Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 319–334

Comparisons of Means using Conditionally Conjugate Priors

Barry C. Arnold, Enrique Castillo[1] and Jose Maria Sarabia[1]
University of California, Riverside, USA

SUMMARY

Conditionally conjugate prior distributions are shown to provide flexible families for modelling prior information in the comparison of normal means, even in the presence of variance heterogeneity. The technique is illustrated using a well known data set.

Key words : Gibbs sampler, Behrens-Fisher problem, Exponential families, Inconsistent prior information.

1. Introduction

It is not uncommon to be faced with a problem of comparing means for independent normal samples. Indeed the problem is almost the canonical introductory problem in statistical methods textbooks. Under the name analysis of variance, one may ask whether or not all the means are equal. Multiple comparisons, estimate contrasts, etc. are performed. And we routinely assume variance homogeneity, to avoid Behrens-Fisher type "problems". Fiducial probabilists were less concerned about variance homogeneity but their viewpoint (despite the weight and influence of R.A. Fisher) never really was accepted by mainstream applied statisticians. Bayesian analysts were undaunted by variance heterogeneity. It just meant, more parameters in the model, more complicated priors and posteriors and a larger computer account in order to process, at least approximately, the data. The current paper accepts this Bayesian thesis. It may be argued however, on Bayesian philosophical grounds, that it is disappointing to resort over frequently to the use of diffuse, vague or non-informative priors. Means are compared because we suspect they might have some simplifying structure; so we do know something about them prior to experimentation.

Informative priors should be used as much as possible. The usual informative priors can be viewed as posteriors obtained from a vague prior utilizing an imaginary (or, rarely, real) data set. It is argued in Arnold, Castillo and Sarabia [1] and earlier in Arnold and Press [2] that such priors force a possibly unnatural dependence structure on the joint prior distribution of parameters in the model. Indeed there is no reason to accept as a dictum that all available prior information is necessarily well summarized by results of an

---

1 University of Cantabria, Santander, Spain

imaginary prior experiment of the same type as that to be now undertaken. A more flexible family of informative priors is called for. In the case of a single normal sample, Arnold, Castillo and Sarabia [1] argued for the use of what may be called conditionally conjugate priors. The present paper advocates the use of such priors in more general comparison of means problems. There is never a nutritious and edible free lunch. You must pay for increased flexibility by eliciting subjective values of more hyperparameters than is usually the case. The elicitation is however feasible, thanks to easily available powerful, cheap and efficient computer programs and assuming a reasonably patient informed expert or experts who will provide a priori "bits" information about the model parameters.

2. *The Data*

Imagine that k independent samples from k normal populations are available. Thus we have independent random variables $\{X_{ij} : i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i\}$ where

$$X_{ij} \sim N(\mu_i, \sigma_i^2) \tag{2.1}$$

Based on the data X we wish to make inferences about the mean vector $\underline{\mu} = (\mu_1, \ldots, \mu_k)$. The variances $(\sigma_i^2, i = 1, 2, \ldots, k)$ are unknown and not assumed to be equal. Convenient sufficient statistics are then

$$\overline{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \quad i = 1, 2, \ldots, k \tag{2.2}$$

and

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2, \quad i = 1, 2, \ldots, k \tag{2.3}$$

All are independent and have (well) known distributions.

$$\overline{X}_i \sim N(\mu_i, \sigma_i^2 / n_i), \quad i = 1, 2, \ldots, k \tag{2.4}$$

$$S_i^2 \sim \Gamma\left(n_i - 1, \frac{2\sigma_i^2}{n_i - 1}\right), \quad i = 1, 2, \ldots, k \tag{2.5}$$

Note that in our notation for gamma distributions the second parameter is a scale parameter i.e. if $X \sim \Gamma(\alpha, \beta)$,

$$f_X(x) = [x^{\alpha-1} e^{-x/\beta}] / [\Gamma(\alpha) \beta^\alpha], \quad x > 0$$

3. *Conditionally Conjugate Priors*

We need to quantify our prior knowledge and beliefs about the 2k parameters of the model (2.1). Since we are concerned with suspected relationships among the μ_i's it is not likely that a joint prior with independent marginals will appropriately describe our prior beliefs. It is comforting that such an option is included as a special case of the class of priors to be subsequently described, so if you or your expert insist on independent priors, such prior beliefs can be readily accommodated.

Our likelihood function forms a 2k-parameter exponential family of distributions. We seek a flexible family of informative priors for the 2k parameters that will lead to reasonably tractable posterior distributions. The conditionally conjugate approach introduced in Arnold, Castillo and Sarabia [1] seems to provide a convenient solution to this problem. We will describe the approach for a general m parameter exponential family likelihood and then return to treat in detail the case of interest involving k mean parameters $\mu_1, \mu_2, \ldots, \mu_k$ and k precision parameters $\delta_1, \delta_2, \ldots, \delta_k$ where $\delta_i = 1/\sigma_i^2$.

