TRANSACTIONS OF SOCIETY OF ACTUARIES
1977 VOL. 29

NOTES ON BAYESIAN GRADUATION

JAMES C. HICKMAN AND ROBERT B. MILLER*

ABSTRACT

Bayesian statistics provide a coherent method for combining prior information and current observations. The parameters of actuarial models usually have been estimated by an informal amalgamation of recent anp past results. Graduation provides examples of alternative methods of utilizing information from several sources to estimate parameters. The Bayesian graduation method of Kimeldorf and Jones is reviewed, and suggestions for managing four technical problems in their method are developed.

I. BAYESIAN STATISTICS

IN 1965 Jones [5] introduced the members of the Society of Actuaries to Bayesian statistics. Perhaps the introduction was unnecessary. Actuaries had already developed a number of specific techniques for modifying the results obtained from recent observations to produce a blend with past and ancillary results for the purpose of making business decisions. Examples of such blending processes include graduation, for smoothing and adjusting results observed from decremental processes, and credibility theory, for modifying premiums as claim experience is obtained. A completely different function for which the same type of blending process has been suggested is the valuation of pension fund assets. For example, Jackson and Hamilton [4] discuss briefly the same type of blending when they suggest using some type of average of market and cost values for the purpose of valuing pension fund assets. Actually, Bayesian statistics provide a coherent method for performing the amalgamation of prior experience and current results that is a characteristic of many actuarial procedures.

This paper deals with certain technical problems in applying the Bayesian approach to graduation. Therefore, it seems wise to review initially the salient elements of Bayesian statistics as stated by Jones. The most important of these elements is the broader range of application of the probability concept permitted in Bayesian statistics. Classical statisti-

* Dr. Miller, not a member of the Society, is associate professor of statistics and business, University of Wisconsin, Madison.

cians have tended to define probability as a long-term relative frequency and to be reluctant to measure on a probability scale uncertainty about matters of fact. Bayesians have adopted a personal definition of probability. This definition permits, even requires, the assignment of numbers that satisfy the axioms of probability, not only to events for which relative frequency is a sensible concept but also to propositions about states of the world that are not known with certainty. For those phenomena for which relative frequency data are available, each school of thought would assign roughly the same probability. The split comes in the way in which prior knowledge enters an investigation. To the classical statistician, prior knowledge may suggest a model and an experiment. To a Bayesian, prior information must also be used in the form of a probability distribution to summarize the degree of certainty that exists about the parameters of the selected model. This past information may be very meager, but to skip this step is to ignore possibly useful data.

Bayes's theorem is a relatively simple consequence of the axioms of probability and the definition of conditional probability. Therefore, the theorem is accepted by all statisticians. However, it is more important for Bayesian statisticians, for it provides a learning machine. That is, Bayes's theorem is an engine by which prior information, quantified as a distribution of probability, is combined with recent experimental or observational results. The output is a posterior distribution that summarizes in a consistent fashion all information available to the decision maker about the parameters under study.

Additional descriptions of the Bayesian approach, with specific insurance examples, may be found in papers on credibility by Mayerson [7] and on premiums by Hickman and Miller [3].

II. GRADUATION

The traditional model for individual life insurance premiums requires estimates of a set of mortality, interest, and expense parameters. In fixing these parameters, the conventional actuarial wisdom is that a study of recent experience is insufficient. Instead, recent experience must be modified following examination of past experience and ancillary information, and adjusted in recognition of the risk created by uncertainty about future parameter values. Although the language is somewhat different, this statement is amenable to a Bayesian interpretation. Within Bayesian statistics, also, parameters are viewed as being uncertain; however, this approach requires the quantification of the uncertainty in the form of a prior probability distribution. Bayes's theorem is used to create a posterior distribution by combining the prior distribution with the distribution used to summarize current results. The posterior distribution may be used explicitly to introduce margins, with probability interpretations, into the selection of insurance premium parameters.

The process of combining past information and recent observations has been most highly developed within actuarial science in the estimation of mortality parameters. Observed mortality probabilities often exhibit irregularities that we tend to attribute to small sample size and errors. These observed mortality probabilities may also lack the margins needed for the conservative management of an insurance enterprise. The process of reducing the irregularities and introducing rational margins is the construction-of-tables process. It is important to observe that, even in this broad description of the process, the identification of an irregularity is itself an application of prior knowledge about mortality. If each set of mortality data is to stand alone, there is little justification for graduation as part of the construction-of-tables process.

The prior information about mortality is most frequently some notion about smoothness. Prior information about smoothness is used in the selection of a graduation method and, within the method, in the selection of a particular model. Information about the level of mortality probabilities is also used in graduation methods employing a standard table, such as methods based on the graduation of ratios of actual to expected mortality. Information on mortality levels is used directly in the Kimeldorf and Jones version of Bayesian graduation [6].

Thus we find no dispute about the importance of prior information in the justification and implementation of the graduation process. Any controversy concerns the extent of the use of prior information and the form in which it is summarized.

III. BAYESIAN GRADUATION

The first section of Jones's 1965 paper [5] traces the history of the difference equation method of graduation. It is indicated that E. T. Whittaker's motivation in developing this method was to produce a "most probable" set of graduated values. In adopting this goal, Whittaker acknowledged that he was following the lead of George King.

In his development, Whittaker used two key attributes of the Bayesian approach to statistics. First, Whittaker was willing to view the quantities being estimated as random variables rather than fixed parameters. Second, he explicitly used Bayes's theorem to combine prior information about smoothness with the results from recent observations.

Kimeldorf and Jones [6] presented a thorough development of a modern theory of Bayesian graduation. They included a numerical example and a primer on the multivariate statistical ideas needed to appreciate the steps in the development. Before turning to some suggestions for extending the usefulness of the method, we will review the steps in performing Bayesian graduation. We will use the notation of the Kimeldorf and Jones paper.

The initial step is to formulate a prior distribution for $\boldsymbol{W}$, the vector of variables (parameters in the traditional model) for which a smooth estimate is sought.[1] The prior distribution is to summarize all information about $\boldsymbol{W}$ except that contained in a vector of recent observations. The vector of observed values will be denoted by $\boldsymbol{U}$. Kimeldorf and Jones restrict the prior distribution of $\boldsymbol{W}$ to the multinormal class. This is done for two reasons. First, a multinormal prior distribution combines with a multinormal distribution of observations to produce, by way of Bayes's theorem, a multinormal posterior distribution. This is the conjugate distribution idea of Bayesian analysis. Second, a prior multinormal distribution may be easier to specify because its parameters have interpretations that are familiar to students of basic statistics. Kimeldorf and Jones denote the two matrices that define the prior multinormal distribution by $\boldsymbol{m}$, the mean vector, and $\boldsymbol{A}$, the covariance matrix.

The second step is to specify a distribution for the observed data, the sampling distribution. The data are assumed to come from a mortality study. Kimeldorf and Jones make the conventional assumption of independent binomial distributions for the mortality process within each age or duration cell group. Then, noting a standard limiting distribution result with respect to the proportion of deaths in a binomial mortality process with a large number of observed lives, they adopt a multinormal distribution for $\boldsymbol{U}$, the vector of observed mortality rates. That is, if $\boldsymbol{W} = \boldsymbol{w}$ is known, it is assumed that $\boldsymbol{U}$ will be normally distributed, with the multivariate parameters needed to specify the distribution given by $\boldsymbol{w}$, the mean vector, and $\boldsymbol{B}$, the covariance matrix. Because of the assumed mutual independence among the elements of $\boldsymbol{U}$, $\boldsymbol{B}$ is a diagonal matrix. If a closed group is studied, that is, the survivors from one group are studied in succeeding years, the assumption that the elements of $\boldsymbol{U}$ are mutually independent is false. For this type of study, $\boldsymbol{B}$ will not be a diagonal matrix. It is not critical to the development in the sequel that $\boldsymbol{B}$ be a diagonal matrix, although the computation would be simplified if it were.

The third and final step is to examine the posterior distribution of $\boldsymbol{W}$, given the observations $\boldsymbol{U} = \boldsymbol{u}$. For the distribution assumptions made

[1] Boldface symbols will denote matrices.

by Kimeldorf and Jones, the posterior distribution is multinormal with mean vector $(A^{-1} + B^{-1})(B^{-1}u + A^{-1}m)$ and covariance matrix $(A^{-1} + B^{-1})^{-1}$. They elect to report the mean vector, which is also the mode (the most probable value), of the posterior distribution as the set of graduated values.

IV. PROBLEMS AND OBJECTIVES

The most persistent impediment to the application of Bayesian methods is the difficulty that arises in specifying a prior distribution. Kimeldorf and Jones manage this problem by using conjugate multinormal distributions for the prior and the sampling distributions. They devote Section II of their paper to defining classes of matrices from which the covariance matrix of the prior distribution might be conveniently selected. Their goal is to define a class of covariance matrices that provides a compromise between compelling the graduator to specify a great many parameters, of which he may have limited understanding, and constraining his ability to represent significant aspects of the prior information. Our first objective will be to provide another type of assistance in the specification of a prior covariance matrix. This assistance will require the specification of more easily interpreted parameters.

Bayes's theorem provides a consistent method for combining prior and recent observations about the variables of interest. For many purposes, it is helpful to have an index of the degree of precision of the information supplied from these two sources. Our second objective will be to suggest an index to help a graduator measure the mix of the precision of the prior and observational information that he is concocting.

Following the path of Whittaker and King, Kimeldorf and Jones sought "most probable" values of the quantities being graduated. Consequently they report the mean vector, which is also the mode, of the posterior multinormal distribution as the vector of graduated values. Other aspects of the posterior distribution are ignored. A modified goal would be to seek a vector of graduated values which has a specified probability of not understating each mortality probability. The posterior distribution of W may be used to make such probability statements. If the posterior distribution is to be used to make probability statements, the degree of approximation of the multinormal distribution becomes of more concern. Our third objective will be to improve the accuracy of the posterior probability statements.

A fourth objective is related to a minor but perplexing problem faced by Kimeldorf and Jones. We adopt a binomial model for the mortality process. The observed rates are $U_i = \theta_i/E_i$; $i = 1, 2, \ldots, \omega$. Using actuarial notation, θ_i is the random number of deaths and E_i is the exposed to risk in terms of number of lives in exposure class i. If $\mathcal{E}[U_i] = w_i$, and E_i is "large," it is known that $(U_i - w_i)\ [w_i(1 - w_i)/E_i]^{-1/2}$ has an approximate standard normal distribution. As pointed out previously, this suggests using a multinormal distribution for $\boldsymbol{U}$, given $\boldsymbol{w}$. The goal of the analysis is the posterior distribution of $\boldsymbol{W}$, given $\boldsymbol{u}$. Unfortunately, the elements of $\boldsymbol{w}$ appear not only in the mean vector of $\boldsymbol{U}$ but also in the variance of each element of $\boldsymbol{U}$. This is inconvenient for the Bayesian analysis of $\boldsymbol{W}$. Kimeldorf and Jones suggest replacing $\{[w_i(1 - w_i)/E_i]^{1/2}\}^2$ by $\{[m_i(1 - m_i)/E_i]^{1/2}\}^2$, $i = 1, 2, \ldots, \omega$, along the principal diagonal of the covariance matrix $\boldsymbol{B}$ of the distribution of $\boldsymbol{U}$, given $\boldsymbol{w}$. Recall that $\mathcal{E}[W_i] = m_i$, using the prior distribution, so the suggestion seems reasonable. However, our fourth objective will be to provide another solution to this problem.

