information about p_1.

3. <u>General r, Density of p_i Proportional to $p_i^{m_i}$</u>.

This section generalizes Section 2 in at least two ways. Firstly, there are now an arbitrary number of components. Secondly, the parameters p_i are assumed to have densities proportional to $p_i^{m_i}$, $i = 1, \ldots, r$. Not only does this include the uniform assumption ($m_1 = m_2 = 0$) thus generalizing Section 2, but now the distributions of the p_i are not necessarily identical. While, mathematically speaking, each of the m_i must be greater than -1, they will be large (the order of r) in most applications, for, taking the m_i small implies that the system is very unreliable.

Assume the m_i are small, say for definiteness $m_1 = \ldots = m_r = 0$, then each of the p_i are initially uniform and the (initial) probability that the system functions is $E_0 p = 1/2)^r$. Furthermore, the Bayes risk, $\mathrm{var}_0 p = (1/3)^r - (1/4)^r$, is small for r large, indicating that few components can be sampled and that, while the estimate of p may be affected by sampling, the initial notion that p is very likely small will not change. For most applications in which r components are connected in series each component is likely to function properly, giving the system reasonable reliability. Thus, while $m_1 = \ldots = m_r = 0$ may have applications for parallel systems, such may not be the case for series systems if r is large.

On the other hand if the m_i are large, and say $m_1 = \ldots = m_r = \alpha r$, then the probability the system functions properly is $E_0 p = (\frac{\alpha r+1}{\alpha r+2})^r$ which is between $e^{-1/\alpha}$ and $\frac{\alpha+1}{\alpha+2}$ for all r. Also, because $var_0 p > e^{-2/\alpha}$ for all r can always be improved (i.e., made smaller) by sampling, such sampling will not be discouraged by the fact that there are many components.

According to the above assumptions concerning p_i, the initial density of $-\ln p_i$ is exponential with expectation $(m_i + 1)^{-1}$. The density of p can take different forms, depending on the equalities among the m_i, the simplest form occurs when $m_1 = \dots = m_r = m$ (since then $-\ln p = \Sigma(-\ln p_i)$ has a gamma density), to wit:

$$f(p) = \frac{(m+1)^r}{(r-1)!} p^m(-\ln p)^{r-1}, \; p \in (0, 1). \tag{3.1}$$

For arbitrary m_i, the initial estimate of p is

$$E_0 p = \prod_{i=1}^{r} \frac{m_i + 1}{m_i + 2}, \tag{3.2}$$

with Bayes risk

$$var_0 p = \prod_{i=1}^{r} \frac{m_i + 1}{m_i + 3} - E_0^2 p. \tag{3.3}$$

Also, for arbitrary $J = (s_1, n_1; \dots; s_r, n_r)$,

$$E(p_i|J) = \frac{s_i + m_i + 1}{n_i + m_i + 2}, \; i = 1, \dots, r, \tag{3.4}$$

and therefore for fixed $n_1, \dots, n_r$,

$$var_0 E(p|J) = \prod_{i=1}^{r} E_0\left(\frac{s_i + m_i + 1}{n_i + m_i + 2}\right)^2 - \prod_{i=1}^{r} E_0^2\left(\frac{s_i + m_i + 1}{n_i + m_i + 2}\right). \tag{3.5}$$

In view of the fact that

$$\begin{aligned} Pr(s_i|J_0, n_i) &= E_0\binom{n_i}{s_i} p_i^{s_i}(1-p_i)^{n_i - s_i} \\ &= (m_i + 1)\binom{n_i}{s_i}\int_0^1 p_i^{s_i + m_i}(1-p_i)^{n_i - s_i} dp_i \\ &= (m_i + 1)\frac{n_i!(s_i - m_i)!}{s_i!(n_i + m_i + 1)!}, \end{aligned} \tag{3.6}$$

for all i (generalizing (2.5)), it follows that for n_i fixed,

$$(3.7) \qquad E_0\left(\frac{s_i + m_i + 1}{n_i + m_i + 2}\right) = \frac{m_i + 1}{m_i + 2} ,$$

$$(3.8) \qquad E_0\left(\frac{s_i + m_i + 1}{n_i + m_i + 2}\right)^2 = \frac{(m_i+1)[n_i(m_i+2) + (m_i+1)(m_i+3)]}{(m_i+2)(m_i+3)(n_i + m_i + 2)} .$$