Suppose that our data X has as its family of likelihoods a m parameter exponential family with parameter vector $\underline{\theta}$. Introduce the notation $\underline{\theta}_{(i)}$ to denote $\underline{\theta}$ with the ith coordinate deleted, $i = 1, 2, \ldots, m$. If $\underline{\theta}_{(1)}$ were known (i.e. if all θ_j's except the first were known) then a p_1 parameter exponential family of conjugate priors for θ_1 will exist of the form

$$f_1(\theta_1) \propto r_1(\theta_1) \exp\left(\sum_{j=1}^{p_1} \eta_{1j} T_{1j}(\theta_1)\right) \tag{3.1}$$

Analogously, for each i, if $\underline{\theta}_{(i)}$ were known, a conjugate prior family for θ_i will be a p_i parameter exponential family of the form

$$f_i(\theta_i) \propto r_i(\theta_i) \exp\left(\sum_{j=1}^{p_i} \eta_{ij} T_{ij}(\theta_i)\right) \tag{3.2}$$

It is reasonable to consider then a joint prior for $\underline{\theta}$ such that for each i, the conditional distribution of θ_i given $\underline{\theta}_{(i)}$ belongs to the appropriate exponential family, i.e. (3.2). The resulting joint density for $\underline{\theta}$ (see e.g. Arnold and Strauss [3]) is of the form

$$f(\underline{\theta}) = \left[\prod_{i=1}^{m} r_i(\theta_i)\right] \exp\left\{\sum_{j1=0}^{p_1} \sum_{j2=0}^{p_2} \cdots \sum_{jm=0}^{p_m} \delta_{j1,j2,\ldots,jm} \left[\prod_{i=1}^{m} q_{i,ji}(\theta_i)\right]\right\} \tag{3.3}$$

where for convenience we have introduced constant functions $q_{i0}(\theta_i) = 1, i = 1, 2, \ldots, m$. The given family of priors thus has $[\Pi_{i=1}^{m}(p_i + 1)] - 1$ hyperparameters ($\delta_{00\ldots0}$ is determined as a function of the other δ's so that the density integrates to 1). The family (3.3) will be a conjugate family for the data and the posterior distribution will retain the conditional structure present in the prior. Thus a posteriori, as well as a priori, θ_i given $\underline{\theta}_{(i)}$ will have a distribution in the p_i parameter exponential family (3.2). As long as we can simulate observations from the densities (3.2) we will, via the Gibbs sampler, be able to simulate observations from the joint density (3.3). The major drawback to the use of (3.3) is the high number of hyper- parameters involved. Note that (3.3) includes as special cases the priors with independent marginals for $\theta_1, \theta_2, \ldots, \underline{\theta}_m$ where θ_i has a density in the family (3.2) for each i. It also includes the standard conjugate prior formulation beginning with $f(\underline{\theta}) \propto 1$ and computing a posterior corresponding to an imaginary sample.

In the normal means example we have a likelihood of the form

$$L(\underline{\mu}, \underline{\delta}) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \frac{\sqrt{\delta_i}}{\sqrt{2\pi}} \exp\left[-\frac{\delta_i}{2}(x_{ij} - \mu_i)^2\right] \tag{3.4}$$

In this setting the conjugate prior for μ_1 assuming all other parameters ($\underline{\mu}_{(1)}$ and $\underline{\delta}$) were known, would be normal. The conjugate prior for δ_1, if $\underline{\delta}_{(1)}$ and $\underline{\mu}$ were known, would be gamma. Similarly each $\mu_i, i = 2, \ldots, k$, would have a normal conjugate prior if all other parameters were known and each δ_i would have a gamma conjugate prior if all other parameters were known. We are thus led to consider a joint (conditionally conjugate) prior for $(\underline{\mu}, \underline{\delta})$ such that for each i, μ_i given $\underline{\mu}_{(i)}$ and $\underline{\delta}$ is normal and, for each i, δ_i given $\underline{\delta}_{(i)}$ and $\underline{\mu}$ is gamma. The resulting joint prior density will be of the form

$$f(\underline{\mu}, \underline{\delta}) = (\delta_1 \delta_2 \ldots \delta_k)^{-1} \exp \sum_{j_1=0}^{2} \sum_{j_2=0}^{2} \cdots \sum_{j_k=0}^{2} \sum_{j'_1=0}^{2} \sum_{j'_2=0}^{2} \cdots$$

$$\sum_{j'_k=0}^{2} c_{\underline{j},\underline{j}'} \left[\prod_{i=1}^{k} q_{ij_i}(\mu_i) \prod_{i'=1}^{k} q'_{i'j'_{i'}}(\delta_{i'})\right] \tag{3.5}$$

