Ann. Inst. Statist. Math.
Vol. 42, No. 1, 1-19 (1990)

A BAYESIAN APPROACH FOR QUANTILE AND RESPONSE PROBABILITY ESTIMATION WITH APPLICATIONS TO RELIABILITY

MOSHE SHAKED[1] AND NOZER D. SINGPURWALLA[2]

[1]*Department of Mathematics, University of Arizona, Tucson, AZ 85721, U.S.A.*
[2]*Department of Operation Research, School of Engineering, The George Washington University, Washington, D.C. 20052, U.S.A.*

(Received May 27, 1987; revised June 20, 1988)

Abstract. In this paper we propose a Bayesian approach for the estimation of a potency curve which is assumed to be nondecreasing and concave or convex. This is done by assigning the Dirichlet as a prior distribution for transformations of some unknown parameters. We motivate our choice of the prior and investigate several aspects of the problem, including the numerical implementation of the suggested scheme. An approach for estimating the quantiles is also given. By casting the problem in a more general context, we argue that distributions which are IHR or IHRA can also be estimated via the suggested procedure. A problem from a government laboratory serves as an example to illustrate the use of our procedure in a realistic scenario.

Key words and phrases: Sensitivity testing, accelerated life testing, damage prediction, bioassay, quantile estimation, potency curve, Dirichlet process, IHR, IHRA, DHR, dose-response experiments, low dose extrapolation.

1. Introduction and summary

In this paper we explore aspects of a problem which commonly occurs in engineering reliability studies and biological life testing experiments. The point of view that we adopt here is Bayesian.

Suppose that several specimens of an item of interest are tested at distinct stress levels which could be chosen in advance of experimentation. At each stress level, one or more specimens are tested and each specimen is observed to either perform or fail to perform its intended function. Failure to perform is referred to as a *response*. The probabilities of response at the different stress levels may be different, and we assume that these do not decrease with an increase in the stress. Given the outcomes of such tests, one goal is to estimate the probability of response as a function of the stress. This, in many engineering applications, is referred to as *damage probability assessment*. Another goal is to estimate the stress level for which the probability of response is less than or equal to a specified value; this, in some applications, is referred to as *quantile estimation*.

The above problem is not new in the statistical literature and has a long history of development. For instance, it is encountered in the context of biological testing under the names "bioassay" and "low dose extrapolation studies", and in the engineering reliability context under the names "sensitivity testing" and "accelerated life testing".

In bioassay, the dose levels of a drug constitute the stress levels and a response implies failure to survive. In a typical bioassay problem the interest mainly centers around the dose for which the probability of response is .5 or under. Probit and logit transforms are the standard tools used in the analysis of data from bioassay experiments. However, as claimed by Ramsey (1972, 1973), these do not work when only one quantal response is available at each stress level (Mantel (1973) has disagreed). A nonparametric method of constrained maximum likelihood estimation involving isotonic regression does provide numerical answers, but these are not satisfactory because the estimated response curve has a tendency to have jumps. Bayesian approaches to the bioassay problem and its related questions have been proposed in the literature; the works of Kraft and van Eeden (1964), Ramsey (1972), Bhattacharya (1981), Ishiguro and Sakamoto (1983) and Kuo (1983, 1988) are noteworthy.

Remark 1. A related scenario in which the above problem re-appears is the estimation of a (continuous) distribution function based on attribute data. To see this, suppose that the stress is the age of the device, say t, and that the response is the failure of the device to survive until t. Then, the response curve is the distribution function of the device. Using the methods of this paper, the estimation of the response curve would correspond to the estimation of the distribution function (see Remark 2 for further details).

In this paper we adopt the Bayesian theme proposed by Ramsey (1972) and consider extensions of it which are motivated by engineering scenarios. Our main extension involves the assumption that the probability of response is a concave or convex function of the stress. (Note that the response probability can be a convex increasing function of the stress when the stress is bounded from above by some constant; see Section 3 for details.) Other extensions, such as the assumption that the response curve is like a distribution function which has an increasing hazard rate (average) (IHR(A)) will also be addressed, albeit briefly—see Section 3.

It should be pointed out that our procedure can be also used for estimating a distribution function which is assumed to be concave (or convex) and when the data are extremely censored, as described in Remark 1. Papers dealing with nonparametric methods of estimating concave distribution functions for other kinds of censored data are by Denby and Vardi (1986), Winter (1987) and the references therein. Schmoyer ((1984), p. 449) considers the estimation of a concave response curve in quantal bioassay. The key difference between this paper and the above reference is that our approach is Bayesian.

2. Notation and preliminaries

Let $R \equiv [0, \infty)$ be the set of all possible stress levels S, and let $p(S)$ be the unknown response curve. We consider response curves which are nondecreasing and are such that $p(0) = 0$ and $p(\infty) = 1$. Furthermore, we assume that p is right-continuous so that $p(S)$ is a distribution function. Let $0 < S_1 < \cdots < S_M < \infty$ be M distinct stress levels at which the specimens are tested. Assume that n_i (≥ 1) specimens are tested at S_i, $i = 1, \ldots, M$. The outcome at S_i is a random variable X_i representing the number of responses at S_i. We shall assume that a specimen which survives at stress S_i will not be tested again at S_j, $j > i$. Let p_i be the probability of response of a specimen at S_i; thus X_i is binomial with parameters n_i and p_i. From the assumed monotonicity of $p(S)$, it follows that

$$0 \equiv p_0 \leq p_1 \leq p_2 \leq \cdots \leq p_M \leq p_{M+1} \equiv 1 . \tag{2.1}$$

Remark 2. As was mentioned in Remark 1, the above setting reappears in the problem of estimating the life distribution of a device based on attribute data. Specifically, suppose that several specimens, say n_i, of the device are put on a life test at location i, $i = 1, 2, \ldots, M$, and suppose that each device can be observed for failure or otherwise only once, and that at a predetermined time S_i. Let X_i be the number of devices which have failed to survive by time S_i at location i, $i = 1, \ldots, M$. Then one has a set of data as described above. The above form of testing is typical in situations involving national security wherein only one person has the necessary clearance to take an observation at location i. For such a set-up, a non-Bayesian procedure for estimating the underlying life distribution, under the assumption that it has an increasing hazard rate average (IHRA), has been discussed by Shaked (1979, Subsection 5.2).

