© Board of the Foundation of the Scandinavian Journal of Statistics 1993. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF and 238 Main Street, Suite 501, Cambridge, MA 02142, USA. Vol. 20: 147–156, 1993

Conjugate Likelihood Distributions

E. I. GEORGE

Department of MSIS, University of Texas at Austin

U. E. MAKOV

Department of Statistics, University of Haifa

A. F. M. SMITH

Department of Mathematics, Imperial College, London

ABSTRACT. Families of probability distributions which arise naturally as parameter likelihoods in conjugate prior distributions for exponential families are identified, described and their relevance to computational issues in Bayes hierarchical models noted.

Key words: Bayes inference, conjugate prior, exponential family, Gibbs sampling, hierarchical model, hyperparameter, likelihood, log-concavity, random variate generation

1. Introduction

Our main purpose in this paper is to describe some families of probability distributions which arise naturally in two (formally related) areas of application of Bayesian statistical methodology, and to illustrate the relevance of the technical machinery we develop. A by-product of the latter is some novel theory relating to exponential families.

The first area is exemplified by the problem of Bayesian inference for the parameters of gamma and beta distributions. The second area relates to inference in Bayesian hierarchical models constructed from exponential families and conjugate priors.

These motivating applications are detailed more fully in section 2. In section 3, we identify a general class of distributions, central to both applications, and provide a formal development. The results are illustrated for the problem of inference for gamma and beta parameters. The relevance of the results to computational problems in Bayesian hierarchical models is illustrated in section 4.

2. Motivating problems

2.1. Bayesian inference for gamma and beta parameters

If we observe an i.i.d. sample $y_1, \ldots, y_p$ from a gamma distribution,

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \qquad \alpha > 0, \quad \beta > 0, \quad y > 0, \tag{1}$$

the induced likelihood family can be written as

$$L^G(\alpha, \beta \mid y_1, \ldots, y_p) \propto \left[\frac{\beta^\alpha}{\Gamma(\alpha)}\right]^p \left[\prod_{i=1}^p y_i\right]^{\alpha-1} \exp\left[-\beta \sum_{i=1}^p y_i\right], \tag{2}$$

where $\alpha > 0, \beta > 0, y_i > 0$, $i = 1, \ldots, p$. Similarly, if we observe an i.i.d. sample $y_1, \ldots, y_p$ from a beta distribution,

$$f(y) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}, \qquad \alpha > 0, \quad \beta > 0, \quad 0 < \theta < 1, \tag{3}$$

the induced likelihood family can be written as

$$L^B(\alpha, \beta \mid y_1, \dots, y_p) \propto \left[\frac{1}{B(\alpha, \beta)}\right]^p \left[\prod_{i=1}^p y_i\right]^{\alpha-1} \left[\prod_{i=1}^p (1-y_i)\right]^{\beta-1}, \tag{4}$$

$\alpha > 0, \beta > 0, 0 < y_i < 1, i = 1, \dots, p$. Both L^G and L^B are here considered as functions of the parameters α and β given fixed $y_1, \dots, y_p$. In both cases, posterior densities for the parameters have mathematical forms proprotional to (2) or (4) multiplied by prior densities. Properties of (2) and (4) then determine many properties of the required Bayesian posteriors.

2.2. *Bayesian hierarchical models*

Hierarchical modeling and analysis is directed towards simultaneous inference for several parameters of the same type. For example in agricultural trials involving many varieties, such parameters would correspond to the varietal means; in actuarial risk assessment such parameters would correspond to the mean claim rates or losses for a class of individual policy holders. Typically, there are relatively many trials or individuals, but relatively little data on each separately. The statistical analysis is therefore aimed at combining the information from the various sources of data (varieties, policy holders) exploiting the "similarity" relationship among the parameters. From a non-Bayesian statistical perspective, well-known formalisms based on the idea of "pulling in" data from related sources include Stein's shrinkage estimation rule (Stein, 1955) and parametric empirical Bayes theory (see, for example, Morris, 1983). We shall focus on the following version of a Bayesian formalism for modeling and analysis (see Deely & Lindley, 1981; Kass & Steffey, 1989).

