Bayesian inference for the ratio of the means of two normal populations with unequal variances

M. Mendoza
Departamento de Estadística
ITAM, Río Hondo 1
01000 México D.F., Mexico
e-mail: mendoza@gauss.rhon.itam.mx

E. Gutiérrez-Peña
Departamento de Probabilidad y Estadística
IIMAS-UNAM, Apartado Postal 20-726
01000 México D.F., Mexico
e-mail: eduardo@sigma.iimas.unam.mx

Summary

The problem of making inferences about the ratio of two normal means has been addressed, both from the frequentist and Bayesian perspectives, by several authors. Most of this work is concerned with the homoscedastic case. In contrast, the situation where the variances are not equal has received little attention. Cox (1985) deals, within the frequentist framework, with a model where the variances are related to the means. His results are mainly based on Fieller's theorem whose drawbacks are well known. In this paper we present a Bayesian analysis of this model and discuss some related problems. An agronomical example is used throughout to illustrate the methods.

Key words and phrases: Bayesian Analysis; Fieller-Creasy Problem; Heteroscedasticity; Ratio of Normal Means; Reference Prior

1 Introduction

The problem of making inferences about the ratio of normal means has been extensively discussed in the statistical literature. A situation where a statistical analysis is required for a parameter of this type arises in a natural way in the context of bioassay (see, for example, Bliss, 1970; Finney, 1978; Srivastava, 1986), although it also appears in other fields. The simplest version of this problem may be stated as follows.

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ be two independent random samples such that

$$\begin{array}{rcll} X_i & \sim & N(x|\mu, \sigma^2); & i = 1, \ldots, n \\ Y_j & \sim & N(y|\eta, \sigma^2); & j = 1, \ldots, m \end{array} \tag{1}$$

where μ, η and σ^2 are unknown. Provided that $\eta \neq 0$, the aim of the analysis is to produce inferences about the parameter $\phi = \mu/\eta$ which describes the relative magnitude of the means. Usually μ and η have the same sign and so it can be assumed without loss of generality that both are positive.

From a frequentist point of view, pointwise estimation of ϕ is rather straightforward. It is well known, however, that several methodological difficulties arise if the objective is to produce interval estimates for this parameter (Fieller, 1954; Creasy, 1954). Somewhat surprisingly, despite these drawbacks Fieller's formula and some generalisations of it are still widely used (see Raftery and Schweder, 1993, for a recent example).

On the other hand, from a Bayesian perspective this problem has been addressed by Kappenman, Geisser and Antle (1970) and Bernardo (1977), among others. Naturally their analyses do not suffer from the difficulties of the frequentist procedure. Some generalisations and related problems have also been successfully analysed within the Bayesian framework. Press (1969) and Zellner (1978) deal with the problem of making inferences about the ratio of two regression parameters; Darby (1980) describes a solution for estimating the relative potency in a parallel line bioassay; Mendoza (1988) discusses a method for analysing the ratio of linear combinations of coefficients in a multiple regression model. A common assumption in all these contributions is, however, that of homoscedasticity.

In this paper, we consider the Bayesian analysis of some heteroscedastic versions of model (1). Section 2 deals with a specific mean-variance relationship introduced by Cox (1985). A hierarchical version of this model is discussed in Section 3. The relative merits of Cox's model are further explored in Section 4, where a general heteroscedastic model is considered. In order to compare the Bayesian results with those obtained by Cox, we use vague priors throughout. Whenever possible we adopt an ordered group reference prior as described in Berger and Bernardo (1992); otherwise we resort to limits of conjugate priors (DeGroot, 1970; Chap. 10). Finally, in Section 4 we offer some concluding remarks.

2 The Basic Model

2.1 Cox's method

As a generalisation to model (1), Cox (1985) assumes that $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ are two independent random samples such that

$$\begin{array}{rcll} X_i & \sim & N(x|\mu, \sigma_1^2); & i = 1, \dots, n \\ Y_j & \sim & N(y|\eta, \sigma_2^2); & j = 1, \dots, m \end{array} \tag{2}$$

with $\sigma_1^2 = \rho^2\,(c+\mu)^k$ and $\sigma_2^2 = \rho^2\,(c+\eta)^k$, where c and k are known constants. Cox argues that c should be chosen so as to guarantee that the variances are positive for any value of k and suggests using $c = 0$ or $c = 1$. In most cases, however, both μ and η can be assumed to be positive and so c is not really necessary. In any event, (1) can be viewed as a particular case of this model with $c = k = 0$.

