TRANSACTIONS OF SOCIETY OF ACTUARIES
1992 VOL. 44

A SIMPLE MONTE CARLO APPROACH TO BAYESIAN GRADUATION

BRADLEY P. CARLIN

ABSTRACT

The problem of graduating a sequence of data values can be cast as a statistical estimation problem. In particular, the Bayesian approach is attractive due to its ability to formally incorporate known ordering and smoothness conditions for the graduated values into the estimation structure. However, this approach has not been widely adopted in practice, primarily because of the arduousness of specifying the prior distributions for the graduated values and carrying out the necessary numerical integrations. This paper presents simple Bayesian graduation models that substantially ease the prior elicitation burden; it also describes a Monte Carlo integration approach that greatly reduces the computational load. The method is presented in generality and subsequently illustrated with two examples, one from the realm of health insurance and the other from the more traditional graduation context of mortality table construction. It is hoped that the method will stimulate greater use of the Bayesian paradigm within the actuarial community.

1. INTRODUCTION

Perhaps no single topic has received more attention in the actuarial literature over the years than the graduation of a sequence of initial mortality rates into smoothed final estimates. Although many practically useful techniques have been available for many years, the recent shift of primary actuarial education from a deterministic to a more stochastic footing (see Bowers et al. [3]) has encouraged the development of more statistical approaches to the graduation problem. These approaches have the advantage of casting the problem as one of estimation of a k-dimensional vector of underlying, unknown rate parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^T$, the values of which help determine the elements of the k-dimensional vector of observed initial rates $\mathbf{y} = (y_1, \ldots, y_k)^T$. Notable papers include the works of Broffitt [4], Chan and Panjer [10], and Hoem [18].

Within the statistical framework, the Bayesian approach to graduation is attractive due to its ability to formally blend the practitioner's prior beliefs about the true rates into the analysis, thus avoiding the informal post-analysis

adjustment sometimes required with other methods because the results "don't look right." The primary references in the Bayesian graduation literature are the papers of Kimeldorf and Jones [20] and Hickman and Miller [17], summaries of which appear in the textbook by London [23]. The basic method posits multivariate normal distributions for the prior distribution of the true rates $\boldsymbol{\theta}$, and also for the conditional distribution of the observed data $\mathbf{y}$ given $\boldsymbol{\theta}$ (often called the *likelihood* for the experiment). These assumptions enable the posterior distribution of the true rates, given the observed data, to emerge as multivariate normal as well, the mean of which is taken as the final, graduated rates. Although the method is computationally simple (only basic matrix manipulations are required), it is frequently hard to implement because of the difficulty in setting the myriad of values required for the prior mean and covariance matrix of the true rates. Further, the simple multivariate normal structure is unable to guarantee prespecified orderings of the rates (such as the requirement that human mortality rates be increasing after age 30), though constraints are frequently imposed on the prior covariance matrix that approximate such conditions.

Of course, the imposition of such natural constraints on the unknown true rates $\boldsymbol{\theta}$ is still *conceptually* easy to handle using the Bayesian framework by including them in the prior distribution for $\boldsymbol{\theta}$; Bayes' Theorem ensures that any constraints present *a priori* must necessarily be present *a posteriori*. However, as with many Bayesian approaches to applied problems, the difficulty in carrying out the associated numerical integration necessary to compute the posterior distribution has prevented such approaches from being seriously contemplated until recently. In a series of excellent papers, Broffitt [5]–[7] has shown that by cleverly choosing prior distributions consistent with a reparameterized version of a model that implicitly includes the relevant constraints, the desired final estimates (in the form of posterior modes $\hat{\theta}_1, \ldots, \hat{\theta}_k$) emerge as the solution to a system of k equations in k unknowns.

Broffitt's method represents a sizable improvement over previous Bayesian approaches, but several criticisms (most typical of many Bayesian data analyses) can be made. Mathematical tractability concerns force the user to adopt a form for the prior distribution that is *conjugate* with the likelihood, that is, a prior that enables the posterior to emerge in a simple closed form. Further, the particular conjugate form chosen depends on the nature of the constraints imposed on $\boldsymbol{\theta}$. For example, requiring the rates to be increasing requires a different prior structure on $\boldsymbol{\theta}$ than if we additionally insist that the rates be convex as well. Finally, and perhaps most limiting from an applied

standpoint, the method is similar to past attempts at Bayesian graduation in the large amount of effort required in specifying the prior parameter values. Combined with the high degree of subjectivity involved, this often serves to discourage the practitioner, who subsequently turns to less subjective, more easily implementable approaches such as Whittaker's method (again, see London [23]).

In this paper we show how a recently developed Monte Carlo integration technique known as the Gibbs sampler can be used to obtain realistic answers in Bayesian graduation contexts without resorting to convenient assumptions about the forms of the model or the prior distribution. The method is easy to implement and does not require extensive numerical analytic expertise. Hardware requirements for most problems are also minimal; generally only a personal computer and a few basic random number generators are required. Estimates of the full posterior distribution or any characteristic thereof for any parameter of interest are easily obtainable. In particular, posterior means, medians, or modes for the rate parameters θ_i can be used as the final rate estimates. Our basic models also streamline the prior elicitation process, historically the real impediment for persons seeking to apply the Bayesian methodology. We develop the methodology for completely general likelihood and prior combinations, and subsequently provide details for two important examples found in practice. First, we consider graduating a series of aging factors in health insurance claim costs, in which an initial series of irregular rates is to be converted into one that is first increasing (up to some age s), then decreasing. Second, we take up the traditional context of human mortality table construction and show how increasing or increasing convex graduated rates can be obtained. Both examples are illustrated with datasets from the literature. Finally, we offer a summary discussion and our suggestions for more general applications of the methodology.

2. BAYESIAN FORMULATION OF THE GRADUATION PROBLEM

We begin with a completely general parametric model for the data $\mathbf{y}$, which depends on our k-dimensional parameter vector $\boldsymbol{\theta}$ that is constrained to lie in a subset S_y of k-dimensional Euclidean space. In this paper, S_y is determined solely by inequalities amongst the θ_i components, so that $S_y = S$ is actually free of $\mathbf{y}$ (though this is unnecessary for the success of the algorithm). For example, for an increasing set of parameters we have $S = \{\boldsymbol{\theta}: \theta_1 < \theta_2 < \cdots < \theta_k\}$. We denote the likelihood [that is, the probability density function (pdf) of the data given $\boldsymbol{\theta}$] by $f(\mathbf{y}|\boldsymbol{\theta})$, and the prior distribution

by $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a vector of parameters (commonly called *hyperparameters*) for $\boldsymbol{\theta}$'s distribution. We refer to the likelihood times prior as the *Bayesian model,* which in this case is given by

$$p(\mathbf{y},\boldsymbol{\theta}|\boldsymbol{\lambda}) = f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})I_S(\boldsymbol{\theta}) \tag{1}$$

where $I_S(\boldsymbol{\theta})$ is the indicator function of the set S, so that $I_S(\boldsymbol{\theta})$ equals 1 when $\boldsymbol{\theta}\in S$ and equals 0 otherwise. (For notational simplicity, we assume that any constraints on the y_i's, such as $y_i>0$, are built into the likelihood $f(\mathbf{y}|\boldsymbol{\theta})$. Also, throughout the paper we use p as a generic symbol for a pdf.) Gelfand, Smith and Lee [15] show that the desired posterior distribution for $\boldsymbol{\theta}$ is proportional to the Bayesian model in Equation (1); one simply computes the normalized version