where

$$q_{i0}(\mu_i) = 1$$

$$q_{i1}(\mu_i) = \mu_i$$

$$q_{i2}(\mu_i) = \mu_i^2$$

and

$$q'_{i'0}(\delta_{i'}) = 1$$

$$q'_{i'1}(\delta_{i'}) = \log \delta_{i'}$$

$$q'_{i'2}(\delta_{i'}) = \delta_{i'}$$

There are thus $3^{2k} - 1$ hyper-parameters (the $c_{\underline{i}, \underline{i}'}$'s) in this prior (the constant $c_{\underline{0}, \underline{0}}$ is determined by the other c's to ensure that the density integrates to 1). The traditional informative prior for this problem has most of these $3^{2k} - 1$ hyperparameters set equal to zero. The only hyperparameters given non-zero values are those 4k hyperparameters which are affected by the data. The traditional prior is thus conjugate but severely restricted in its ability to match prior beliefs. To elicit appropriate values for the array of $3^{2k} - 1$ hyperparameters, we propose to request the informed expert to provide values for prior conditional means and precisions of each μ_i given a spectrum of specific values of $\underline{\mu}_{(i)}$ and $\underline{\delta}$ and of each δ_i' given a spectrum specific values of $\underline{\delta}_{(i')}$ and $\underline{\mu}$. These, in a manner parallel to that described in Arnold, Castillo and Sarabia [1] for the case $k = 1$, yield a collection of linear relations that should hold among the hyperparameters. Typically no solution exists, since our expert is not infallible and will usually give inconstant a priori values for conditional moments. Choose hyperparameters to be minimally discrepant from the given information in the sense of being a least squares solution. As mentioned earlier, only 4k of these parameters will have different values in the posterior distribution from those values held in the prior distributions.

We are currently developing a computer program which will accept elicited conditional moments as inputs and output the best choice of hyperparameters for the conditionally conjugate prior (3.5). At the moment our implementation is limited to the case of two populations (i.e. $k = 2$). This will be illustrated in Section 4. In the more general case ($k > 2$), assuming that appropriate prior hyperparameters can be obtained and that 4k of them can be updated using the data to obtain posterior hyperparameters, we propose using the Gibbs sampler to generate realizations $(\underline{\mu}^{(k)}, \underline{\delta}^{(k)})$, $k = 1, 2, \ldots, N$, from the posterior distribution, after discarding the initial iterations. One can then study the approximate posterior distribution of $\Sigma_{i=1}^{k} (\mu_i - \overline{\mu})^2$ in order to decide whether there is evidence for differences among the μ_i's etc.

4. *An Example Involving Two Populations*

Our data set is a much analysed data set described in Snedecor and Cochran ([5], p. 118), based on a 1940 Ph.D. Thesis of Charlotte Young, and reproduced in Table 1. The goal is to compare basal metabolism of college women under two different sleep regimes.

Table 1. Basal metabolism of 26 college women
(Calories per square meter per hour)

7 or More hours of sleep				6 or Less hours of sleep			
1.	35.3	9.	33.3	1.	32.5	7.	34.6
2.	35.9	10.	33.6	2.	34.0	8.	33.5
3.	37.2	11.	37.9	3.	34.4	9.	33.6
4.	33.0	12.	35.6	4.	31.8	10.	31.5
5.	31.9	13.	29.0	5.	35.0	11.	33.8
6.	33.7	14.	33.7	6.	34.6		
7.	36.0	15.	35.7				
8.	35.0	ΣX_{1j}	= 516.8			ΣX_{2j}	= 369.3
$n_1 = 15, X_1 = 34.45$ cal./sq.m./hr.				$n_2 = 11, X_2 = 33.57$ cal./sq.m./hr.			

We wish to specify a conditionally conjugate joint prior for $(\mu_1, \mu_2, \delta_1, \delta_2)$, utilize the data in Table 1 to obtain the corresponding (still conditionally conjugate) posterior for $(\mu_1, \mu_2, \delta_1, \delta_2)$ and then we wish to consider the approximate posterior distribution of the difference between means $v \triangleq \mu_1 - \mu_2$. In addition we will look at the approximate posterior distribution of $\xi \triangleq \delta_1 / \delta_2$ to verify whether we are indeed in a Behrens-Fisher setting, i.e., a setting in which $\xi \neq 1$. Our conditionally conjugate prior family of joint densities for $(\mu_1, \mu_2, \delta_1, \delta_2)$ is of the following form (cf. equation (3.5)).

$$f(\mu_1, \mu_2, \delta_1, \delta_2) \propto (\delta_1 \delta_2)^{-1} \exp\,[\,c_{1000}\,\mu_1 + c_{0100}\,\mu_2 + c_{0010} \log \delta_1 + c_{0001} \log \delta_2 + \ldots + c_{2222}\,\mu_1^2\,\mu_2^2\,\delta_1 . \delta_2\,] \tag{4.1}$$

involving $3^4 - 1 = 80$ hyperparameters. Only the 8 hyperparameters $c_{0010}, c_{0001}, c_{0020}, c_{0002}, c_{1020}, c_{0102}, c_{2020}$ and c_{0202} will be changed from prior to posterior by the likelihood of the data set in Table 1. The classical Bayesian analysis of this data set would give non-zero values to some or all of these 8 hyperparameters and set the remaining 72 equal to 0. We have the additional flexibility provided by the 80 hyperparameters family.