Suppose that data $x_1, \dots, x_p$ are assumed conditionally independent, given scalar parameters $\theta_1, \dots, \theta_p$, with joint density

$$f(x_1, \dots, x_p \mid \theta_1, \dots, \theta_p) = \prod_{i=1}^p f(x_i \mid \theta_i), \tag{5}$$

with each of the product terms from the same exponential family. Suppose further that the parameters $\theta_1, \dots, \theta_p$ are conditionally independent, given "hyperparameters" α, β, with joint density

$$\pi(\theta_1, \dots, \theta_p \mid \alpha, \beta) = \prod_{i=1}^p \pi(\theta_i \mid \alpha, \beta), \tag{6}$$

with each of the factors from the conjugate family (parametrized by α, β) corresponding to the exponential family. The hierarchical model is then completed by specifying a prior density for α, β.

Examples of (5) and (6) include: for $i = 1, \dots, p$, x_i having a Poisson (θ_i) distribution and θ_i a Gamma (α, β) distribution (or, similarly, x_i exponential, given θ_i); for $i = 1, \dots, p$, x_i having a Bernoulli (θ_i) distribution and θ_i a beta (α, β) distribution (or, similarly, x_i binomial, negative-binomial or geometric, given θ_i); for $i = 1, \dots, p$, x_i having a normal $(\theta_i, 1)$ distribution and θ_i a normal (α, β) distribution.

Marginal posterior inferences based on (5), (6) and a prior specification for α, β may be required both for the θ_i ("individual" inference) and for α, β ("population" inference). In some cases, however, whatever prior form is adopted for α, β the integrations required to obtain marginals from the joint posterior density for $\theta_1, \dots, \theta_p, \alpha, \beta$ cannot be performed analytically and some form of numerical procedure is required. A simulation-based Bayesian computational method that is currently attracting considerable attention is the so-called Gibbs sampler algorithm (Geman & Geman, 1984; Gelfand & Smith, 1990). In the context of the above model, one version of this algorithm proceeds as follows.

From the form of the joint posterior density for $\theta_1, \ldots, \theta_p, \alpha, \beta$ identify the forms of the conditional densities for each of $\theta_1, \ldots, \theta_p, (\alpha, \beta)$, given $x = (x_1, \ldots, x_p)$ and given all the other unknown parameters. With $\pi(\cdot \mid \cdot)$ generically denoting posterior forms of interest, and θ_{-i} denoting $\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p$, we therefore require

$$\pi(\theta_i \mid x, \theta_{-i}, \alpha, \beta) = \pi(\theta_i \mid x_i, \alpha, \beta) \qquad i = 1, \ldots, p \tag{7}$$

and

$$\pi(\alpha, \beta \mid x, \theta_1, \ldots, \theta_p) = \pi(\alpha, \beta \mid \theta_1, \ldots, \theta_p), \tag{8}$$

the reduced forms being simple consequences of the conditional independence structure underlying (5) and (6). A variant is to replace (8) by the two marginal conditional forms $\pi(\alpha \mid \theta_1, \ldots, \theta_p, \beta)$ and $\pi(\beta \mid \theta_1, \ldots, \theta_p, \alpha)$. In either case, the algorithm proceeds by iterated simulation from the conditional forms, always conditioning on the most recent realization of the given parameters. Ergodic averages of appropriate functions of simulated values then provide consistent estimates of features of interest of posterior distributions.

We note that the above scheme motivates the study of the form $\pi(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$, (which, in turn, characterizes the marginal conditional forms for α and β). From (6) we see that this is proportional to the conjugate likelihood form defined by i.i.d. $\theta_1, \ldots, \theta_p$ from the conjugate distribution (parametrized by α, β) to the exponential family defining (5) multiplied by a joint prior form for α, β. If the exponential family is the Poisson, the conjugate likelihood form is precisely $L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$ as in (2), (with y_i replaced by θ_i); if the exponential family is Bernoulli related, the conjugate likelihood form is precisely $L^B(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$ as in (4), (with y_i replaced by θ_i).

This completes our motivation of the study of conjugate likelihood families and establishes the link between the two types of motivating applications. Our formal development will establish that families such as L^G and L^B have many appealing properties. For example, for $p \geqslant 2$, both families are normalizable and so can be treated as joint probability distributions for the parameters α, β. This allows one to treat L^G and L^B as formal posterior distributions arising from an improper uniform prior for α, β. Another appealing property is that both L^G and L^B are log-concave. This opens the way to highly efficient generation of random variates from these parameter distributions by methods such as the adaptive rejection sampling algorithm of Gilks & Wild (1992).

