A Bayesian Approach to Environmental Stress Screening

R.E. Barlow
University of California at Berkeley

C.A.B. Pereira and S. Wechsler
University of São Paulo

The article presents a Bayesian analysis for the environmental stress screening problem. The decision problem of deriving optimal stress screen durations is solved. Given a screen duration, the optimal stress level can also be determined. Indicators of the quality of a screen of any duration are derived. A statistical model is presented which allows a posterior density for the rate of early failures of the production process to be calculated. This enables the user to update his opinion about the quality of the process. © 1994 John Wiley & Sons, Inc.

1. INTRODUCTION

Environmental stress screening (ESS) has been given high priority by industry, as it is able to reduce costs by enhancing the reliability of equipment. However, a rational approach to ESS design or Bayesian models to handle data generated by ESS experiments have not yet been suggested.

Environmental stress screening has been used, for instance, to purge populations from parts having hidden latent defects not detectable by quality control. This is done by submitting the parts to stress (compression of time) which will precipitate early failures. Defective parts are eliminated from the population, and the surviving parts have a much smaller proportion of defective items. The economical benefits of such procedures are clear, since the cost of stressing items is usually much smaller than the cost of having failures occur after the equipment is in the field.

An excellent description of the problem is given by Perlstein, Littlefield, and Bazovsky [12] where there is a derivation of stress screen durations that will leave the failure rate of the surviving parts within a chosen distance of the failure rate of the "good" parts. In this article the determination of optimal durations is approached from a Bayesian point of view. In addition to its philosophical advantages, the Bayesian approach pays more attention to economic considerations and is more realistic than the available classical-inspired plans which are conditional on unknown quantities.

The ESS experiment also provides data that statisticians may use to learn

Naval Research Logistics, Vol. 41, pp. 215–228 (1994)

CCC 0894-069X/94/020215-14

more about the production process. In the last part of this article Bayesian inference for the rate of early failures is developed.

2. NOTATION AND ASSUMPTIONS

We are considering a lot of N parts having a proportion p of substandard parts. These substandard parts have hidden latent defects which will very likely provoke the early failures. They cannot be detected by quality control (even if applied 100%) because they appear to be good parts. (It is assumed that every part could, at least conceptually, be opened. One could then verify precisely whether the part was substandard or not.)

This parts purging situation is essentially the setup used by Perlstein et al. [12]. Environmental Stress Screening can also be applied to populations of assembled parts or of finished products. The items failing the screen are then repaired or improved, instead of replaced, as in the parts case. Such models are known as complex systems. We shall discuss the parts populations case, although the same Bayesian approach is applied to the complex systems situation without any further conceptual complication.

The failure rate of the substandard parts *under stress* is λ_e, while the failure rate of the good parts under stress is λ_g, $\lambda_e > \lambda_g$. There are two points which should be carefully considered on the applicability of this model. First, the constant failure rate assumption, which often is just a mathematical simplification, plays a crucial role in ESS. The assumption that the failure times are exponentially distributed guarantees that the parts surviving the stress screen will not be aged by stressing. The user of a stress screen plan should be attentive to the fact that if the exponential (or perhaps a decreasing failure rate) assumption is not adequate, there might end up being a lot of mostly good parts surviving the screen which will, nevertheless, be aged. (On the other hand, a screen of small duration undertaken soon after manufacture can still be of interest for parts having bathtub failure rates.)

The consideration of an exponential distribution for the lifetime of parts is equivalent to the supposition that parts do not age under stressing. An equivalent way of describing predictively the no-aging assumption is by assuming that the associated process Y, of the number of failures Y_i during the ith consecutive unit of time, is Poisson. Either formulation induces the parametric exponential model. A judgment of *partial exchangeability*, that is, of exchangeability of parts having the same quality with respect to behavior under stress screening, is also made. De Finetti's representation theorem [5] for partially exchangeable sequences then induces consideration of the two failure rates λ_g and λ_e. The prior density for (λ_g, λ_e) on the plane reflects the statistician's opinion about the failure rates. For example, a prior distribution totally concentrated on the diagonal $\lambda_g = \lambda_e$ is a judgment of complete exchangeability. Since in the statistician's opinion $\lambda_g < \lambda_e$, such a prior will be totally concentrated on the region $0 < \lambda_g < \lambda_e$.

We have derived the parametric modeling of the lifetime of parts from exchangeability and invariance considerations. Such judgments of indifference are the most natural subjective judgments to consider.

Another point that needs to be mentioned is the quantification of stress levels. In Perlstein et al. there are reminders of the fact that the stress level should be kept under a "threshold which could precipitate failure modes that would never occur in normal operation or which could damage the part" [9, Sect. 2.2]. In addition to this assumption of nondamaging stress levels, it is assumed that the stress levels used can be translated into a factor of acceleration of time. In both Bayesian and non-Bayesian derivations, the optimal duration of the stress screen turns out to be the inverse of the acceleration of time factor multiplied by acceleration-invariant constants.

In practice the user of these plans needs to be able to express the stress level as a constant factor of acceleration of time. We will use this assumption and denote such a factor by l. We then have $\lambda_g = l\lambda_g'$ and $\lambda_e = l\lambda_e'$, where λ_g' and λ_e' denote the failure rates under normal operating conditions.