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\lambda}) &= \frac{p(\mathbf{y},\boldsymbol{\theta}|\boldsymbol{\lambda})}{p(\mathbf{y}|\boldsymbol{\lambda})} = \frac{p(\mathbf{y},\boldsymbol{\theta}|\boldsymbol{\lambda})}{\int p(\mathbf{y},\boldsymbol{\theta}|\boldsymbol{\lambda})\,d\boldsymbol{\theta}} \\ &= \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})I_S(\boldsymbol{\theta})}{\int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})I_S(\boldsymbol{\theta})\,d\boldsymbol{\theta}}\,. \end{aligned} \tag{2}$$

If $\boldsymbol{\lambda}$ is known, this posterior distribution is fully specified, and one simply takes some appropriate summary (such as the posterior mean) to be the set of graduated values. Notice that by using Equation (2) we are circumventing the intermediate step of standardizing the prior to satisfy the constraint set S. In other words, the *proper* distribution (that is, one that integrates to 1 over S) that actually characterizes our prior beliefs about $\boldsymbol{\theta}$ is

$$\pi^*(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \frac{\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})I_S(\boldsymbol{\theta})}{\int \pi(\boldsymbol{\psi}|\boldsymbol{\lambda})I_S(\boldsymbol{\psi})\,d\boldsymbol{\psi}}\,.$$

Replacing π with π^* in Equation (2) produces the same result for $p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\lambda})$, because the $d\boldsymbol{\psi}$ integrals will cancel in the numerator and denominator. Thus working with π directly (instead of π^*) simplifies the necessary formulas and also makes prior specification (choice of $\boldsymbol{\lambda}$) easier, because we may think about $\boldsymbol{\lambda}$ separately from the constraint set S. Section 4 contains further guidance on this issue in the context of two examples.

As an alternative to specifying a vector of fixed values for $\boldsymbol{\lambda}$, a second-stage prior distribution (sometimes called a *hyperprior*) can be selected for

it. If we denote this distribution by $h(\boldsymbol{\lambda})$, the desired posterior for $\boldsymbol{\theta}$ is now obtained by marginalizing over $\boldsymbol{\lambda}$,

$$p(\boldsymbol{\theta}\,|\,\mathbf{y}) = \frac{p(\mathbf{y},\boldsymbol{\theta})}{p(\mathbf{y})} = \frac{\int p(\mathbf{y},\boldsymbol{\theta},\boldsymbol{\lambda})\,d\boldsymbol{\lambda}}{\int\int p(\mathbf{y},\boldsymbol{\theta},\boldsymbol{\lambda})\,d\boldsymbol{\lambda}\,d\boldsymbol{\theta}}$$

$$= \frac{\int f(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})h(\boldsymbol{\lambda})I_S(\boldsymbol{\theta})\,d\boldsymbol{\lambda}}{\int\int f(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})h(\boldsymbol{\lambda})I_S(\boldsymbol{\theta})\,d\boldsymbol{\lambda}\,d\boldsymbol{\theta}}. \tag{3}$$

Unfortunately, the integrations depicted in Equations (2) and (3) typically required sophisticated numerical techniques because of the presence of the constraint set S.

Before introducing our computational method, we wish to consider the *complete conditional* distribution of the i-th component of $\boldsymbol{\theta}$, that is, the distribution of θ_i given values for the data and all the other parameters in the model. We notate this distribution as $p(\theta_i|\mathbf{y},\boldsymbol{\lambda},\theta_{j\neq i})$. Corresponding to this distribution is a cross-section of the constraint set S, which we notate as S_i. For example, looking again at our increasing parameter case above, we would have $S_i=\{\theta_i:\theta_{i-1}<\theta_i<\theta_{i+1}\}$. Using logic similar to that preceding Equation (2), we can write

$$p(\theta_i\,|\,\mathbf{y},\boldsymbol{\lambda},\theta_{j\neq i}) = \frac{f(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})I_{S_i}(\theta_i)}{\int f(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})I_{S_i}(\theta_i)\,d\theta_i}.$$

Because the denominator of the right-hand side of this expression is a constant with respect to θ_i, a convenient shorthand representation of this complete conditional distribution is given by

$$p(\theta_i\,|\,\mathbf{y},\boldsymbol{\lambda},\theta_{j\neq i}) \propto f(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}\,|\,\boldsymbol{\lambda})I_{S_i}(\theta_i), \tag{4}$$

where we remember to view the right-hand side of (4) as a function of θ_i for given values of $\theta_{j\neq i}$. Notice that if π is chosen to be conjugate with the likelihood f, so that the unconstrained complete conditional distribution emerges as a familiar standard form, then the constrained version given in (4) is simply the same standard distribution restricted to S_i. This fact will be key in our sampling implementation, which we describe in the next section.

3. REVISED ESTIMATE CALCULATION VIA THE GIBBS SAMPLER

The Gibbs sampler is a Monte Carlo integration method, developed formally by Geman and Geman [16] in the context of image restoration. In the

Bayesian framework, Tanner and Wong [25] used essentially this algorithm in their substitution sampling approach. Most recently, Gelfand and Smith [13] developed the Gibbs sampler for general Bayesian settings; that paper contains a more complete discussion of the method and its properties.

To summarize the method, suppose we have a collection of n random variables $\mathbf{U} = (U_1, \ldots, U_n)$ whose complete conditional distributions, denoted generically by $g_i(U_i|U_{j\neq i})$, $i = 1, \ldots, n$, are available for sampling. Here, "available" means that samples can be generated by some method, given values of the appropriate conditioning random variables. Under mild conditions (see Besag [2]), these complete conditional distributions uniquely determine the full joint distribution, $p(U_1, \ldots, U_n)$, and hence all marginal distributions, $p(U_i)$, $i = 1, \ldots, n$. The Gibbs sampler generates samples from these marginal distributions as follows: Given an arbitrary starting set of values $U_{1(0)}, \ldots, U_{n(0)}$, we draw $U_{1(1)}$ from $g_1(U_1|U_{2(0)}, \ldots, U_{n(0)})$, then $U_{2(1)}$ from $g_2(U_2|U_{1(1)}, U_{3(0)}, \ldots, U_{n(0)})$, and so on up to $U_{n(1)}$ from $g_n(U_n|U_{1(1)}, \ldots, U_{n-1(1)})$ to complete one iteration of the scheme. After t such iterations, we obtain $(U_{1(t)}, \ldots, U_{n(t)})$. Geman and Geman [16] show under mild conditions that this n-tuple converges in distribution to a random observation from $p(U_1, \ldots, U_n)$ as $t \to \infty$. For this reason, in what follows we suppress the (t) subscript, assuming that t is sufficiently large for the generated sample to be thought of as a realization from the joint distribution.