3. Conjugate likelihood distributions

In this section we present a unified treatment of a general class of distributions, of which L^G and L^B are special cases. These distributions arise as conjugate prior forms for particular exponential families. By considering the class of all such conjugate distributions, links are made between the likelihood distributions and the corresponding exponential family structure. Furthermore, this level of generality allows for the application of our main results to other familiar conjugate priors for exponential families such as the Wishart or the Dirichlet, see DeGroot (1970).

3.1. General development

We begin our treatment of conjugate likelihood distributions by reviewing aspects of the development of conjugate prior distributions for exponential families which appears in Diaconis & Ylvisaker (1979) and Brown (1986). Let v be a fixed σ-finite measure on the Borel sets of R^k. For $\theta \in R^k$, the *natural parameter space* is defined to be

$$N = \left\{\theta \,\middle|\, \int \exp(x \cdot \theta)\, dv(x) < \infty\right\}. \tag{9}$$

For $\Theta \subset N$, a *k-dimensional exponential family* $\{P_\theta \mid \theta \in \Theta\}$ of probability measures through ν is defined to be

$$dP_\theta(x) = \exp[x \cdot \theta - \psi(\theta)]\, d\nu(x), \qquad \text{for } \theta \in \Theta \tag{10}$$

where the *cumulant generating function*

$$\psi(\theta) \equiv \ln \int \exp(x \cdot \theta)\, d\nu(x), \tag{11}$$

serves to normalize the measure. As is well-known, N is a convex set, and ψ is a convex function on N. The *conjugate prior measure* for P_θ is defined to be

$$d\Pi(\theta \mid x_0, n_0) \propto \exp[x_0 \cdot \theta - n_0\psi(\theta)] I_\Theta(\theta)\, d\theta, \qquad x_0 \in R^k, \quad n_0 \geqslant 0 \tag{12}$$

where $I_\Theta(\theta)$ is the indicator function of $\theta \in \Theta$, and $d\theta$ is Lebesgue measure on R^k. The following result, due to Diaconis & Ylvisaker (1979), provides necessary and sufficient conditions for the conjugate prior to be proper. For related results, see Barndorff-Nielsen (1970).

Lemma 3.1
The measure $\Pi(\theta \mid x_0, n_0)$ *is finite, i.e.* $\int_\Theta \exp[x_0 \cdot \theta - n_0\psi(\theta)]\, d\theta < \infty$, *iff*

$$x_0/n_0 \in K^\circ \quad \text{and} \quad n_0 > 0, \tag{13}$$

where K° *is the interior of the convex support of* ν.

Thus, for x_0 and n_0 satisfying (13), Π may be expressed as a proper conjugate prior distribution on R^k, namely

$$d\Pi(\theta \mid x_0, n_0) = \exp[x_0 \cdot \theta - n_0\psi(\theta) - \varphi(x_0, n_0)] I_\Theta(\theta)\, d\theta, \tag{14}$$

where

$$\varphi(x_0, n_0) = \ln \int_\Theta \exp[x_0 \cdot \theta - n_0\psi(\theta)]\, d\theta. \tag{15}$$

The following result is obtained by noting that the conjugate priors (14) also form an exponential family.

Lemma 3.2
$\varphi(x_0, n_0)$ *is convex.*

Proof. Defining

$$\eta_\theta \equiv (\theta, -\psi(\theta)) \in R^{k+1}, \tag{16}$$

$d\Pi(\theta \mid x_0, n_0)$ may be rewritten as

$$d\Pi(\eta_\theta \mid x_0, n_0) = \exp[(x_0, n_0) \cdot \eta_\theta - \varphi(x_0, n_0)]\, d\mu(\eta_\theta), \tag{17}$$

where

$$\varphi(x_0, n_0) = \ln \int \exp[(x_0, n_0) \cdot \eta_\theta]\, d\mu(\eta_\theta), \tag{18}$$

$$d\mu(\eta_\theta) = I_H(\eta_\theta)\, d\theta, \qquad H \equiv \{\eta_\theta \mid \theta \in \Theta\} \subset R^{k+1}, \tag{19}$$

and $d\theta$ is Lebesgue measure on R^k. This expression reveals that $\varphi(x_0, n_0)$ is the cumulant generating function of the exponential family $\Pi(\eta_0 \mid x_0, n_0)$, and so must be convex. □