We illustrate to some extent this flexibility by analysing the metabolism data using 3 prior specification paradigms, all of which are encompassed by the family (4.1) of priors. The three examples are : (i) Diffuse prior information (all c's set equal to zero in (4.1)). (ii) Independent conjugate priors for each parameter (the only non-zero c's in (4.1) are $c_{1000}, c_{0100}, c_{0010}, c_{0001}, c_{2000}, c_{0200}, c_{0020}$, and c_{0002}) and (iii) A classical analysis that assumes that only the hyperparameters that will be affected by the data are non-zero (i.e. $c_{0010}, c_{0001}, c_{0020}, c_{0002}, c_{1020}, c_{0102}, c_{2020}$ and c_{0202}).

(i) *Diffuse prior.* Here no prior eliciation is required. The prior is of the form

$$f(\underline{\mu}, \underline{\delta}) \propto (\delta_1 \delta_2)^{-1} \text{ if } \delta_1, \delta_2 > 0 \text{ and } -\infty < \mu_1, \mu_2 < \infty \tag{4.2}$$

The posterior distribution becomes :.

$$f(\underline{\mu}, \underline{\delta} \mid \text{Data}) \propto (\delta_1 \delta_2)^{-1} \exp\left(\frac{n_1}{2} \log \delta_1 + \frac{n_2}{2} \log \delta_2 - \delta_1 \frac{1}{2} \sum_{j=1}^{n_1} x_{1j}^2 \right.$$
$$\left. - \delta_2 \frac{1}{2} \sum_{j=1}^{n_2} x_{2j}^2 + \mu_1 \delta_1 \sum_{j=1}^{n_1} x_{1j} + \mu_2 \delta_2 \sum_{j=1}^{n_2} x_{2j} - \frac{n_1}{2} \mu_1^2 \delta_1 - \frac{n_2}{2} \mu_2^2 \delta_2 \right) \tag{4.3}$$

The posterior conditional distributions to be used in the Gibbs sampler are :

$$\mu_1 \mid \delta_1 \sim N\left(\mu = \frac{1}{n_1} \sum_{j=1}^{n_1} x_{1j} ;\ \sigma^2 = \frac{1}{n_1 \delta_1} \right)$$

$$\mu_2 \mid \delta_2 \sim N\left(\mu = \frac{1}{n_2} \sum_{j=1}^{n_2} x_{2j} ;\ \sigma^2 = \frac{1}{n_2 \delta_2} \right)$$

$$\delta_1 \mid \mu_1 \sim \Gamma\left(\frac{n_1}{2} ; \left[\frac{1}{2} \sum_{j=1}^{n_1} x_{1j}^2 - \mu_1 \sum_{j=1}^{n_1} x_{1j} + \mu_1^2 \frac{n_1}{2} \right]^{-1} \right)$$

$$\delta_2 \mid \mu_2 \sim \Gamma\left(\frac{n_2}{2} ; \left[\frac{1}{2} \sum_{j=1}^{n_2} x_{2j}^2 - \mu_2 \sum_{j=1}^{n_2} x_{2j} + \mu_2^2 \frac{n_2}{2} \right]^{-1} \right)$$

Using the data from Table 1, the non-zero posterior hyperparameters are

$$c_{0010} = 15/2$$

$$c_{0001} = 11/2$$

$$c_{0020} = -8937.4$$

$$c_{0002} = -6206$$

$$c_{1020} = 516.8$$

$$c_{0102} = 369.3$$

$$c_{2020} = -15/2$$

$$c_{0202} = -11/2$$

Using these posterior hyperparameters, simulated approximate posterior distributions of $v = \mu_1 - \mu_2$ and of $\xi = \delta_1 / \delta_2$ were obtained using the Gibbs sampler with 800 iterations discarding the first 300. To display the results of these and subsequent simulations we have used kernel density estimates to obtain smooth curves.

We have used the kernel estimation expression

$$\hat{f}(x) = \frac{1}{hn} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-(x - x_i)^2 / (2h^2)\right)$$

where $$h = 1.06 \frac{\sigma}{n^{1/5}}$$

in the symmetric case and $$h = 0.9 \frac{\sigma}{n^{1/5}}$$

in the non-symmetric case, as suggested in Silverman [4]. The resulting approximate posterior densities for $v = \mu_1 - \mu_2$ and $\xi = \delta_1 / \delta_2$ are shown in Figures 1 and 2.