Let T be the optimal duration for the screen (under a given stress level l) and t denote time. We will now introduce the notation and discuss the costs involved in stressing and stopping the screen. Let c_1 represent the cost of having a substandard part escape the screen and c_2 represent the cost of having a good part destroyed by the screen. The costs c_1 and c_2 are "decision" costs in the sense that they describe the cost of wrong "decisions" regarding a part. Since the major concern in ESS is to eliminate poor parts, the cost c_1 is usually much larger than c_2. If one defines a substandard part as a part failing the screen, then $c_2 = 0$. The stressing costs as opposed to decision costs are now considered. Let c_3 be the cost of stressing and failing a part, and c_4 be the cost of stressing and releasing a part. We are assuming, for simplicity, that c_3 and c_4 depend neither on t nor on the quality of the part. These assumptions are of course not suitable in many situations. For example, the cost of stressing can be modeled as a linear function of time of stress of the part until failure or release. Under these more realistic assumptions, the adaptation of the derivation presented in this section is straightforward.

The total cost depends on λ_g, λ_e and on p, the "parameters" of this model. The proportion p of substandard parts in the lot is structurally distinct from the failure rates since it is, at least conceptually, observable. One could open every part and count the number of substandard ones. The *verifiability* [6] of the failure rate values depends on "observing" limits related to infinite sequences of failure times. It is immaterial whether p is treated as a random quantity or as a parameter.

Since p will not be observed, we will model it as a parameter. A joint prior density for $(p, \lambda_g, \lambda_e)$ will be given. The design problem to be solved is the determination of $t = T$ in such a way that the expected total cost with respect to the prior distribution for $(p, \lambda_g, \lambda_e)$ is minimized.

We will consider families of prior densities which are convenient and large enough to accommodate different opinions. A family of joint prior densities for $(p, \lambda_g, \lambda_e)$ which is weakly conjugate will be introduced.

We suggest the use of the beta family of prior densities for p as a natural approximation for a strictly coherent discrete prior on $\{jN^{-1}: j = 0, 1, \ldots, N\}$. The beta family is indexed by positive numbers a and b and the beta(a, b) prior density is given by

$$f(p) = [\Gamma(a + b)/(\Gamma(a)\Gamma(b))]p^{a-1}(1 - p)^{b-1}.$$

The uniform density on (0, 1) is the particular case $a = b = 1$. The expected value of a beta (a, b) distribution is given by $\mathbf{E}(p) = a/(a + b)$ and its variance by $\mathbf{V}(p) = ab(a + b)^{-2}(a + b + 1)^{-1}$.

We will assume the following joint prior density for $(p, \lambda_g, \lambda_e)$:

$$f(p, \lambda_g, \lambda_e) = \theta\tau[\Gamma(a + b)/(\Gamma(a)\Gamma(b))]p^{a-1}(1 - p)^{b-1}e^{-\lambda_g(\theta-\tau)}e^{-\lambda_e\tau},$$

for $0 < p < 1$ and $0 < \lambda_g < \lambda_e$. This is equivalent to assuming p independent of (λ_g, λ_e), a beta(a, b) density for p, an exponential (θ) density for λ_g, and a conditional exponential (τ) density for λ_e shifted by λ_g. It follows that $\mathbf{E}(\lambda_g) = \theta^{-1}$, $\mathbf{E}(\lambda_e|\lambda_g) = \lambda_g + \tau^{-1}$, and $\mathbf{E}(\lambda_e) = (\theta + \tau)/(\theta\tau)$. The marginal distribution of λ_e has a mode at $[\ln(\tau/\theta)]/[\tau - \theta]$. The statistician will elicit values a, b, θ, and τ and it will typically be the case in ESS that $a < b$ and $\theta > \tau$.

In practical situations, where λ_g is much smaller than λ_e, the value of θ will be chosen much larger than the value of τ. The conditional exponential density for λ_e then becomes practically flat if compared to the prior exponential density for λ_g. The prior uncertainty about λ_g can be expressed through prior densities that make use of the knowledge about production process standards, as contained in publications such as the military standards series. On the other hand, the statistician is able to express his relatively much larger ignorance about λ_e given λ_g through an almost flat prior conditional density which is nevertheless proper. In addition to satisfying coherence requirements, proper priors can be very helpful when deriving marginal posterior densities as in Section 5. For values of θ much larger than the value of τ, the following relation shows how small the prior probability of having λ_e close to λ_g is—even if the mode of the marginal prior density $f(\lambda_g, \lambda_e)$ is the origin.

FACT: With the prior $f(p, \lambda_g, \lambda_e)$ given above,

$$\mathbf{P}(\lambda_e < M\lambda_g) = 1 - \theta[\theta + \tau(M - 1)]^{-1}, \qquad \text{for every } M > 1.$$

The proof follows immediately by conditioning on λ_g.

3. OPTIMIZING THE STRESS SCREENING DURATION

There is a proportion p of substandard parts in a lot of size N. But inspection of a part does not reveal whether it is substandard or not. This fact makes all parts look similar and entails a judgment of exchangeability of the parts with respect to quality and behavior under the screening stress experiment. In particular, for any part in the lot, the statistician's probability that it is substandard is $\mathbf{E}(p) = a/(a + b)$, where $\mathbf{E}$ stands for integration with respect to the beta(a, b) prior for p.