Now let $\theta_1, \ldots, \theta_p$ be an i.i.d. sample from the conjugate prior distribution $d\Pi(\theta \mid x_0, n_0)$ in (14), so that

$$d\Pi(\theta_1, \ldots, \theta_p \mid x_0, n_0) = \exp\left[x_0 \sum_{i=1}^p \theta_i - n_0 \sum_{i=1}^p \psi(\theta_i) - p\varphi(x_0\, n_0)\right]\left[\prod_{i=1}^p I_\Theta(\theta_i)\, d\theta_i\right]. \tag{20}$$

We define the *conjugate likelihood distribution* induced by this prior to be

$$L(x_0, n_0 \mid \theta_1, \ldots, \theta_p) \propto \exp\left[x_0 \sum_{i=1}^p \theta_i - n_0 \sum_{i=1}^p \psi(\theta_i) - p\varphi(x_0, n_0)\right] I_{K^\circ}(x_0/n_0) I_{(0,\infty)}(n_0). \tag{21}$$

Theorem 3.3

Let $\theta_1, \ldots, \theta_p$ be such that $\theta_i \in \Theta$. Then $L(x_0, n_0 \mid \theta_1, \ldots, \theta_p)$ is log-concave in (x_0, n_0). Furthermore, if Θ is convex and $\psi(\theta)$ is strictly convex, then letting dx_0 and dn_0 be Lebesgue measures on R^k and R respectively,

$$\int_{R^k} L(x_0, n_0 \mid \theta_1, \ldots, \theta_p)\, dx_0 < \infty \quad \textit{for all } p, \tag{22}$$

and

$$\int_{R^{k+1}} L(x_0, n_0 \mid \theta_1, \ldots, \theta_p)\, dx_0\, dn_0 < \infty \quad \textit{iff } p \geqslant 2. \tag{23}$$

Proof. The first assertion follows directly from applying lemma 3.2 to the log of (21). For $p \geqslant 2$, (22) will follow from (23). To establish (22) for $p = 1$, note that

$$\varphi(x_0, n_0) = \ln \int \exp(x_0 \cdot \theta)\, d\mu_n(\theta), \quad \text{where } d\mu_n(\theta) \equiv \exp[-n_0 \psi(\theta)] I_\Theta(\theta)\, d\theta. \tag{24}$$

It then follows from lemma 3.1 that

$$\int_{R^k} \exp[x_0 \cdot \theta - \varphi(x_0, n_0)] I_{K^\circ}(x_0/n_0)\, dx_0 < \infty \tag{25}$$

for $\theta \in \Theta$ since Θ is in the convex support of μ_n for all n. Since the left hand side of (22) is proportional to the left hand side of (25), (22) follows.

To establish (23), note that using the notation (16)–(19), (20) may be rewritten as

$$d\Pi(\eta_{\theta_1}, \ldots, \eta_{0_p}) = \exp\left[(x_0, n_0) \cdot \sum_{i=1}^p \eta_{0_i} - p\varphi(x_0, n_0)\right]\left[\prod_{i=1}^p d\mu(\eta_{0_i})\right]. \tag{26}$$

This in turn induces the measure

$$d\Pi(\eta_+ \mid x_0, n_0) = \exp[(x_0, n_0) \cdot \eta_+ - p\varphi(x_0, n_0)]\, d\mu_+(\eta_+), \tag{27}$$

where

$$\eta_+ \equiv \sum_{i=1}^p \eta_{0_i}, \tag{28}$$

and μ_+ is the measure on R^{k+1} such that

$$\mu_+(A) = \int_{\eta_+ \in A} \prod_{i=1}^{p} d\mu(\eta_{0_i}) \int_{\eta_+ \in A} \prod_{i=1}^{p} I_H(\eta_{0_i})\, d\theta_i. \tag{29}$$

Because (21) may be expressed as

$$L(x_0, n_0 \mid \theta_1, \ldots, \theta_p) \equiv \exp\left[(x_0, n_0) \cdot \eta_+ - p\varphi(x_0, n_0)\right] I_{K^\circ}(x_0/n_0) I_{(0,\infty)}(n_0), \tag{30}$$

by lemma 3.1 it suffices to show that η_+/p is contained in the interior of the convex support of μ_+ which is just the interior of the convex hull of H, denoted $(\text{co}\,(H))^\circ$. Since Θ is convex and $\psi(\theta)$ is strictly convex, it follows that $\eta_+/p \in (\text{co}\,(H))^\circ$ iff $p \geqslant 2$. □

