

BAYESIANS LEARN WHILE WAITING *

WILLIAM S. JEWELL

In many estimation problems, incomplete as well as complete samples are available for Bayesian prediction. After developing the theory for a special, but useful family of distributions, examples are given in life testing, renewal risk processes, life contingencies, and the problem of estimating a defective distribution.

1. INTRODUCTION

In Bayesian prediction problems, one is interested in using observed values of a given process to update the prior knowledge about the process parameters, and thence to make better predictions about the process itself. Most of the theory concerns itself either with exact calculations using so-called natural-conjugate families of prior and likelihood distributions¹, or with best linear least-squares approximations, referred to in the actuarial literature as credibility theory². However, both approaches consider only the use of complete data samples.

The purpose of this paper is to show that there are many situations in which *incomplete* observations also provide updating information, that is, Bayesians can learn while waiting for the finish of the sampling experiment. After developing the necessary theory and introducing the gamma-proportional-hazard family of distributions most appropriate for incomplete data formulations, examples are given from life testing, renewal risk processes, and life contingency reserving. It is shown in what sense an individual life (or cohort of similar lives) can learn about his (their) own remaining lifetime distributions with the passage of time. The paper concludes with the problem of estimating the parameters and the defect in a defective distribution.

2. MODEL

As is usual in Bayesian models, we assume that \tilde{x} , the random lifetime of interest, has a *likelihood distribution function*, $P(x|\theta)$, which depends upon an unknown random parameter $\tilde{\theta}$ which has a *prior distribution function*, $P(\theta)$. We use $P^c = 1 - P$ to denote the complementary distribution (or survival) function, and we assume that (continuous or discrete) densities exist, denoted by $p(x|\theta)$, $p(\theta)$ etc.

* An earlier version of this paper was presented at the 13th ASTIN Colloquium, Washington, D.C., May 1977.

¹ see AITCHINSON and DUNSMORE (1975),

² see JEWELL (1978).

The basic problem is to use observational data, sampled from the likelihood distribution with fixed, but unknown parameter, in Bayes' law to find the *posterior-to-data distribution* of the parameter, and thence to predict various moments and economic functions of the underlying lifetime process.

To illustrate the natural way in which incomplete samples arise, we consider a life-testing scheme in reliability, as in JEWELL (1977), in which:

1. N items, all with lifetimes drawn as samples from $P(x|\theta)$ with common and fixed θ , are put "on test" at epochs $\{t_i\}$, and removed from test at epochs $\{t_i + T_i\}$, ($i, 1, 2, \dots, N$);
2. C of these items (with indices in the set S) will have failed before removal with observed lifetimes $\{\tilde{x}_i = x_i \leq T_i\}$ ($i \in S$);
3. The remaining lifetimes are not completely observed, since the items are still operating at removal, so it is known only that $\{\tilde{x}_i > T_i\}$ ($i \notin S$).

Depending upon the experimental protocol, the $\{T_i\}$ may be fixed in advance, giving then a random C ; or, C may be fixed in advance for a simultaneous test, giving a common, random time-on-test, T . Considering for a moment that the $\{T_i\}$ are fixed, and denoting the observed data by $D = \{y_1, y_2, \dots, y_N; S\}$, where

$$(2.1) \quad \tilde{y}_i = \begin{cases} x_i & (i \in S) \\ T_i & (i \notin S) \end{cases},$$

we can easily argue that the likelihood density of this data set, given θ , is:

$$(2.2) \quad p(D|\theta) = \prod_{i \in S} p(x_i|\theta) \prod_{i \notin S} P_c(T_j|\theta).$$

Bayes' law then gives the predictive density for continued testing of items $j \notin S$, or for future experiments on other items with the same parameter value:

$$(2.3) \quad p(x|D) = \int p(x|\theta) \left[\frac{p(D|\theta) p(\theta)}{\int p(D|\theta') p(\theta') d\theta'} \right] d\theta.$$

The ratio in square brackets is the posterior-to-data parameter density, $p(\theta|D)$.

(2.2) is also useful for many other life testing protocols. Suppose that all items are put on test at the same epoch; the common testing interval T need not be fixed in advance, but may be a continuously-evaluated *stopping rule*, a possibly random decision to stop experimenting that depends upon the values $\{x_1, x_2, \dots, x_C; S\}$ observed up to and including T , but not *directly* upon θ . In this case, the likelihood includes additional terms relating to the stopping rule that cancel out of the ratio in (2.3); the stopping rule is *non-informative*, and the likelihood kernel (2.2) is *sufficient for* θ . For instance, one could stop after the fifth failure, or at T equal to twice the first-observed complete lifetime.