Now, replicating the entire process in parallel G times provides independent and identically distributed (i.i.d.) n-tuples $(U_1^{(g)}, \ldots, U_n^{(g)})$, $g = 1, \ldots, G$ from the joint distribution. (Typically the same starting values are used for each parallel sampling chain, though this is a just a matter of convenience and is unnecessary for the convergence of the algorithm.) These observations can then be used for estimation of any of the marginal densities or any features thereof. In particular, the marginal mean of U_i can be estimated by

$$\hat{E}(U_i) = \overline{U}_i = \frac{1}{G}\sum_{g=1}^{G} U_i^{(g)}. \tag{5}$$

Moreover, since the $U_i^{(g)}$ are i.i.d., the Central Limit Theorem implies that $\overline{U}_i$ is approximately normally distributed with mean $E(U_i)$ and variance $Var(U_i)/G$. Thus a simple standard error estimate for the point estimator in (5) is given by

$$\hat{se}(\overline{U}_i) = \sqrt{\frac{\sum_{g=1}^{G}(U_i^{(g)} - \overline{U}_i)^2}{(G-1)G}},$$

the square root of the sample variance of the $U_i^{(g)}$'s divided by G. Gelfand and Smith [14] give more sophisticated methods for estimating marginal moments and computing density estimates, but the simple methods given above suffice for our purposes.

In our application, we have $\mathbf{U}=\boldsymbol{\theta}$, or $\mathbf{U}=(\boldsymbol{\theta},\boldsymbol{\lambda})$ if the hyperparameters are also unknown and thus have distributions as well. Notice that $\mathbf{y}$ is *not* a component of $\mathbf{U}$; that is, we generate from $p(\theta_i|\boldsymbol{\theta}_{j\neq i},\boldsymbol{\lambda},\mathbf{y})$, $i=1, \ldots, k$, and from $p(\lambda_i|\boldsymbol{\lambda}_{j\neq i},\boldsymbol{\theta},\mathbf{y})$, $i=1, \ldots, dim(\boldsymbol{\lambda})$, but *not* from $p(y_i|\mathbf{y}_{j\neq i},\boldsymbol{\theta},\boldsymbol{\lambda})$, $i=1, \ldots, k$. This is because what is desired at convergence of the algorithm is not the marginal *prior* $p(\theta_i)$, but the marginal *posterior* $p(\theta_i|\mathbf{y})$. Because the $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ complete conditionals are by definition conditioned on the data $\mathbf{y}$, by leaving these components out of the sampling process, the algorithm converges to the desired posterior distributions.

Looking again at Equation (4), we see sampling from the θ_i complete conditionals could be naively accomplished simply by sampling from the unconstrained distribution and then accepting the generated variate only if it satisfies the constraint $\theta_i^{(g)}\in S_i$. This, however, could lead to quite inefficient generation, because most variates generated might be rejected. A much more efficient approach for generating from invertible truncated distributions is given by Devroye [11]. Suppose X is a random variable having cumulative distribution function (cdf) F, and Y is a truncated version of this random variable with support restricted to the interval $[a,b]$. Then Y has cdf

$$G(y) = \begin{cases} 0, & y < a \\ \dfrac{F(y) - F(a)}{F(b) - F(a)}, & a \leq y \leq b \\ 1, & y > b \end{cases}.$$

Then Y can be generated as $F^{-1}\{F(a)+V[F(b)-F(a)]\}$, where V is a $U(0,1)$ random variate and U denotes the uniform distribution. This enables "one-for-one" generation from truncated distributions, eliminating the need for rejection algorithms. We make great use of this fact in the examples of our next section.

4. EXAMPLES AND NUMERICAL ILLUSTRATIONS

In this section two examples illustrate the methodology for some common distributional models in familiar actuarial settings.

Example 1: Health Insurance Aging Factors

With the recent publication by the Financial Accounting Standards Board of *FAS 106*, "Employer's Accounting for Postretirement Benefits Other Than Pensions," the development of methods for projecting postretirement health costs has received increased attention. Such projections are highly sensitive to health insurance "aging factors," or factors that project the increase in health-care cost as an individual ages. These quantities can be computed from a claim cost distribution as the annualized percentage increase in the monthly cost per person at a given age bracket over the monthly cost per person from the immediately preceding age bracket. That is, suppose $C_{[i-n_1,i]}$ and $C_{[i,i+n_2]}$ are the true monthly costs per person for two consecutive age brackets having lengths n_1 and n_2 years, respectively. Then representing the intervals by their midpoints, the i-th "true" aging factor θ_i satisfies the equation

$$(1 + \theta_i)^{\frac{n_1+n_2}{2}} = \frac{C_{[i,i+n_2]}}{C_{[i-n_1,i]}}.$$

We can compute estimates y_i of these θ_i's by replacing the true monthly costs C by estimated (observed) $\hat{C}$ values obtained from employee claim cost data. Solving, we obtain

$$y_i = \left(\frac{\hat{C}_{[i,i+n_2]}}{\hat{C}_{[i-n_1,i]}}\right)^{\frac{2}{(n_1+n_2)}} - 1. \tag{7}$$

Like ungraduated mortality rates, these "raw" aging factors y_i may form a noisy, irregular sequence. Because our goal is to project increases in medical costs due to the aging of a population, we would naturally prefer a smooth sequence of aging factors. Although the post-65 aging factors are of primary interest, we must also be concerned with pre-65 factors because of the presence of younger spouses, disabled workers and early (pre-Medicare) retirees. Further, health-care professionals believe that aging factors should be near zero for the youngest ages and also for the oldest ages. Finally, we would likely prefer a series without any negative aging factors, because, excluding infancy, we would not expect any age range over which costs are *decreasing*. This collection of prior opinions suggests a sequence of true aging factors that are small (but positive) for the early ages, increase smoothly with age up to some age s near retirement, and then decrease smoothly until

the end of the table, where the aging factor is again small yet positive. We may also want to impose some upper bound B that no θ_i may exceed.

In the notation of Section 2, suppose we specify the likelihood function f by assuming that the observed (data-based) aging factors y_i arise from independent normal populations having means θ_i and variances σ^2 (that is, we assume homogeneous variances across ages). We specify our prior π by adopting the familiar product of independent conjugate normal distributions for the θ_i *before imposing the constraints*. That is, we assume that if we ignore the constraints, the θ_i's constitute a random sample from a single (prior) normal population. The Bayesian model in Equation (1) may be written as

$$p(\mathbf{y},\boldsymbol{\theta}|\boldsymbol{\lambda}) = \prod_{i=1}^{k} N(y_i|\theta_i,\sigma^2) \prod_{i=1}^{k} N(\theta_i|\mu,\tau^2) I_{S_i}(\theta_i) \tag{8}$$

where again I denotes an indicator function, N denotes the normal distribution, and $\boldsymbol{\lambda}=(\sigma^2,\tau^2,\mu)$. In fact, we go one step further by assuming that $\boldsymbol{\lambda}$ is unknown as well and assign hyperprior distributions to its components. A sufficiently general class of hyperpriors is offered by standard conjugate forms, which have the decided advantage of allowing closed forms for the $\boldsymbol{\lambda}$ complete conditional distributions, though as we remark in Section 5, this is not necessary for the implementation of the method. Thus we assume $\sigma^2 \sim IG(a_1,b_1)$, $\tau^2 \sim IG(a_2,b_2)$, and $\mu \sim N(c,d^2)$ where IG denotes the inverse (reciprocal) gamma distribution. (The IG distribution gets its name from the fact that if X is distributed as a gamma random variable $G(a,b)$, that is, having pdf $f(x)=x^{a-1}\exp(-x/b)I_{(0,\infty)}(x)/[\Gamma(a)b^a]$, then $Y=1/X$ is distributed $IG(a,b)$ with pdf $f(y)=\exp(-1/by)I_{(0,\infty)}(y)/[\Gamma(a)b^a y^{a+1}]$.)