Therefore, in view of (2.4) and generalizing (2.7),

$$(3.9) \qquad B(n_1,\ldots, n_r) = \mathrm{var}_0 p - \mathrm{var}_0 E(p|J) + \sum_{i=1}^{r} c_i n_i$$

$$= \prod_{i=1}^{r} \frac{m_i + 1}{m_i + 3} - \prod_{i=1}^{r} \frac{(m_i+1)[n_i(m_i+2) + (m_i+1)(m_i+3)]}{(m_i+2)(m_i+3)(n_i + m_i + 2)} + \sum_{i=1}^{r} c_i n_i .$$

Taking the derivatives with respect to the n_i in (3.9) and equating them to zero yields the system of equations:

$$(3.10) \qquad -\frac{m_i + 1}{(m_i+2)(m_i+3)(n_i^0 + m_i + 2)^2} \prod_{j \neq i} \frac{m_j + 1}{(m_j+3)(n_j^0 + m_j + 2)} \left[n_j^0 + \frac{(m_j+1)(m_j+3)}{m_j + 2}\right] + c_i = 0,$$

for $i = 1,\ldots, r$. The solution of (3.10) involves finding roots of a polynomial of degree $r + 1$ and of $r - 1$ polynomials of degree r. Analogous to the techniques of Section 2, however, approximate n_i^0 can be obtained from (3.10) and improved in one iteration. For convenience, let

$$(3.11) \qquad \gamma = E_0 \prod_{i=1}^{r} p_i^2 = \prod_{i=1}^{r} \frac{m_i + 1}{m_i + 3} .$$

Since for each i,

$$(3.12) \qquad n_i^0 + \frac{(m_i + 1)(m_i + 3)}{m_i + 2} = n_i^0 + m_i + 2 - \frac{1}{m_i + 2} \doteq n_i^0 + m_i + 2 ,$$

equations (3.10) become (approximately)

$$(3.13) \qquad -\frac{\gamma}{(m_i + 2)(n_i^0 + m_i + 2)^2} + c_i \doteq 0, \; i = 1,\ldots, r,$$

and therefore,

$$(3.14)\qquad n_i^0 + m_i + 2 \doteq \sqrt{\frac{\gamma}{c_i(m_i+2)}} \quad , \; i = 1,\ldots, r,$$

which are exact in the limit as $\max c_i \to 0$. Using these approximations for n_j^0 in (3.10) for $j \neq i$ yields the improvements:

$$(3.15)\qquad n_i^0 \doteq \left[\frac{m_i+1}{c_i(m_i+2)(m_i+3)} \prod_{j\neq i} \frac{m_j+1}{m_j+3}\left(1 - \sqrt{\frac{c_j}{\gamma(m_j+2)}}\right)\right]^{\frac{1}{2}} - (m_i+2), i=1,\ldots,r.$$

Before turning to special cases and examples, it should be noted from (3.14) that the n_i^0 get large without bound as the costs tend to zero, and in the limit,

$$(3.16)\qquad (n_1^0,\ldots, n_r^0) \propto \left(\frac{1}{\sqrt{c_1(m_1+2)}}, \ldots, \frac{1}{\sqrt{c_r(m_r+2)}}\right).$$

As has been seen in the special case of Section 2, and will be seen in upcoming examples, the asymptotic relation (3.16) is accurate for moderate as well as small values of the c_i. According to (3.16), among components with similar sampling costs a component is sampled less if it is deemed to be more reliable. Of course, this statement is tied to assumptions of this section, according to which a component deemed reliable (by making the corresponding m_i large) also has a small variance associated with p_i, thus necessitating fewer observations on that component.

As in Section 2 (cf. (2.14)), $B(n_1^0,\ldots, n_r^0) \to 0$ as $\max c_i \to 0$ so that sampling, even considering the costs involved, reduces the expected losses to zero in the limit. Furthermore, in view of (3.14), (3.9) becomes (approximately)

$$(3.17)\qquad B(n_1^0,\ldots, n_r^0) \doteq \gamma - \gamma \prod_{i=1}^{r}\left(1 - \sqrt{\frac{c_i}{\gamma(m_i+2)}}\right) + \sum_{i=1}^{r} c_i\sqrt{\frac{\gamma}{c_i(m_i+2)}}$$

$$\doteq \gamma - \gamma + \gamma \sum_{i=1}^{r} \sqrt{\frac{c_i}{\gamma(m_i+2)}} + \sum_{i=1}^{r}\sqrt{\frac{\gamma c_i}{m_i+2}}$$

$$= 2\sqrt{\gamma} \sum_{i=1}^{r} \frac{\sqrt{c_i}}{\sqrt{m_i+2}}$$

The Case $m_1 = \dots = m_r = 0$.