3. THE PROPORTIONAL-HAZARD FAMILY

The calculation of (2.3) can, of course, be carried out by computer for any given prior and likelihood distributions. However, for model-building, it is desirable to use a family of distributions in which the calculations are especially tractable so that parametric behavior can be analyzed theoretically. Unfortunately, the Koopman-Pitman-Darmois exponential family of distributions so useful in credibility theory has no simple form for P^c ; see JEWELL (1974).

However, a special case of the exponential family, the *proportional-hazard family*, has useful properties:

$$(3.1) \quad P^c(x | \theta) = e^{-\theta Q(x)}; \quad p(x | \theta) = \theta q(x) e^{-\theta Q(x)}, \quad (x \geq 0)$$

where $Q(x)$ is a monotone non-decreasing function ($Q(0) = 0$), and $q(x) = dQ(x)/dx$. We note:

1. $\theta Q(x)$ is the *cumulative hazard (failure) function*, making $q(x)$ a *unit- or prototype failure rate*;
2. If \tilde{w} is a random variable with absolute failure rate, $q(w)$, and θ is an integer, the original lifetime, \tilde{x} , has a physical interpretation as

$$\tilde{x} = \min(\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_\theta);$$

3. This family includes the exponential, Weibull, and Gumbel (extreme-value) distributions.

The data likelihood (2.2) becomes:

$$(3.2) \quad p(D | \theta) = \prod_{i \in S} q(x_i) [\theta^C e^{-\theta TQT(D)}],$$

where TQT is a statistic,

$$(3.3) \quad TQT(D) = \sum_{i=1}^N Q(x_i) = \sum_{i \in S} Q(x_i) + \sum_{j \notin S} Q(T_j),$$

referred to in JEWELL (1977) as the *total-Q-on-test-statistic*, a generalization of the "total-time-on-test" concept of reliability life-testing. Note that if item k was already age S_k (and still working) when placed on test, then $Q(S_k)$ should be *subtracted* from the TQT .

A convenient natural conjugate prior for $\tilde{\theta}$ is the gamma density,

$$(3.4) \quad p(\theta) = p(\theta | C_0, Q_0) = \frac{Q_0 (Q_0 \theta)^{C_0 - 1} e^{-\theta Q_0}}{\Gamma(C_0)}, \quad (\theta \geq 0)$$

with hyperparameters C_0, Q_0 ; the usefulness of (3.4) in modelling uni-modal densities is well known. It is easy to see that Bayes' law then gives a posterior-

to-data density of the parameter, $p(\theta | D)$, that is also gamma, with updated parameters:

$$(3.5) \quad C_0 \leftarrow C_0 + C; Q_0 \leftarrow Q_0 + TQT(D).$$

Furthermore, the updated means of $\tilde{\theta}$ and $\tilde{\theta}^{-1}$ obey the exact credibility formulae:

$$(3.6) \quad [E\{\tilde{\theta} | D\}]^{-1} = (1 - Z_1) [E\{\tilde{\theta}\}]^{-1} + Z_1 \left[\frac{C}{TQT(D)} \right]^{-1},$$

$$(3.7) \quad E\{\tilde{\theta}^{-1} | D\} = (1 - Z_2) E\{\tilde{\theta}^{-1}\} + Z_2 \left[\frac{TQT(D)}{C} \right],$$

with credibility factors:

$$(3.8) \quad Z_1 = C / (C_0 + C); Z_2 = C / (C_0 - 1 + C).$$

The posterior-to-data variances are also easily obtained:

$$(3.9) \quad V\{\tilde{\theta} | D\} = \left[\frac{1}{C_0 + C} \right] [E\{\tilde{\theta} | D\}]^2,$$

$$(3.10) \quad V\{\tilde{\theta}^{-1} | D\} = \left[\frac{1}{C_0 - 2 + C} \right] [E\{\tilde{\theta}^{-1} | D\}]^2;$$

the first terms decrease with increasing C , and so, ultimately, with probability one, do the variances. This makes precise the difference between incomplete and complete samples; two different data sets could lead to the same mean forecast, but we would have more “confidence” in the result with the larger number of complete samples.