Corollary 3.4

Let $\theta_1, \ldots, \theta_p$ be such that $\theta_i \in \Theta$. Let $L_{n_0}(x_0 \mid \theta_1, \ldots, \theta_p)$ and $L_{x_0}(n_0 \mid \theta_1, \ldots, \theta_p)$ denote respectively, $L(x_0, n_0 \mid \theta_1, \ldots, \theta_p)$ considered as a function of $x_0(n_0)$ for fixed $n_0(x_0)$. Then $L_{n_0}(x_0 \mid \theta_1, \ldots, \theta_p)$ and $L_{x_0}(n_0 \mid \theta_1, \ldots, \theta_p)$ are log-concave in their respective arguments.

Proof. Trivial.

3.2. The gamma and beta cases

The results of section 3.1 are general and apply to any conjugate likelihood distribution for an exponential family. Here we show how these results may be applied to the gamma likelihood (2) and beta likelihood (4) discussed in section 2. In a similar manner, these results can be applied to likelihoods obtained from other conjugate priors such as the Wishart or the Dirichlet.

We note first that the gamma distribution arises as the conjugate prior for the Poisson member of the exponential family

$$f(x \mid \theta) = \frac{\theta^x \exp(-\theta)}{x!} I_{\{0, \ldots, \infty\}}(x) = \exp\left[x \cdot \theta^* - \psi(\theta^*)\right] d\nu(x), \tag{31}$$

where

$$\theta^* = \ln\theta, \quad \psi(\theta^*) = \theta, \quad \text{and} \quad d\nu(x) = \frac{1}{x!} I_{\{0, \ldots, \infty\}}(x).$$

(θ and θ^* are different parametrizations). Thus the familiar form of the gamma may be expressed in the form (14) as,

$$\begin{aligned}\pi(\theta \mid \alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta) I_{(0,\infty)}(y) \\ &= \exp\left[x_0 \cdot \theta^* - n_0\psi(\theta^*) - \varphi(x_0, n_0)\right] I_{\Theta^*}(\theta^*)\end{aligned} \tag{32}$$

where

$$x_0 = \alpha - 1, \quad n_0 = \beta, \quad \varphi(x_0, n_0) = -\ln\frac{\beta^\alpha}{\Gamma(\alpha)} \quad \text{and} \quad \Theta^* = (-\infty, \infty).$$

Because α and β are linear functions of x_0 and n_0, we have the following corollary of theorem 3.3.

Corollary 3.5
Let $\theta_1, \ldots, \theta_p$ be such that $\theta_i \in (0, \infty)$. Then $L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$ in (2) is log-concave in (α, β). Furthermore

$$\int_0^\infty L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\, d\alpha < \infty \qquad \text{for all } p, \tag{33}$$

and

$$\int_0^\infty \int_0^\infty L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\, d\alpha\, d\beta < \infty \qquad \text{iff } p \geqslant 2. \tag{34}$$

We next consider the beta distribution which arises as the conjugate prior for the binomial member of the exponential family

$$\begin{aligned} f(x \mid \theta) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} I_{\{0, \ldots, n\}}(x) \\ &= \exp\,[x \cdot \theta^* - \psi(\theta^*)]\, dv(x), \end{aligned} \tag{35}$$

where

$$\theta^* = \ln\left(\frac{\theta}{1-\theta}\right), \qquad \psi(^*) = n \ln(1-\theta), \quad \text{and} \quad dv(x) = \binom{n}{x} I_{\{0, \ldots, n\}}(x).$$

Thus the familiar form of the beta may be expressed in the form (14) as,

$$\begin{aligned} \pi(\theta \mid \alpha, \beta) &= \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} I_{[0,1]}(y) \\ &= \exp\,[x_0 \cdot \theta^* - n_0 \psi(\theta^*) - \varphi(x_0, n_0)] I_{\Theta^*}(\theta^*) \end{aligned} \tag{36}$$

where

$$x_0 = \alpha - 1, \quad n_0 = \frac{\alpha + \beta - 2}{n}, \quad \varphi(x_0, n_0) = -\ln \frac{1}{B(\alpha, \beta)} \quad \text{and} \quad \Theta^* = (-\infty, \infty).$$

Because α and β are linear functions of x_0 and n_0, we have the following corollary of theorem 2.3.