Given $\boldsymbol{X} = (X_1, \ldots, X_M)$, our one goal is to estimate the unknown p_i's subject to the inequalities (2.1). Another goal would be to estimate $q \equiv p(S)$ for any specified $S > 0$. Clearly, if $S_i < S < S_{i+1}$, then the estimate should satisfy $p_i \leq q \leq p_{i+1}$, $i = 0, \ldots, M$; this corresponds to estimating the probability of response at a stress where no specimen was tested. Yet a third goal would be to estimate the largest stress, say S_α, for which $p(S_\alpha) \leq \alpha$, where

$\alpha \in (0, 1)$ is specified; that is,

$$S_\alpha = \sup\{S: p(S) \le \alpha\}, \qquad 0 < \alpha < 1 . \tag{2.2}$$

Ramsey's (1972) approach is based on that of Kraft and van Eeden (1964)—namely that a prior distribution is assigned to the entire response function. This is done by assigning the Dirichlet as a prior distribution for the successive differences $p_1, p_2 - p_1, \ldots, p_M - p_{M-1}, 1 - p_M$, and then using the modal value of the resulting (joint) posterior distribution—with respect to some convenient measure—as an estimate of $(p_1, \ldots, p_M)$. The modal value is computed with the inequalities (2.1) satisfied. Having estimated the p_i's, the estimation of q and S_α is undertaken via a linear interpolation.

Ramsey's (1972) approach is quite general, since the only assumption made is that $p(S)$ be nondecreasing in S. In some situations there may be additional information about the behaviour of $p(S)$ that is available to an analyst. The nature of such information and some motivation for it is given below. When such is the case, it is desirable to incorporate this prior information into the analysis. A purpose of this paper is to suggest a possible approach for undertaking the above.

2.1 *Motivation for the assumption of concavity*

Assuming that $p(S)$ is concave is equivalent to treating $p(S)$ as the distribution function of a nonincreasing density. To the best of our knowledge, Grenander (1956) was the first to encounter concave distribution functions in his work on the theory of mortality measurement. Kiefer and Wolfowitz (1976, 1979) and the references therein discuss concave distributions in reliability theory. Szekli (1986) has derived the concavity of the waiting time distribution in some $GI/G/1$ queues.

In the context of a submarine hull damage prediction problem (see Section 5), McDonald (1979) chooses $p(S) = 1 - \exp\{-(\lambda S)^\beta\}$, $\lambda > 0$, $\beta > 0$, as his response curve; this function, which is like the distribution function of Weibull, is concave when $\beta \le 1$. McDonald's choice of the Weibull distribution parallels the choice of a Gaussian distribution in probit analysis, the logistic distribution in logit analysis, or the rectangular distribution in rankit analysis (cf. Elandt-Johnson and Johnson (1980), p. 204). The model based on the Weibull distribution is often referred to as the "complementary log-log" model—that is, $\log[-\log(1 - p(S))]$ is a linear function of the parameters. The "half-normal" (with density $\sqrt{2}\,\pi^{-1/2}\sigma^{-1}\exp\{-2^{-1}\sigma^{-2}x^2\}$, $x \ge 0$), the half-logistic (similarly defined) and the rectangular (with left endpoint 0) distributions are concave on $[0, \infty)$. The set-up of this paper would apply if the analyst does not wish to specify any particular form for $p(S)$ except that it is concave.

Similarly, in the reliability theory setting of Remarks 1 and 2, $p(S)$ (which is the distribution function of each of the devices) is concave whenever the devices have DHR.

Another motivation for the concavity of $p(S)$ is due to Denby and Vardi (1986). Consider a repairable system with many identical components having independent lifelengths with a common distribution F having a mean μ. Suppose that we start observing the system after it has been working for a long time. Then the time to failure of each of the working components, as measured from the time we begin our observation, has the distribution function

$$p(S) = \mu^{-1} \int_0^S [1 - F(s)] ds, \qquad S > 0 .$$

Since $1 - F(s)$ is nonincreasing in s it follows that $p(S)$ is concave.

In this paper we will introduce a Bayesian procedure which can take into account various forms or prior information about $p(S)$ such as:

(i) $p(S)$ is a nondecreasing concave function of S.
(ii) $p(S)$ is a nondecreasing convex function of S.
(iii) $p(S)$ is an IHR (DHR) distribution.
(iv) $p(S)$ is an IHRA (DHRA) distribution.

Case (i) will be worked out in detail and a numerical example will be given. An indication as to how to take assumptions (ii), (iii) or (iv) into account will be given in Section 3.

3. Developments of the prior distribution

Consider the situation wherein $p(S)$ is assumed to be concave and nondecreasing. Let $S_0 \equiv 0$ and let S_{M+1} be the right endpoint of the support of p, i.e., $S_{M+1} = \sup\{S: p(S) < 1\}$; S_{M+1} may be ∞. Thus $p_0 \equiv p(S_0) = 0$ and $p_{M+1} \equiv p(S_{M+1}) = 1$. The likelihood of the response probabilities at the observed stresses $S_1 < S_2 < \cdots < S_M$ is

$$L = \prod_{i=1}^{M} \binom{n_i}{X_i} p_i^{X_i} (1 - p_i)^{n_i - X_i} . \tag{3.1}$$

Let $\Delta_i = S_i - S_{i-1}$, $i = 1, \ldots, M + 1$; note that $\Delta_{M+1} = \infty$ if $S_{M+1} = \infty$. Let $Z_i = (p_i - p_{i-1})/\Delta_i$, $i = 1, \ldots, M + 1$; note that if $\Delta_{M+1} = \infty$, then $Z_{M+1} = 0$; in such cases we define $\Delta_{M+1} Z_{M+1} = 1 - p_M$. The monotonicity of $p(S)$ ensures that $Z_i \geq 0$, $i = 1, \ldots, M + 1$. Also $\sum_{i=1}^{M+1} \Delta_i Z_i = 1$. Furthermore, if $Y_i = Z_i - Z_{i+1}$, $i = 1, \ldots, M + 1$ (here $Z_{M+2} \equiv 0$), then the concavity of $p(S)$ implies that $Y_i \geq 0$, $i = 1, \ldots, M + 1$. Since $Z_i = Y_i + \cdots + Y_{M+1}$, and $\sum_{i=1}^{M+1} \Delta_i Z_i = 1$, we have