The terms in square brackets in (3.6) (3.7) are the classical maximum-likelihood estimators got from the term in square brackets in (3.2). If the experiment gives a large number of complete observations, relative to C_0 , then the Bayesian and maximum-likelihood estimators coincide. However, for relatively incomplete tests, more weight is given the prior means, $E\{\tilde{\theta}\} = C_0 / Q_0$, or $E\{\tilde{\theta}^{-1}\} = Q_0 / (C_0 - 1)$.

Classical estimators are often obtained from Bayesian formulae when the prior knowledge becomes “diffuse”; in our model this corresponds to keeping $E\{\tilde{\theta}\}$ or $E\{\tilde{\theta}^{-1}\}$ fixed, and letting the corresponding variances (the prior uncertainty) increase without limit. From (3.9) (3.10) we see this corresponds to letting $C_0 \rightarrow 0$ or $C_0 \rightarrow 2$, respectively (with corresponding adjustments in Q_0). Thus, *with very uncertain prior knowledge*, we get:

$$(3.11) \quad E\{\tilde{\theta} | D\} = \left[\frac{C}{TQT(D)} \right],$$

$$(3.12) \quad E \{ \tilde{\theta}^{-1} \mid D \} = \frac{E \{ \hat{\theta}^{-1} \} + TQT(D)}{1 + C}.$$

Thus, when estimating $\tilde{\theta}$, we place “full credibility” in the maximum-likelihood estimator, and ignore all prior information; but, when estimating $\tilde{\theta}^{-1}$, a Bayesian would always insist on keeping the prior mean as an initial data point, because the prior is still informative and proper in this case.

The *mixed*, or *predictive distribution* of \tilde{x} , averaged over all possible values of $\tilde{\theta}$, is:

$$(3.13) \quad P^c(x \mid C_0, Q_0) = [Q_0 / (Q_0 + Q(x))]^{C_0}, \quad (x \geq 0)$$

with density

$$(3.14) \quad p(x \mid C_0, Q_0) = (C_0 q(x) / Q_0) [Q_0 / (Q_0 + Q(x))]^{C_0+1},$$

a generalization of the shifted Pareto distribution. If the prototype failure function is Gumbel, we get exponential tails for large x in (3.13), while if the underlying failures are Weibull, we get the “more dangerous” algebraic tails. Posterior-to-the-data, predictive density is of the same form, but with updated parameters.

The cumulative hazard function of the mixed distribution is:

$$(3.15) \quad R(x \mid C_0, Q_0) = -\ln P^c(x \mid C_0, Q_0) = C_0 \ln [1 + (Q(x) / Q_0)].$$

One can show that this mixing tends to decrease the rate of failure; in fact, the mixed population may have approximately constant or decreasing hazard rate, even with increasing $q(x)$.

Life testing applications are covered in more detail in JEWELL (1977), and the problem of model identification of the form of Q is also considered. We turn now to applications of these ideas in risk theory.

4. RENEWAL PROCESSES

In one model of the collective risk process, claims are assumed to follow a renewal process. If, during an exposure interval T , C events (accidents, claims, equipment failures, etc.) are observed, this means there are C complete interval samples $\{x_i\}$, and the final interval-in-progress, $T - \sum_{i=1}^C x_i$. If all intervals are sampled from (3.1) with fixed θ , the parameter updating becomes:

$$(4.1) \quad C_0 \leftarrow C_0 + C; \quad Q_0 \leftarrow Q_0 + \sum_{i=1}^C Q(x_i) + Q(T - \sum_{i=1}^C x_i).$$

Note that not only the random number of events in $(0, T]$, but also the actual lengths of the intervals provide information in the general case.

An important special case in risk processes occurs when $Q(x) = x$, leading

to exponentially-distributed intervals, and a Poisson counting process, for each θ . However, here

$$(4.2) \quad Q_0 \leftarrow Q_0 + \sum x_i + (T - \sum x_i) = Q_0 + T,$$

so we conclude that the Poisson process is special in that only the number of events in $(0, T]$, not the epochs of events, provides predictive information!

5. INDIVIDUAL LEARNING ABOUT REMAINING LIFE

We turn now to the interesting question of whether or not a Bayesian can learn about his own remaining lifetime distribution function (rldf). For a mixed population with average tail distribution P^c ,

$$(5.1) \quad Pr \{ \tilde{x} > T + u \mid \tilde{x} > T \} = \frac{P^c(T + u)}{P^c(T)} = P_T^c(u)$$

represents the fraction of those individual components alive (operating) at age T which will survive until age $T + u$.