V. THE TRANSFORMATION

The following theorem [12, sec. 5e] will help achieve our four objectives.

THEOREM. *If X_n is a sequence of statistics such that $n^{1/2}(X_n - \mu)$ has a limiting normal distribution with mean zero and variance σ^2, then for any continuously differentiable function t the limiting distribution of $n^{1/2}[t(X_n) - t(\mu)]$ is normal with mean zero and variance $\sigma^2[t'(\mu)]^2$.*

In our application, $U_i = \theta_i/E_i$, and each $E_i^{1/2}(U_i - w_i)$, $i = 1, 2, \ldots, \omega$, has a limiting normal distribution with mean zero and variance $w_i(1 - w_i)$. If $t(U_i) = \text{arc sin } U_i^{1/2}$, then $E_i^{1/2}(\text{arc sin } U_i^{1/2} - \text{arc sin } W_i^{1/2})$, $i = 1, 2, \ldots, \omega$, has a limiting normal distribution with mean zero and variance $\frac{1}{4}$. The variance-stabilizing property of the arc sine transformation is important in the sequel. Therefore, we will present the details here.

$$
\begin{aligned}
\text{Var } &[E_i^{1/2}(\text{arc sin } U_i^{1/2} - \text{arc sin } W_i^{1/2})] \\
&\simeq \sigma^2[t'(\mu)]^2 \\
&= w_i(1 - w_i)[(\text{arc sin } w_i^{1/2})']^2 \\
&= w_i(1 - w_i)[(1 - w_i)^{-1/2} w_i^{-1/2}/2]^2 \\
&= \tfrac{1}{4}.
\end{aligned}
$$

Consequently, applying the arc sine transformation to the elements of $\boldsymbol{U}$ produces a random vector $t(\boldsymbol{U}) = \{\text{arc sin } U_i^{1/2}\}$. This vector will have an approximate multinormal distribution with mean vector $t(\boldsymbol{u}) = \{\text{arc sin}$

$w_i^{1/2}\}$ and a diagonal covariance matrix $\mathbf{t}(\mathbf{B})$ with elements $1/4E_i$ along the principal diagonal.[2]

Novick, Lewis, and Jackson [11] also examine the Bayesian analysis of the results from observing a binomial process using the arc sine transformation. However, their goal and the methods they use to achieve it are very different from ours.

VI. THE PRIOR DISTRIBUTION

Our plan is to perform Bayesian graduation on the transformed vector of observations $\mathbf{t}(\mathbf{U})$ in order to achieve the four goals outlined in Section IV. Consequently, the prior distribution for $\mathbf{t}(\mathbf{W})$ will have to be in terms of a metric corresponding to that used for the observations.

In actuarial graduation there is usually a fairly obvious choice for $\mathcal{E}[\mathbf{W}] = \mathbf{m}$, the mean vector of the prior distribution in the original metric. Therefore, it is natural to set $\mathcal{E}[\mathbf{t}(\mathbf{W})] = \{\text{arc sin } m_i^{1/2}\} = \mathbf{t}(\mathbf{m})$. It is also plausible for the graduator to conceive that his prior information about $\mathbf{W}$ has been derived from actual or perhaps hypothetical past observations of the mortality process. As a result of the variance-stabilizing property of the arc sine transformation, the graduator does not need to specify a variance for each $t(W_i)$. Instead, he needs only to fix the actual or hypothetical size of the past sample, to be denoted by n_i', $i = 1, 2, \ldots, \omega$, which will be associated with the values of the elements of $\mathbf{t}(\mathbf{W})$.

Kimeldorf and Jones point out that the correlation matrix defining the interrelations among the variables W_i is the principal mechanism for defining the smoothness inherent in past knowledge. The remaining question is the impact of the transformation on correlation coefficients formulated in the original metric.

Ryder [13] considers this question. We let $X = R_1/n_1$, where R_1 has a binomial distribution with parameters n_1 and π_1, and let $t(x) = \text{arc sin } x^{1/2}$. Using Taylor series expansions, it is easy to show, confirming results stated earlier, that

$$\lim_{n_1 \to \infty} \mathcal{E}[t(X)] = t(\pi) ,$$

and

$$\lim_{n_1 \to \infty} \text{Var}\,[n_1^{1/2} t(X)] = \tfrac{1}{4} .$$

If the random variable Y has a binomial distribution with parameters n_2 and π_2 and a coefficient of correlation r with X, we obtain, using Taylor

[2] In this paper we will be doing Bayesian analysis by transforming both the observations ($\mathbf{U}$) and the random parameters of interest ($\mathbf{W}$). In order to avoid a profusion of notation, a matrix in the new metric, in which the analysis is performed, will be denoted by a boldface $\mathbf{t}$ operating on the corresponding matrix in the original metric.

series expansions once more,

$$\lim_{n_1\to\infty\; n_2\to\infty} \frac{\mathcal{E}[t(X)t(Y) - \mathcal{E}[t(X)]\mathcal{E}[t(Y)]]}{\{\mathrm{Var}\,[t(X)]\,\mathrm{Var}\,[t(Y)]\}^{1/2}} = r\,.$$

In summary, we are asking the graduator to fix his prior distribution of $\boldsymbol{W}$ by picking the mean vector $\boldsymbol{m}$ and a set of equivalent past sample or exposure sizes n_i', $i = 1, 2, \ldots, \omega$. The graduator must also specify a correlation matrix reflecting his prior information concerning the interrelations, that is, the smoothness of $\boldsymbol{W}$. The prior information is thought of as coming from past mortality observations with an approximate multinormal distribution, with the degree of approximation improving with the size of past samples. If we subject the prior distribution to the arc sine transformation, the theorem of Section V and the demonstration concerning the correlation coefficient lead us to note that $t(\boldsymbol{W})$ has an approximate normal distribution with mean vector $t(\boldsymbol{m})$ and covariance matrix $t(\boldsymbol{A}) = \{r_{ij}/4(n_i'n_j')^{1/2}\}$.

In the interest of brevity, the details of several of the developments in this section have been omitted, so it seems wise to present the following observations:

1. The method developed by Kimeldorf and Jones asks the graduator to specify a multinormal prior distribution for the vector $\boldsymbol{W}$ by fixing the mean vector $\boldsymbol{m}$ and the covariance matrix $\boldsymbol{A}$.
2. The arc sine transformation of the random variables is approximately linear in the neighborhood of the mean vector.
3. With large prior sample sizes, the distribution of $\boldsymbol{W}$ will be concentrated near the mean.
4. An approximate linear transformation of $\boldsymbol{W}$, $t(\boldsymbol{W})$, will produce an approximate multinormal distribution in the new metric with mean vector $t(\boldsymbol{m})$ and covariance matrix $t(\boldsymbol{A}) = \{r_{ij}/4(n_i'n_j')^{1/2}\}$.

VII. CORRELATION MATRICES

In their example, Kimeldorf and Jones used a correlation matrix $\boldsymbol{R}$ of the form $\{r^{|i-j|}\}$. They had previously shown that, when combined with positive standard deviations, the result is an admissible (positive definite) covariance matrix. This class of correlation matrices will play a role in the sequel.

The same class of correlation matrices has appeared in several recent actuarial discussions. In particular, it was used by Shur [14] in the context of credibility theory. Shur shows that, for a $k \times k$ correlation matrix of the form $\{r^{|i-j|}\}$, the determinant of the matrix is $(1 - r^2)^{k-1}$.

In their numerical example, Kimeldorf and Jones selected a large positive value (0.942809) for r. They reasoned that, given certain information about a single mortality probability, the prior distribution for neighboring probabilities, conditioned on this information, will become significantly less diffuse.

This reasoning has appeal except at young ages. Especially among men, monotone smoothness is not a characteristic of mortality probabilities below age 35. See the graphs of mortality probabilities contained in Myers' paper "United States Life Tables for 1969–71" [9] to support this statement. These graphs illustrate the hump that seems to characterize the mortality of young adult males. Consequently, for these ages, it would seem inappropriate to expect a high positive correlation among adjacent mortality probabilities. Apparently the random accident hazard dominates the smoother impact of aging. At young ages, knowledge of one mortality probability may not reduce the variance of the conditional prior distributions of neighboring mortality probabilities.

For this reason, we suggest prior correlation matrices of the form

$$D = \begin{pmatrix} I & 0 \\ 0 & R \end{pmatrix},$$

where $R = \{r^{|i-j|}\}$ and I is the identity matrix with dimensions chosen to correspond to the number of age intervals where monotone smoothness is not a characteristic of the prior information. Within the interval of ages where the prior distribution has zero correlations, a rather simple result for the mean of the transformed posterior distribution may be obtained, if the observations are also mutually independent. Under these conditions

$$\mathcal{E}[t(V_i)] = \frac{n_i t(u_i) + n_i' t(m_i)}{n_i + n_i'} \qquad i = 1, 2, \ldots, \omega - k .$$

Not only does this formula resemble traditional credibility formulas, but it also resembles formulas for graduation with respect to standard tables.

This suggestion illustrates how admissible matrices may be combined to produce a more appropriate prior covariance matrix. The key point is that the prior covariance matrix should not be arbitrarily chosen but should reflect a serious study of past experience.

VIII. PRECISION MEASURES

Graduation is a multidimensional process. Whether or not Bayesian methods are formally used, the process involves combining a vector of observation u with multidimensional prior information. Indeed, without prior information, smoothing is an unjustified process.

It is natural to seek a one-dimensional measure of degree of precision of each of the inputs to the process. Clearly there are various ways to measure the precision of a multidimensional array of data. In the case of Bayesian graduation using conjugate multinormal distributions, it is suggested that the determinants of the inverses of the covariance matrices of the prior distribution and the distribution of the observations are tractable and meaningful measures of the precision of the information entering the graduation process from the two sources.

In a certain general sense, if one views the covariance matrix as describing the dispersion of a distribution, then the inverse of the covariance matrix is a reasonable measure of the precision or concentration of the distribution. The precision matrix B^{-1} describes the degree of concentration of the vector of observations of a mortality process. The precision matrix A^{-1} describes the degree of precision of prior information about the mortality process.