As previously pointed out, this case corresponds to the one considered in Section 2 if $r = 2$ and seems unrealistic for most series systems if r is large, though it may have applications for parallel systems. The principal reason for considering this case here is to illustrate further its unrealistic nature and to demonstrate a danger arising from the blind use of a Bayesian approach. To this end fix c_1 and c_2 and let $\max\{c_3, \dots, c_r\} \to 0$. According to (3.15),

$$n_1^0 \doteq (\tfrac{1}{3})^{r/2 - 1} \sqrt{\frac{2 - \sqrt{2 \cdot 3^r c_2}}{36 c_1}} - 2, \tag{3.18a}$$

$$n_2^0 \doteq (\tfrac{1}{3})^{r/2 - 1} \sqrt{\frac{2 - \sqrt{2 \cdot 3^r c_1}}{36 c_2}} - 2, \tag{3.18b}$$

and $n_3^0, \dots, n_r^0 \to \infty$ giving effectively complete information about $p_3, \dots, p_r$.

Since the initial distributions of p_1 and p_2, the only effectively unknown parameters, coincide with those of Section 2 where $r = 2$, one may expect that the optimal sample sizes should be the same as in Section 2. In fact, since the estimates of p given $J = (s_1, n_1; \dots ; s_r, n_r)$ is $\left(\frac{s_1+1}{n_1+2}\ \frac{s_2+1}{n_2+2}\right) p_3 \dots p_r$ rather than $\left(\frac{s_1+1}{n_1+2}\ \frac{s_2+1}{n_2+1}\right)$ and the corresponding Bayes risks are different, the optimal sample sizes when there are two components are approximately $3^{r/2 - 1}$ times as large as when there are r components. This is illustrated for $r = 3$ (and positive but small c_3) for four examples in Table 3.1. When compared with Table 2.1 it will be seen that the values of n_1^0 and n_2^0 given there are about $\sqrt{3}$ times those of Table 3.1. Incidentally, in each example the true n_i^0 has been verified to be the values in Table 3.1 rounded to the nearest integer. In the first, for example, $(n_1^0, n_2^0, n_3^0) = (0, 2, 5)$.

TABLE 3.1

c_1	c_2	c_3	n_1^0(app)	n_2^0(app)	n_3^0(app)
4×10^{-3}	10^{-3}	$10^{-3}/4$	- .04	1.66	5.09
4×10^{-4}	10^{-4}	$10^{-4}/4$	4.62	10.98	23.71
4×10^{-5}	10^{-5}	$10^{-5}/4$	19.33	40.41	82.56
4×10^{-6}	10^{-6}	$10^{-6}/4$	65.85	133.46	268.67

The Case $m_1 = \dots = m_r = \alpha r$.

In this case the probability that the system functions is $\left(\frac{\alpha r+1}{\alpha r+2}\right)^r \doteq e^{-1/\alpha}$ and $\gamma = \left(\frac{\alpha r+1}{\alpha r+3}\right)^r \doteq e^{-2/3\alpha}$ (where the approximations assume r large). According to (3.15),

$$n_i^0 \doteq \left[\frac{\gamma}{c_i(\alpha r+2)} \prod_{j\neq i} \left(1 - \sqrt{\frac{c_i}{\gamma(\alpha r+2)}}\right)\right]^{\frac{1}{2}} - (\alpha r+2),\ i = 1,\dots,r. \tag{3.19}$$

For $r = 3$ and $\alpha = 2$ these approximations are given for four examples in Table 3.2. In view of (3.16) the (n_1^0, n_2^0, n_3^0) in this table are approximately proportional to the corresponding numbers in Table 3.1. However, the numbers are smaller since in this case much more information is present in J_0, i.e., before sampling--it is as though the information $m_1 = m_2 = m_3 = 0$ has been modified by 6 successes in 6 observations on of the three component types.