We now give the collection of complete conditional distributions necessary for the implementation of the Gibbs sampler. Applying Equation (4) and the standard conjugate normal theory (see, for example, Berger [1]), we have

$$p(\theta_i|\mathbf{y},\boldsymbol{\lambda},\theta_{j\neq i}) \propto \begin{cases} N\left(\theta_i \,\middle|\, \dfrac{\sigma^2\mu+\tau^2 y_i}{\sigma^2+\tau^2}, \dfrac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right) I_{(\theta_{i-1},\theta_{i+1})}(\theta_i), & i=1,\ldots,s-1 \\ N\left(\theta_i \,\middle|\, \dfrac{\sigma^2\mu+\tau^2 y_i}{\sigma^2+\tau^2}, \dfrac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right) I_{(\max(\theta_{s-1},\theta_{s+1}),B)}(\theta_i), & i=s \\ N\left(\theta_i \,\middle|\, \dfrac{\sigma^2\mu+\tau^2 y_i}{\sigma^2+\tau^2}, \dfrac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right) I_{(\theta_{i+1},\theta_{i-1})}(\theta_i), & i=s+1,\ldots,k \end{cases} \tag{9}$$

where $\theta_0 = \theta_{k+1} \equiv 0$, and s indexes the age corresponding to the largest aging factor. The remaining complete conditionals (those for the λ components) also emerge from standard hierarchical Bayes calculations as

$$p(\sigma^2 \mid \mathbf{y},\boldsymbol{\theta},\mu,\tau^2) = p(\sigma^2 \mid \mathbf{y},\boldsymbol{\theta}) = IG\left(\sigma^2 \,\middle|\, a_1 + \frac{k}{2}, \left\{b_1^{-1} + \frac{1}{2}\sum_{i=1}^{k}(y_i - \theta_i)^2\right\}^{-1}\right),$$

$$p(\tau^2 \mid \mathbf{y},\boldsymbol{\theta},\mu,\sigma^2) = p(\tau^2 \mid \boldsymbol{\theta},\mu) = IG\left(\tau^2 \,\middle|\, a_2 + \frac{k}{2}, \left\{b_2^{-1} + \frac{1}{2}\sum_{i=1}^{k}(\theta_i - \mu)^2\right\}^{-1}\right),$$

and

$$p(\mu \mid \mathbf{y},\boldsymbol{\theta},\sigma^2,\tau^2) = p(\mu \mid \boldsymbol{\theta},\tau^2) = N\left(\mu \,\middle|\, \frac{\tau^2 c + kd^2\bar{\theta}}{\tau^2 + kd^2}, \frac{\tau^2 d^2}{\tau^2 + kd^2}\right), \tag{10}$$

where $\bar{\theta} = \sum_{i=1}^{k} \theta_i/k$.

As a numerical illustration, we analyze the raw aging factors y_i presented in Table 1, which were computed from average claim cost data presented by Hutchings and Ullman [19]. These observed $\hat{C}$ values were used with Equation (7) to obtain the raw aging factors in the table. The cost experience is based on 676,000 Blue Cross and Blue Shield of Greater New York contracts for the calendar year 1978. Our factors are unisex, the result of simply averaging the sex-specific monthly per-person costs. Our smoothed aging factors are thus appropriate for use with a population that has roughly the same number of males as females. Because these are private-carrier cost data, there is a sharp drop in the $\hat{C}$ values at age 65, due to the impact of Medicare funds offsetting the total cost. As a result of this discontinuity, no aging factor was calculated centered at age 65. Notice that the initial estimates y_i are very rough, violating our order restrictions several times and even becoming negative for one older age (90) where our information is less reliable.

In order to fit the above model, we need to complete the specification of the hyperpriors by choosing values for the constants s, B, a_1, b_1, a_2, b_2, c, and d. First, we took $s = 60$ as the age corresponding to the maximum aging factor. This selection is admittedly somewhat data-based, as the observed aging factors y_i in Table 1 seem to offer overwhelming evidence in

TABLE 1

OBSERVED HEALTH CLAIM COST AGING FACTORS

i	age_i	y_i	i	age_i	y_i
1	17.50	0.0047	8	70.00	0.0454
2	30.00	0.0341	9	75.00	0.0390
3	38.75	0.0403	10	80.00	0.0244
4	45.00	0.0503	11	85.00	0.0240
5	50.00	0.0467	12	90.00	−0.0021
6	55.00	0.0184	13	95.00	0.0001
7	60.00	0.0713			

support of this s value. (We remark to practitioners uncomfortable with this selection that s could be regarded as another component of λ and included in the sampling algorithm at little increase in complexity.) We then took $B = 0.15$, believing that no aging factor should exceed this value.

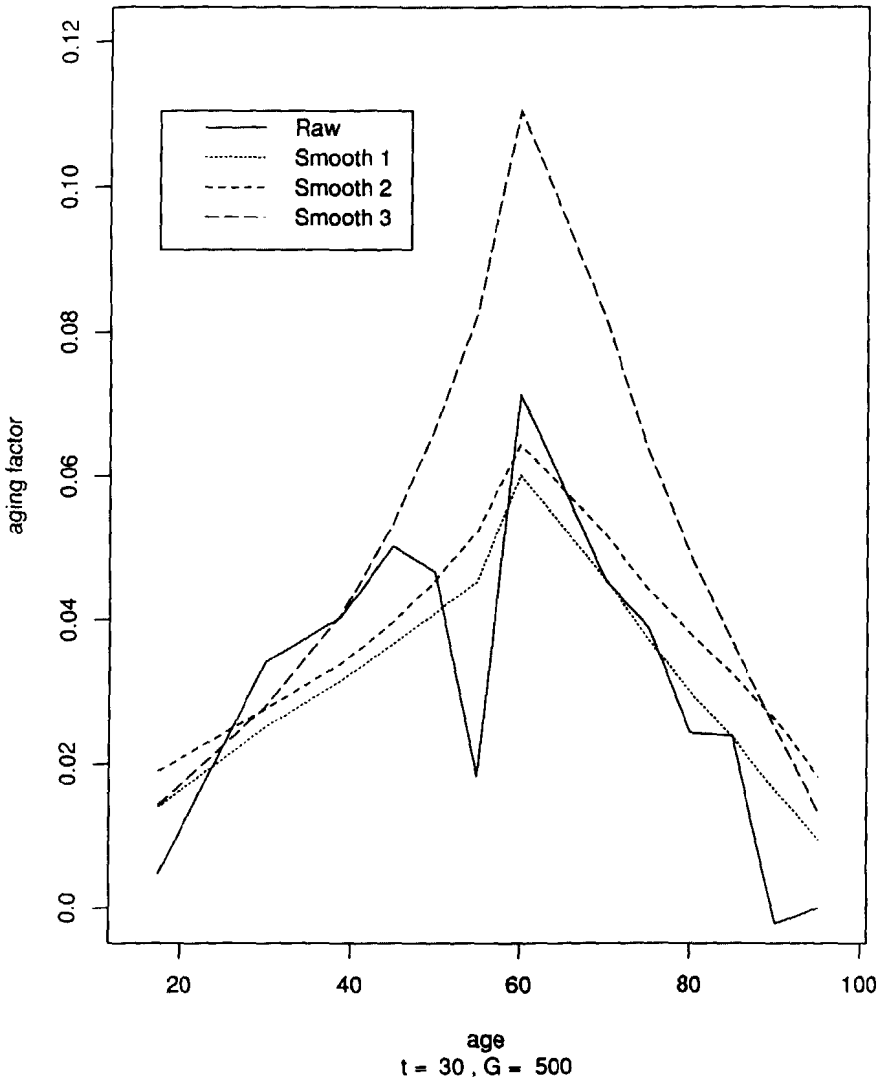
Next recall that c and d^2 are the mean and variance of μ, which in turn is the prior mean of the *collection* of true aging factors (that is, ignoring their ordering for the moment). As such, we chose the hyperprior mean $c = 0.035$, but also chose the hyperprior standard deviation $d = 0.05$, a very large value (relative to the mean) intended to allow the data to have a greater impact on our results.