There are, of course, several ways of measuring a matrix. In the design of experiments, one frequently chosen goal is to maximize the determinant of the precision matrix. The determinant of the covariance matrix (the inverse of the precision matrix) is called the generalized variance of the distribution. Additional insights will be gained by observing that the determinant of the precision matrix is inversely proportional to the square of the volume of the ellipsoid of concentration of a multinormal distribution. The volume of the ellipsoid of concentration [2, p. 301] is a traditional measure of the degree of concentration of a probability density within a multinormal distribution. A distribution with a smaller volume of the ellipsoid of concentration (larger determinant of the precision matrix) may be viewed as being more concentrated than the distribution with which it is compared.

For the specific covariance matrices considered, the computation of the determinant of the precision matrix is easy. We will denote the determinant of the inverse of matrix A by det A^{-1}. Then, in general, for covariance matrix A, det $A^{-1} = (\sigma_1^2\sigma_2^2 \ldots \sigma_\omega^2)^{-1}(\det R)^{-1}$, where R is the correlation matrix. As noted in Section VII, if $R = \{r^{|i-j|}\}$, det $R = (1 - r^2)^{\omega-1}$. If the correlation matrix is given by D, as displayed in Section VII, det $D = (1 - r^2)^{k-1}$, where $k \leq \omega$ is the dimension of the square matrix R.

The precision index for the transformed observations, assuming mutual independence among the numbers of deaths in the groups under study, will be

$$\det t(B)^{-1} = 4^{\omega}\prod_1^{\omega} E_i\,.$$

The ratio $h = \{\det[t(A)^{-1}]/\det[t(B)^{-1}]\}^{1/2}$ may be used to summarize the relative precision of the two inputs to the graduation process. If $1 < h$, the precision of prior information is greater than that of the data. If $1 > h$, the precision of the data is greater.

For example, if $t(A) = \{r^{|i-j|}/4(n_i'n_j')^{1/2}\}$ and $t(B)$ is a diagonal matrix with diagonal elements $1/4E_i$, we have

$$h = \{\det[t(A)^{-1}]/\det[t(B)^{-1}]\}^{1/2}$$

$$= \left[\prod_1^{\omega}(n_i'/E_i)\Big/(1-r^2)^{\omega-1}\right]^{1/2}.$$

The interpretation of h has much in common with the conventional interpretation of h in Whittaker-Henderson difference equation graduations. That is, h measures the relative importance of fit and smoothness. A small value of h is associated with greater stress on fit; a large value of h means emphasis on smoothness.

In fact, in Whittaker's development [15, p. 306], the constant h in the loss function $F + hS$ (fit + $h\cdot$smoothness) is sometimes interpreted as the ratio of (i) the precision of the normal distribution that measures the certainty of smoothness (S) to (ii) the assumed common precision of the observations. The descriptive number h defined in this section is an obvious generalization, but we have chosen not to work in terms of squared units.

The gain in precision from the prior state to the posterior state (after the observations) may be summarized by the ratio

$$\{\det[t(A)^{-1} + t(B)^{-1}]/\det[t(A)^{-1}]\}^{1/2}.$$

IX. REPORTING RESULTS

In the transformed metric, the posterior distribution of $t(W)$, given $t(u)$, is multinormal with mean vector $t(v) = [t(A)^{-1} + t(B)^{-1}]^{-1} \times [t(B)^{-1}t(u) + t(A)^{-1}t(m)]$ and covariance matrix $[t(A)^{-1} + t(B)^{-1}]^{-1}$. To restore the mean vector to the original metric, form $v = \{(\sin[t(v_i)])^2\}$. Kimeldorf and Jones suggest reporting the posterior mean, in the original metric, as the set of graduated values.

However, an interesting alternative is available. The arc sine transformation has the effect of stabilizing the variance of the elements of U. The transformation has the additional feature that it improves the accuracy of the normal approximation. This feature of the transformation, and additional modification of it to improve the approximation still further, was developed by Anscombe [1]. In general, we can expect the normal approximation in the transformed metric to be more accurate than the normal approximation to the untransformed observations. We do not want to overemphasize this point. For the sample sizes typical of actuarial studies, large sample theory usually provides ample justification for the multinormal assumption.

A warning is required. If the number of deaths in a mortality investigation is small, the arc sine transformation introduces a distortion in the range of probability values. A commonsense rule is to limit the use of the arc sine transformation to experience cell groups with five or more deaths.

Smoothed mortality probability estimates are used in many business decision problems. Therefore, there may be reason to report "safe" graduated values rather than mean or modal values. The posterior distribution combines all available information about the variables being estimated. Each variable $t(W_i)$ has a posterior distribution that is approximately normal, with a mean that may be obtained from the mean vector and a variance from the principal diagonal of the posterior covariance matrix. Thus it would be possible to produce a set of graduated values v_i such that $P[t(W_i) \leq t(v_i)] = P[W_i \leq v_i] = p$; $i = 1, 2, \ldots, \omega$. For safety purposes, p might be 0.75 in our mortality example. Such a choice might perhaps be directly relevant to the situation where the mortality probabilities are to be used in fixing group term life premiums. Because of the many uses of mortality probabilities, and the differing degrees of conservatism that may be appropriate, it is important to report the posterior distribution of the mortality probabilities at the conclusion of a study.

X. EXAMPLE

As the title of the paper suggests, the purpose is to augment the work of Kimeldorf and Jones. Thus the abbreviated numerical example will be in the same style as in the earlier paper. It should be clear that the example is designed to illustrate the ideas developed in this paper rather than to extend Bayesian graduation.

A. *Prior Distribution*

The mean vector is taken from the 1955–60 Male Select Basic Tables [10]. The equivalent past sample sizes, n'_i, $i = 1, 2, \ldots, 13$, are shown in Table 1. The correlation matrix is of the form of matrix $\boldsymbol{D}$ of Section V, where I has dimension 4×4 and the correlation matrix $\boldsymbol{R} = \{0.942809^{|i-j|}\}$. This is the same value of r used by Kimeldorf and Jones.

B. *Observations*

The data are for policies issued in 1954 to standard medically examined male lives, observed during the eighth policy year (between policy anniversaries in 1961 and 1962) [8]. The data were adjusted by decreasing

TABLE 1

MORTALITY EXAMPLE

i	Age at Issue	$E_i\times10^{-6}$	$u_i\times10^3$	$m_i\times10^3$	n_i'	$v_i\times10^3$
1.	10–14	51.071	1.68	0.99	2,000	1.15
2.	15–19	102.352	0.73	1.07	2,000	0.92
3.	20–24	278.026	1.07	0.90	3,000	0.99
4.	25–29	729.655	1.20	1.05	4,000	1.16
5.	30–34	998.975	1.44	1.72	5,000	1.69
6.	35–39	919.910	3.07	2.94	5,000	3.01
7.	40–44	698.755	5.37	4.86	5,000	5.05
8.	45–49	422.317	8.41	7.66	5,000	7.97
9.	50–54	214.521	14.45	12.10	5,000	12.49
10.	55–59	93.843	15.75	17.10	4,000	17.46
11.	60–64	30.619	19.63	22.70	3,000	23.05
12.	65–69	7.528	20.46	31.51	2,000	31.91
13.	70 and over	0.686	71.42	65.60	2,000	66.14

each exposure by a factor of one-tenth. The data were chosen from the same era as were those of Kimeldorf and Jones. Therefore, it is assumed that the equivalent average amount used in computing the variance of observed mortality probability is $7,500.

C. Graduated Values

The graduated values were derived by applying the inverse arc sine transformation to the posterior mean vector in the transformed metric.

D. *Relative Precision*

$$h=\left[\prod_{1}^{13}\left(\frac{n_i'}{E_i/7,500}\right)\Big/(1-0.942809^2)^8\right]^{1/2}=896,875$$

E. *Discussion*

The principal point in the example is that monotone smoothness is not a characteristic of mortality experience of young males. Covariance matrices of a type like matrix D in Section V are useful in preventing oversmoothing. In constructing this example, various sets of past sample sizes and values of r were used. The chi-square statistic was computed to measure fit, and the sum of third differences squared was computed to measure smoothness. This was done not to accept or reject the graduation but to gain insight into the influence of changes in the parameters of the prior distribution. As expected, large values of r produce greater smoothness, and small values of n_i', $i=1, 2, \ldots, \omega$, tend to produce better fit.

As was indicated in the discussion of the examples in the Kimeldorf and Jones paper, there is a time element in our prior knowledge that has not been captured in the prior distribution used in this example. That is, our knowledge of mortality probabilities in the eighth policy year consists of (1) information about the level of smoothness of mortality probabilities as a function of age within the eighth policy year and (2) information about the level of smoothness of mortality probabilities in the eighth policy year with respect to those in the seventh and ninth policy years. In addition, there is correlation among the observed mortality probabilities in successive durations.

What is needed is a grid of estimates of select mortality probabilities rather than a vector of smoothed estimates for a single policy year. This is an important problem that has not been solved satisfactorily by any of the traditional approaches to graduation. Within the Bayesian framework, the problem will probably be approached with very large models and with a partitioned prior covariance matrix built up of blocks, each of which will be admissible and each selected to capture an aspect of prior knowledge.

XI. SUMMARY

We have met the four objectives listed in Section IV as follows:

1. By operating in a new metric, that is, by transforming the observations and the prior distribution by application of the arc sine transformation, the graduator does not have to specify individual variances for the prior distribution. Instead, more easily interpreted past sample sizes are specified.
2. The determinants of the precision matrices of the prior distribution and the sampling distribution of the observations provide convenient measures of the precision of the inputs of the graduation process.
3. Because the multinormal distribution should be a better approximation in the transformed metric, the reporting of "safe" graduated values with posterior probability measures of the degree of safety becomes more feasible.
4. The use of the arc sine transformation eliminates the necessity to approximate the variance of each observed mortality rate in the sampling distribution.

In addition to these specific results, a continuing theme of the paper has been the necessity to look at the data carefully to avoid the arbitrary selection of a prior distribution. In particular, we found that imposing smoothness on the mortality experience of young males by way of the prior distribution may be avoided easily.