In his paper, Cox proposes a pivotal statistic to obtain confidence intervals for the quantity $(c+\mu)/(c+\eta)$; the latter corresponds to the ratio of the means only if $c = 0$. On the other hand, he provides no guidelines for choosing k, but recognises that for $k > 2$ further research is required in order to find useful intervals. He goes on to discuss in detail only the case $k = 2$. It is interesting to note that in this situation model (2) has a nice interpretation. Specifically, under this model the two populations have the same coefficient of variation, ρ. It should be pointed out that the procedure proposed by Cox is entirely based on Fieller's formula.

Here we focus on the ratio of the means, $\phi = \mu/\eta$, so in the sequel we assume $c = 0$. Also, we only consider the case $k = 2$ since, as we shall argue in Section 3, the case $k > 2$ is typically of no practical interest and is actually more difficult to analyse than the general model considered there.

2.2 A Bayesian analysis of Cox's model

It should be emphasised that for this problem the parameter of interest is $\phi = \mu/\eta$, even though the model is initially parametrised in terms of μ, η and ρ. Thus the likelihood function for (ϕ, η, ρ) is given by

$$\begin{array}{rcl} L(\phi, \eta, \rho) & \propto & \rho^{-(n+m)}\,\phi^{-n}\,\eta^{-(n+m)} \\ & \times & \exp\left\{-\frac{1}{2\rho^2\phi^2\eta^2}\left(\sum_{i=1}^n (x_i - \phi\eta)^2 + \phi^2 \sum_{j=1}^m (y_j - \eta)^2\right)\right\}. \end{array} \tag{3}$$

Bayesian inference for ϕ should be based on the corresponding marginal posterior distribution. In order to obtain a joint posterior distribution for the whole vector of parameters, we must assign a joint prior on (ϕ, η, ρ). For the sake of comparison with related frequentist results it will be convenient to use a noninformative prior. Several procedures are now available that produce priors which may be regarded as 'noninformative'. Among these, the reference prior method proposed by Berger and Bernardo (1992) distinguishes the role of the parameters of interest as opposed to nuisance parameters. Moreover, reference priors have successfully dealt with a number of common difficulties arising in connection with noninformative prior distributions, such as lack of invariance and marginalisation paradoxes.

Application of the reference prior method requires the ordering of the parameters according to inferential importance. Here we consider the ordered parametrisation (ϕ, η, ρ), thus taking into account the role of ϕ as the quantity of interest, while treating η and ρ as nuisance parameters. After some involved calculations, it can then be shown that the resulting reference prior is given by

$$\pi(\phi, \eta, \rho) \propto \phi^{-1}\, \eta^{-1}\, \rho^{-1}; \quad \phi > 0, \eta > 0, \rho > 0, \tag{4}$$

yielding the posterior

$$\begin{aligned} \pi(\phi, \eta, \rho | \boldsymbol{x}, \boldsymbol{y}) &\propto \rho^{-(n+m+1)}\, \phi^{-(n+1)}\, \eta^{-(n+m+1)} \\ &\times \exp\left\{-\tfrac{1}{2\rho^2\phi^2\eta^2}\left(\textstyle\sum_{i=1}^{n}(x_i - \phi\eta)^2 + \phi^2 \sum_{j=1}^{m}(y_j - \eta)^2\right)\right\}. \end{aligned} \tag{5}$$

Unfortunately, this posterior distribution is useless due to the fact that it is improper for *any* sample sizes n and m. Nevertheless this result suggests a pseudo-reference prior of the form

$$\pi(\phi, \eta, \rho) \propto \phi^{-a}\, \eta^{-b}\, \rho^{-c}; \quad \phi > 0, \eta > 0, \rho > 0,$$

where a, b, and c are suitable constants. There are many possible choices for a, b and c which lead to a proper posterior distribution. In particular, it suffices to take $a = 1$, $b < 1$ and $c = 1$. In what follows we shall use $b = 1/2$, although any other value of b close to one may be considered. The posterior distribution corresponding to this choice is given by

$$\begin{aligned} \pi(\phi, \eta, \rho | \boldsymbol{x}, \boldsymbol{y}) &\propto \rho^{-(n+m+1)}\, \phi^{-(n+1)}\, \eta^{-(n+m+1/2)} \\ &\times \exp\left\{-\tfrac{1}{2\rho^2}\left[\tfrac{n}{\phi^2\eta^2}\ \{s_x^2 + (\phi\eta - \bar{x})^2\} + \tfrac{m}{\eta^2}\ \left\{s_y^2 + (\eta - \bar{y})^2\right\}\right]\right\}. \end{aligned} \tag{6}$$