Corollary 3.6
Let $\theta_1, \ldots, \theta_p$ be such that $\theta_i \in (0, 1)$. Then $L^B(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$ in (4) is log-concave in (α, β). Furthermore

$$\int_0^\infty L^B(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\, d\alpha < \infty \quad \text{and} \quad \int_0^\infty L^B(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\, d\beta < \infty \quad \text{for all } p, \tag{37}$$

and

$$\int_0^\infty \int_0^\infty L^B(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\, d\alpha\, d\beta < \infty \quad \text{iff } p \geqslant 2. \tag{38}$$

As an interesting aside, we note that when $p = 1$, $L^B(\alpha, \beta)$ considered as a function of α for fixed β, or β for fixed α, gives rise to a proper (conditional conjugate) distribution, and yet the joint distribution is improper.

4. An illustrative hierarchical model analysis

In this section, we illustrate the results of section 3 by providing a complete Bayesian hierarchical analysis of the pump failure data previously analyzed by Gaver & O'Muircheartaigh (1987). The data, which appear in Table 1, give for each of $p = 10$ power plant pumps, the number of failures x_i, and the length of operation time t_i (in thousands of hours).

The gamma–Poisson hierarchical model may be applied to these data as follows. Conditional on θ_i, the failure rate of pump i, the number of failures x_i is distributed Poisson $(\theta_i t_i)$,

$$f(x_i \mid \theta_i) = \frac{(\theta_i t_i)^{x_i} \exp(-\theta_i t_i)}{x_i!}, \quad x_i = 0, 1, \ldots, \tag{39}$$

and independent of $x_j, j \neq i$. Conditional on α, β, the failure rates $\theta_1, \ldots, \theta_p$ are independent and identically distributed as gamma (α, β),

$$\pi(\theta_i \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_i^{\alpha-1} \exp(-\beta\theta_i), \qquad \alpha > 0, \quad \beta > 0, \quad \theta_i > 0. \tag{40}$$

For this setup, Gaver & O'Muircheartaigh proceeded by treating α and β as fixed at empirical Bayes estimates. Gelfand & Smith (1990) treated α as fixed at an empirical Bayes estimate, and then put a gamma (0.1, 1) prior on β. We shall illustrate complete Bayesian analyses by putting priors on both α and β, and perform the calculations with the Gibbs sampler. For the purposes of illustration, we considered four different priors of the form $\pi(\alpha, \beta) = \pi(\alpha)\pi(\beta)$ on $\alpha, \beta > 0$ where $\pi(\alpha) = \exp\{-\alpha\}$ or $\pi(\alpha) \propto \exp\{-\alpha/100\}$, and $\pi(\beta) =$ gamma (0.1, 1) or $\pi(\beta) =$ gamma (0.1, 100).

To apply the Gibbs sampler, the conditional posterior forms for $\theta_1, \ldots, \theta_p, \alpha, \beta$ described in section 2 are needed. Because we have used a conjugate prior, the posterior for θ_i is also a gamma distribution,

$$\pi(\theta_i \mid x_i, \alpha, \beta) = \text{gamma}\,(\alpha + x_i, \beta + t_i), \tag{41}$$

which can be simulated by standard methods. The joint posterior for α, β is of the form

$$\pi(\alpha, \beta \mid \theta_1, \ldots, \theta_p) \propto L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p)\pi(\alpha, \beta) \tag{42}$$

where

$$L^G(\alpha, \beta \mid \theta_1, \ldots, \theta_p) \propto \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^p \left(\prod_{i=1}^p \theta_i\right)^{\alpha-1} \exp\left(-\beta \sum_{i=1}^p \theta_i\right) \tag{43}$$

is the gamma conjugate likelihood in (2). By corollary 3.5, (43) is log-concave in (α, β), and is integrable for $p \geq 2$. Thus $\pi(\alpha, \beta \mid \theta_1, \ldots, \theta_p)$ will be a proper distribution under all four of our priors on α, β.