REFERENCES

1. Anscombe, F. J. "The Transformation of Poisson, Binomial and Negative Binomial Data," *Biometrika*, XXXV (1948), 246.
2. Cramér, H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

3. HICKMAN, J. C., and MILLER, R. B. "Insurance Premiums and Decision Analysis," *Journal of Risk and Insurance*, XXXVII (1970), 567.
4. JACKSON, P. H., and HAMILTON, J. A. "The Valuation of Pension Fund Assets," *TSA*, XX (1968), 386.
5. JONES, D. A. "Bayesian Statistics," *TSA*, XVII (1965), 33.
6. KIMELDORF, G. S., and JONES, D. A. "Bayesian Graduation," *TSA*, XIX (1967), 66.
7. MAYERSON, A. L. "A Bayesian View of Credibility," *PCAS*, LI (1964), 85.
8. "Mortality under Standard Ordinary Insurance Issues between 1961 and 1962 Anniversaries." *TSA*, *1963 Reports*, p. 1.
9. MYERS, R. J. "United States Life Tables for 1969–71," *TSA*, XXVIII (1976), 93.
10. "1955–1960 Basic Tables." *TSA*, *1962 Reports*, p. 44.
11. NOVICK, M., LEWIS, C., and JACKSON, P. "The Estimation of Proportions in Groups," *Psychometrika*, XXXVIII (1973), 19.
12. RAO, C. R. *Advanced Statistical Methods in Biometric Research*. New York: John Wiley & Sons, 1952.
13. RYDER, J. M. "A Review of Graduation and Bayesian Methods," *Transactions of the Institute of Actuaries of Australia and New Zealand*, 1972, p.1.
14. SHUR, W. "Discussion of M. Margolin's 'Credibility for Group Insurance Claims Experience,' " *TSA*, XXIII (1971), 240.
15. WHITTAKER, E. T., and ROBINSON, G. *The Calculus of Observations*. 4th ed. London and Glasgow: Blackie & Son, Ltd., 1944.

DISCUSSION OF PRECEDING PAPER

DONALD A. JONES:

Professors Hickman and Miller have taken a giant step in the direction of applying Bayesian graduation in the real world. The use of the arc sine transformation and the development of the measure of relative precision should be especially helpful in applications.

I find their suggested form for the prior correlation matrix to be at odds with my prior distribution for mortality rates. The stochastic independence of mortality rates at the young ages would imply that the variance of the marginal prior for a single rate is equal to the variance of the conditional prior for the single rate, given the values of the two adjacent rates. This is not a property of my prior. Thus I need one of two things: either some positive correlation at the younger ages or some more general information about male mortality so that I can hold a prior distribution in the future that will exhibit the required independence.

Positive correlation at the younger ages can be incorporated by the adoption of an appropriate member of the class of admissible matrices defined by formula (27) in the Kimeldorf and Jones paper referenced by the authors. That formula is

$$a_{ij} = \prod_{k=1}^{j-1} r_k \quad \text{for} \quad i < j$$

$$= 1 \qquad \text{for} \quad i = j ,$$

assuming $p = 1$ and symmetry to save printing costs.

The authors' suggested form falls in this class by setting the first r_i's equal to zero and the balance equal to a constant r. If the hump in male mortality rates requires a smaller positive correlation at younger ages, then the first r_i's could be set equal to a small positive constant—smaller than the one at older ages.

Another possibility in this direction would be to use covariance matrices that correspond to the measures of roughness involving higher-order differences. Kimeldorf developed the general scheme of these in his dissertation at the University of Michigan (see reference to Kimeldorf and Jones paper).

I hope that the authors will have the resources to include in their response some graduated rates based on other prior covariance matrices for the example in their excellent paper.

STUART A. KLUGMAN:

Professors Hickman and Miller are to be thanked for adding some practical touches to the method of Bayesian graduation. The graduator is now provided with an excellent set of guidelines for selecting those crucial quantities, the parameters defining the prior distribution. By introducing a variance-stabilizing transformation, the graduator can transform easily his prior notions of fit, smoothness, and shape into a working formula.

I would like to take issue with the authors' interpretation of one of the parameters, r, the correlation coefficient. They note correctly that decreasing r will improve the fidelity of the graduation to the observed rates (as evidenced by the role of r in the measure of relative precision). However, increasing r does not ensure increased smoothness of the posterior means; it merely indicates that greater reliance is being placed on the prior distribution. It might be more appropriate to call r a shape parameter, since values close to 1 tend to transfer the shape of the prior to the posterior. In the example used in the paper, if $r = 0.942809$ is used at all thirteen age groups, the first four posterior means ($\times 10^3$) are 1.06, 1.13, 0.95, and 1.10. They are not much smoother than the rates in Table 1 of the paper, but the dip is now in the third age group, as it was in the prior. Even though monotone smoothness is not a characteristic of these rates, a moderately high correlation is still reasonable; if one rate is increased, the others are likely to follow.

One way to allow the graduation parameters to account for differing mortality characteristics at younger ages would be to select a prior correlation matrix of the form

$$D = \begin{pmatrix} R_1 & S \\ S' & R_2 \end{pmatrix},$$

where $R_1 = \{r_1^{|i-j|}\}$ is $k \times k$, and $R_2 = \{r_2^{|i-j|}\}$ is $(n-k) \times (n-k)$. To make the resulting covariance matrix admissible, the S-matrix could have $r_1^{(k-i+0.5)} r_2^{(j-0.5)}$ as its ijth element. Admissibility follows from the fact that D is a member of class a_2 of Kimeldorf and Jones (*TSA*, XIX, 81). By selecting $r_1 < r_2$, allowance can be made for the more erratic behavior of rates at the younger ages.

As a final note, Stephen Portnoy in a letter to the *American Statistician* (XXXI [No. 1], 54) remarks that the identity $2 \arcsin \sqrt{p} = \arcsin (2p - 1) + \pi/2$ can be used to simplify the calculations in stabilizing the variance. The function $t(p) = \arcsin (2p - 1)$ will achieve the desired results, the only difference being that the variance is 1 instead of $\frac{1}{4}$.

THOMAS G. KABELE:

I believe that Hickman and Miller have made a significant contribution to the actuarial literature with their presentation of the arc sine statistical distribution to smooth the ungraduated values, and with their presentation of a varying smoothness criterion to apply to different parts of the mortality curve. These devices were discussed in terms of the Bayesian form of graduation, but they are in fact applicable to other types of graduation. Before giving more specific comments, I will outline the general theory of Whittaker-Henderson graduation, of which the Kimeldorf-Jones or Bayesian form is a special case [10].

Whittaker-Henderson Graduation

In a 1957 paper [5], and in his discussion of the Kimeldorf-Jones paper [7], Greville defined the general Whittaker-Henderson method. In this method the graduator is given a vector $\boldsymbol{u} = (u_1, \ldots, u_n)$ of ungraduated values and seeks to find a vector $\boldsymbol{v} = (v_1, \ldots, v_n)$ of graduated values that minimizes the quadratic form

$$(\boldsymbol{v} - \boldsymbol{u})'\boldsymbol{E}(\boldsymbol{v} - \boldsymbol{u}) + \boldsymbol{v}'\boldsymbol{Z}\boldsymbol{v}\,, \tag{1}$$

where the prime denotes a matrix transpose and where $\boldsymbol{E}$ is a positive-definite matrix and $\boldsymbol{Z}$ a positive-semidefinite matrix. The matrix $\boldsymbol{E}$ is invertible or nonsingular, but the matrix $\boldsymbol{Z}$ may be singular. The first term in expression (1) measures the fit or "fidelity" of the graduated values, and the second term measures the roughness. The solution that minimizes expression (1) is

$$\boldsymbol{v} = (\boldsymbol{E} + \boldsymbol{Z})^{-1}\boldsymbol{E}\boldsymbol{u}\,, \tag{2}$$

and the proof is given in Greville's study note for Part 5 of the Society of Actuaries examinations.

In our *Transactions* I have found three specific forms of the Whittaker-Henderson method. In the classical form, which was first considered by Whittaker in 1919 (see the Dover reprint of [15], p. 303), the matrix $\boldsymbol{E}$ is diagonal with positive entries, and $\boldsymbol{Z} = h\boldsymbol{K}'\boldsymbol{K}$, where h is a constant and $\boldsymbol{K}$ is a rectangular matrix such that, for every vector $\boldsymbol{w}$, $\boldsymbol{K}\boldsymbol{w} = \Delta^z \boldsymbol{w}$. Here $\Delta^z \boldsymbol{w}$ is a column vector with entries $(\Delta^z w_1, \ldots, \Delta^z w_{n-z})$, and z is usually 2, 3, or 4.

In 1950 and 1955 the classical form was generalized to what we shall call the "Camp form" after its inventor [2]. In the Camp form the matrices $\boldsymbol{E}$ and $\boldsymbol{Z}$ are defined as in the classical form, but $\boldsymbol{K}$ is defined so that

$$(\boldsymbol{K}\boldsymbol{v})_{i1} = s_i(\Delta^z v_i) - (\Delta^z v_{i+1}) \qquad (i = 1, \ldots, n-1)\,,$$

where the s_i's are "shape constants." Camp used his new form to graduate the 1951 Group Annuity Mortality Table. He took $z = 1$ and let $s_i = c$, where c is the Makeham constant obtained by a preliminary graphic Makeham graduation. In his discussion of the Kimeldorf-Jones paper, Greville used the Camp form with $z = 0$ and $s_i = 1.5$ [7].

In 1967 Kimeldorf and Jones presented a third form of the Whittaker-Henderson method, which we shall call "Kimeldorf's form" rather than "Bayesian graduation," the term used in the 1967 paper. The authors first transformed the ungraduated values $\boldsymbol{u}$ by subtracting a vector $\boldsymbol{m}$ of "prior means." They then graduated $\boldsymbol{u} - \boldsymbol{m}$. Their $\boldsymbol{Z}$-matrix, however, was taken to be a positive-definite matrix rather than the singular matrix used in both the classical and Camp forms. The matrix $\boldsymbol{Z}$ was tridiagonal:

$$Z_{ii} = \frac{1 - r_{i+1}^2 r_i^2}{(1 - r_{i+1}^2)(1 - r_i^2)p_i^2},$$

$$Z_{i-1,i} = Z_{i,i+1} = \frac{-r_i}{p_i p_{i+1}(1 - r_{i+1}^2)},$$

$$Z_{ij} = 0 \qquad \text{if } j \neq i-1, i, i+1,$$

and ([10], p. 81)

$$\begin{aligned} Z_{ij}^{-1} &= p_i p_j \prod_{k=1}^{j-1} r_k \qquad (i < j) \\ &= p_i^2 \qquad\qquad\quad (i = j) \\ &= p_i p_j \prod_{k=j}^{i-1} r_k \qquad (i > j). \end{aligned}$$

Here the p_i's are the "prior standard deviations," and the r_i's ($i = 1, \ldots, n - 1$) are the "prior adjacent correlations." For convenience we define $r_0 = r_n = 0$. In their paper Kimeldorf and Jones ([10], p. 89) defined

$$r_i = 0.942809, \qquad p_i = 0.01415\sqrt{m_i}.$$

In Hickman and Miller's notation, $\boldsymbol{A} = \boldsymbol{Z}^{-1}$ and $\boldsymbol{B} = \boldsymbol{E}^{-1}$. The above definition of $\boldsymbol{Z}$ is a slight generalization of the class $\mathfrak{a}_2$ matrix defined in the Kimeldorf-Jones paper and was given by Halmstad ([8], pp. 23–25).