The corresponding approximate posterior means and variances are

$$E(v) = 0.910$$
$$Var(v) = 0.591$$
$$E(\xi) = 0.372$$
$$Var(\xi) = 0.096$$

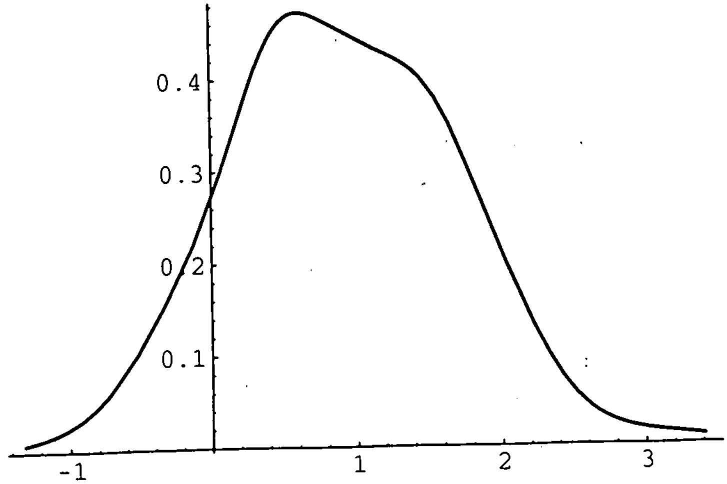

Figure 1. Diffuse priors: Simulated density of $\mu_1 - \mu_2$ using the Gibbs sampler with 500 replications and 300 starting runs

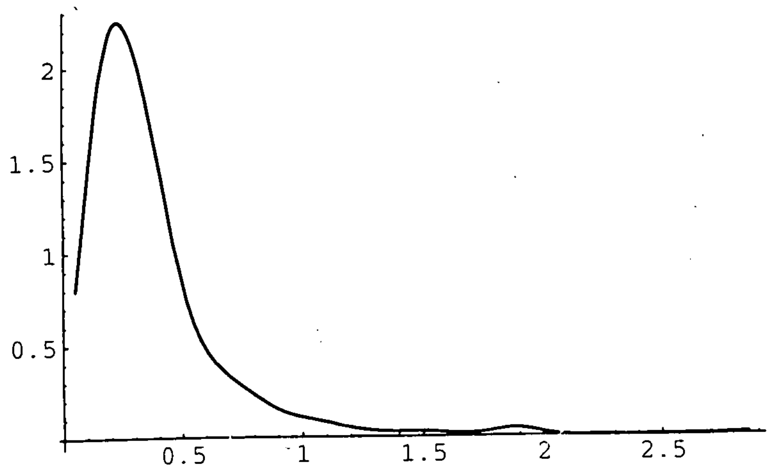

Figure 2. Diffuse priors: Simulated density of δ_1 / δ_2 using the Gibbs sampler with 500 replications and 300 starting runs

(ii) *Independent priors for each parameter.* In this case the prior distribution is

$$f(\underline{\mu}, \underline{\delta}) \propto (\delta_1, \delta_2)^{-1} \exp(c_{1000}\mu_1 + c_{2000}\mu_1^2 + c_{0100}\mu_2 + c_{0200}\mu_2^2$$
$$+ c_{0010} \log \delta_1 + c_{0020}\, \delta_1 + c_{0001} \log \delta_2 + c_{0002}\, \delta_2) \tag{4.4}$$

where the hyperparameters must satisfy the constraints

$c_{2000} < 0$; $c_{2020} < 0$; $c_{0200} < 0$; $2_{0202} < 0$;

$c_{0010} > 0$; $c_{0001} > 0$; $c_{0020} < 0$; $c_{1020}^2 < 4c_{2020}c_{0020}$;

$c_{0002} < 0$; $c_{0102}^2 < 4c_{0202}c_{0002}$

Assuming that marginal moments will be assessed we will use the following expressions relating hyperparameters in (4.2) to the assessed moments.

$$c_{1000} = \frac{E(\mu_1)}{\text{Var}(\mu_1)}$$

$$c_{2000} = \frac{1}{2\,\text{Var}(\mu_1)}$$

$$c_{0100} = \frac{E(\mu_2)}{\text{Var}(\mu_2)}$$

$$c_{0200} = -\frac{1}{2\,\text{Var}(\mu_2)}$$

$$c_{0020} = -\frac{E(\delta_1)}{\text{Var}(\delta_1)}$$

$$c_{0010} = \frac{[E(\delta_1)]^2}{\text{Var}(\delta_1)}$$

$$c_{0002} = -\frac{E(\delta_2)}{\text{Var}(\delta_2)}$$

$$c_{0001} = \frac{[E(\delta_2)]^2}{\text{Var}(\delta_2)}$$

Suppose that elicited prior moments were

$$E(\mu_1) = 35$$

$$Var(\mu_1) = 100$$

$$E(\mu_2) = 30$$

$$Var(\mu_2) = 100$$

$$E(\delta_1) = 0.2$$

$$Var(\delta_1) = 1.0$$

$$E(\delta_2) = 1.1$$

$$Var(\delta_2) = 2.0$$

The corresponding prior values of non-zero hyperparameters in (4.1) are shown in Table 2.