The posterior for α,

$$\pi(\alpha, \mid \theta_1, \ldots, \theta_p, \beta) \propto \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^p \left(\prod_{i=1}^p \theta_i\right)^\alpha \pi(\alpha) \tag{44}$$

Table 1. *Pump failure data*

Obs #	1	2	3	4	5	6	7	8	9	10
t_i	94.3	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22

Table 2. *Posterior mean estimates*

Prior	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	α	β
MLE	0.053	0.064	0.080	0.111	0.573	0.604	0.954	0.954	1.908	2.099	1.27	0.82
π_1	0.060	0.102	0.089	0.116	0.602	0.609	0.891	0.894	1.588	1.994	0.7	0.9
π_2	0.061	0.106	0.090	0.117	0.603	0.609	0.884	0.886	1.560	1.981	0.8	1.0
π_3	0.061	0.107	0.091	0.117	0.584	0.605	0.806	0.808	1.453	1.936	0.8	1.4
π_4	0.062	0.113	0.093	0.118	0.585	0.604	0.791	0.789	1.398	1.905	1.0	1.6

Note: (1) $\pi_1(\alpha, \beta) \propto \beta^{-0.9} \exp[-(\alpha+\beta)]$, $\pi_2(\alpha, \beta) \propto \beta^{-0.9} \exp[-(\alpha/100)-\beta]$, $\pi_3(\alpha, \beta) \propto \beta^{-0.9} \exp[-\alpha(\beta/100)]$, $\pi_4(\alpha, \beta) \propto \beta^{-0.9} \exp[-(\alpha+\beta)/100]$. (2) Estimates are based on Gibbs samples of size 100,000.

is log-concave in α by corollary 3.4 and the log-concavity of $\pi(\alpha) \propto \exp\{-\alpha\}$ and $\pi(\alpha) \propto \exp\{-\alpha/100\}$. Thus, this distribution can be easily simulated using the adaptive rejection sampling algorithm of Gilks & Wild (1992). The posterior for β,

$$\pi(\beta \mid \theta_1, \ldots, \theta_p, \alpha) \propto \beta^{p\alpha} \exp\left(-\beta \sum_{i=1}^{p} \theta_i\right)\pi(\beta) \tag{45}$$

is a gamma distribution for both $\pi(\beta) = \text{gamma}\,(0.1, 1)$ and $\pi(\beta) = \text{gamma}\,(0.1, 100)$, and so can be simulated by standard methods.

Ergodic averages based on successive simulation from (41), (44) and (45) were used to estimate features of the posterior distribution of $\theta_1, \ldots, \theta_p, \alpha, \beta$ for the pump failure data using each of our four priors. Posterior means estimates of these parameters based on 100,000 values are given in Table 2. Also given for comparison are the maximum likelihood estimates, (the MLE's for α, β are the empirical Bayes estimates used by Gaver & O'Muircheartaigh).

Even this limited summary suffices to show that the MLE or empirical Bayes approximations differ from the complete Bayesian inferences. The Bayes estimates of $\theta_1, \ldots, \theta_p$ are less dispersed from the MLE's, having been "pulled in" towards a central value. Furthermore, these estimates are relatively similar across the four priors, especially compared to the MLE's. Finally, we note the contrast between the posterior mean pairs for α, β and the MLE's.

Acknowledgements

We thank Larry Brown, Larry Wasserman and reviewers for useful comments. This material formed part of the first Morris H. DeGroot Memorial Lecture at Carnegie–Mellon University by A. F. M. Smith on September 28, 1991.

References

Barndorff-Nielsen, O. (1970). *Exponential families: exact theory*, Various Publication Series No. 19. Aarhus University, Aarhus.

Brown, L. D. (1986). *Fundamentals of statistical exponential families*, IMS Lecture Notes, Series (ed. S. Gupta). Hayward, California.

DeGroot, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill; New York.

Deely, J. J. & Lindley, D. V. (1981). Bayes empirical Bayes, *J. Amer. Statist. Assoc.* **76,** 833–841.

Diaconis, P. & Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7,** 269–281.

Gaver, D. P. & O'Muircheartaigh, I G. (1987). Robust empirical Bayes analysis of event rates. *Technometrics* **29,** 1–15.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85,** 398–409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6,** 721–741.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. Roy. Statist. Soc. Ser C* **41,** 337–348.

Kass, R. E. & Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Amer. Statist. Assoc.* **84,** 717–726.

Morris, C. (1983). Parametric empirical Bayes inference: theory and applications, (with discussion). *J. Amer. Statist. Assoc.* **78,** 47–65.

Stein, C. M. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. 3rd Berkeley Symp. Math. Statist. Probab.* **1,** 197–207. University of California Press.

Received October 1991, in final form July 1992

E. I. George, Department of MSIS, University of Texas at Austin, Austin, TX 78712-1175, USA.