Statistical Justification

Whittaker and Kimeldorf both justify their graduation formulas by using statistical arguments, and Kimeldorf's argument is practically identical with the earlier one by Whittaker (see [15], pp. 304–6; [10], pp. 70–71; [9], pp. 34–36).

Both authors start with Bayes's theorem, which may be written as

$$\text{Prob}\,[\boldsymbol{w} < \boldsymbol{W} < \boldsymbol{w} + \sigma \mid \boldsymbol{U} = \boldsymbol{u}] = \frac{\text{Prob}\,[\boldsymbol{w} < \boldsymbol{W} < \boldsymbol{w} + \sigma]\,\text{Prob}\,[\boldsymbol{U} = \boldsymbol{u} \mid \boldsymbol{w} < \boldsymbol{W} < \boldsymbol{w} + \sigma]}{\text{Prob}\,[\boldsymbol{U} = \boldsymbol{u}]},$$

where $\boldsymbol{U}$ is the random vector of observed rates, $\boldsymbol{W}$ is the random vector of "true" rates, $\boldsymbol{u}$ is a particular observation of $\boldsymbol{U}$, and σ is a constant.

Both authors define the graduated values to be the vector $\boldsymbol{w} = \boldsymbol{v}$ that maximizes the left-hand side. Both authors assume that

$$\text{Prob}\,[\boldsymbol{U} = \boldsymbol{u} \mid \boldsymbol{w} < \boldsymbol{W} < \boldsymbol{w} + \sigma] = k\sigma^n \exp(-\boldsymbol{F}),$$

where $\boldsymbol{F} = (\boldsymbol{u} - \boldsymbol{w})'\boldsymbol{E}(\boldsymbol{u} - \boldsymbol{w})$ and k is a constant. Both authors assume that Prob $[\boldsymbol{w} < \boldsymbol{W} < \boldsymbol{w} + \sigma]$ can be represented in the form

$$k\sigma^n \exp(-\lambda^2 \boldsymbol{S}).$$

The only difference is in the definition of $\boldsymbol{S}$. Whittaker defines $\boldsymbol{S}$ as $\boldsymbol{w}'\boldsymbol{K}'\boldsymbol{K}\boldsymbol{w}$, whereas Kimeldorf defines it as $(\boldsymbol{w} - \boldsymbol{m})'\boldsymbol{Z}(\boldsymbol{w} - \boldsymbol{m})$, where $\boldsymbol{Z}$ is defined as above. The matrix $\boldsymbol{K}'\boldsymbol{K}$ used by Whittaker is singular (i.e., not invertible), and the statistical distribution of the random variable $\boldsymbol{W}$ is called a singular normal distribution (see [3], p. 287).

Comments

TERMINOLOGY

Since all forms of Whittaker-Henderson graduations can be justified using Bayes's theorem, I do not believe we should use the term "Bayesian graduation" for only one of them. Thus I suggest that the form of graduation presented in Kimeldorf's Ph.D. thesis and his 1967 paper be called "Kimeldorf's form."

I believe that terminology is important. For example, in a recent article [12], Ryder criticized all Bayesian methods, using as an example the lack of smoothness of the graduation in Kimeldorf's paper. I do not believe, however, that Mr. Ryder would be willing to condemn all Whittaker-Henderson graduations even though all are equally "Bayesian."

SIMILARITY OF KIMELDORF'S FORM WITH THE CLASSICAL

As remarked by Kimeldorf and Jones ([10], pp. 74–75), Kimeldorf's form of the Whittaker-Henderson method is very similar to the classical form applied to $\boldsymbol{u} - \boldsymbol{m}$ with perfect smoothness, meaning that first differences are zero or, equivalently, that $\boldsymbol{v} - \boldsymbol{m}$ is a constant. This particular version of the classical Whittaker-Henderson method seems to me to be a very mediocre choice because it constrains the graduated values v to lie parallel to the "prior means," and it may overgraduate the ends of the table. Kimeldorf's form, however, may not be even as good as the classical form. In fact, since the matrix Z in Kimeldorf's form is invertible, the "smoothness" term $(v - m)'Z(v - m)$ equals zero if and only if $v = m$ (see [7]).

Let us suppose that the number of terms to be graduated is $n = 4$ and that the p_i's and r_i's in Kimeldorf's form are constant. Then the Z-matrix for Kimeldorf's form and the matrix $K'K$ for the classical form are

$$Z = \frac{1}{p^2(1 - r^2)} \begin{pmatrix} 1 & -r & 0 & 0 \\ -r & 1 + r^2 & -r & 0 \\ 0 & -r & 1 + r^2 & -r \\ 0 & 0 & -r & 1 \end{pmatrix},$$

$$K'K = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

The coefficient $1/p^2(1 - r^2)$ for the Z-matrix can be regarded as the smoothing constant h. If $r = 0.942809$, as suggested by Kimeldorf and Jones, then we see that the above matrices are practically identical.

CREDIBILITY FORMULAS

Both Kimeldorf and Jones ([10], p. 71) and Hickman and Miller (Sec. VII) imply that Kimeldorf's form of the Whittaker-Henderson method is a generalized credibility formula because the graduated values v can be written in the form

$$v = (Z + E)^{-1}(Eu + Zm),$$

that is, v is the "weighted average" of u and m. However, for the Hickman and Miller example, excluding the first four values, about half the entries in the product matrix Zm are negative (see Table 1). Professor Hickman assures me that there is nothing inherently wrong with this, but I have an uneasy feeling about using negative weights.

POSSIBLE IMPROVEMENTS IN KIMELDORF'S FORM

I agree with comments Professor Hickman made to me over the phone that Kimeldorf's form is interesting theoretically, but in its present form it is of limited practical use. I would like to suggest some possible improvements. First, I believe that if we are aiming for a generalized credibility-smoothing formula, we should choose the matrices $\boldsymbol{Z}$ and $\boldsymbol{E}$ to have the same form. Either both should be diagonal with positive entries, or else both should be "smoothing matrices" of the type used in the classical Whittaker-Henderson method. That is, both $\boldsymbol{E}$ and $\boldsymbol{Z}$ should be of the form $(\boldsymbol{D} + \boldsymbol{S})^{-1}\boldsymbol{D}$, where $\boldsymbol{D}$ is diagonal with positive entries and $\boldsymbol{S} = h\boldsymbol{K}'\boldsymbol{K}$.

TABLE 1

INTERMEDIATE VALUES IN KIMELDORF GRADUATION APPLIED TO ARC SINE TRANSFORMED VALUES

i	$Z_{ii} \times 10^{-3}$	$Z_{i,i+1} \times 10^{-3}$	$E_{ii} \times 10^{-3}$	$(Wtu)_{i1}$	$(Ztm)'_{i1}$
1	8	0	2.724	112	252
2	8	0	5.460	148	262
3	12	0	14.818	485	360
4	16	0	38.916	1,348	519
5	180	−169.706	53.280	2,022	−1,739
6	340	−169.706	49.060	2,720	− 436
7	340	−169.706	37.268	2,733	− 356
8	340	−169.706	22.524	2,068	− 750
9	340	−151.789	11.440	1,379	2,698
10	272	−117.575	5.004	630	1,158
11	204	− 83.138	1.632	230	597
12	136	− 67.882	0.400	58	−5,886
13	72	xxxx	0.036	10	6,535

Second, we could eliminate most of the negative entries in the product matrix $\boldsymbol{Zm}$ by reducing the adjacent correlations from about 0.9 to, say, 0.6. The latter figure is closer to the posterior values of the adjacent correlations (see Table 2).

Third, we could multiply the matrix $\boldsymbol{Z}$ in Kimeldorf's form by a smoothing constant. In fact, this is essentially what Hickman and Miller did with their illustrative data. They multiplied $\boldsymbol{Z}$ by 10 or, equivalently, divided $\boldsymbol{E}$ by 10. I believe that for most Whittaker-Henderson graduations a reasonable value of the smoothing constant is from 1 to 15 times the average value of E_{ii}/Z_{ii}.

Fourth, instead of restricting $\boldsymbol{Z}$ to be tridiagonal (which limits its smoothing power), we could define $\boldsymbol{Z}$ to have 5, 7, or 9 nonzero diagonals. If we wanted 7 nonzero diagonals, we could define

$$\boldsymbol{Z} = 0.01\boldsymbol{I} + \boldsymbol{K}'\boldsymbol{K}\,,$$

where $\boldsymbol{Kw} = \Delta^3 \boldsymbol{w}$.

TABLE 2

GRADUATION OF ARC SINE TRANSFORMED VALUES

i	POSTERIOR VALUES			UNGRADUATED VALUES	WEIGHTS	A PRIORI VALUES		
	tv_i	p_i	r_i	tu_i	E_{ii}	tm_i	p_i	r_i
1	.033890	.009657	.000000	.040999	2,724	.031469	.011180	.000000
2	.030407	.008620	.000000	.027022	5,459	.032717	.011180	.000000
3	.031504	.006105	.000000	.032717	14,828	.030005	.009129	.000000
4	.033996	.004267	.000000	.034648	38,915	.032409	.007906	.000000
5	.041122	.002803	.674184	.037956	53,279	.041485	.007071	.942809
6	.054857	.002598	.664960	.055436	49,062	.054248	.007071	.942809
7	.071185	.002704	.700829	.073346	37,267	.069770	.007071	.942809
8	.089385	.003002	.751957	.091835	22,524	.087634	.007071	.942809
9	.111993	.003397	.798054	.120500	11,441	.110223	.007071	.942809
10	.132507	.004263	.833691	.125831	5,005	.131143	.007906	.942809
11	.152396	.005455	.859961	.140570	1,633	.151241	.009129	.942809
12	.179594	.007280	.878770	.143531	401	.178456	.011180	.942809
13	.260089	.007807	.000000	.270533	37	.259011	.011180	.000000

INTERPOLATION

One advantage of Kimeldorf's method is that it allows us to interpolate for a missing value. Since $\boldsymbol{Z}$ is invertible in Kimeldorf's form, we can allow the weight matrix $\boldsymbol{E}$ to be positive semidefinite by inserting a zero in the diagonal where there is no ungraduated value. (I found this fact mentioned in some of David Halmstad's notes.) One can also do "Whittaker-Henderson interpolation" using splines as described in Greville's book ([6], p. 19).

PURPOSE OF GRADUATION

In Section IV, Hickman and Miller report that "following the path of Whittaker and King, Kimeldorf and Jones sought 'most probable' values of the quantities being graduated." It is true that the two pairs of authors both sought most probable values, but I believe the emphasis was entirely different. The emphasis of King and Whittaker was on *smoothing*. They believed that the true values could be obtained from the ungraduated values by eliminating, as Whittaker puts it ([15], p. 303), the "irregularities due to accidental causes." The complete title of the chapter on graduation in Whittaker and Robinson is "Graduation, or the Smoothing of Data." On the other hand, Kimeldorf and Jones's paper emphasizes adjusting the *level* of mortality by using "prior means."