Finally, recall that σ^2 is the variance of the *observed* aging factors y_i, while τ^2 is the prior variance of the *true* aging factors θ_i (again, ignoring the constraints). For the inverse gamma priors on these two parameters, we computed that setting the prior mean and prior standard deviation equal to $(0.1)^2$ (again, a rather vague specification) implies the values $a = 3.0$, $b = 50$. Setting these values equal to the smaller value $(0.02)^2$ (a more restrictive hyperprior), we instead obtain $a = 3.0$, $b = 1250$. A limited sensitivity analysis suggested that a change in the b's had the greatest impact on our graduated $\boldsymbol{\theta}$ values, and so in Figure 1 we compare the three graduations ($b_1 = 1250$, $b_2 = 1250$), ($b_1 = 50$, $b_2 = 1250$), and ($b_1 = 50$, $b_2 = 50$). These hyperprior specifications reflect an increasing amount of variability in our prior beliefs and translate into increasingly variable graduated rates. We denote the three graduations by Smooth 1, Smooth 2, and Smooth 3, respectively.

We implemented the Gibbs sampler by using the complete conditionals given in (9) and (10) above, and generated values in the following order: $\theta_1, \theta_2, \ldots, \theta_k, \sigma^2, \tau^2, \mu$. We selected $\theta_{i(0)}^{(g)} = c = 0.035$ for $i = 1, \ldots, k$, $\sigma_{(0)}^{(g)} = \tau_{(0)}^{(g)} = 0.1$, and $\mu_{(0)}^{(g)} = c = 0.035$ as reasonable starting values for the algorithm, and used these starting values for each of our G independent parallel

FIGURE 1

ORIGINAL AND GRADUATED HEALTH INSURANCE AGING FACTORS
$s=60$, $B=0.15$, $a_1=a_2=3$, $c=0.035$, $d=0.05$



sampling chains. Initially we took $G=20$, but at iteration 16 we increased G to 100 by generating five independent $\theta_{i(16)}^{(g)}$ values for each of the 20 $\mu_{(15)}^{(g)}$ values on hand; we used the same trick to increase G to 500 at iteration 21. Using Equation (6) for each $i=1, \ldots, k$, the largest value of $\hat{se}(\theta_i)$ obtained was approximately 0.00025. Thus, using the $\pm 2\hat{se}$ error range suggested by the asymptotic normality of $\bar{\theta}_i$, it seems that $G=500$ is large enough to ensure three decimal place accuracy in this example. Regarding the choice of a sufficiently large t, convergence of Gibbs sampler algorithms is typically judged by monitoring empirical quantiles or entire density estimates from iterations that are far enough apart (say, five iterations) to be thought of as independent. In this example, stabilization of the empirical $\theta_i^{(g)}$ quantiles indicated convergence of the algorithm by iteration $t=25$.

We computed our final θ_i estimates using Equation (5) for the three hyperprior specifications, and plotted these along with the raw y_i aging factors in Figure 1. Notice that Smooth 2 (corresponding to a strict σ^2 specification but a more vague τ^2 specification) gives graduated rates that are almost exactly parallel to those of the more restrictive Smooth 1, but roughly 0.005 larger. One of these graduations might be preferred to the other in the interest of conservatism, though which one would depend on the use of the final estimates. For example, if our client population comprised primarily older aged persons and our goal was to project backwards for the younger ages, choosing the larger values from Smooth 2 might tend to understate these actual costs. Allowing even more variability by decreasing b_1 as well (Smooth 3), we get even more sharply peaked rates than in the previous two graduations and a much larger maximum rate as well. Of course we do not wish to claim that one of the above graduations is "correct," but simply to give the reader an impression of the range of results that can be obtained using our simple prior specification.

Example 2: Graduation of Human Mortality Rates

Suppose we are interested in human mortality rates between ages x and $x+k$, where $x \geq 30$. Data available from mortality studies of a group of independent lives typically include values for d_i, the number of deaths observed in the unit age intervals $[x+i-1, x+i]$, and e_i, the *exposure* for this age interval, or the total number of person-years the lives were under observation in the interval. By assuming that the force of mortality in the unit

age interval $[x+i-1, x+i]$ is equal to the constant θ_i, Broffitt [7] obtains a simple expression for the likelihood function,

$$f(\mathbf{y} \mid \boldsymbol{\theta}) \propto \prod_{i=1}^{k} \theta_i^{d_i} \exp(-e_i\theta_i), \tag{11}$$

where $\mathbf{y}=(d_1, \ldots, d_k, e_1, \ldots, e_k)$. (Note that if we view the exposures e_i as fixed quantities, this likelihood is equivalent to that obtained by assuming that $d_i|\theta_i$ has a Poisson distribution with mean $e_i\theta_i$.) An initial estimate of θ_i is provided by $r_i=d_i/e_i$, the unrestricted maximum likelihood estimate of θ_i, more commonly known as the "raw" (ungraduated) mortality rate. We wish to produce a graduated sequence of θ_i's that conform to the *increasing* condition

$$\boldsymbol{\theta} \in S^{INC} = \{\boldsymbol{\theta} : 0 < \theta_1 < \cdots < \theta_k < B\} \tag{12}$$

or perhaps to the more restrictive *increasing convex* condition

$$\boldsymbol{\theta} \in S^{INCCON} = \{\boldsymbol{\theta} : \theta_1 > 0, \theta_k < B, 0 < \theta_2-\theta_1 < \cdots < \theta_i-\theta_{k-1}\} \tag{13}$$