Table 2. Independent case : Prior and posterior values for the hyperparameters

Parameter	Prior	Posterior
c_{1000}	7/20	7/20
c_{2000}	–1/200	–1/200
c_{0100}	3/10	3/10
c_{0200}	–1/200	–1/200
c_{0010}	0.04	7.54
c_{0020}	–0.2	–8937.6
c_{0001}	0.605	6.105
c_{0002}	–0.55	–6206.6
c_{1020}	0	516.8
c_{2020}	0	–7.5
c_{0102}	0	369.3
c_{0202}	0	–5.5

The conditional distributions used in the Gibbs sampler are :

$$\mu_1 \mid \delta_1 \sim N\left(\mu = \frac{c_{1000} + c_{1020}\,\delta_1}{2\,(c_{2000} + c_{2020}\,\delta_1)};\ \sigma^2 = -\frac{1}{2\,(c_{2000} + c_{2020}\,\delta_1)}\right)$$

$$\mu_2 \mid \delta_2 \sim N\left(\mu = \frac{c_{0100} + c_{0102}\,\delta_2}{2\,(c_{0200} + c_{0202}\,\delta_2)};\ \sigma^2 = -\frac{1}{2\,(c_{0200} + c_{0202}\,\delta_2)}\right)$$

$$\delta_1 \mid \mu_1 \sim \Gamma\left(c_{0010}; -[c_{0020} + c_{1020}\mu_1 + c_{2020}\mu_1^2]^{-1}\right)$$

$$\delta_2 \mid \mu_2 \sim \Gamma\left(c_{0001}; -[c_{0002} + c_{0102}\,\mu_2 + c_{0202}\,\mu_2^2]^{-1}\right)$$

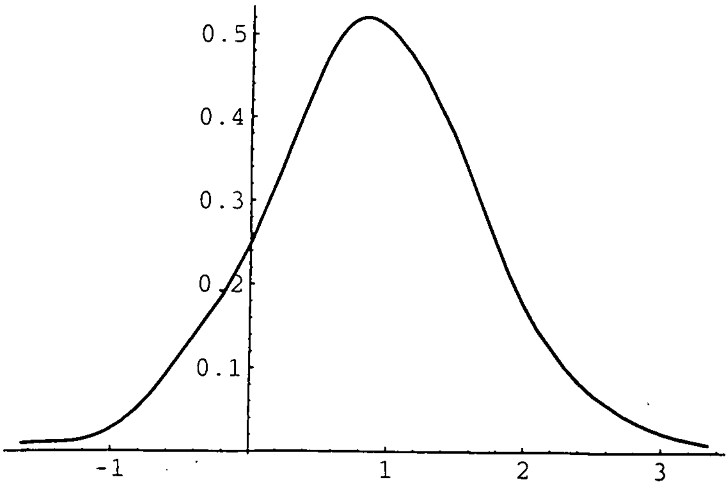

Figure 3. Independent priors: Simulated density of $\mu_1 - \mu_2$ using the Gibbs sampler with 500 replications and 300 starting runs

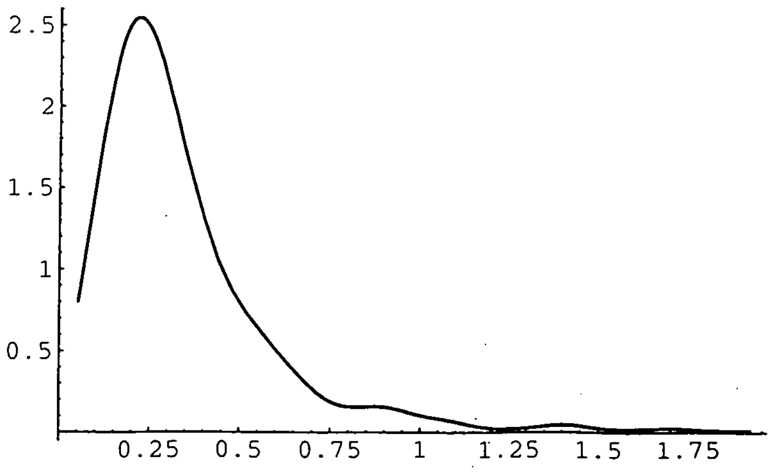

Figure 4. Independent priors: Simulated density of δ_1 / δ_2 using the Gibbs sampler with 500 replications and 300 starting runs