USE OF STANDARD TABLES

In Section II of their paper Hickman and Miller say that "information about the *level* [emphasis added] of mortality probabilities is also used in graduation methods employing a standard table, such as methods based on the graduation of ratios of actual to expected mortality." Actually, the mortality ratios are used to specify the *shape* of the curve, not the *level*. If we multiply all the standard values by a constant c, the graduated ratios will be $1/c$ times the former values, and we will still get the same graduated values.

The United States Life Tables use a standard table to compute the ${}_5q_x$ values from the ${}_5m_x$ values. But again the standard table is used to calculate a *shape* parameter g rather than to specify the *level* of mortality. The formula for ${}_5q_x$ is

$$ {}_5q_x = \frac{5\ {}_5m_x}{1 + g\ {}_5m_x} , $$

where g is calculated by replacing the ${}_5q_x$ and ${}_5m_x$ on both sides by values from the standard table (see Sirken's paper [13] or Spiegelman's outline of it in [14], p. 134).

Benjamin and Haycocks in their book define a graduation method based on two standard tables ([1], pp. 309–11). They define the graduated values by

$$v_x = aq_x^{(1)} + bq_x^{(2)} ,$$

where $q^{(1)}$ and $q^{(2)}$ are the standard table values and the constants a and b are found by equating the moments of v_x and the ungraduated values u_x. Again the standard tables are used to specify the *shape* of the graduated values, not the *level*. Thus, if we multiply both the standard table values $q^{(1)}$ and $q^{(2)}$ by 2, the constants a and b are reduced by one-half and the graduated values remain unchanged.

In calculating insurance premiums, I believe it is better to try to predict the *shape* of the mortality curve rather than the *level*. If the shape is wrong we may overcharge some and undercharge others, but if the level is wrong we may overcharge everyone or undercharge everyone.

Of course, if the original data are sparse, the actuary may have to estimate the level of mortality by using a "credibility formula," where the estimated mortality values are a weighted average of the crude and the standard table values.

NUMERICAL EXAMPLES

In Tables 1 and 2, I list some intermediate values calculated in my attempt to reproduce the graduated values in the preprint of Hickman and Miller's paper. I ran into some minor difficulties. First, I found out by a letter from Professor Hickman that the diagonal elements of E had been divided by 10. Second, I believe they used a "canned" FORTRAN subroutine to calculate the inverse of $I + E^{-1}Z = E^{-1}(E + Z)$, and this subroutine works only for symmetric matrices. The matrix $I + E^{-1}Z$ is not symmetric. Of course these minor inconsistencies in no way flaw the conclusions of the paper, since the data are merely illustrative.

In Table 3, I illustrate seven different Whittaker-Henderson graduations of the ungraduated values, the third being the graduation used by Hickman and Miller (with the discrepancies noted above). I excluded the first four values from the graduation in order to be consistent with the paper, which calculated these values as weighted averages of the crude and standard table values (called "prior mean").

In the graduations in columns 1 and 2 of Table 3, I used the formulas suggested by Kimeldorf and Jones [10]. In the first graduation I used the prior means given by Hickman and Miller, while in the second I multiplied the prior means by $\frac{1}{2}$. As one can see from Table 2, both graduations produced results almost equal to the ungraduated values. Although Kimeldorf's form is supposed to allow the graduator to predict the level

TABLE 3

Comparison of Various Whittaker-Henderson Graduations

i	Graduations							Data	
	Kimeldorf with No Transformation		Kimeldorf with Arc Sine Transformation		Classical with First Differences Zero	Camp	Classical with Second Differences Zero	Ungraduated ($u_i \times 10^3$)	Prior Means ($m_i \times 10^3$)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
5	1.441	1.441	1.69	1.892	1.72	1.684	1.519	1.44	1.72
6	3.07	3.07	3.006	3.037	2.94	3.046	2.973	3.07	2.94
7	5.37	5.37	5.059	4.71	4.862	5.128	5.312	5.37	4.86
8	8.413	8.414	7.968	6.86	7.663	8.096	8.545	8.41	7.66
9	14.42	14.42	12.49	9.84	12.1	12.77	13.04	14.45	12.10
10	15.79	15.78	17.46	13.33	17.1	18.0	16.44	15.75	17.10
11	19.51	19.52	23.05	17.32	22.7	23.88	19.61	19.63	22.70
12	22.22	22.14	31.91	23.83	31.51	33.14	26.53	20.46	31.51
13	59.38	51.66	66.13	43.29	65.6	68.99	62.61	71.42	65.60

of mortality rates, the graduated values were not changed very much by multiplying the prior means by $\frac{1}{2}$.

In the graduations in columns 3 and 4, I used the arc sine transformation. For the fourth graduation I multiplied the prior means by $\frac{1}{2}$, which in this case produced a significant change in the graduated values. The third graduation seems to me to be "smoother" than those in column 1 or column 2, but the graduated values are too close to the prior means.

In the graduation in column 5, I used the classical form of the Whittaker-Henderson method with the weights in the $\boldsymbol{E}$-matrix chosen so that the quotients E_{ii}/Z_{ii} would be equal to the corresponding quotients of the graduation in column 3. Perfect smoothness was defined to mean that first differences are zero. The graduated values were virtually identical with the prior means, that is, the classical graduation reproduced the prior means better than the Kimeldorf graduation. This is in spite of the fact that for Kimeldorf graduations perfect smoothness means that the graduated values are *equal* to the prior means. This strange result leads me to believe that the invertibility of the $\boldsymbol{Z}$-matrix is not as important as one would think.

In the sixth graduation I used the Camp form with perfect smoothness, meaning that

$$\sum_{i=1}^{n-1} (s_i u_i - u_{i+1})^2 = 0$$

where $s_i = m_{i+1}/m_i$. In the last graduation I used the classical form applied to the ratios u_i/m_i, with perfect smoothness meaning that second differences are zero. The results of the sixth and seventh graduations were smoother than those of the first and second and had a better fit than those of the third.

The greater smoothness of the graduation in column 7 of Table 3 could be due to the $\boldsymbol{Z}$-matrix having five nonzero diagonals rather than three. The $\boldsymbol{Z}$-matrix of the Camp graduations, however, is only tridiagonal, and the results are fairly smooth. I believe that the Camp form should be given more prominence in our literature as a unique form of the Whittaker-Henderson method. The old distinction (before Greville wrote his Part 5 study note for the Society of Actuaries) between Whittaker-Henderson A and B should be played down.

TRANSFORMATIONS OF THE UNGRADUATED VALUES

In their 1967 paper Kimeldorf and Jones transformed the ungraduated values by subtracting the prior means. In the present paper Hickman and Miller employ the arc sine transformation. I believe that transformations can be quite valuable, and I am happy to see them emphasized in the present paper. I hope that the next CSO, CSG, or annuity valuation table is graduated using a transformation.

I believe that transformations should be chosen so as to increase the smoothness, where smoothness is defined by the graduation method itself. Thus, if one were using the classical Whittaker-Henderson method with perfect smoothness meaning that second differences are zero, then the transformed values should lie almost on a straight line. The logarithmic transformation used by King,

$$u_x = \log (0.1 + q_x)$$

(see [11], p. 58), is particularly good. In fact, above age 40, the mortality curve for the 1958 CSO Table is almost a straight line when drawn on logarithmic paper.

Instead of graduating mortality ratios q_x/q^s_x, the Lidstone transformation,

$$u_x = \log (p^s_x/p_x)$$

can be used (see [1], p. 308).

The arc sine transformation,

$$u_x = \text{arc sin } q_x^{1/2} ,$$

does indeed smooth the ungraduated values. The smoothing has little to do with the arc sine function (which acts like the identity operator near the origin) but depends on the square-root function. In fact, the square-root curve has the same general shape as the logarithmic function used by King. The arc sine distribution function is used to find the probability of the number of times the lead could change in coin-tossing games (see [4], chap. 3).

VARIABLE SMOOTHING CRITERION

Hickman and Miller have furnished what I believe is the first example of the use of a variable smoothing criterion. In their graduation, the first four rows of the "smoothness term" $(\boldsymbol{v} - \boldsymbol{m})'\boldsymbol{Z}(\boldsymbol{v} - \boldsymbol{m})$ are zero if $\boldsymbol{v} - \boldsymbol{m} = 0$, while the other rows are approximately zero if the first differences of $\boldsymbol{v} - \boldsymbol{m}$ are zero. Further generalizations are possible. Thus one could employ the classical Whittaker-Henderson form with perfect smoothness meaning that fourth differences are zero for the first few values and that third differences are zero for the other values.

Miscellaneous Remarks

I noticed that the last value in Hickman and Miller's data did not seem to fit the pattern of the other ungraduated values. I believe this may be caused by the Society of Actuaries' method of collecting the data. The Society collects all data in the five-year age groups 30–34, 35–39, 40–44, and so on. In place of this rigid method, I believe that the Society should ask companies to report data in a format corresponding to their own underwriting breakpoints. For example, my company, Manhattan Life, and many other companies use 0–30, 31–35, 36–40, and so on, as age groups for underwriting classification. Data with different breakpoints can be graduated easily using splines, or the data can be adjusted to the Society's pattern by using ratios (as is currently done in combining age-nearest- and age-last-birthday data).

With regard to graduating select mortality tables, the Social Security Administration has just published an Actuarial Report in which the Whittaker-Henderson method was modified to handle select data (Bayo, F., and Wilkin, J. C., *Actuarial Study No. 74*, 1977, DHEW Publication No. (SSA) 77-11521). The paper "Modified 1965–70 Select and Ultimate Basic Tables" in these *Transactions* by O. David Green III gives another method of graduating select data.

REFERENCES

1. BENJAMIN B., and HAYCOCKS, H. *Analysis of Mortality and Other Actuarial Statistics.* London: Cambridge University Press, 1970.
2. CAMP, K. "New Possibilities in Graduation," *TSA*, VII (1955), 6.
3. CRAMÉR, H. *Mathematical Methods of Statistics.* Princeton, N.J.: Princeton University Press, 1946.
4. FELLER, W. *Introduction to Probability Theory and Its Applications.* Vol. I. 2d ed. New York: John Wiley & Sons, Inc., 1950.
5. GREVILLE, T. N. E. "On Smoothing a Finite Table: A Matrix Approach," *Journal of the Society for Industrial and Applied Mathematics*, V (1957), 137.
6. ———. *Theory and Applications of Spline Functions.* New York: Academic Press, 1969.
7. ———. Discussion of G. S. Kimeldorf's and D. A. Jones's "Bayesian Graduation," *TSA*, XIX (1967), 118.
8. HALMSTAD, D. G. "Act * Plus." Working Memorandum No. 103, Scientific Time Sharing Corporation.
9. JONES, D. A. "Bayesian Statistics," *TSA*, XVII (1965), 33.
10. KIMELDORF, G. S., and JONES, D. A. "Bayesian Graduation," *TSA*, XIX (1967), 66.
11. MILLER, M. D. *Elements of Graduation.* Philadelphia: Actuarial Society of America–American Institute of Actuaries, 1949.
12. RYDER, J. M. "Subjectivism—a Reply in Defense of Classical Actuarial Methods," *JIA*, CIII (1976), 59.