The conditional cost per part of a screen of duration t at stress level l is therefore easily derived as

$$p[(1 - e^{-\lambda_e t})c_3 + e^{-\lambda_e t}(c_1 + c_4)] + (1 - p)[(1 - e^{-\lambda_g t})(c_2 + c_3) + e^{-\lambda_g t}c_4].$$

The conditional cost is the *expected* total screening cost of a part. The assumption of exponentiality of the lifetime distribution is used in the derivation of the expression above, since

$$\mathbf{P}(X > x | p, \lambda_g, \lambda_e) = pe^{-\lambda_e x} + (1 - p)e^{-\lambda_g x},$$

where X is the lifetime of a part from the lot. Integration over the sample space for X is correct for the Bayesian since there is no violation of the likelihood principle [3] when the choice of stress duration has to be made before the observation $\min(x, T)$ becomes available. Preposterior integration is a common feature in decision problems concerned with design.

The optimization problem can be easily visualized by means of an *influence diagram* [1]. Figure 1 is the influence diagram for the problem of finding the duration T minimizing the expected value of the cost with respect to the relevant random quantities. The influence diagram is for a single part since the expected cost for the lot is N times the expected cost per part when parts are exchangeable.

After rearranging the terms, we obtain the cost per part expressed as

$$c_3 + c_2(1 - p) + (c_1 + c_4 - c_3)pe^{-\lambda_e t} + (c_4 - c_2 - c_3)(1 - p)e^{-\lambda_g t}.$$

The dependence of the conditional cost on the stress level l is expressed through the failure rates which are average numbers of failures per unit of time under stress. Failure rates under stress level l are related to those under normal operating conditions, λ_g' and λ_e', through the relations stated in Section 2:

$$\lambda_g = l\lambda_g' \quad \text{and} \quad \lambda_e = l\lambda_e'.$$

The last expression for the conditional cost can be integrated with respect to the joint prior $f(p, \lambda_g, \lambda_e)$. We will then have the expected cost per part (in the statistician's opinion) or the *risk* of a screen plan of duration t and stress level l, denoted by $R(t, l)$. The risk obviously depends also on the cost structure of the experiment, but we will omit this from the notation. By using the joint prior $f(p, \lambda_g, \lambda_e)$ presented in Section 2, one obtains

$$R(t, l) = c_3 + c_2 b(a + b)^{-1} + (c_1 + c_4 - c_3)a(a + b)^{-1}\theta\tau(\tau + t)^{-1}(\theta + t)^{-1}$$
$$+ (c_4 - c_2 - c_3)b(a + b)^{-1}\theta(\theta + t)^{-1}.$$

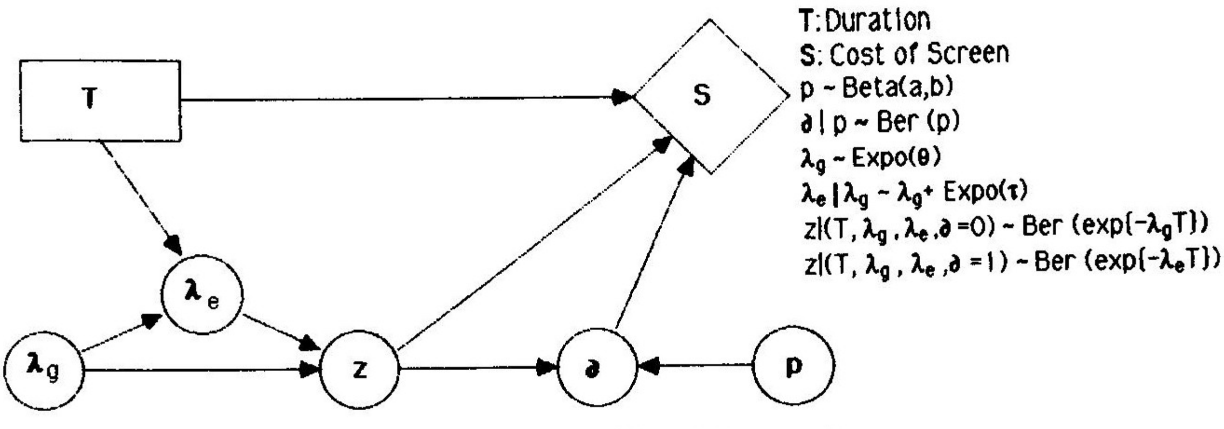

Figure 1. Decision problem influence diagram.

The prior values for normal operating conditions, θ' and τ', must satisfy $\theta' = l\theta$ and $\tau' = l\tau$ for coherence. By using these values, the risk can also be written explicitly in terms of l by simply replacing θ' for θ, τ' for τ, and lt for t in the expression of $R(t, l)$ above.

In order to obtain the optimal T: i.e., the value of t minimizing the risk or expected cost, one can minimize $R(t, l)$ by elementary differentiation methods. Let us set

$$K = \tau^{-1}(b/a)[(c_4 - c_2 - c_3)/(c_3 - c_1 - c_4)].$$

If K is either negative or zero, the optimal T will typically be equal to 0, no screening, or to ∞, screening until failure of all parts. We will examine the more interesting situation where K is strictly positive and finite. Under this assumption, one obtains

$$T = K^{-1}[1 - K\tau + (1 - K(\tau - \theta))^{1/2}]$$

as the optimal duration for the screen. If the value of T in the expression above is negative, then the optimal decision is to not stress the population. Notice that T can be written in the form $T = l^{-1}D$, with D being a constant relative to acceleration of time; that is, D has the same value for all l. The minimal total expected cost of the screening experiment of optimal duration T is $R(T, l)$ multiplied by N, by the exchangeability assumption.

The optimal solution obtained is not a strict Bayes solution, as it is not sequential. A sequential solution would dynamically incorporate the information resulting from the failure of any part in a new (conditional) derivation of an optimal time T. Such a genuine Bayesian solution is nevertheless too complicated to obtain. We derived, instead, the preposterior solution, which is sufficiently close to the sequential one.