$$\Delta_1 Y_1 + (\Delta_1 + \Delta_2) Y_2 + \cdots + (\Delta_1 + \cdots + \Delta_i) Y_i + \cdots$$
$$+ (\Delta_1 + \cdots + \Delta_{M+1}) Y_{M+1} = 1 ,$$

or, equivalently,

$$S_1 Y_1 + S_2 Y_2 + \cdots + S_{M+1} Y_{M+1} = 1 . \tag{3.2}$$

Motivated by (3.2), we define

$$U_i = S_i Y_i = S_i(- \Delta_i^{-1} p_{i-1} + (\Delta_i^{-1} + \Delta_{i+1}^{-1}) p_i - \Delta_{i+1}^{-1} p_{i+1}) ,$$
$$i = 1,\ldots, M + 1 ,$$

and note that

$$U_i \geq 0, \quad i = 1,\ldots, M + 1 , \tag{*}$$

and

$$U_1 + U_2 + \cdots + U_M + U_{M+1} = 1 . \tag{**}$$

The restrictions (*) and (**) imply that $\boldsymbol{U} = (U_1,\ldots, U_{M+1})$ can be regarded as a vector of probabilities, and so a meaningful prior distribution on $\boldsymbol{U}$ would be a Dirichlet distribution. Specifically, we assume that for some constants $\alpha_i > 0$, $i = 1,\ldots, M + 1$, and $\beta \geq 0$, with $\sum_{i=1}^{M+1} \alpha_i = 1$, the random vector $\boldsymbol{U}$ has the prior density at $\boldsymbol{u} = (u_1, u_2,\ldots, u_{M+1})$

$$f(u_1,\ldots, u_{M+1}) = \frac{\Gamma(\beta\alpha_1 + \cdots + \beta\alpha_{M+1})}{\prod_{i=1}^{M+1} \Gamma(\beta\alpha_i)} u_1^{\beta\alpha_1 - 1} \cdots u_M^{\beta\alpha_M - 1} u_{M+1}^{\beta\alpha_{M+1} - 1} , \tag{3.3}$$

where $u_i \geq 0$, $i = 1,\ldots, M + 1$ and $\sum_{i=1}^{M+1} u_i = 1$. It is well known that

$$EU_i = \alpha_i, \quad i = 1,\ldots, M + 1 , \tag{3.4}$$

and that

$$\mathrm{Var}\,(U_i) = \frac{\alpha_i(1 - \alpha_i)}{\beta + 1}, \quad i = 1,\ldots, M + 1 . \tag{3.5}$$

By writing $\boldsymbol{p} = (p_1, \ldots, p_{M+1})$ as a transformation of $\boldsymbol{U}$ (details in Subsection 3.1) we can obtain the prior density of $\boldsymbol{p}$.

Remark 3. It is of interest to point out the limiting behavior of our prior as the Δ_i's go to 0. If S_i is fixed and Δ_i and Δ_{i+1} go to 0, then, for a "smooth" $p(S)$, Y_i looks like $-p''(S_i)\Delta(S_i)$, where p'' denotes the second derivative of p and $\Delta(S_i)$ has an obvious interpretation. The constraint (3.2) corresponds then to the relation

$$-\int_0^\infty Sp''(S)dS = 1$$

provided $\lim_{S\to 0} Sp'(S) = 0$ and $\lim_{S\to\infty} Sp'(S) = 0$, where p' denotes the first derivative of p. Thus, our approach, "in the limit", amounts to assuming a Dirichlet prior for the function

$$g(S) = -\int_0^S tp''(t)dt = Sp'(S) - p(S)\,. \tag{3.6}$$

That is, the quantities

$$V_i = -\int_{S_{i-1}}^{S_i} tp''(t)dt = S_{i-1}p'(S_{i-1}) - S_ip'(S_i) + p_i - p_{i-1} \tag{3.7}$$

have a joint Dirichlet distribution. Our method can be regarded as a particular discretised approximation to (3.7), but bearing in mind that (3.6) and (3.7) lack a natural interpretation.

The definition of the U_i's (particularly the fact that they satisfy (*) and (**) which enables us to assign a meaningful prior to them) illustrates the idea underlying the above transformation. Specifically, given prior information of the type (i)–(iv), we look for a transformation $\boldsymbol{U}$ of $\boldsymbol{p}$ which satisfies (*) and (**) (or even just (*)), assign a meaningful prior on $\boldsymbol{U}$ and then use the inverse transformation $\boldsymbol{U} \to \boldsymbol{p}$ to obtain the prior of $\boldsymbol{p}$.

For example, suppose that $p(S)$ is convex. Then the right endpoint of its support, say A, must be finite. Often, when $p(S)$ is convex, the value of A is known (for example, the statistician may know from past experience the maximal value that the underlying random life may take on). In this case $S_{M+1} = A$. Let Δ_i and Z_i be as defined before, but instead of considering the Y_i's as $Z_i - Z_{i+1}$, let $\hat{Y}_i = Z_i - Z_{i-1}$, $i = 1, \ldots, M+1$ (here $Z_0 \equiv 0$). The convexity of $p(S)$ implies that $\hat{Y}_i \geq 0$, $i = 1, \ldots, M+1$. If we set $\hat{U}_i \equiv (\Delta_i + \cdots + \Delta_{M+1})\hat{Y}_i = (S_{M+1} - S_{i-1})\hat{Y}_i$, $i = 1, \ldots, M+1$, then

$$\hat{U}_1 + \cdots + \hat{U}_{M+1} = 1\,. \tag{3.8}$$

Thus the $\hat{U}$'s also satisfy (*) and (**) and so we can assign to them the Dirichlet prior density (3.3). By transforming $\hat{\boldsymbol{U}} \to \boldsymbol{p}$ one can obtain the prior density of $\boldsymbol{p}$.

The above strategy (of finding a transformation of $\boldsymbol{p}$ which satisfies (*) and (**)) has also been used in the more general context of what is known as "non-parametric Bayesian problems", in which the estimation of a general distribution function is of prime concern. To see the above, recall that Ramsey (1972) assigns a Dirichlet prior on (using our notation) $\Delta_1 Z_1, \ldots, \Delta_M Z_M, \Delta_{M+1} Z_{M+1}$; here the $\Delta_i Z_i$'s satisfy (*) and (**). Ferguson (1974) and also Antoniak (1974) observe that this is equivalent to choosing a Dirichlet process prior for the response curve $p(S)$. It should be emphasized, however, that what Ramsey (1972) ends up with is the posterior distribution of $\boldsymbol{p}$ (using our notation)—not the posterior distribution of the whole curve $p \equiv \{p(S), S \geq 0\}$. Ramsey then obtains an estimate of $p(S)$ by first obtaining point estimates of p_i, $i = 1, \ldots, M$, and then using an interpolation procedure.