Broffitt [7] observed that a natural conjugate prior family for the likelihood in (11) is offered by the gamma distribution (see pdf given in Example 1). Unfortunately, the intractability of the posterior after imposing the order constraints (12) or (13) forces him to choose a reparametrization of the model that depends on the constraint set chosen, and subsequently to impose the gamma prior structure on the resulting new set of parameters. Using the Gibbs sampler, such constraint-dependent reconstruction of the model is unnecessary. We simply assume that the θ_i's are an independent identically distributed sample from a $G(\alpha,\beta)$ distribution prior to imposing the constraints, so that the Bayesian model (1) becomes

$$p(\mathbf{y},\boldsymbol{\theta} \mid \boldsymbol{\lambda}) \propto \prod_{i=1}^{k} \theta_i^{d_i} \exp(-e_i\theta_i) \prod_{i=1}^{k} \theta_i^{\alpha-1} \exp(-\theta_i/\beta) I_{S_i}(\theta_i). \tag{14}$$

We complete the model specification by assuming that α is a known constant but that β has an $IG(a,b)$ distribution, so that $\boldsymbol{\lambda}=\beta$ in Equation (14). Hence in order to implement the Gibbs sampler, we need to find the complete conditionals for the θ_i and β. The former are given by

$$p(\theta_i \mid \mathbf{y},\beta,\theta_{j\neq i}) \propto G(\theta_i \mid \alpha^*,\beta^*) I_{(\theta_{i-1},\theta_{i+1})}(\theta_i), \qquad i = 1, \ldots, k,$$

for rates which satisfy the increasing condition, and

$$p(\theta_i \mid \mathbf{y},\boldsymbol{\lambda},\boldsymbol{\theta}_{j\neq i}) \propto \begin{cases} G(\theta_i \mid \alpha^*,\beta^*)I_{(max(0,2\theta_2-\theta_3),\theta_2)}(\theta_i), & i = 1, \\ G(\theta_i \mid \alpha^*,\beta^*)I_{(max(\theta_1,2\theta_3-\theta_4),\frac{\theta_1+\theta_3}{2})}(\theta_i), & i = 2, \\ G(\theta_i \mid \alpha^*,\beta^*)I_{(max(2\theta_{i-1}-\theta_{i-2},2\theta_{i+1}-\theta_{i+2}),\frac{\theta_{i-1}+\theta_{i+1}}{2})}(\theta_i), & \\ & i = 3, \ldots, k-2, \\ G(\theta_i \mid \alpha^*,\beta^*)I_{(2\theta_{k-2}-\theta_{k-3},\frac{\theta_{k-2}+\theta_k}{2})}(\theta_i), & i = k-1, \\ G(\theta_i \mid \alpha^*,\beta^*)I_{(2\theta_{k-1}-\theta_{k-2},B)}(\theta_i), & i = k, \end{cases}$$

for rates satisfying the increasing convex condition. In both cases $\alpha^* = \alpha + d_i$, $\beta^* = (\beta^{-1} + e_i)^{-1}$, $\theta_0 \equiv 0$, and $\theta_{k+1} \equiv B$. The complete conditional for β is readily available as

$$p(\beta \mid \mathbf{y},\boldsymbol{\theta}) = IG\left(a + k\alpha, \left\{b^{-1} + \sum_{i=1}^{k} \theta_i\right\}^{-1}\right).$$

TABLE 2

RAW MORTALITY DATA

i	age$_i$	d_i	e_i	r_i	i	age$_i$	d_i	e_i	r_i
1	35	3	1771.5	0.0016935	16	50	4	1516.0	0.0026385
2	36	1	2126.5	0.0004703	17	51	7	1371.5	0.0051039
3	37	3	2743.5	0.0010935	18	52	4	1343.0	0.0029784
4	38	2	2766.0	0.0007231	19	53	4	1304.0	0.0030675
5	39	2	2463.0	0.0008120	20	54	11	1232.5	0.0089249
6	40	4	2368.0	0.0016892	21	55	11	1204.5	0.0091324
7	41	4	2310.0	0.0017316	22	56	13	1113.5	0.0116749
8	42	7	2306.5	0.0030349	23	57	12	1048.0	0.0114504
9	43	5	2059.5	0.0024278	24	58	12	1155.0	0.0103896
10	44	2	1917.0	0.0010433	25	59	19	1018.5	0.0186549
11	45	8	1931.0	0.0041429	26	60	12	945.0	0.0126984
12	46	13	1746.5	0.0074435	27	61	16	853.0	0.0187573
13	47	8	1580.0	0.0050633	28	62	12	750.0	0.0160000
14	48	2	1580.0	0.0012658	29	63	6	693.0	0.0086580
15	49	7	1467.5	0.0047700	30	64	10	594.0	0.0168350

Table 2 gives a dataset of male ultimate (duration $\geq$ 16) experience originally presented and analyzed by Broffitt [7]. Before graduating these rates, we must specify values for the constants α, a and b. In this regard, the paper by Gaver and O'Muircheartaigh [12] is helpful. These authors recommend a method of moments approach to determining α and β. In other words, we

equate the first two sample moments of the raw rates, $\bar{r} = \Sigma_{i=1}^{k} r_i/k$ and $s_r^2 = \Sigma_{i=1}^{k} (r_i - \bar{r})^2/(k-1)$, to the corresponding population moments in the (unconstrained) marginal family $m(r_i|\alpha,\beta)$,

$$E(r_i) = \frac{1}{e_i}E(d_i) = \frac{1}{e_i}E[E(d_i|\theta_i)] = \frac{1}{e_i}E(e_i\theta_i) = E(\theta_i) = \alpha\beta,$$

and

$$\begin{aligned} Var(r_i) &= \frac{1}{e_i^2}Var(d_i) = \frac{1}{e_i^2}\{Var[E(d_i|\theta_i)] + E[Var(d_i|\theta_i)]\} \\ &= \frac{1}{e_i^2}[Var(e_i\theta_i) + E(e_i\theta_i)\}] = \frac{1}{e_i^2}[e_i^2(\alpha\beta^2) + e_i(\alpha\beta)] \\ &= \alpha\beta^2 + e_i^{-1}\alpha\beta, \end{aligned}$$

using the Poisson distribution of $d_i|\theta_i$ and the gamma distribution of $\theta_i|\alpha,\beta$. This results in the system of two equations and two unknowns

$$\begin{aligned} \bar{r} &= \alpha\beta \\ s_r^2 &= \alpha\beta^2 + \alpha\beta \sum_{i=1}^{k} e_i^{-1}/k. \end{aligned}$$

Solving, we obtain $\hat{\alpha} = \bar{r}^2/(s_r^2 - \bar{r}\Sigma_{i=1}^{k} e_i^{-1}/k)$ and $\hat{\beta} = \bar{r}/\hat{\alpha}$. Since in our specification we actually have an $IG(a,b)$ hyperprior on β, we might choose $a = 3.0$ and $b = \hat{\alpha}/(2\bar{r})$, corresponding to a hyperprior having mean and standard deviation both equal to $\hat{\beta}$ (again, a rather vague hyperprior specification). Notice that by using the data to help complete the hyperprior specification, our approach now has an empirical Bayes flavor (see, for example, Morris [24]), as opposed to a strict Bayesian interpretation. But for practitioners willing to accept the above data-based formulas, there is no prior specification burden at all.