We now derive goodness measures of a screen of duration t. The probability that a substandard part will escape from the screening is

$$\mathbf{E}(e^{-\lambda_e t}) = \theta\tau/[(\tau + t)(\theta + t)].$$

The expected number of substandard parts that will escape from the screening, also called the *remaining defect density*, is therefore

$$D_R(t) = N\mathbf{E}(p)\theta\tau/[(\tau + t)(\theta + t)] = N[a/(a + b)]\theta\tau/[(\tau + t)(\theta + t)].$$

The use of the assumption of prior independence between p and λ_e should be noticed.

The probability that a substandard part will *not* escape from the Screen is sometimes called the *screening strength* (SS) and we have

$$SS(t) = 1 - \theta\tau/[(\tau + t)(\theta + t)].$$

On the other hand, the probability that a good part will survive the screening is

$$\mathbf{E}(e^{-\lambda_g t}) = \theta/(\theta + t)$$

and the expected number of good parts remaining in the lot after the Screen is

$$N\mathbf{E}(1 - p)\theta/(\theta + t) = N[b/(a + b)]\theta/(\theta + t).$$

Another measure of interest in the military standards literature is the *yield*, defined as the prior probability of having zero substandard parts remaining in the lot after the Screen. The yield is derived as

$$\begin{aligned} Y(t) &= \mathbf{E}([1 - e^{-\lambda_e t}]^{pN}) = \mathbf{E}([p(1 - e^{-\lambda_e t}) + (1 - p)]^N) \\ &= \theta\tau[\Gamma(a + b)/\Gamma(a)] \sum_{j=0}^{N} \binom{N}{j} (-1)^{N-j}\Gamma(a + N - j) \\ &\quad /[\Gamma(a + b + N - j)(\tau + (N - j)t)(\theta + (N - j)t)] \\ &= \theta\tau[\Gamma(a + b)/(\Gamma(a)\Gamma(b)\Gamma(a + b + N))] \\ &\quad \times \sum_{j=0}^{N} \sum_{i=0}^{j} \binom{N}{j}\binom{j}{i}(-1)^{j-i}\Gamma(a + j)\Gamma(b + N - j) \\ &\quad /[(\tau + (j - i)t)(\theta + (j - i)t)]. \end{aligned}$$

The approximation above is obtained by assuming a prior beta(a, b) density for a propensity p in place of the proportion p. Such "infinite population" models are discussed in Section 5. The identity then results from the equivalence between a beta(a, b) assumption for the propensity p and a beta-binomial (N, a, b) assumption for the number of substandard parts initially present in the lot: By using the equivalence, we obtain

$$\mathbf{E}([p(1 - e^{-\lambda_e t}) + (1 - p)]^N) = \sum_{j=0}^{N} q_j \mathbf{E}[(1 - e^{-\lambda_e t})^j],$$

where q_j are the beta-binomial probability function values.

For fixed values of the yield $Y(t)$, it is very difficult to solve for t, using either of the expressions above. Hence, we suggest the following heuristic approximation, which is a Poisson term for zero occurrences

$$Y(t) \approx e^{-\theta\tau[Na/(a+b)]/[(\theta+t)(\tau+t)]} = e^{-D_R(t)}.$$

The parameter of the approximating Poisson distribution is the remaining defect density $D_R(t)$. Notice that $D_R(t) = (1 - SS(t))D_{\text{in}}$, where D_{in} is the expected number of substandard parts before the screen, or the *incoming defect density* $[Na/(a + b)]$, while $1 - SS(t) = \theta\tau/[(\theta + t)(\tau + t)]$.

4. NUMERICAL EXAMPLES

We will first examine the same numerical example of Section 5 of Perlstein et al. [12]. They consider a population of 3750 electronic parts and derive a duration of 215 hours in order to have 96% "power screen" value. Let us now consider a cost structure given by $c_1 = 100$, $c_2 = 20$, $c_3 = 1$, and $c_4 = 0.01$. Suppose the statistician chooses $a = 1$ and $b = 3999$, and (knowing l) $\theta = 2(10^6)$ and $\tau = 67$. We obtain $K = 12.6535$ and $T = 330.64$ hours. The expected cost when using this duration is 0.0179 per part or 67.07 for the whole lot. The expected cost when using a duration of 215 hours is 0.0184 per part or 68.94 for the whole lot. The expected cost per part if no screening at all is performed ($t = 0$) is $0.0350 - 0.01 = 0.0250$ or 93.75 for the lot. Notice that c_4 is to be subtracted from $R(0, l)$ in order to obtain the correct risk at $t = 0$ (there is no stressing cost when $t = 0$). The relative stability of the expected cost—per part—for small and moderate durations is caused, in this particular example, by the assumptions of an extremely small proportion of defectives in the lot (as expressed by the prior beta values of a and b) and independence of c_3 and c_4 from t. The graph of $R(t, l)$ as a function of t is on Figure 2.