13. SIRKEN, M. G. *Comparison of Two Methods of Constructing Abridged Life Tables by Reference to a Standard Table.* National Center for Health Statistics, Ser. 2, No. 4, February, 1964.
14. SPIEGELMAN, M. *Introduction to Demography.* Rev. ed. Cambridge: Harvard University Press, 1968.
15. WHITTAKER, E. T., and ROBINSON, G. *The Calculus of Observations.* 4th ed. London and Glasgow: Blackie & Son, Ltd., 1944.

APPENDIX

I list in Exhibit I the APL computer programs that I ran on Manhattan Life's minicomputer to perform the various Whittaker-Henderson graduations. One advantage of APL is that commonly used actuarial functions such as matrix inverse (⌹) are built in and one does not have to resort to "canned" FORTRAN subroutine packages. In printing Table 3, I interchanged the $v2$ and $v3$ calculated by the program MAIN. The formulas in the program WHKIMEL for the posterior adjacent correlations and standard deviations come from some notes by David Halmstad for the Scientific Time Sharing Corporation.

EXHIBIT I

```
      ∇WHKIMEL[⎕]∇
    ∇ VPR←UW WHKIMEL MPR;M;P;R
[1]   ⍝ ***** CALCULATE THE TRIDIAGONAL  MATRICIES  Z, Z+W ****
[2]     NN←(⍴MPR)[⎕IO]
[3]     NUMER←1-(1↓RSQ)×¯1↓RSQ←0,(R×R←¯1↓MPR[;3]),0
[4]     ZDIAG←NUMER÷DENOM←(1-1↓RSQ)×(1-¯1↓RSQ)×P×P←MPR[;2]
[5]     ZOFF←(-R)÷(¯1↓P)×(1↓P)×(1-¯1↓1↓RSQ)
[6]     ZPLUSW←(0,ZOFF),(UW[;2]+ZDIAG),[1.5](ZOFF,0)
[7]     ZPLUSW← 0 1 ↓(NN,(1+NN))⍴ZPLUSW,(NN,(NN-1))⍴0
[8]   ⍝ ******* CALCULATE OUTPUT ******
[9]     INVERSE←⌹ZPLUSW
[10]    TEMPW←UW[;2]×UW[;1]
[11]    TEMPZ←(ZDIAG×M)+((ZOFF×1↓M),0)+(0,ZOFF×¯1↓M←MPR[;1])
[12]    VPR←INVERSE+.×TEMPZ+TEMPW
[13]    P←(1 1 ⍉INVERSE)*0.5
[14]    VPR←VPR,P,[1.5] R←(1 1 ⍉ 0 1 ↓INVERSE,0)÷P×1↓P,1
    ∇
      ∇WHCLASS[⎕]∇
    ∇ V←WT WHCLASS U;NN;NROW;STRING;K;W
[1]   ⍝ GLOBAL INPUT  KWH=SMOOTHING CONSTANT,   Z=DIFFERENCE
[2]     NROW←(NN←(⍴U)[⎕IO])-Z
[3]     STRING←ABIN Z
[4]     K←(NROW,NN)⍴STRING,NROW⍴0
[5]     W←(NN,NN)⍴WT,(NN,NN)⍴0
[6]     A←W+KWH×(⍉K)+.×K
[7]     V←(U×⍉(⌽⍴U)⍴WT)⌹A
    ∇
```

EXHIBIT I—*Continued*

```
    ∇WHCAMP[⎕]∇
  ∇ V←SW WHCAMP U
[1]  ⍝ GLOBAL INPUT  KWH=SMOOTHING CONSTANT,  Z=DIFFERENCE
[2]   NROW←(NN←(ρU)[⎕IO])-1+Z
[3]   STRING←ABIN Z
[4]   KK←((NROW,Z+2)ρ0,STRING)-(NROW↑SW[;1])∘.×STRING,0
[5]   K←(NROW,NN)ρKK,(NROW,NROW)ρ0
[6]   W←(NN,NN)ρSW[;2],(NN,NN)ρ0
[7]   A←W+KWH×(⍉K)+.×K
[8]   V←(U×⍉(⌽ρU)ρSW[;2])⌹A
  ∇
    ∇ABIN[⎕]∇
  ∇ STRING←ABIN Z
[1]  ⍝ CALCULATES ALTERNAING BINOMIAL COEFFICIENTS
[2]   STRING←,1
[3]  L1:→((ρSTRING)>Z)ρ0
[4]   STRING←(0,STRING)-(STRING,0)
[5]   →L1
  ∇
```

EXHIBIT I—*Continued*

```
      ∇TESTAHICKMAN[⎕]∇
    ∇ TESTAHICKMAN MFRACT
[1]   U←1E¯5× 168 73 107 120 144 , 307 537 841 1445 1575 , 1963 2046 7142
[2]   NP←1000× 2 2 3 4 5 , 5 5 5 5 4 , 3 2 2
[3]   M←1E¯5× 99 107 90 105 172 , 294 486 766 1210 1710 , 2270 3151 6560
[4]   M←M×MFRACT
[5]   AMT← 51071 102352 278026 729655 998975 , 919910 698755 422317 214521 93843
[6]   LIVES←(100÷7500)×AMT←AMT, 30619 7528 686
[7]  ⍝ ****** KIMELDORF FORM    *******
[8]   WT←LIVES÷M×1−M
[9]   R←(4⍴0),(8⍴0.942809),0
[10]  P←0.7×M*0.5
[11]  V1←, 0 ¯2 ↓VPR←(UW←U,[1.5] WT) WHKIMEL(MPR←M,P,[1.5] R)
[12] ⍝ ****** KIMELDORF FORM APPLIED TO ARC SINE VALUES *****
[13]  UA←¯1○U*0.5
[14]  MA←¯1○M*0.5
[15]  PA←0.5×NP*¯0.5
[16]  WT←4×LIVES
[17]  VPR←(UW←UA,[1.5] WT) WHKIMEL(MPR←MA,PA,[1.5] R)
[18]  V2←(1○VPR[;1])*2
    ∇
```

EXHIBIT I—*Continued*

```
      ∇TESTAHICKMAN2[⎕]∇
    ∇ TESTAHICKMAN2
[1]  ⍝ ****** CLASSICAL METHOD APPLIED TO  ARC SINE VALUES ****
[2]   WT←(4↓UW[;2])×(1,(7⍴2),1)÷(4↓ZOTAG)
[3]   Z←KWH←1
[4]   V5A←WT WHCLASS 4↓UA-MA
[5]   V5←(4↓M)+(1○V5A)*2
[6]  ⍝ ******  CAMP METHOD    SHAPE VECTOR = RATIOS *******
[7]   WT←4↓LIVES÷M×1-M
[8]   SW←(((1↓M4)÷¯1↓M4←4↓M),0)[;1,5] WT
[9]   KWH←(4÷9)×+/SW[;2]
[10]  Z←0
[11]  V6←SW WHCAMP 4↓U
[12] ⍝ **** CLASSICAL METHOD    APPLIED TO   RATIOS *****
[13]  Z←2
[14]  KWH←1
[15]  V7←(4↓M)×VANS←(9⍴1) WHCLASS RATIO←4↓U÷M
    ∇
```

EXHIBIT I—*Continued*

```
      ∇MAIN[⎕]∇
    ∇ MAIN
[1]   TESTAHICKMAN 0.5
[2]   V3←V1
[3]   V4←V2
[4]   TESTAHICKMAN 1
[5]   TESTAHICKMAN2
[6]   TEMP←V1,V2,V3,[1.5] V4
[7]   (4 0 ↓TEMP),V5,V6,[1.5] V7
    ∇
      )SAVE
SAVED 1001 GRAD
```

STEVEN F. MCKAY:

I would like to discuss the problem of graduating a grid of estimates of select probabilities, mentioned by Drs. Hickman and Miller near the end of their paper. We at the Social Security Administration faced this problem when graduating disability termination rates. Our solution was to use a Whittaker-Henderson type B graduation that smoothed in two directions at once. Since a Whittaker-Henderson type B merely minimizes a function of smoothness and fit, the generalization to a grid of probabilities only amounts to adding another term to the function to be minimized. Complications arise in the computations involved, but we were aided greatly by the Part 5 study notes on graduation written by Dr. T. N. E. Greville.

The procedure is outlined in the appendix to *Experience of Disabled-Worker Benefits under OASDI, 1965–74* by F. R. Bayo and J. C. Wilkin (Actuarial Study No. 74) and may be obtained by writing to: Office of the Actuary, Social Security Administration, 6401 Security Boulevard, Baltimore, Maryland 21235. A listing of the computer program is also available. I would be interested in any comments anyone may have on the graduation procedure.

HARWOOD ROSSER:

This reviewer is supposed to be an expert on graduation; at least, he has served as consultant to the Part 5 examination committee on this subject for over ten years. Bayesian graduation involves statistical techniques, which is made quite clear in some of the earlier papers. For example, the paper by Kimeldorf and Jones (ref. [6] in the paper under discussion) states at its outset: "The graduation problem is stated in the context of multivariate statistical estimation and analyzed according to Bayesian procedures."

Faye Albert, the chairman of the current Part 5 committee, also agrees. Speaking for the committee as a whole, and making recommendations to the general officers of the Education Committee, she wrote recently: "Consideration of Bayesian graduation should be included in the syllabus, but this depends on Bayesian statistics being covered." Just in passing, most, if not all, of the Education and Examination part committees use the Bayesian approach in setting pass marks.

Many years ago, another Education and Examination committee, with the aid of a syllabus change and a congressional investigation, gave me a permanent inferiority complex on the subject of statistics. The paper by Kimeldorf and Jones did nothing to alleviate that complex (although I found the courage to offer a discussion of it), nor does the current paper by Hickman and Miller. Understandably, therefore, I will refrain from comment on the more technical aspects of it.

Up to this point, these remarks have been directed toward essentially background information. Looking at the paper itself, and considering the possibility that some or all of the material in the paper might ultimately find its way into the syllabus, one is compelled to suggest that some clarification would be required first. For example, in the abstract of the paper, it is stated that "suggestions for managing four technical problems in their method are developed." Are these four "technical problems" the same as the four "objectives" discussed in Sections IV and XI of the paper? Also, are these four, whatever they are, illustrated in Table 1? If either answer is "yes," then I—and perhaps quite a few other actuaries—would require several more readings, without further assistance, before we could "go and do likewise" in an actual situation. Dr. Hickman, in discussing the earlier paper, characterized one section as "a primer on the mathematics needed to use the new approach." The current paper, in my view, is no primer.