For this Gibbs sampling run, the starting values $\theta_{i(0)}^{(g)} = 0.0000222i^2$ (suggested by a quadratic regression of the r_i's on i) and $\beta_{(0)}^{(g)} = \bar{r}/\hat{\alpha} = 0.00435$ were used for each chain, $g = 1, \ldots, G$. As in Example 1, empirical quantiles of the $\theta_i^{(g)}$ were checked every five iterations, and we again increased G from 20 to 100 at iteration 16 and from 100 to 500 at iteration 21. Figure 2 shows the raw rates r_i and graduated rates $\bar{\theta}_i$ obtained from the $G = 500$ independent replications of the Gibbs sampler at iteration $t = 25$. In this example, this G value was large enough to produce a maximal standard error in Equation (6) of roughly 0.0000045, which in turns suggests at least four

good decimal places in our rate estimates. The rates given in the figure as Bayes 2 are those obtained using the prior specification technique described in the previous paragraph (for this data, $\hat{\alpha}=1.49$, $a=3.0$ and $b=115$), an upper bound B of 0.025, and the increasing constraint set. In Bayes 1 we reduce the impact of the prior by setting $\alpha=0$ and $b=0.0005$ (recall that this implies $\alpha^*=d_i$ and an extremely vague prior on β). The figure shows that the Bayes 1 rates are indeed a bit more variable, being smaller than those for Bayes 2 for younger ages but larger at the older ages. The Bayes 3 rates are obtained by choosing the same α, a and b values as in Bayes 2, taking $B=0.020$, and imposing the increasing convex constraint set. The results are clearly much smoother but exhibit less fidelity to the original rate sequence.

Actuaries less familiar with Bayesian graduation methods might well wonder how they compare with more familiar approaches. Klugman [21], [22] gives a formal Bayesian derivation of Whittaker's method, and one with a relatively light prior specification burden. Figure 2 offers a less formal comparison in the context of the Table 2 data by showing the results of two Whittaker Type B graduations along the Bayesian results. Both Whittaker graduations use the standardized exposure values $e_i/\bar{e}$ as weights, and take a third-degree polynomial as the standard of smoothness. The rates plotted as Whittaker 1 use the smoothing constant $h=500$, which apparently is large enough to produce rates that are increasing, but not necessarily convex. The Whittaker 2 rates use $h=5000$, which does result in a convex sequence; still larger values of h did not significantly alter the graduated results. The Whittaker results are fairly similar to the Bayes results, though the Whittaker rates tend to be influenced more by the unusually low rate at age 63. Of course, the two convex sequences (Whittaker 2 and Bayes 3) could be made even more similar, either by altering the Whittaker weights or modifying the Bayes prior. We remark that a very large value of the smoothing constant must be employed to ensure Whittaker rates that are convex, and that the choice of this constant basically comes down to guesswork. The Bayesian approach formalizes this uncertainty into prior distributions and, via our Monte Carlo approach, enables direct imposition of the desired shape constraints.

Finally, Figure 3 presents histograms of the raw rates and all three Bayes rate sequences plotted in Figure 2. Except for a bit too much mass in the right tail, the histogram of the raw rates seems to validate our gamma distributional assumption. The fact that $\alpha=0$ in Bayes 1 is borne out by its heavier left tail as compared to Bayes 2. Although $\alpha\neq 0$ in Bayes 3, its

FIGURE 2

ORIGINAL AND GRADUATED FORCES OF MORTALITY

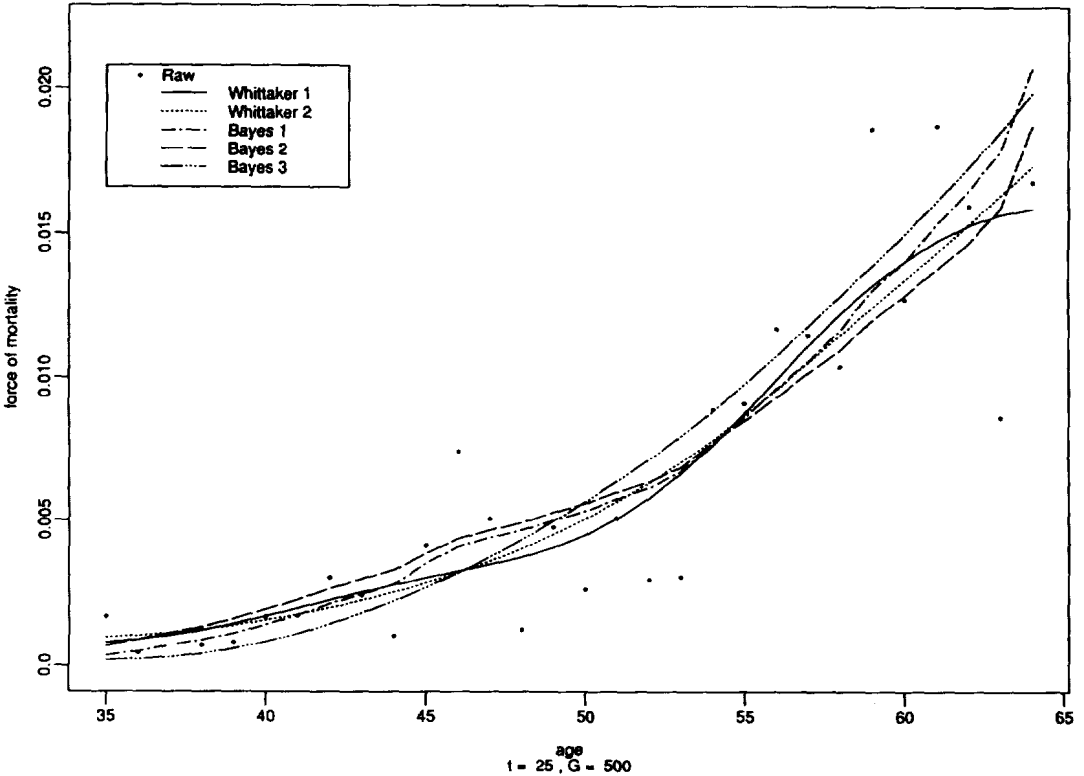
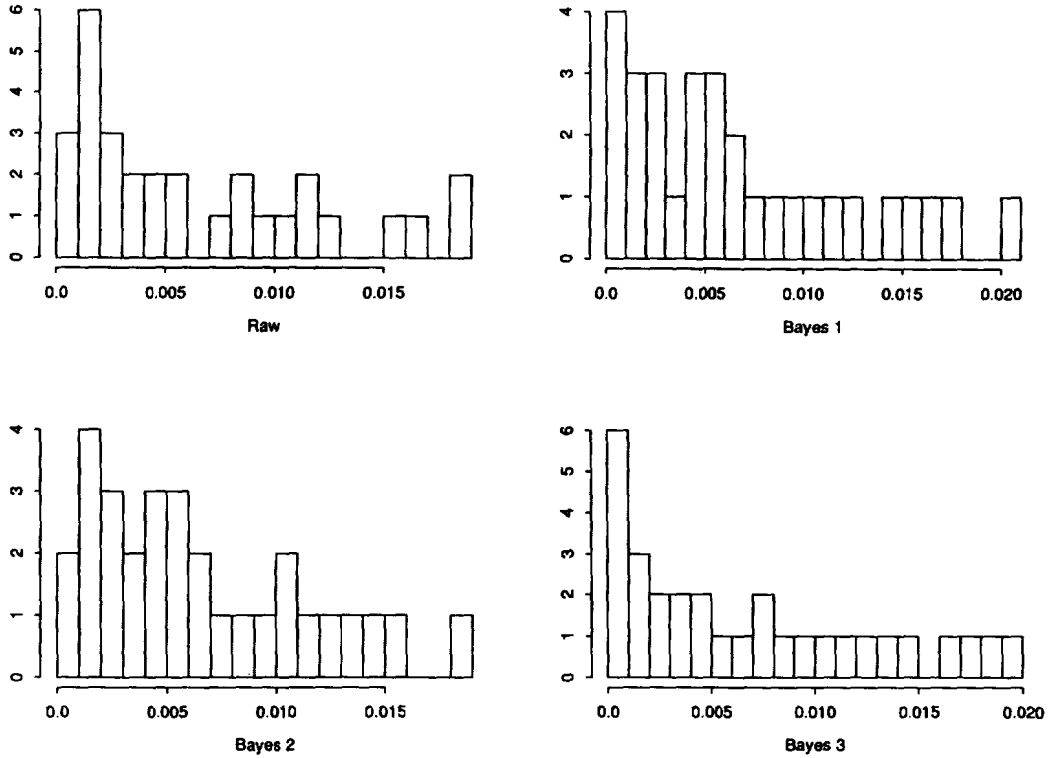