The observation of Ferguson (1974) and Antoniak (1974), mentioned above, indicates a method of assigning a convenient prior distribution to $\boldsymbol{p}$ in cases (iii) and (iv) as follows.

If $p(S)$ is assumed to be an IHR distribution (case (iii)), then the prior on it can be assigned via the following consideration: Let $\{w(S), S \geq 0\}$ be a right-continuous stochastic process, nondecreasing and nonnegative and which satisfies $\int_0^\infty w(S)\,dS = \infty$ a.s. The process $\{w(S), S \geq 0\}$ can be thought of as a random hazard rate and $p(S)$ is obtained from the transformation

$$p(S) = 1 - \exp\left\{ -\int_0^S w(v)\,dv \right\}, \quad S \geq 0\,. \tag{3.9}$$

The joint prior distribution of $p_1 = p(S_1), \ldots, p_M = p(S_M)$ can then be found from (3.9). See Dykstra and Laud (1981), Padgett and Wei (1981), Burridge (1981) and Ammann (1984) for treatment of (3.9) and further references. In a similar manner we can assign a prior on $p(S)$ if it is known that $p(S)$ is a DHR distribution.

If $p(S)$ is known to be an IHRA distribution (case (iv)), then we can get a prior on it by letting $\{w(S), S \geq 0\}$ be a right-continuous stochastic process which satisfies:

$$w(S) \text{ is nondecreasing in } S \text{ a.s.}$$

$$\lim_{S \to 0} S w(S) = 0 \quad \text{a.s.}$$

$$\lim_{S \to \infty} S w(S) = \infty \quad \text{a.s.}$$

and then letting

$$p(S) = 1 - \exp\{-Sw(S)\}, \qquad S \geq 0 .$$

Similarly, one can assign a prior if it is known that $p(S)$ is a DHRA distribution.

In this paper, only case (i) will be discussed in detail.

3.1 *Interpretation of the prior*

Whereas Ramsey (1972) assigns a Dirichlet prior distribution on the differences of the response probabilities $p_i - p_{i-1}$, $i = 1,\ldots, M+1$, we, by assuming concavity of $p(S)$, assign a Dirichlet prior distribution on

$$U_i = S_i(-\Delta_i^{-1}p_{i-1} + (\Delta_i^{-1} + \Delta_{i+1}^{-1})p_i - \Delta_{i+1}^{-1}p_{i+1}), \qquad i = 1,\ldots, M+1 .$$

By transforming $\boldsymbol{U} \to \boldsymbol{Y}$, $\boldsymbol{Y} \to \boldsymbol{Z}$ and $\boldsymbol{Z} \to \boldsymbol{p}$, a straightforward computation yields the prior density of $\boldsymbol{p} = (p_1,\ldots,p_M)$ as

$$(3.10) \qquad f(p_1,\ldots,p_M) = \frac{\Gamma(\beta)}{\prod_{i=1}^{M+1} \Gamma(\beta\alpha_i)} \prod_{i=1}^{M+1} \left(\frac{S_i}{\Delta_i}\right) \times \prod_{i=1}^{M+1} \{S_i[-\Delta_i^{-1}p_{i-1} + (\Delta_i^{-1} + \Delta_{i+1}^{-1})p_i - \Delta_{i+1}^{-1}p_{i+1}]\}^{\beta\alpha_i - 1} ,$$

where $p_0 \equiv 0$, $p_{M+1} = p_{M+2} = 1$ and (3.10) is defined for $\boldsymbol{p} = (p_1,\ldots,p_M)$ with the p_i's satisfying

$$(3.11) \qquad \frac{S_i}{S_{i+1}} p_{i+1} \leq p_i \leq \frac{\Delta_{i+1} + \Delta_{i+2}}{\Delta_{i+2}} p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}} p_{i+2}, \qquad i = 1,\ldots, M .$$

To interpret this prior we find it convenient to use the notation "$X \sim$ Beta (α, β, a, b)" which denotes the fact that $X = a + (b-a)Z$, $a < b$, where Z is a standard beta distribution on $(0, 1)$ with parameters α and β; that is, for $\alpha > 0$, $\beta > 0$ the density of Z is

$$f(z; \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha-1}(1-z)^{\beta-1} & 0 < z < 1 , \\ 0 & \text{otherwise} . \end{cases}$$

Let $a_i = \sum_{j=1}^{i} \alpha_j$. A lengthy, but straightforward calculation shows that

(3.10) and (3.11) are equivalent to saying that

$$(3.12)\quad p_M \sim \text{Beta}\left(\beta a_M, \beta\alpha_{M+1}, \frac{S_M}{S_{M+1}}, 1\right),$$

$$(3.13)\quad p_{M-1}|p_M$$

$$\sim \text{Beta}\left(\beta a_{M-1}, \beta\alpha_M, \frac{S_{M-1}}{S_M} p_M, \frac{\Delta_M + \Delta_{M+1}}{\Delta_{M+1}} p_M - \frac{\Delta_M}{\Delta_{M+1}}\right),$$

and

$$(3.14)\quad (p_i|p_{i+1},\ldots,p_M)$$

$$\sim \text{Beta}\left(\beta a_i, \beta\alpha_{i+1}, \frac{S_i}{S_{i+1}} p_{i+1}, \frac{\Delta_{i+1} + \Delta_{i+2}}{\Delta_{i+2}} p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}} p_{i+1}\right),$$

$$i = M-1, M-2,\ldots, 1\,;$$

($(p_i|p_{i+1},\ldots,p_M)$ denotes the random variable p_i conditioned on $p_{i+1},\ldots,p_M$ being given).