If the optimal duration $T = 330$ is used, the expected number of good parts surviving the test is

$$3750 \times (3999/4000) \times 2(10^6)/[2(10^6) + 330] = 3748.44$$

and the expected number of defective parts surviving the test is

$$3750 \times (1/4000) \times 2(10^6) \times 67/[\{2(10^6) + 330\}\{67 + 330\}] = 0.16.$$

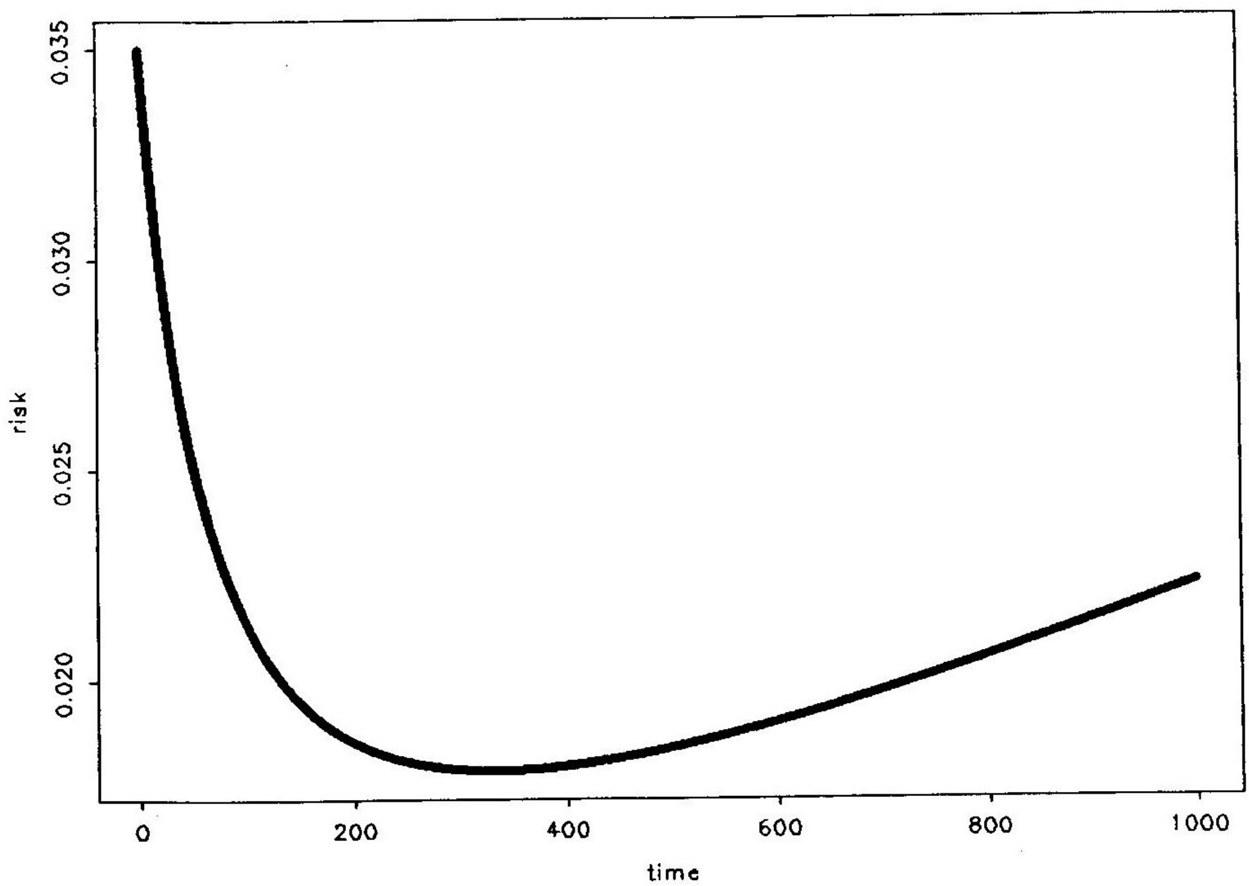

Figure 2. The risk function.