TABLE 3.2

c_1	c_2	c_3	n_1^0(app)	n_2^0(app)	n_3^0(app)
4×10^{-3}	10^{-3}	$10^{-3}/4$	-4.21	-0.49	6.96
4×10^{-4}	10^{-4}	$10^{-4}/4$	4.08	16.10	40.13
4×10^{-5}	10^{-5}	$10^{-5}/4$	30.30	68.53	145.01
4×10^{-6}	10^{-6}	$10^{-6}/4$	113.21	234.36	476.66

4. Distributions on p.

If the initial information is in the form of a distribution specified for p, but not for the individual p_i, then this can be modified by making observations on the system, but the character of the information is such that there is no value in observing which components failed. There could be value in making observations on individual components, however, under assumptions about the way in which the components contribute to the system. For example, under the (rather strong) assumption that the p_i are independent and identically distributed these distributions can be found (in theory if not in practice), thus specifying J_0, and the earlier sections apply. If this assumption is made and, for example, $r = 2$ and the density of p is given by (2.1) then Section 2 applies. However, for arbitrary density of p can lead to practical difficulties as this brief section is designed to suggest.

Seemingly the simplest distribution is p uniform on (0, 1). Incidentally, this treats series and parallel systems equally. Assuming the p_i are independent and identically distributed, it can be seen that $-\ln p_i$ has a gamma density with expectation and variance r^{-1}. Therefore the density of p_i is

$$(4.1)\qquad (-\ln x)^{1/r - 1}/\Gamma(\tfrac{1}{r}), \quad x \in (0, 1).$$

$E_0 p_i$ and $E_0 p_i^2$ are easily found from (4.1), or from symmetry, to be $(1/2)^{1/r}$ and $(1/3)^{1/r}$. However, the probability of s_i successes in n_i observations on component type i is not easily found. After some calculation,

$$(4.2)\qquad \Pr(s_i | J_0, n_i) = E_0\binom{n_i}{s_i} p_i^{s_i}(1-p_i)^{n_i s_i} = \binom{n_i}{s_i}\sum_{h=0}^{n_i - s_i}\binom{n_i - s_i}{h}(-1)^h(h+s_i+1)^{-1/r},$$

which cannot be further reduced though it can be approximated by

$$(4.3)\qquad \binom{n_i}{s_i}\frac{d^{n_i-s_i}}{ds_i^{n_i-s_i}}(s_i+1)^{-1/r} = \binom{n_i}{s_i}(s_i+1)^{-(\frac{1}{r}+n_i-s_i)}\Gamma(\tfrac{1}{r}+n_i-s_i)/\Gamma(\tfrac{1}{r}).$$

Evidently, the mathematics become complicated even using approximate methods.

Two approaches readily suggest themselves. One is to perform all calculations numerically. The second is to approximate the density of p with one more tractible. For example, the uniform density on p can be fitted reasonably well (the first two moments exactly) by a density of the form (3.1) by setting

$$(4.4)\qquad m = \left(3^{-\frac{1}{r}+1} - 2^{-\frac{1}{r}+1}\right)\Big/\left(2^{-\frac{1}{r}} - 3^{-\frac{1}{r}}\right).$$

This problem is solved in Section 3.

5. Comments.

While this paper has been concerned with estimating the reliability of systems of independent components connected in series or in parallel, there are obvious extensions for more general systems. The purpose of this paper is not to exhaust the possibilities but to illustrate an approach, one which can be fruitfully used in estimating reliabilities. The main selling point of a Bayesian decision theoretic approach is that empirical information is handled in a unified way--accumulating data affects current knowledge according to Bayes' theorem, and the value of data or prospective data can be assessed on that basis. The problem considered here is one for which accumulating data on the reliability of individual components affects the state of knowledge about the parameter of interest, the reliability of the system, in a very interesting way.

It is for this reason that a sequential treatment of the data collection problem--the subject of Part II of this paper--is so appealing.

Acknowledgment.

My interest in this problem was stimulated by a seminar presented by Professor Robert J. Buehler.

REFERENCES

Degroot, M. (1970). Optimal Statistical Decisions. McGraw-Hill, New York.

Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian Statistical Inference for Psychological Research. Psychological Review 70 193-242.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II. Wiley, New York.

Hwang, D.-S., and Buehler, R. J. (1971). Confidence Intervals for Some Functions of Several Bernoulli Parameters with Reliability Applications. Technical Report No. 161, School of Statistics, University of Minnesota.

Myhre, J. M. and Saunders, S. C. (1971). Approximate Confidence Intervals for Complex Systems with Exponential Component Lives. Ann. Math. Statist. 42 343-348.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard University, Boston, Massachusetts.