Equations (3.12)–(3.14) give a new interpretation to the prior density (3.10), and they also suggest an inductive procedure of developing the prior for $\boldsymbol{p}$. Furthermore, from (3.12)–(3.14), we see that

$$Ep_M = \frac{S_M}{S_{M+1}} + \frac{\Delta_{M+1}}{S_{M+1}} a_M$$

(if $\Delta_{M+1} = \infty$, then $Ep_M = a_M$),

$$E[p_{M-1}|p_M] = \frac{S_{M-1}}{S_M} p_M + \frac{\Delta_M}{\Delta_{M+1}}\left(\frac{S_{M+1}}{S_M} p_M - 1\right)\frac{\alpha_M}{a_M}$$

(if $\Delta_{M+1} = \infty$, then $E[p_{M-1}|p_M] = (S_{M-1}/S_M)p_M + (\Delta_M/S_M)p_M(\alpha_M/a_M)$), and

$$E[p_i|p_{i+1},\ldots,p_M] = \frac{S_i}{S_{i+1}} p_{i+1} + \frac{\Delta_{i+1}}{\Delta_{i+2}}\left(\frac{S_{i+2}}{S_{i+1}} p_{i+1} - p_{i+2}\right)\frac{\alpha_{i+1}}{a_{i+1}},$$

$$i = M-2,\ldots, 1\,.$$

Also,

$$(3.15)\quad \text{Var}\,(p_M) = \left(\frac{\Delta_{M+1}}{S_{M+1}}\right)^2 \frac{a_M\alpha_{M+1}}{\beta+1} = \left(\frac{\Delta_{M+1}}{S_{M+1}}\right)^2 \frac{(1-\alpha_{M+1})\alpha_{M+1}}{\beta+1}$$

(if $\Delta_{M+1} = \infty$, then $\operatorname{Var}(p_M) = a_M\alpha_{M+1}/(\beta+1)$),

$$\operatorname{Var}(p_{M-1}|p_M) = \left[\frac{\Delta_M}{\Delta_{M+1}}\left(\frac{S_{M+1}}{S_M}p_M - 1\right)\right]^2 \frac{a_{M-1}\alpha_M}{\alpha_M^2(\beta a_M + 1)}$$

and

$$\operatorname{Var}(p_i|p_{i+1},\ldots,p_M) = \left[\frac{\Delta_{i+1}}{\Delta_{i+2}}\left(\frac{S_{i+2}}{S_{i+1}}p_{i+1} - p_{i+2}\right)\right]^2 \frac{a_i\alpha_{i+1}}{a_{i+1}^2(\beta a_{i+1} + 1)}, \quad i = M-2,\ldots,1 .$$

3.2 *Choosing the prior parameters*

Let $p^*(S)$, $S \geq 0$, be the best prior guess about $p(S)$, $S \geq 0$, such that $p^*(S)$ is a concave distribution function. Let $p_0^* = 0$, $p_{M+1}^* = p_{M+2}^* = 1$ and $p_i^* = p^*(S_i)$, $i = 1,\ldots, M$. In choosing the prior parameters α_i, $i = 1,\ldots, M+1$, we make use of equation (3.4) and the fact that $U_i = S_i[\Delta_i^{-1}(p_i - p_{i-1}) - \Delta_{i+1}^{-1}(p_{i+1} - p_i)]$, $i = 1,\ldots, M+1$. A natural strategy, then, is to set

$$\text{(3.16)} \qquad \alpha_i = S_i[\Delta_i^{-1}(p_i^* - p_{i-1}^*) - \Delta_{i+1}^{-1}(p_{i+1}^* - p_i^*)], \qquad i = 1,\ldots, M+1 .$$

If, in a particular application, it is not possible to elicit the prior guesses p_i^*, then one approach would be to let $p^*(S) = 1 - \exp\{-S\}$. A justification for this arbitrary choice is the empirically claimed result of Ramsey (1972), that often the posterior distribution is not very sensitive to the choice of α_i, $i = 1,\ldots, M+1$.

In order to choose β we need to have some idea about the uncertainty associated with our choice of p_M^*. This, in practice, can be expressed in one of the following ways:

(a) Suppose that in addition to p_M^*, our best guess about the variance of p_M is $\operatorname{Var}(p_M)$. Then, substituting $\alpha_{M+1} = \Delta_{M+1}^{-1}S_{M+1}(1 - p_M^*)$ in (3.15), we have

$$\operatorname{Var}(p_M) = \left(\frac{\Delta_{M+1}}{S_{M+1}}\right)^2 \frac{(1-\alpha_{M+1})\alpha_{M+1}}{\beta+1},$$

so that

$$\beta = \frac{(S_{M+1}p_M^* - S_M)(1 - p_M^*)}{S_{M+1}\operatorname{Var}(p_M)} - 1$$

(if $S_{M+1} = \infty$, then $\beta = (\operatorname{var}(p_M))^{-1}p_M^*(1 - p_M^*) - 1$).

(b) Often in practice (see, e.g., McDonald (1979)), associated with the very best guess value p_M^*, a user is able to specify two numbers,

$a_M^* > S_M/S_{M+1}$ and $b_M^* < 1$, such that for some γ_M (specified by the user), $0 < \gamma_M < 1$,

$$p\{a_M^* < p_M < b_M^*\} = 1 - \gamma_M .$$

Since $p_M \sim$ Beta $(\beta(1-\alpha_{M+1}), \beta\alpha_{M+1}, S_M/S_{M+1}, 1)$, given p_M^*, we set $\alpha_{M+1} = \Delta_{M+1}^{-1} S_{M+1}(1-p_M^*)$ and find the value of β such that

$$(3.17)\qquad \int_{a_M^*}^{b_M^*} \left(\frac{\Delta_{M+1}}{S_{M+1}}\right)^{-\beta+1} \frac{\Gamma(\beta)}{\Gamma(\beta\alpha_{M+1})\Gamma(\beta(1-\alpha_{M+1}))} \times\left(p_M - \frac{S_M}{S_{M+1}}\right)^{\beta(1-\alpha_{M+1})-1}(1-p_M)^{\beta\alpha_{M+1}}dp_M = 1-\gamma_M .$$

Suppose further that for any one or more of the indices i, $i = M-1,\ldots,1$, a user is also able to specify two numbers $a_i^* > S_{i+1}^{-1}S_i p_{i+1}^*$ and $b_i^* < \Delta_{i+2}^{-1}((\Delta_{i+1}+\Delta_{i+2})p_{i+1}^* - \Delta_{i+1}p_{i+2}^*)$ such that for some γ_i (specified by the user), $0 < \gamma_i < 1$,

$$P\{a_i^* < p_i^* < b_i^* \mid p_{i+1}^*,\ldots,p_M^*\} = 1 - \gamma_i .$$

Then, using the fact that the density of p_i given $p_{i+1},\ldots,p_M$ is (3.14), denoted by $f(p_i|p_{i+1},\ldots,p_M,\beta,\alpha_1,\ldots,\alpha_i)$ say, we can find, for a fixed i, a value of β, β_i say, which satisfies

$$(3.18)\qquad \int_{a_i^*}^{b_i^*} f(p_i|p_{i+1},\ldots,p_M,\beta,\alpha_1,\ldots,\alpha_i)\,dp_i = 1-\gamma_i ,$$

with α_i, $i = 1,\ldots,M$, given in (3.16).