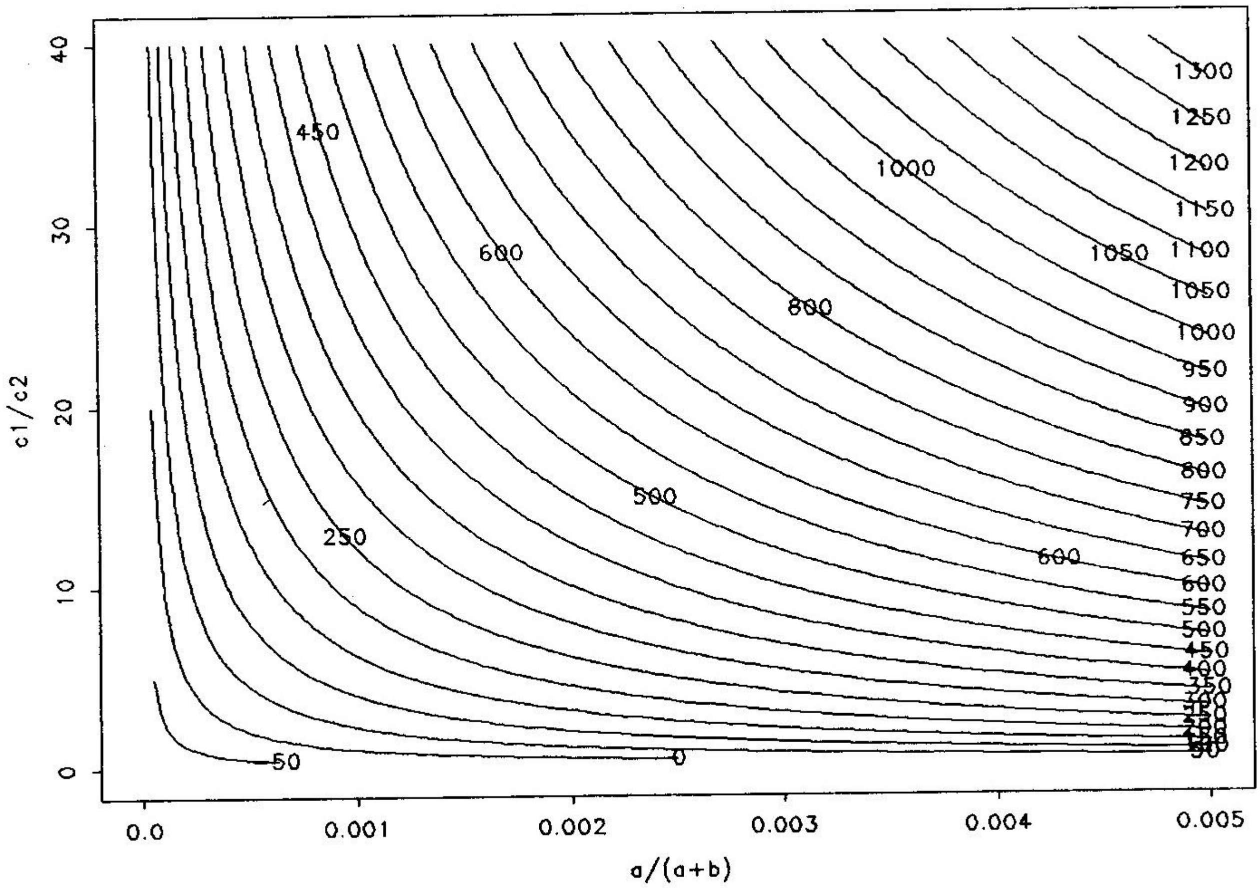

Figure 3. Contour curves of optimal duration.

There is no precise way of comparing the Bayesian approach to the power screen approach of Perlstein et al. [12], since they do not consider any cost structure in their numerical example. In addition, their non-Bayesian approach is conditioned on known values of the parameters.

We now consider the particular situation where $c_3 = c_4 = 0$; that is, stressing is free. In Figure 3, we present a plot showing contour curves of optimal duration: The horizontal axis has values of $a/(a + b)$, the prior expected proportion of substandard parts. The vertical axis has values of c_1/c_2, the ratio between the decision costs. The optimal duration values were computed by taking $\theta = 10^5$ and $\tau = 10^2$. Notice, however, that since T is always proportional to the inverse stress level, l^{-1}, the plot is useful for any other values of θ and τ such that $\theta/\tau = 10^3$. For prechosen screen durations, optimal stress levels can actually be obtained from such plots, i.e., from the relation between T and l^{-1} derived in Section 3.

Similar plots can be constructed for different values of θ/τ, c_3, and c_4. Notice also the "*negative*" values of T for the small values of c_1/c_2 that make it optimal to not start screening at all.

5. STATISTICS: THE POSTERIOR DENSITY

In the previous sections, we used the Bayesian approach to decision-making in the problem of determining the optimal duration of a stress screening test.

It was assumed that the user has the figures for the cost structure and values determining the joint prior density for the three parameters involved in the problem. Now, we will consider the use of *data*. In fact, an ESS experiment provides data consisting of failure times and the number of truncations at the duration. The genuine Bayesian obtains a joint posterior (to the data) density which is his updated (by the data) opinion about the parameters. He will use it as the prior in the derivation of the duration of a possible second screen. Alternatively, he might use the updated opinion in order to suggest changes in the production process. Conceptually, the production process could be fine tuned up to the point where future screenings would not be necessary. This is of course an idealistic goal, but it illustrates the dynamics of Bayesian statistical control: The screening experiment purges the lot; in addition to this original aim, it provides data which are informative about the production process. Changes in the production process might improve it, making future screenings shorter and less expensive.

There are computational difficulties in the derivation and use of such joint posteriors—they are not as simple as the joint prior that we presented in Section 2. The problem has been approached by Bernardo and Girón [4] who consider a very particular case where the only parameter is p, the other two assumed to be known. They point out the absence of conjugate densities for this kind of model and present some approximations to the posterior methods. The practical difficulty of deriving multidimensional posterior densities has been recently discussed by several authors, such as Kass, Kadane, and Tierney [8]. The problem of using data provided by the screen has also been discussed, from a non-Bayesian point of view, by Mendenhall and Hader [10], who suggest an iterative method to solve simultaneous equations for maximum likelihood estimates. However, they assume that all failed parts have their quality revealed; Rider [13] uses the classical method of moments for estimation of the parameters. However, he assumes no truncation of the observations. In addition, the classical method of moments very often provides negative estimates for the failure rates. Coherent procedures never obtain such conclusions. An efficient sequential design of ESS screens should make use of the (truncated) data arising from the screens. Such a procedure would not need any estimation experiment and would be better derived as a Bayesian sequential design procedure.

We suggest a "restriction to interest" approach: We will slightly summarize the data, sacrificing some of the information obtained about the failure rates, in order to obtain a computable marginal posterior density for p. It is important to realize that the original model is being replaced by another one. Specifically, the proportion *per se* of substandard parts in the original lot is no longer a parameter of interest. Actually, after the screening this proportion is drastically reduced. At this point, what we are interested in is the proportion of substandard parts in the *next* lot. Hence, we are now considering a model where the production process generates parts which are substandard with propensity p and from which the first lot was actually a sample of size N, with a proportion of substandard parts no more necessarily equal to p. This modeling is naturally induced by an exchangeability judgment the user has about the parts with respect to their quality [9].