FIGURE 3

HISTOGRAMS OF ORIGINAL AND GRADUATED FORCES OF MORTALITY



histogram also exhibits a heavy left tail. This suggests that the very large number of constraints placed on the θ_i's in this case may be overwhelming the shape of the original (unconstrained) distribution.

5. SUMMARY AND DISCUSSION

In this paper we have shown how a simple Monte Carlo integration procedure known as the Gibbs sampler enables routine implementation of Bayesian graduation without the need for artificial distributional assumptions or extensive numerical analytic expertise on the user's part. We also purposely avoided models requiring large amounts of prior elicitation, in response to the many practitioners who object to Bayesian graduation on the grounds that it lacks objectivity and practicality. As an unfortunate side effect, however, our models may seem an oversimplification to some. Indeed, satisfying a large constraint set like that described in (13) with only three free parameters may lead to unsatisfactory results. But this in no way diminishes the value of the sampling-based strategy we have described; many generalizations are possible. First, the (unconstrained) identical distribution assumption on the θ_i may be dropped. For instance, in Equation (14) we might replace α and β by α_i and β_i, as Broffitt [7] does. This increases the prior elicitation burden k-fold, but not the computational burden because closed forms for the complete conditionals are still available. However, if even this does not give acceptable results, we can drop the conjugate prior assumption as well. This does increase the computational burden, because now the necessary Gibbs samples $\theta_i^{(g)}$ must be generated using some sort of rejection algorithm (see, for example, Devroye [11]).

Other generalizations are possible using the Gibbs sampling approach. One could graduate male and female rates simultaneously, perhaps with an unknown crossover point in the rate patterns (see Carlin, Gelfand and Smith [8]). Covariates (such as health status, sex, and the like) could be included in a parametric graduation version (see London [23]) of our Bayesian model. The implementation of formal Bayesian model choice techniques for comparing several competing graduations is also possible using the Gibbs sampler (Carlin and Polson [9]). In short, the added flexibility offered by Monte Carlo methods suggests a rosy future in the applications world.

ACKNOWLEDGMENTS

The author thanks Caroline Carlin for suggesting the dataset and providing the background for Example 1, Dr. James Broffitt for helpful discussions,

and a team of anonymous referees for several comments that led to substantial improvements in the presentation.

REFERENCES

1. BERGER, J.O. *Statistical Decision Theory and Bayesian Analysis*. 2d ed. New York: Springer-Verlag, 1985.
2. BESAG, J. "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society, Series B*, 36 (1974): 192–236.
3. BOWERS, N.L., JR., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathematics.* Itasca, Ill.: Society of Actuaries, 1986.
4. BROFFITT, J.D. "Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates," *TSA* XXXVI (1984): 77–122.
5. BROFFITT, J.D. "A Bayes Estimator for Ordered Parameters and Isotonic Bayesian Graduation," *Scandinavian Actuarial Journal* (1984): 231–47.
6. BROFFITT, J.D. "Isotonic Bayesian Graduation with an Additive Prior," *Advances in the Statistical Sciences.* Vol. 6, *Actuarial Science*. Edited by I.B. MacNeill and G.J. Umphrey, 19–40. Boston: D. Reidel Publishing Co., 1986.
7. BROFFITT, J.D. "Increasing and Increasing Convex Bayesian Graduation" (with discussion), *TSA* XL (1988): 115–48.
8. CARLIN, B.P., GELFAND, A.E., AND SMITH, A.F.M. "Hierarchical Bayesian Analysis of Change Point Problems," *Journal of the Royal Statistical Society, Series C,* 41 (1992): 389–405.
9. CARLIN, B.P., AND POLSON, N.G. "Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler," *Canadian Journal of Statistics* 19 (1991): 399–405.
10. CHAN, L.K., AND PANJER, H.H. "A Statistical Approach to Graduation by Mathematical Formula," *Insurance: Mathematics and Economics* 2 (1983): 33–47.
11. DEVROYE, L. *Non-Uniform Random Variate Generation.* New York: Springer-Verlag, 1986.
12. GAVER, D.P., AND O'MUIRCHEARTAIGH, I.G. "Robust Empirical Bayes Analyses of Event Rates," *Technometrics* 29 (1987): 1–15.
13. GELFAND, A.E., AND SMITH, A.F.M. "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85 (1990): 398–409.
14. GELFAND, A.E., AND SMITH, A.F.M. "Gibbs Sampling for Marginal Posterior Expectations," *Communications in Statistics, Part A—Theory and Methods* 20 (1991): 1747–66.
15. GELFAND, A.E., SMITH, A.F.M., AND LEE, T-M. "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling," *Journal of the American Statistical Association* 87 (1992): 523–32.
16. GEMAN, S., AND GEMAN, D. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984): 721–41.

17. HICKMAN, J.C., AND MILLER, R.B. "Notes on Bayesian Graduation," *TSA* XXIX (1977): 1–21.
18. HOEM, J.M. "On the Statistical Theory of Analytic Graduation," *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability,* 569–600. Berkeley, Calif: University of California Press, 1970.
19. HUTCHINGS, P.L., AND ULLMAN, R.E. "Prepaid Hospital Care Age/Sex and Hospital Continuation Study," *TSA* XXXV (1983): 623–55.
20. KIMELDORF, G.S., AND JONES, D.A. "Bayesian Graduation," *TSA* XIX (1967): 66–112.
21. KLUGMAN, S. *Bayesian Statistics in Actuarial Science with Emphasis on Credibility.* Boston, Mass.: Kluwer Academic Publishers, 1992.
22. KLUGMAN, S. "Hierarchical Bayesian Whittaker Graduation," *ARCH* 1992. 1: 161–68.
23. LONDON, R.L. *Graduation: The Revision of Estimates.* Winsted, Conn.: ACTEX Publications, 1985.
24. MORRIS, C.N. "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association* 78 (1983): 47–59.
25. TANNER, M.A., AND WONG, W.H. "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association* 82 (1987): 528–50.