Denote the smallest β_i by β^* and let this be the choice of β. A computer program which determines the smallest value β^* is available; the details of this program are given by Mazzuchi and Soyer (1982).

Our reason for choosing the smallest value of β stems from the fact that large values of β tend to make the mode of the posterior distribution not too different from the mode of the prior distribution. That is, a large value of β causes us to put more "faith" in the prior, with the result that even a large amount of failure data will not change our prior distribution by much.

4. The posterior distribution

The joint density function of the posterior distribution of $p_1,\ldots,p_M$ is proportional to the product of the prior density (3.10) and the likelihood function (3.1). Thus,

(4.1) $$f(p_1,\dots,p_M|X_1,\dots,X_M)$$
$$\propto \prod_{i=1}^{M}\binom{n_i}{X_i}p_i^{X_i}(1-p_i)^{n_i-X_i}\frac{\Gamma(\beta)}{\prod_{i=1}^{M+1}\Gamma(\beta\alpha_i)}\prod_{i=1}^{M+1}\left(\frac{S_i}{\Delta_i}\right)$$
$$\times\prod_{i=1}^{M+1}\left\{S_i\left[-\frac{p_{i-1}}{\Delta_i}+\left(\frac{1}{\Delta_i}+\frac{1}{\Delta_{i+1}}\right)p_i-\frac{p_{i+1}}{\Delta_{i+1}}\right]\right\}^{\beta\alpha_i-1}$$

for

(4.2) $$\frac{S_i}{S_{i+1}}p_{i+1}\le p_i\le\left(1+\frac{\Delta_{i+1}}{\Delta_{i+2}}\right)p_{i+1}-\frac{\Delta_{i+1}}{\Delta_{i+2}}p_{i+2},\qquad i=1,\dots,M\,,$$

where $p_0\equiv 0$ and $p_{M+1}\equiv p_{M+2}\equiv 1$.

It should be noticed that the posterior density (4.1) is a finite mixture of densities of the form (3.10). This can be shown by using the methodology of Antoniak (1974) and Bhattacharya (1981) who have obtained analogous results by showing that, for the model of Ramsey (1972) which was described in Section 2, the posterior density of $\boldsymbol{p}$ is a mixture of Dirichlet distributions.

We have not tried to write explicitly the posterior marginal densities of p_i, $i=1,\dots,M$, nor have we tried to obtain the posterior marginal moments. Consequently, we obtain the mode of the posterior density (4.1) and use it to estimate the response curve. Thus we need to find the M-dimensional point $(\hat{p}_1,\dots,\hat{p}_M)$ which maximizes (4.1) subject to the constraints (4.2).

A way to accomplish the above maximization is using the "Sequential Unconstrained Minimization Technique" (SUMT) of Fiacco and McCormic (1968). A computer code which adopts SUMT for the specific problem considered here is described by Mazzuchi and Soyer (1982). This code has been used on several sets of data taken from various sources, and has proved to be successful. It should be mentioned that in order to be able to implement the SUMT algorithm, and also in order to ensure that (4.1) has a unique mode, we need all the $(\beta\alpha_i-1)$'s to be nonnegative. To achieve this, Mazzuchi and Soyer (1982) followed the scheme proposed by Ramsey (1972) by replacing $\beta\alpha_i-1$ by $\beta\alpha_i$. Thus their code finds the $\boldsymbol{p}$ which maximizes, not (4.1), but some modification of (4.1).

A result of Fahmi *et al.* (1982) can be used in the present setting to analyze the influence of the sample on the posterior distribution. Rewrite (4.1) as

$$f(\boldsymbol{p}|\boldsymbol{X})=h(\boldsymbol{X})\left[\prod_{i=1}^{M}g_i(X_i,p_i)\right]C(\boldsymbol{p})\,,$$

where $g_i(x,p) = p^x(1-p)^{n_i-x}$, $i = 1,\dots, M$. It is easy to verify that, for $i = 1,\dots, M$,

$$g_i(x,p)g_i(x',p') \geq g_i(x',p)g_i(x,p')$$

whenever $x \leq x'$ and $p \leq p'$, that is, g_i is a totally positive of order 2 (TP_2) function. Hence, specializing Theorem 4 and Remark 2 of Fahmi *et al.* (1982), it follows that, marginally, p_i is stochastically increasing in X_i, $i = 1,\dots, M$.

We note in passing that the conclusion of Theorem 5 or even of Remark 4 of Fahmi *et al.* (1982) (that is, that p_i is stochastically decreasing in X_j, $j \neq i$) need not be true in the present setting.

4.1 *Interpolation procedure and the estimation of quantiles*

The M-dimensional point $\hat{\boldsymbol{p}} = (\hat{p}_1,\dots,\hat{p}_M)$ is our estimate of $\boldsymbol{p} = (p_1,\dots,p_M)$. The following problem, regarding our method, should be pointed out. Suppose two persons with the same prior beliefs about the function $p(S)$ use our procedure, but with different choices of the stresses $\boldsymbol{S} = (S_1,\dots, S_{M+1})$. Then it is possible that they may end up with different estimates of $\boldsymbol{p} = (p_1,\dots,p_M)$. At first glance this does not look surprising, because they will have different sets of data. But even if they have the same set of data, but different choices of stresses (for example, if the second person has all the S_i's of the first person, and also some additional "dummy" S_i's, that is, S_i's which have no observations associated with them), then they may end up with different estimates of $\boldsymbol{p} = (p_1,\dots,p_M)$. That means that in our method an approximation is involved somewhere.

Suppose we wish to estimate $q \equiv p(S)$ or some specific S where, for some $i_0 \in \{0, 1,\dots, M\}$, $S_{i_0} < S < S_{i_0+1}$. Two approaches to this problem are possible.