We will now derive the joint posterior density. The full data provided by the experiment are a list of failure times and the number of surviving parts. A crucial assumption here is that an autopsy on all the failed parts will be performed, with some of them failing to respond, that is, not revealing whether they were of substandard quality or not. This enables one to use the methodology of Basu and Pereira [2] to handle data showing nonresponse.

Once the autopsies are performed, we organize the data as x = number of failed parts that were (revealed by autopsy) substandard, y = number of failed parts that were (revealed by autopsy) good, z = number of failed parts with quality not revealed by the autopsy, and $m = N - (x + y + z)$ number of surviving parts. Notice that the actual failure times are collapsed.

A factorable nuisance parameter has to be now introduced. Let α denote the probability that a failed part does not respond to the autopsy. Once again, the introduction of α in the model depends on a judgment of exchangeability of the failed parts with respect to responsiveness.

Let us recall the prior density introduced in Section 2:

$$f(p, \lambda_g, \lambda_e) = \theta\tau[\Gamma(a + b)/(\Gamma(a)\Gamma(b))]p^{a-1}(1 - p)^{b-1}e^{-\lambda_g(\theta-\tau)}e^{-\lambda_e\tau},$$

for $0 < p < 1$ and $0 < \lambda_g < \lambda_e$. We will use Bayes theorem in order to obtain the joint posterior $f(p, \lambda_g, \lambda_e|x, y, z, m)$ and the marginal $f(p|x, y, z, m)$. The probabilistic dependencies among parameters and data can be easily visualized in the influence diagram for this inference problem (see Figure 4). All the operations performed below correspond to reversals and removals of arcs in the influence diagram aiming toward the final diagram, which has node "data" as the only predecessor of node p.

The likelihood $L(\alpha, p, \lambda_g, \lambda_e|x, y, z, m)$ is proportional to the multinomial probability term:

$$\begin{aligned} &L(\alpha, p, \lambda_g, \lambda_e|x, y, z, m)\\ &\quad = [p(1 - e^{-\lambda_e T})(1 - \alpha)]^x \times [(1 - p)(1 - e^{-\lambda_g T})(1 - \alpha)]^y\\ &\quad\quad \times[\alpha p(1 - e^{-\lambda_e T}) + \alpha(1 - p)(1 - e^{-\lambda_g T})]^z\\ &\quad\quad \times [pe^{-\lambda_e T} + (1 - p)e^{-\lambda_g T}]^m. \end{aligned}$$

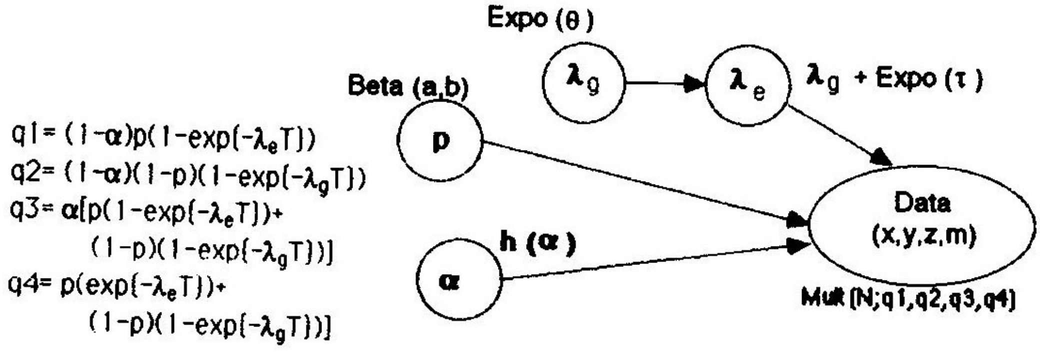

Figure 4. Inference problem influence diagram.

Assuming that α is a priori independent of $(p, \lambda_g, \lambda_e)$, one derives [for any prior density $h(\alpha)$] the posterior density $f(p|x, y, z, m)$ by first using Bayes' theorem to obtain the joint posterior density

$$f(\alpha, p, \lambda_g, \lambda_e|x, y, z, m) = Ch(\alpha)f(p, \lambda_g, \lambda_e)L(\alpha, p, \lambda_g, \lambda_e|x, y, z, m).$$

Successive applications of Newton's binomial formula then yield

$$f(\alpha, p, \lambda_g, \lambda_e|x, y, z, m) = C[h(\alpha)\alpha^z(1-\alpha)^{x+y}]$$
$$\times\, p^{a+x-1}(1-p)^{b+y-1}\sum_{i=0}^{x}\sum_{j=0}^{y}\sum_{l=0}^{m}\sum_{k=0}^{z}\sum_{r=0}^{k}\sum_{s=0}^{z-k}\binom{x}{i}\binom{y}{j}\binom{m}{l}\binom{z}{k}\binom{k}{r}\binom{z-k}{s}$$
$$\times\, (-1)^{x+y+z-i-j-r-s} \times p^{k+l}(1-p)^{m-l+z-k}e^{-\lambda_e T(x-i+l+k-r+\tau/T)}$$
$$\times\, e^{-\lambda_g T(y-j+m-l+z-k-s+(\theta-\tau)/T)}.$$