However, for a single life component with known parameter θ , the appropriate rldf is:

$$(5.2) \quad Pr \{ \tilde{x} > T + u \mid \tilde{x} > T; \theta \} = \frac{P^c(T + u \mid \theta)}{P^c(T \mid \theta)} = P_T^c(u \mid \theta).$$

If we have to estimate this single life behavior as averaged over the population (i.e., without Bayesian learning), we get the prior expected rldf:

$$(5.3) \quad E \{ P_T^c(u \mid \tilde{\theta}) \} = \int \frac{P^c(T + u \mid \theta)}{P^c(T \mid \theta)} p(\theta) d\theta,$$

which is clearly *not* identical with (5.1).

Now let us adopt the Bayesian point of view, and estimate the remaining life of a single individual who has lived to age T ; since he is still alive, we have the single datum $D = \{x > T\}$, which must update the parameter density to:

$$(5.4) \quad p(\theta \mid D) = \frac{P^c(T \mid \theta)p(\theta)}{\int P^c(T \mid \phi)p(\phi)d\phi} = \frac{P^c(T \mid \theta)p(\theta)}{P^c(T)}.$$

So the Bayesian-updated rldf will be

$$(5.5) \quad E_{\tilde{\theta} \mid D} \{ P_T^c(u \mid \theta) \} = \int \frac{P^c(T + u \mid \theta)p(\theta)}{P^c(T)} d\theta,$$

which is *exactly* the same as the population rldf in (5.1)! Stated another way, a single life (or component) cannot, on the average, gather any additional information about his remaining lifetime distribution by the mere passage of time, other than that given for the population as a whole—even though he can learn about his parameter! A surprising, but satisfying result.

6. COHORT LEARNING ABOUT REMAINING LIFE

This does not mean, however, that *several* incomplete samples cannot provide information about other lifetimes with the same θ , nor that a group of lives with the same θ cannot learn from the passage of time. Consider a cohort of N lives with the same parameter which are put "on test" at the same epoch. From Section 2, with $T_i = T$ for all i , we see that the data $D = \{x_i \leq T \ (i \in S)\}$; S changes (5.5) to:

$$(6.1) \quad E_{\tilde{\theta}|D} \{P^c(u | \tilde{\theta})\} = \int P^c(T + u | \theta) \frac{[P^c(T | \theta)]^{N-C-1} \prod p(x_i | \theta) p(\theta)}{P^c(D)} d\theta,$$

where learning would clearly take place.

For the proportional-hazard family,

$$(6.2) \quad P^c_T(u | \theta) = e^{-\theta[Q(T+u) - Q(T)]}.$$

If the prior at $T = 0$ is gamma with hyperparameters C_0 and Q_0 , the posterior-to-data density of θ at T is gamma with hyperparameters $C_0 + C$ and $Q_0 + (N - C)Q(T) + \sum Q(x_i)$, giving finally the special cohort-experienced remaining-lifetime distribution function:

$$(6.3) \quad E_{\tilde{\theta}|D} \{P^c_T(u | \tilde{\theta})\} = \left[\frac{Q_0 + (N - C)Q(T) + \sum Q(x_i)}{Q_0 + (N - 1 - C)Q(T) + \sum Q(x_i) + Q(T + u)} \right]^{C_0 + C}.$$

It is easy to see how learning vanishes when $N = 1$ and $C = 0$.

7. LIFE CONTINGENCIES AND RESERVES

To apply the results above, consider that we are determining the net single premium for a continuous life annuity of \$ 1/year, at force of interest δ , for an individual aged x . Given θ , this would be (we omit the usual overbar notation):

$$(7.1) \quad a_x(\theta) = \int e^{-\delta u} P^c_x(u | \theta) du = \int e^{-\delta u - [Q(x+u) - Q(x)]} du.$$

Let us suppose that the prior on θ is gamma with hyperparameters C_1, Q_1 at the moment of underwriting (age x). The population-average annuity fair premium is then:

$$(7.2) \quad a_x(C_1, Q_1) = Q_1^{C_1} \int e^{\delta - u} [Q_1 + Q_x(u)]^{-C_1} du,$$

where

$$(7.3) \quad Q_x(u) = Q(x + u) - Q(x)$$

is the prototype cumulative failure function for the remaining life, beginning at age x .