(a) If S has been determined *in advance* of experimentation, then it can be added to the set $\{S_1,\dots, S_M\}$ with the interpretation that this *S is not an observational stress level.* The q, then, has a best guess q^*, the dimension of $\boldsymbol{p}$, then, will increase to $M+1$ and the previous analysis applies to the new setting just as before with M replaced by $M+1$, and then n and the X, which are associated with q, are identically 0.

(b) If we want to estimate q *after* we have obtained $\boldsymbol{p}$, then $\hat{q}$, the estimate of q, can be obtained by linear interpolation. Similarly, S_α, the α-th quantile ($0 < \alpha < 1$), can be estimated by linear interpolation.

Roughly speaking, approach (b) corresponds to the estimation of the function $p(\cdot)$ by a piecewise linear function over the interval $[0, S_M]$. This function is obtained by connecting the points $(S_i,\hat{p}_i)$ and $(S_{i+1},\hat{p}_{i+1})$ by straight lines, $i = 0, 1,\dots, M-1$. On the other hand, approach (a) corresponds to a recomputation of the $\hat{p}_i$'s by adding one more stress level. Approach (a) requires a redetermination of the prior parameter α's and β

(which can be done, e.g., as described in Subsection 3.2); using this approach, $M + 1$ α's are to be determined as compared to the original M α's. It is seen at once that the two approaches usually yield different estimates of $p(S)$ for any specific S.

Approach (b) is easier to perform since it requires only a simple computation of linear interpolation. But approach (b) does not take into account any prior information which one may have regarding the specific S. When such prior information is available, it can be incorporated into the computations by having an additional α. In that case approach (a) should be used. Roughly speaking, approach (b) is an approximation to the "correct" approach (a).

It should be pointed out that a desirable property of an estimation procedure for the problem at hand is that it should enable extra points to be included *without* recalculating the prior. But our method, described in Subsection 3.2, requires this. This shows that the method of Subsection 3.2 is necessarily an approximation.

5. An illustrative example

The following description of what is known as the "submarine pressure hull damage problem" is based on McDonald (1979).

A diminutive model of a submarine pressure hull is subjected to an underwater shock wave created by an explosion. The explosion is caused by either a nuclear device or a more conventional chemical device. The strength of the shock wave is determined by, among other things, the magnitude of the change in the explosion and the location of the epicenter of the explosion from the model of the hull. It is common to refer to the strength of the shock wave as the "stress" to which the model is subjected. The main items of interest in the submarine damage assessment problem are: the stress which the model can withstand without any damage to it, and the stress which assumes damage to the model. More generally, we are interested in an assessment of the effects of the various stress levels on the probability of damage to the model.

In order to achieve the above, a copy of the model is subjected to a particular stress and a record is made of whether the model is damaged or not damaged. This procedure is then repeated over a range of appropriately chosen stress values. Because of economic as well as practical considerations, it is possible to test only one copy of the model at each stress level. The results of the test are given in Table 1. The notation used in Table 1 is explained in Sections 2, 3.2 and 4; the data have been altered for reasons of confidentiality.

Following the discussion given in Section 2, it was felt reasonable to assume that the probability of response $p(S)$ is a concave function of the stress S. This requirement is therefore satisfied by the best guess values p_i^*.

Table 1. Results of tests on submarine hulls at stress levels S_i, best guess values of the probabilities of damage p_i with their 90% coverage probabilities, and their posterior model values $\hat{p}_i$.

Index i	1	2	3	4	5	6
Stress levels S_i	1.20	2.10	2.90	4.20	4.70	4.90
Response X_i	0	0	1	1	1	1
Best guess values p_i^*	.30	.34	.36	.385	.393	.395
Prior 90% probability of coverage values for p_i^*, $[a_i^*, b_i^*]$	[.12, .74]	[.17, .86]	[.20, .92]	[.25, .96]	[.26, .97]	[.27, .97]
Posterior modal values $\hat{p}_i$ assuming concave response	.384	.445	.474	.505	.514	.516
First moment of the marginal posterior distribution of p_i, assuming response is an increasing function of the stress	.378	.441	.482	.518	.524	.528

The best guess values p_i^*, together with the 90% probability of coverage intervals $[a_i^*, b_i^*]$, $a_i^* < b_i^*$, $i = 1, \ldots, M = 6$, were given to us by McDonald (1979) who obtained them using engineering and other subjective prior considerations. These are given in Table 1 and also shown displayed in Fig. 1.

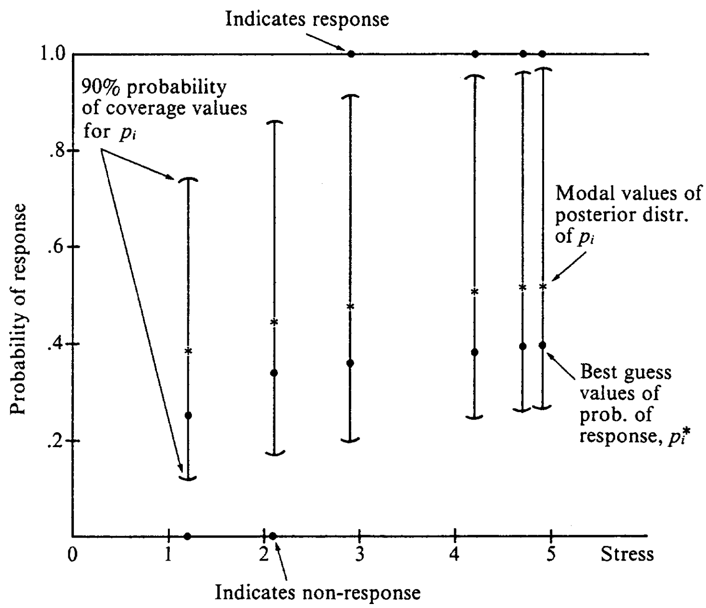

Fig. 1. Best guess values of the probability of damage p_i, their 90% coverage probabilities, and their posterior values.

We note that these values appear to be inconsistent with the data. This is not too surprising, because the best guess values are given prior to observing the data and furthermore, in this case, are based on an engineering analysis.

The prior parameters α_i, $i = 1,\ldots,6$, were computed via the relationship (3.16) and the smoothing parameter β was determined to be 9 using the computer code described in Subsection 3.2.

The mode of the joint posterior distribution $(\hat{p}_1,\ldots,\hat{p}_6)$ was determined using the computer code described in Section 4. The values $\hat{p}_i$ are given in Table 1 and are also displayed in Fig. 1. Not surprisingly, we note that the data have caused us to revise (increase) our prior guesses of the probabilities of failure by an appreciable amount consistent with the constraint of concavity.