with $ns_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ and $ms_y^2 = \sum_{j=1}^{m}(y_j - \bar{y})^2$, where $\bar{x}$ and $\bar{y}$ denote the sample means. Here ρ can be integrated out analytically, leading to the marginal posterior

$$\begin{aligned} \pi(\phi, \eta | \boldsymbol{x}, \boldsymbol{y}) &\propto \phi^{-(n+1)}\, \eta^{-(m+n+1/2)} \\ &\times \left[\tfrac{n}{\phi^2\eta^2}\ \{s_x^2 + (\phi\eta - \bar{x})^2\} + \tfrac{m}{\eta^2}\ \left\{s_y^2 + (\eta - \bar{y})^2\right\}\right]^{-(n+m)/2} \end{aligned} \tag{7}$$

for ϕ and η. Notice that η cannot be eliminated from this expression in a similar way. Thus, the marginal posterior distribution $\pi(\phi|\boldsymbol{x}, \boldsymbol{y})$ must be approximated numerically.

In order to show the kind of results that can be obtained using this approach, we now analyse a data set previously considered by Cox (1985).

Example. Table 1 shows a data set concerning the percentage of fine gravel in two types of surface soil. For these data, $\bar{x} = 10.91$, $\bar{y} = 3.94$, $s_x^2 = 34.39$ and $s_y^2 = 5.96$.

Table 1: Fine gravel in surface soils

Good	5.9	3.8	6.5	18.3	18.2	16.1	7.6
Poor	7.6	0.4	1.1	3.2	6.5	4.1	4.7

A sample of size 2000 was drawn from the posterior distribution (7) by means of the Metropolis-Hastings algorithm (see, for example, Gilks, Richardson and Spiegelhalter, 1996). We worked in terms of $\psi = \log\phi$ and $\lambda = \log\eta$ so as to get a more normal-like shape for the posterior. We then used the independence sampler with proposal distribution given by a bivariate normal approximation to $\pi(\psi, \lambda|\boldsymbol{x}, \boldsymbol{y})$ with mean at the mode and covariance matrix equal to minus 1.5 times the inverse Hessian matrix evaluated at the mode.

Figure 1 shows the marginal posterior density of ϕ. The resulting 95% HPD interval (Bayesian I) is reported in Table 2, together with the corresponding 95% confidence interval (Cox). Note that, although the two intervals have about the same lower limit, the Bayesian interval is narrower.

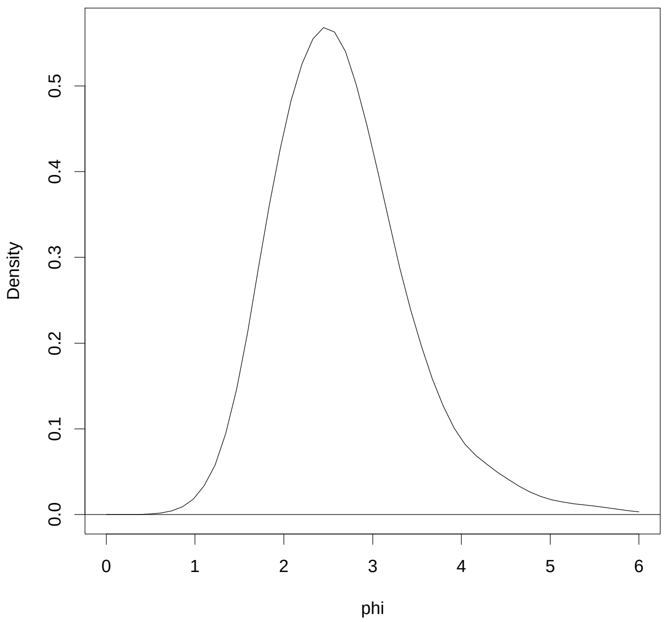

Figure 1: *Posterior density of ϕ for Cox's model*

2.3 A Bayesian sequential analysis of Cox's model

We note in passing the following alternative formulation of Cox's model. Let

$$\begin{array}{rcll} Y_j & \sim & N(y|\eta, \sigma_2^2); & j = 1, \ldots, m \\ X_i & \sim & N(x|\phi\eta, \phi^2\sigma_2^2); & i = 1, \ldots, n \end{array}$$

where $\sigma_2^2 = \eta^2\rho^2$. This is equivalent to (2) when $k = 2$ and $c = 0$. Thus we can make inferences about ϕ sequentially, first using the sample $Y_1, \ldots, Y_m$ in order to obtain a posterior for (η, σ_2), and then using this distribution as a prior, along with the sample $X_1, \ldots, X_n$, to get a joint posterior for (ϕ, η, σ_2) from which the marginal posterior distribution of ϕ can be obtained.