Now, suppose we have insured a cohort of N lives aged x , all of whom

have the same parameter, and let us follow the cohort for t additional years. During this time the data provided by the C expirations at additional ages $\{t_i\}$, together with the fact that $N - C$ lives are still in existence at age $x + t$, would update the hyperparameters to:

$$(7.4) \quad C_2 = C_1 + C; \quad Q_2(t) = Q_1 + (N - C) Q_x(t) + \sum_{t \in S} Q_x(t_i).$$

Although it is too late to change the premium, this additional knowledge could be useful in *adaptive modification* of the reserves on the $N - C$ outstanding policies; for a single-premium annuity of \$ 1/year still outstanding at age $x + t$, the correct adaptive reserve would be:

$$(7.5) \quad {}_tV(a_x) = a_{x+t}(C_2, Q_2(t)).$$

We remind the reader that C_2 and $Q_2(t)$ will be random outcomes, depending upon actual cohort experience during ages $(x, x + t]$. Only when there is a single incomplete life under observation ($C_2 = C_1; Q_2(t) = Q_1 + Q_x(t)$) will no learning take place, and the reserves will follow the classic result for an average member of the mixed population:

$$(7.6) \quad a_{x+t} = \int e^{-\delta u} \frac{P_x^c(t+u)}{P_x^c(t)} du = a_{x+t}(C_1, Q_1 + Q_x(t)).$$

A similar development could be given in terms of the net single premium for a life assurance of \$ 1, at force of interest δ , payable at the instant of death of an individual now aged x ,

$$(7.7) \quad A_x(0) = \int e^{-\delta u} p_x(u | 0) du.$$

The appropriate formulae follow from the previous results by the universal relation $A_x = 1 - \delta a_x$.

It is of interest to follow through the actual stochastic behavior of a "learning reserve" of the type (7.5). First of all, we note that adaptive annuity reserves do not decrease as quickly as the corresponding a_{x+t} , for small t and $C = 0$, which can be seen from:

$$(7.8) \quad \begin{aligned} \frac{da_{x+t}(C_2, Q_2(t))}{dt} &= \delta a_{x+t}(C_2, Q_2(t)) - 1 \\ &+ \left[\frac{C_2 q(x+t)}{Q_2(t)} \right] \{ (N - C) a_{x+t}(C_2, Q_2(t)) \\ &- (N - C - 1) a_{x+t}(C_2 + 1; Q_2(t)) \} \end{aligned}$$

as compared with the well-known classical result

$$(7.9) \quad \frac{da_{x+t}}{dt} = \delta a_{x+t} - 1 + \left[\frac{C_2 q(x+t)}{Q_1 + Q_x(t)} \right] a_{x+t}.$$

The term in square brackets is, of course, the failure rate at $x + t$ for the mixed population in the proportional-hazard family, *i.e.*, the derivative of (3.15). When the first and subsequent deaths occur, there is an instantaneous drop in (7.5), since C_2 increases by unity, but $Q_2(t)$ is continuous. In general, if fewer (more) lives than expected terminate during $(x, x + t]$, the reserves on the remaining lives are larger (smaller) than usual, since this indicates that the value of θ is smaller (larger) than average for this cohort. A complementary effect occurs for life assurance learning reserves.

It should be mentioned that a gamma-mixed proportional-hazard model should be used with care for human mortality. If, for example, the prototype failure rate is assumed to follow Makeham’s law, $q(t) = A + Be^{\alpha t}$, we find that the mixed hazard rate (the derivative of (3.15)) is asymptotically constant, due to the failure-rate-decreasing properties of mixing! One would have to assume that, given θ , individuals follow a much stronger “wear-out” (say, Gumbel), in order to obtain a population Makeham-type law. It is interesting to speculate as to whether or not this occurs for closely-matched humans, where θ would have to include health, genetic, and environmental effects.

8. DEFECTIVE DISTRIBUTIONS

Component and human lives are finite, with probability one; however, defective distributions arise in a variety of other operational situations. Consider, for example, the estimation of the time it takes for a number of requests for bids, mailed survey responses, etc., to be returned. Some responses are received rather quickly; in other cases, an answer is never received.