The resulting smoothed approximate posterior densities for $v=\mu_1-\mu_2$ and $\xi=\delta_1/\delta_2$ are shown in Figures 3 and 4. The corresponding approximate posterior means and variances are

$$E(v)=0.892$$
$$Var(v)=0.608$$
$$E(\xi)=0.344$$
$$Var(\xi)=0.066$$

(iii) *Classical analysis.* 8 hyperparameters are assumed to be non-zero in the prior (4.1) (namely those that *will* be affected by the data). The prior distribution in the classical case is

$$f(\underline{\mu},\underline{\delta}) \propto (\delta_1\delta_2)^{-1}\exp(c_{0010}\log\delta_1+c_{0001}\log\delta_2+c_{0020}\delta_1+c_{0002}\delta_2$$
$$+c_{1020}\mu_1\delta_1+c_{0102}\mu_2\delta_2+c_{2020}\mu_1^2\delta_1+c_{0202}\mu_2^2\delta_2) \quad (4.5)$$

where the hyperparameters must satisfy the constraints

$$c_{2020}<0; \quad c_{0202}<0; \quad c_{0010}>0; \quad c_{0001}>0;$$
$$c_{0020}<0; \quad c_{1020}^2<4c_{2020}c_{0020}; \quad c_{0002}<0;$$
$$c_{0102}^2<4c_{0202}c_{0002}$$

The prior (4.5) has the following conditional distributions

$$\mu_1/\delta_1 \sim N\left(\mu=-\frac{c_{1020}}{2c_{2020}};\ \sigma^2=-\frac{1}{2c_{2020}\delta_1}\right)$$
$$\mu_2/\delta_2 \sim N\left(\mu=-\frac{c_{0102}}{2c_{0202}};\ \sigma^2=-\frac{1}{2c_{0202}\delta_2}\right)$$
$$\delta_1\,|\,\mu_1 \sim \Gamma\left(c_{0010};-[c_{0020}+c_{1020}\mu_1+c_{2020}\mu_1^2]^{-1}\right)$$
$$\delta_2\,|\,\mu_2 \sim \Gamma\left(c_{0001};-[c_{0002}+c_{0102}\mu_2+c_{0202}\mu_2^2]^{-1}\right)$$

One may use conditional moments to assess appropriate values for the hyperparameters. The values of the hyperparameters will be related to the assessed conditional moments as follows :

$$c_{1020}\delta_1=\frac{E(\mu_1\,|\,\delta_1)}{Var(\mu_1\,|\,\delta_1)}$$

$$c_{2020}\,\delta_1 = \frac{1}{2\,\mathrm{Var}\,(\mu_1 \mid \delta_1)}$$

$$c_{0102}\,\delta_2 = \frac{E\,(\mu_2 \mid \delta_2)}{\mathrm{Var}\,(\mu_2 \mid \delta_2)}$$

$$c_{0202}\,\delta_2 = -\frac{1}{2\,\mathrm{Var}\,(\mu_2 \mid \delta_2)}$$

$$c_{0020} + c_{1020}\,\mu_1 + c_{2020}\,\mu_1^2 = -\frac{E\,(\delta_2 \mid \mu_1)}{\mathrm{Var}\,(\delta_1 \mid \mu_1)}$$

$$c_{0010} = \frac{[E\,(\delta_1 \mid \mu_1)]^2}{\mathrm{Var}\,(\delta_1 \mid \mu_1)}$$

$$c_{0002} + c_{0102}\,\mu_2 + c_{0202}\,\mu_2^2 = -\frac{E\,(\delta_2 \mid \mu_2)}{\mathrm{Var}\,(\delta_2 \mid \mu_2)}$$

$$c_{0001} = \frac{[E\,(\delta_2 \mid \mu_2)]^2}{\mathrm{Var}\,(\delta_2 \mid \mu_2)}$$

We assume that the expert supplies the following information:

$$E\,(\mu_1 \mid \delta_1 = 0.22) = 34.4$$

$$\mathrm{Var}\,(\mu_1 \mid \delta_1 = 0.22) = 100.$$

$$E\,(\mu_2 \mid \delta_2 = 0.80) = 33.6$$

$$\mathrm{Var}\,(\mu_2 \mid \delta_2 = 0.80) = 90$$

$$E\,(\delta_1 \mid \mu_1 = 35) = 0.30$$

$$\mathrm{Var}\,(\delta_1 \mid \mu_1 = 35) = 0.02$$

$$E\,(\delta_2 \mid \mu_2 = 32) = 0.80$$

$$\mathrm{Var}\,(\delta_2 \mid \mu_2 = 32) = 0.04$$

The resulting prior and posterior values of non-zero hyperparameters are shown in Table 3.