By recalling the (propriety of prior) fact

$$\int_0^\infty\int_{\lambda_g}^\infty e^{-\lambda_g(\theta-\tau)}e^{-\lambda_e\tau}\,d\lambda_e\,d\lambda_g = (\theta\tau)^{-1},$$

one can integrate out λ_g, λ_e (and α), obtaining

$$f(p|x, y, z, m) = C\sum_{*} G(i, j, l, k, r, s)\,p^{a+x+k+l-1}(1-p)^{b+y+m-l+z-k-1},$$

where Σ_* denotes

$$\sum_{i=0}^{x}\sum_{j=0}^{y}\sum_{l=0}^{m}\sum_{k=0}^{z}\sum_{r=0}^{k}\sum_{s=0}^{z-k}$$

and

$$G(i, j, l, k, r, s) = \binom{x}{i}\binom{y}{j}\binom{m}{l}\binom{z}{k}\binom{k}{r}\binom{z-k}{s}(-1)^{x+y+z-i-j-r-s}$$
$$/[(x-i+l+k-r+\tau/T)(y-j+m+z-s+\theta/T+x-i-r)].$$

Finally, by recalling that

$$\int_0^1 p^{a-1}(1-p)^{b-1}\,dp = \Gamma(a)\Gamma(b)/\Gamma(a+b) = B(a, b)$$

we can solve for C, thus determining the computable analytic expression of the posterior density of p:

$$1/C = \sum_{*} G(i, j, l, k, r, s)B(a+x+k+l,\, b+y+m-l+z-k).$$

Analogous computations provide the posterior moments of p and integration of the posterior density $f(p|\text{Data})$ determines probability intervals for p. We have, for example,

$$\mathbf{E}(p|x, y, z, m) = S_1/S_0,$$

$$\mathbf{E}(p^2|x, y, z, m) = S_2/S_0.$$

where

$$S_\delta = \sum_{*} G(i, j, l, k, r, s)B(a + x + k + l + \delta, b + y + m - l + z - k)$$

for $\delta = 0, 1, 2$.

The joint posterior density $f(p, \lambda_g, \lambda_e|x, y, z, m)$ is seen to be a mixture of densities of the same kind as the joint prior $f(p, \lambda_g, \lambda_e)$. In this sense, they are weakly conjugate. As a consequence, estimates and probability intervals for the failure rates are obtained in an analogous way, even if the experimental design is restricted to p.

6. CONCLUSION

The existing literature for planning, monitoring, and controlling ESS experiments is entirely non-Bayesian. The present work introduces a predictivist Bayesian treatment to the ESS problem which possesses the usual advantages that the Bayesian formulation has over other methods. The duration design problem is solved in an economically optimal way. The difficulty of obtaining an exact separation—i.e., a tractable posterior $f(p, \lambda_g, \lambda_e|\text{Data})$—reflects the general computational problems of handling nonconjugate multidimensional posterior densities. New developments on hardware and software are expected to ease the problem. We approach this situation by summarizing the data and using techniques for data having nonresponse which proved to be successful in medical research area [11]. This enables one to obtain a useful analytic one-dimensional posterior probability density for p. Probability intervals and Bayesian point estimates of p can therefore be obtained. We believe this article indicates a correct way of handling more complex ESS situations which also exist.

ACKNOWLEDGMENTS

The authors wish to thank an anonymous referee for suggestions which greatly improved the paper. This research was partially supported by CNPQ grants 304692/89-1, 404388/90-6 and 352925/92-2 (501864/91-1).

REFERENCES

[1] Barlow, R.E., and Pereira, C.A.B., The Bayesian operation and probabilistic influence diagrams, Report No. ESRC 87-7, Engineering Systems Research Center, University of California, Berkeley, 1987.

[2] Basu, D., and Pereira, C.A.B., On the Bayesian Analysis of Categorical Data: The Problem of Nonresponse, *Journal of Statistical Planning and Inference*, **6**, 345–362 (1982).
[3] Berger, J.O., and Wolpert, R.L., Institute of Mathematical Statistics, Lecture Notes—Monograph Series, Vol. 6, 1984.
[4] Bernardo, J.M., and Girón, J., "A Bayesian Analysis of Simple Mixture Models," *Proceedings of the Third Valencia International Meeting on Bayesian Statistics*, Altea, Spain, June 1987.
[5] De Finetti, B., "Sur la Condition d'Équivalence Partielle (Colloque Genève, 1937), in *Actualités Scientifiques et Industrielles*, No. 739. English translation in Jeffrey, R. (Ed.), *Studies in Inductive Logic and Probability*, 1980, Vol. 2.
[6] De Finetti, B., "Le Vrai et le Probable," *Dialectica*, **3**, 78–92 (1949).
[8] Kass, R.E., Kadane, J.B., and Tierney, L., "Asymptotics in Bayesian Computation," in *Proceedings of the Third Valencia International Meeting on Bayesian Statistics*, Altea, Spain, June 1987.
[9] Lindley, D. V., and Novick, M. R., "The Role of Exchangeability in Inference," *Annals of Statistics*, **9**, 45–58 (1981).
[10] Mendenhall, W., and Hader, R.J., "Estimation of Parameters of Mixed Exponentially Distributed Failure Time Distributions from Censored Life Test Data," *Biometrika*, **45**, 504–520 (1958).
[11] Pereira, C.A.B., and Barlow, R.E., "Medical Diagnosis Using Influence Diagrams," *Networks*, **20**, 565–577 (1990).
[12] Perlstein, H.J., Littlefield, J.W., Bazovsky, I., Sr., "The Quantification of Environmental Stress Screening," *Proceedings of the Institute of Environmental Sciences*, May 1987.
[13] Rider, P.R., "The Method of Moments Applied to a Mixture of Two Exponential Distributions," *Annals of Mathematical Statistics*, **32**, 143–147 (1961).

Manuscript received April 11, 1991
Revised manuscript received June 23, 1993
Accepted July 15, 1993