Along similar lines, perhaps the greatest weakness of the paper is in the area of numerical examples, as to both quantity and documentation. For a paper as technical as this, a single example seems rather sparse, especially when several proposals are made. By documentation I mean such things as column headings and sources of information. The authors are, I am afraid, guilty of insularity, and of assuming a restricted readership. These two are not completely identical. Their Table 1 assumes that readers know that the u column means ungraduated values, and the v column graduated ones. This may not be fully obvious to British and European readers and could be clarified by proper labeling. The fact that papers on graduation, among other subjects, are read on both sides of the Atlantic should be obvious from the phrase "Whittaker-Henderson graduation process." Also, papers in the *Transactions* are not read only by actuaries, just as actuaries read nonactuarial publications as well. Better labeling in Table 1 would have helped such other readers also. For all readers, some of the intermediate calculations supporting Table 1 might be of value.

At this point, one is inclined to ask whether or not certain precautions, suggested in connection with the earlier paper by Kimeldorf and Jones, have been taken here. This reviewer asked how the traditional approach of testing the results of a graduation before accepting it could be satisfied under Bayesian graduation. The authors suggested, in reply, a rather involved technical procedure. It is not clear whether or not this has been carried out in connection with Table 1.

There may be some implied lessons here for the Committee on Papers. Television programs are regularly subjected to surveys to ascertain percentages of viewership. Readers of public periodicals such as newspapers and magazines vote by way of newsstand purchases and new or renewal subscriptions. The recipients of the *Transactions*, however, are largely a captive group; very few people who are not Society members subscribe. A number of institutions subscribe, but they are not readers as such.

It thus might be worthwhile for the Society to conduct periodic surveys on readership. The number of members who submit written discussions of papers is, of course, readily ascertainable. (In the old days, there were also those offering oral discussions of papers.) It might be interesting to explore for selected published papers such questions as these:

1. Did you read the entire paper?
2. Did you merely skim it?
3. Did you wait for the discussions to appear before looking at it? (I am frequently in this category).
4. Did you ignore it altogether?
5. Did you find in it some practical applications to your own work?
6. Would you like to see this on the examination syllabus, either in its original form or after adaptation, perhaps in a study note?
7. Would you recommend it for a triennial prize, if it is eligible?
8. Would you like to see more papers on this subject?
9. Would you like to see other papers of this type (historical, explanatory, technical, polemic, etc.)?
10. Is the *Transactions* the most appropriate forum for this paper?
11. Was the paper
    a) Too long?
    b) Too short?
    c) Too technical?
    d) Too elementary?
    e) Badly organized?
    f) Other? (Explain.)

The tabulation of the results could be furnished privately to the authors. Perhaps, with identification suppressed, the results could be published. This might give some guidance to the Committee on Papers and also to potential authors of future papers.

In conclusion, this is an excellent paper on balance, although it is technically somewhat difficult. The subject would seem to be in the mainstream of actuarial trends that are developing, and the authors are to be congratulated on an admirable contribution to actuarial knowledge.

(AUTHORS' REVIEW OF DISCUSSION)

JAMES C. HICKMAN AND ROBERT B. MILLER:

We are grateful to Dr. Kabele for pointing out an error in some of the graduated values reported in Table 1 of the preprint of this paper. The error has been corrected in the final version. The corrected values are somewhat smoother than those that appeared in the preprint.

Our reply to the discussions will be organized under topic headings.

1. *The Hump*

Jones, Klugman, and Kabele each comment on the problem of using a correlation matrix from the class α_1 of the Kimeldorf-Jones paper when the nature of the prior information is different for various sections of the vector of observed values to be graduated. This problem becomes especially pressing when male mortality data are being graduated. The hump in young adult ages appears to be a significant aspect of the mortality data, yet a thoughtless graduation may remove this feature. We join Jones and Klugman in stating that complete independence among adjacent probabilities, even at young male ages, is not a property of our prior knowledge about mortality probabilities. However, because we had not made an intensive study of mortality patterns at these ages, we had real difficulty in specifying the form of the dependence. The approach taken in the paper was designed to show that we recognize the problem of the hump and to suggest that a partitioned prior covariance matrix might be used to, in Dr. Kabele's words, "vary the smoothing criteria."

Each of the discussants who touch on this point makes a valuable suggestion. Dr. Jones suggests that the correlation matrix be selected from the class of matrices denoted by α_2 in the Kimeldorf-Jones paper. Dr. Klugman's suggestion is closely related. He shows how to use a special member of the class of matrices α_2. Unfortunately, we have not gained enough experience with members of this class to permit us to make any practical suggestions for their use.

2. *Bivariate Graduation*

At about the time the preprint of this paper appeared, we saw Actuarial Study No. 74, published by the Office of the Actuary, Social Security Administration. We found that the technical appendix by Steven F. McKay and John C. Wilkin contained a development of a two-dimensional Whittaker type B graduation formula that had much in common with our tentative ideas for graduating select mortality or demographic data organized in a Lexis diagram. Our goal is somewhat different from that of McKay and Wilkin. In addition to smoothing, we hope to extend our model to make projections with associated probability statements. We thank Mr. McKay for indicating to the readers of this paper the interesting work on graduation going on in the Office of the Actuary.

3. *Terminology*

Dr. Kabele reminds us that E. T. Whittaker provided a Bayesian motivation for the graduation method that bears his name. It is regrettable that his original development is not well known. At least to Whittaker, there was no mystery about the interpretation of k in $F + kS$; it was the ratio of two variances. We agree that Bayesian graduation includes much more than the particular methods illustrated in the Kimeldorf-Jones paper.

Dr. Klugman suggests that the parameter r, the basic parameter if the correlation matrix is selected from class $\mathcal{a}_1$ of the Kimeldorf-Jones paper, should not be interpreted solely as a smoothness parameter. Rather, it also measures the weight attached to the vector of prior means, which may or may not be smooth. We agree, and would point out that the index h in our paper, which measures the relative precision of the two inputs into the graduation process, is extremely sensitive to the choice of the parameter r when the prior correlation matrix belongs to the class $\mathcal{a}_1$ of matrices of the Kimeldorf-Jones paper. In the narrative we placed too much stress on the smoothing properties of r.

4. *Standard Tables*

Dr. Kabele chides us gently for asserting that graduation using standard tables involves the use of information about the level of mortality from the standard table. He displays an example from the construction of abridged tables using demographic data in which information from the standard table is used only to perform an approximate integration.

Our statement was probably too sweeping. Nevertheless, we expect that the difficulty may be largely semantic. One of the uses of standard tables we had in mind involves minimizing

$$S(\boldsymbol{\theta}) = \sum_{x=1}^{n} W_x[MR_x - f(x, \boldsymbol{\theta})]^2 ,$$

where W_x is a set of positive weights, MR_x is a set of mortality ratios, and $f(x,\boldsymbol{\theta})$ is a prescribed function of x which depends on a vector of parameters $\boldsymbol{\theta}$. For example, if $f(x, \boldsymbol{\theta})$ is a polynomial

$$\theta_0 + \theta_1 x + \ldots + \theta_m x^m ,$$

the graduated values will be given by

$$q_x^s(\hat{\theta}_0 + \hat{\theta}_1 x + \ldots + \hat{\theta}_m x^m) .$$

In this expression, $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_m$ denote weighted least-squares estimates of the parameters as determined from the data. Thus, both the "level" $q_x^s\hat{\theta}_0$, the "slope" $q_x^s\hat{\theta}_1x_1$, and the other characteristics of the graduated sequence are in part determined by the standard table.

Dr. Kabele's remark about the effect of multiplying the standard table by a constant is correct. However, it seems that this transformation might be called a change in scale rather than a change of level.

5. *Transformations*

Transformations are powerful statistical tools. By their judicious use the range of application of standard statistical models may be expanded. As Dr. Kabele points out, the logarithmic transformation has a long history as a device for reducing the variability of mortality data. Mortality ratios, loss ratios, and pure premiums (losses divided by exposure units) are other examples of routine uses of transformations in actuarial work.

Dr. Kabele remarks very perceptively that the most important aspect of the transformation used in the paper is the square-root operation performed on the original data. It is well known that for small values of q_x the Poisson distribution rather than the binomial distribution may be used to model the mortality process. The square-root transformation is used to stabilize the variance of Poisson-distributed data in much the same fashion as the arc sine transformation is used with binomial data. (See the paper by Anscombe cited in ref. [1] in the paper for a development of this statement.) Consequently, one would expect essentially identical results for most mortality data using the square-root and the arc sine transformations. We elected to develop the arc sine transformation, combined with the Bayesian analysis of mortality data with a binomial distribution, because of the greater familiarity of members of the Society of Actuaries with these topics. We could have illustrated equally well our ideas on the use of transformations in the Bayesian analysis of data by using Poisson-distributed general insurance claim data coupled with the square-root transformation. (See R. B. Miller and J. C. Hickman, "A Prior Distribution Arising in Risk Theory," *ARCH*, 1973, No. 4.)

6. *Technical Papers*

We must accept Mr. Rosser's criticism of the exposition in this paper. It was intended as a supplement to the Kimeldorf-Jones paper, and it was not planned that it should be self-contained.

Mr. Rosser also suggests that technical and professional journals engage in more market research. We cannot argue against an organization's being sensitive to its members' interests and needs. However, to restrict scientific publications in any way to those topics of current interest to practitioners of the science would almost guarantee the intellectual death of the field of activity.

7. Credibility

Dr. Kabele reminds us of the similarity between graduation formulas and credibility formulas. This observation provides a theme that can unify many of the procedures used in actuarial science.

Dr. Kabele goes on to suggest a class of matrices that might be used in a generalized approach to credibility and smoothing. Because of the embryonic state of the art of specifying covariance matrices, all suggestions for helping with this perplexing task are welcome. Since the actuary selects matrix $\boldsymbol{A}$ (Hickman-Miller notation) or $\boldsymbol{Z}^{-1}$ (Kabele's notation), the suggested class of matrices might be very useful in carrying out this task. However, matrix $\boldsymbol{B}$ (Hickman-Miller notation) or $\boldsymbol{E}^{-1}$ (Kabele's notation) is restricted by the nature of the process being observed and the design of the experiment for observing the process. Therefore, it seems that the actuary's freedom in specifying a matrix $\boldsymbol{B}$ is not complete. For example, consider the situation where several years of observation are used in a mortality investigation and the survivors of one age are observed at the next age. It is clear that, because of dependence among observed mortality probabilities, the covariance matrix $\boldsymbol{B}$ will no longer be diagonal and might not fall into the proposed class.

The discussions by Jones, Kabele, Klugman, McKay, and Rosser are very provocative. They contain ideas and suggestions that deserve much more comment than is appropriate for a discussion. We thank them for contributing to actuarial science in this fashion.