The results of the simulation using the Gibbs sampler are shown in Figures 5 and 6, which show the smoothed histograms for the posterior distribution of $\upsilon = \mu_1 - \mu_2$ and $\xi = \delta_1 / \delta_2$. The corresponding approximate posterior means and variances are

Table 3. Prior and posterior values of non-zero hyperparameters for the classical case

Parameter	Prior	Posterior
c_{0010}	4.5	12
c_{0001}	16	21.5
c_{0020}	−41.89	−8979.31
c_{0002}	−27.82	−6233.9
c_{1020}	1.564	518.36
c_{0102}	0.467	369.8
c_{2020}	−0.0227	−7.523
c_{0202}	−0.00694	−5.507

$E(v) = 0.8521$

$\mathrm{Var}(v) = 0.4684$

$E(\xi) = 0.3131$

$\mathrm{Var}(\xi) = 0.0139$

It is clear from the Figures 5, 6 that, for this data set, $\mu_1 - \mu_2$ is slightly positive (more sleep associated with higher metabolism) although the treatment difference might well be considered to be not significant (95% intervals for all three analyses would include $v = 0$). It is also clear, since δ_1 / δ_2 appears

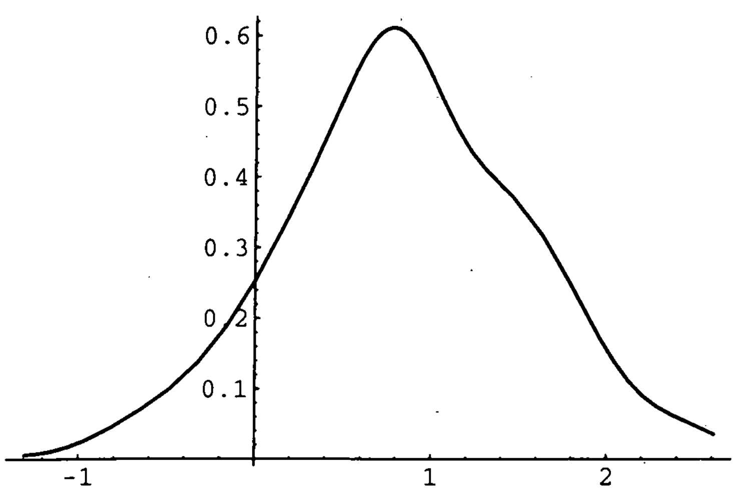

Figure 5. Classical priors: Simulated density of $\mu_1 - \mu_2$ using the Gibbs sampler with 500 replications and 300 starting runs

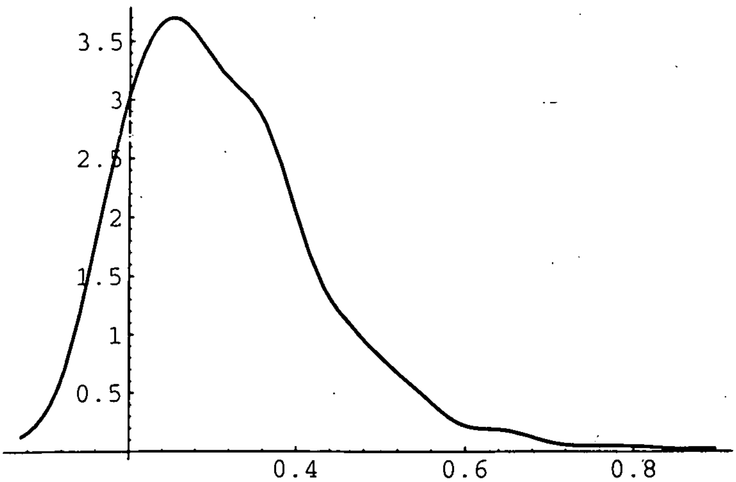

Figure 6. Classical priors: Simulated density of δ_1 / δ_2 using the Gibbs sampler with 500 replications and 300 starting runs.

to be clearly less than 1, that indeed we were right in not assuming equal variances. We were indeed confronted by a Behrens-Fisher situation.

ACKNOWLEDGEMENTS

Research partially supported by the DGICYT (Dirección General de Investigación Científica y Tecnica), under project PB94-1060.

REFERENCES

[1] Arnold, B.C., Castillo, E. and Sarabia, J.M., 1996. Bayesian analysis for classical distributions using conditionally specified priors. (*Submitted for publication*).

[2] Arnold, B.C. and Press, S.J., 1983. Bayesian inference for Pareto populations. *Journal of Econometrics*, **2**, 287-306.

[3] Arnold, B.C. and Strauss, D., 1991. Bivariate distributions with conditionals in prescribed exponential families. *J. Roy. Statist. Soc.*, **B53**, 365-375.

[4] Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[5] Snedecor, G.W. and Cochran, W.G., 1967. *Statistical Methods*. 6th ed. Iowa State University Press, Ames, Iowa.