5.1 *Discussion and critique of the data analysis*

It may be of interest to compare the $\hat{p}_i$'s to the estimate of the p_i's when we *delete* the constraint that the probability of response be concave. Using the prior best guesses of Table 1 we can determine the prior parameters for the joint distribution of $p_1, p_2 - p_1, \ldots, p_M - p_{M-1}, 1 - p_M$ as described in Mazzuchi (1982). Then the moments of the marginal posterior distributions of p_i can be obtained using the formulas of Mazzuchi (1982). In Table 1 we give the first moments of these marginal posterior distributions. An inspection of the last two rows of Table 1 shows that the posterior probabilities of response with and without the assumption of concavity are not too different. The little difference between the two (in particular at the higher stress levels) is systematic in the sense that concavity causes the response curve to have a decreasing slope. The above exercise addresses the question as to how much could be lost by not assuming concavity and by working instead with priors on the wider class of all distributions. The exercise shows that, at least for the example considered here, not much seems to be lost. This implies that a justification for using concavity (and the technical paraphernalia that accompanies it) is that when appropriate—and opinion about this should be strong—incorporating concavity in the analysis is the proper thing to do.

Another issue pertains to the fact that the proposed method appears to give a lot of emphasis to the prior inputs, since all the posterior model values $\hat{p}_i$ are less than .52, whereas the data contain 1's for the last 4 observations. The maximum likelihood estimates of the p_i's under the assumption of concavity turn out to be .29, .50, .69, 1, 1, 1 and these appear more palatable than the p_i's. The implication here is that there should be a strong justification for a particular choice of the best guess values p_i^*. This latter point is particularly germane since we have not been able to provide a measure of precision (i.e. interval estimates) of our estimates $\hat{p}_i$.

Finally, analyses of the type discussed here should be cast in a decision theoretic framework in order to consider the consequences of actions based upon their results. If the problem involves the safety of personnel within the submarine hulls, then the consequences of a bad decision could be severe. In that case, the expected payoff in terms of increased precision due to using correct prior information may not compensate for the expected loss if the prior information turns out to be misguided. If the problem is to assess the failure characteristics of consumer items, such as refrigerators, then the situation may well be reversed. Thus, different settings may call for different degrees of detail, attention, and procedures, even if the data were to be identical.

Acknowledgement

The authors would like to acknowledge the several suggestions of a referee of this paper, in particular the material given in Remark 3.

REFERENCES

Ammann, L. P. (1984). Bayesian nonparametric inference for quantal response data, *Ann. Statist.*, **12**, 636–645.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Ann. Statist.*, **2**, 1152–1174.

Bhattacharya, P. K. (1981). Posterior distribution of a Dirichlet process from quantal response data, *Ann. Statist.*, **9**, 803–811.

Burridge, J. (1981). Empirical Bayes analysis of survival time data, *J. Roy. Statist. Soc. Ser. B*, **43**, 65–75.

Denby, L. and Vardi, Y. (1986). The survival curve with decreasing density, *Technometrics*, **28**, 359–367.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability, *Ann. Statist.*, **9**, 356–367.

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis*, Wiley, New York.

Fahmi, S., Pereira, C. A. B., Proschan, F. and Shaked, M. (1982). The influence of the sample on the posterior distribution, *Comm. Statist. A—Theory Methods*, **11**, 1757–1768.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures, *Ann. Statist.*, **2**, 615–629.

Fiacco, A. V. and McCormic, G. P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Technique*, Wiley, New York.

Grenander, U. (1956). On the theory of mortality measurement, Part II, *Skand. Aktuarietidskr.*, **39**, 125–153.

Ishiguro, M. and Sakamoto, Y. (1983). A Bayesian approach to binary response curve estimation, *Ann. Inst. Statist. Math.*, **35**, 115–137.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions, *Z. Wahrsch. Verw. Gebiete*, **34**, 73–83.

Kiefer, J. and Wolfowitz, J. (1979). Asymptotically minimax estimation of concave and convex distribution functions, II, *Statistical Decision Theory and Related Topics* II, (eds. S. S. Gupta and D. S. Moore), 193–211, Academic Press, New York.

Kraft, C. H. and van Eeden, C. (1964). Bayesian bio-assay, *Ann. Math. Statist.*, **35**, 886–890.

Kuo, L. (1983). Bayesian bioassay design, *Ann. Statist.*, **11**, 886–895.

Kuo, L. (1988). Linear Bayes estimators of the potency curve in bioassay, *Biometrika*, **75**, 91–96.

Mantel, N. (1973). Quantal bioassay with one animal at a dose level (correspondence), *Biometrics*, **29**, 225–226.

Mazzuchi, T. A. (1982). Bayesian nonparametric estimation of the failure rate, D. Sc. Dissertation, George Washington University, Washington, D.C.

Mazzuchi, T. A. and Soyer, R. (1982). Computer programs for 'A Bayesian approach to quantile and response probability estimation using binary response data'—a user's guide, Technical Paper Serial GWU/IRRA/TR-82/5, George Washington University, Washington, D.C.

McDonald, W. (1979). A statistical treatment of the explosion damage problem, with applications to the case of damage to submarine pressure hulls by very low yield nuclear weapons, Tech. Report, Naval Surface Weapon Center, White Oak, Maryland.

Padgett, W. J. and Wei, L. J. (1981). A Bayesian nonparametric estimator of survival probability assuring increasing failure rate, *Comm. Statist. A—Theory Methods*, **10**, 49–63.

Ramsey, F. L. (1972). A Bayesian approach to bioassay, *Biometrics*, **28**, 841–858.

Ramsey, F. L. (1973). Correction, *Biometrics*, **29**, 830.

Schmoyer, R. L. (1984). Sigmoidally constrained maximum likelihood estimation in quantal bioassay, *J. Amer. Statist. Assoc.*, **79**, 448–453.

Shaked, M. (1979). Estimation of starshaped sequences of Poisson and normal means, *Ann. Statist.*, **7**, 729–741.

Szekli, R. (1986). On the concavity of the waiting-time distribution in some $GI/G/1$ queues, *J. Appl. Probab.*, **23**, 555–561.

Winter, B. B. (1987). Nonparametric estimation with censored data from a distribution with nonincreasing density, *Comm. Statist. A—Theory Methods*, **16**, 93–120.