A reasonable model for this situation would add an unknown *defect parameter*, ϕ , to the usual lifetime distribution, as follows:

$$(8.1) \quad P^c(x | \theta, \phi) = \phi + (1 - \phi) P^c(x | \theta); \quad p(x | \theta, \phi) = (1 - \phi) p(x | \theta).$$

ϕ is then the probability that the lifetime is “infinite”.

Under the life testing scheme of Section 2, the likelihood of the date set D becomes:

$$(8.2) \quad p(D | \theta, \phi) = \binom{N}{C} (1 - \phi)^C \prod_{i \in S} p(x_i | \theta) \prod_{j \notin S} [\phi + (1 - \phi) P^c(T_j | \theta)].$$

Assuming all the intervals T_j have common value T , we find the posterior-to-data density of θ and ϕ by a binomial expansion:

$$(8.3) \quad p(\theta, \phi | D) = K \sum_{j=0}^{N-C} \binom{N-C}{j} \phi^j (1 - \phi)^{N-j} \prod_{i \in S} p(x_i | \theta) P^c(T | \theta)^{N-C-j} p(\theta, \phi)$$

where K is a normalizing constant to make $\iint p(\theta, \phi | D) d\theta d\phi = 1$. To illustrate the calculations further, assume that the “honest” part of (8.1), $P^c(x | \theta)$,

is from the proportional-hazard family (3.1), with gamma prior on θ (3.4). For simplicity, assume ϕ is, *a priori*, independent of θ , and has a beta prior density:

$$(8.4) \quad p(\phi) = p(\phi | a_0, b_0) = B^{-1}(a_0, b_0) \phi^{a_0-1} (1 - \phi)^{b_0-1} \quad (0 \leq \phi \leq 1).$$

$B(a_0, b_0)$ is the beta function, $\Gamma(a_0) \Gamma(b_0) / \Gamma(a_0 + b_0)$. After straightforward calculations with these special forms, we find the mixed beta-gamma:

$$(8.5) \quad p(\theta, \phi | D) = \sum_{j=0}^{N-C} \Pi_j(D) p(\phi | a_0 + j, b_0 + N - j) p(\theta | C_0 + C, Q_0 + TQT_j(D)),$$

where

$$(8.6) \quad TQT_j(D) = \sum_{i \in S} Q(x_i) + (N - C - j) Q(T),$$

and the mixing probabilities are given by:

$$(8.7) \quad \Pi_j(D) = K \cdot B(a_0 + j, b_0 + N - j) [Q_0 + TQT_j(D)]^{-(c_0 + c)},$$

where, again, K is a normalizing factor to make $\sum \Pi_j = 1$. It is important to note that, posterior-to-data, the estimates of θ and ϕ are dependent, unless all of the observations are complete. For estimating the mean defect, we have

$$(8.8) \quad E\{\phi | D\} = \sum_{j=0}^{N-C} \Pi_j(D) \left[\frac{a_0 + j}{a_0 + b_0 + N} \right],$$

where we recognize the term in square brackets as the mean of ϕ , given only that we observe j defects out of N trials. For $N = 1$ and no failure:

$$(8.9) \quad E\{\phi | D\} = \frac{a_0}{a_0 + b_0 + 1} \left[\frac{(a_0 + 1) [1 + Q(T)/Q_0]^{c_0 + b_0}}{a_0 [1 + Q(T)/Q_0]^{c_0 + b_0}} \right],$$

which shows clearly how the mean defect increases from its original estimate of $a_0 / (a_0 + b_0)$ towards $(a_0 + 1) / (a_0 + b_0 + 1)$ as $T \rightarrow \infty$ with no failure. Of course, if the lifetime ever terminates, $E\{\phi | D\}$ jumps to $(b_0 + 1) / (a_0 + b_0 + 1)$. Other mixing models are given in JEWELL (1977).

REFERENCES

AITCHINSON, J. and J. R. DUNSMORE (1975). *Statistical Prediction Analysis*, Cambridge University Press, New York.
 JEWELL, W. S. (1974). "Exact Multidimensional Credibility", *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, **74**, 193-214.
 JEWELL, W. S. (1978). A Survey of Credibility Theory, *Scandinavian Actuarial Journal*, to appear.
 JEWELL, W. S. (1977). *Bayesian Life Testing Using the Total Q on Test*, chapter of Tsokos and Shimi (1977).
 TSOKOS, C. P. and I. N. SHIMI (eds.) (1977). *The Theory and Application of Reliability*, Vol. 1, Academic Press, New York.