Provided that ϕ and (η, σ_2) are assumed independent *a priori*, and if they are given proper priors, the sequential analysis described above is totally equivalent to the Bayesian analysis discussed in the previous section, as the following argument will show. Proceeding sequentially, we first obtain the posterior of (η, σ_2) as

$$p(\eta, \sigma_2|\boldsymbol{y}) = \frac{p(\eta, \sigma_2)\, p(\boldsymbol{y}|\eta, \sigma_2)}{p(\boldsymbol{y})},$$

which in turn yields the integrated likelihood

$$p(\boldsymbol{x}|\phi, \boldsymbol{y}) = \int\int p(\boldsymbol{x}|\phi, \eta, \sigma_2)\, p(\eta, \sigma_2|\boldsymbol{y})\, d\eta\, d\sigma_2.$$

Note the $\boldsymbol{X}$ and $\boldsymbol{Y}$ are conditionally independent given (ϕ, η, σ_2). This, together with the prior independence of ϕ and (η, σ_2), implies that $p(\phi) = p(\phi|\boldsymbol{y})$. Hence

$$p(\phi|\boldsymbol{x}, \boldsymbol{y}) = \frac{p(\phi)\, p(\boldsymbol{x}|\phi, \boldsymbol{y})}{p(\boldsymbol{x}|\boldsymbol{y})}.$$

On the other hand, from the analysis of Section 2.2, and assuming that ϕ and (η, σ_2) are independent *a priori*, it follows that

$$\begin{aligned} p(\phi|\boldsymbol{x}, \boldsymbol{y}) &= \frac{\int\int p(\phi, \eta, \sigma_2)\, p(\boldsymbol{x}, \boldsymbol{y}|\phi, \eta, \sigma_2)\, d\eta\, d\sigma_2}{p(\boldsymbol{x}, \boldsymbol{y})} \\ &= \frac{p(\phi)\ \int\int p(\boldsymbol{x}|\phi, \eta, \sigma_2)\, p(\eta, \sigma_2|\boldsymbol{y})\, d\eta\, d\sigma_2}{p(\boldsymbol{x}|\boldsymbol{y})}. \end{aligned}$$

Therefore the marginal posterior distribution of ϕ does not depend on the sampling scheme.

The sequential model discussed in this section has the following nice interpretation. Suppose we observe a signal Y having an unknown normal distribution. We then observe an 'amplified' signal X, with the same distribution as Y except for an amplification factor ϕ. The problem is to make inferences about ϕ.

The equivalence between the (sequential) analysis of this section and that of Section 2.2 relies on the propriety of the priors $p(\phi)$ and $p(\eta, \sigma_2)$. We note that, as an alternative to using improper noninformative priors, if necessary we can always consider proper conjugate priors with hyperparameters chosen so as to approximately describe vague prior knowledge.

3 A Bayesian Hierarchical Alternative

The heteroscedastic model proposed by Cox is interesting for a number of reasons. In particular, it deals with the problem of unequal variances by assuming a specific structure relating the variances to the means, one which is intended to generalise a relationship commonly observed in practice. Unfortunately, as far as the parameter k (which is supposed to provide such a generalisation) is concerned, only the case $k = 2$ has been analysed. Therefore, from a modelling perspective, Cox's proposal seems to be rather restrictive. One possibility is to consider a general model where the variances are different and no further constraints are imposed on them. This approach will be explored in the next section. An intermediate alternative is to keep the restriction $k = 2$ and introduce a hierarchical structure into the model. In fact, as we shall see, this idea provides a robust version of the analysis presented in Section 2.2.

Let us assume that, at the first level of the specification,

$$X_i \sim N(\mu, \tau_1^{-1}); \quad i = 1, \ldots, n$$

$$Y_j \sim N(\eta, \tau_2^{-1}); \quad j = 1, \ldots, m$$

where, for convenience, we have parametrised the model in terms of the precisions $\tau_1 = 1/\sigma_1^2$ and $\tau_2 = 1/\sigma_2^2$ instead of the corresponding variances. For the second level we assume that (μ, τ_1) and (η, τ_2) are independent and that each of these vectors follows a truncated *Gamma-Normal* distribution, *i.e.*,

$$\tau_1|\mu \sim \text{Ga}(\alpha + 1, \mu^2), \qquad \mu \sim N_T(\mu_0, \lambda_1^{-1}),$$

$$\tau_2|\eta \sim \text{Ga}(\alpha + 1, \eta^2), \qquad \eta \sim N_T(\eta_0, \lambda_2^{-1}),$$

where μ_0, λ_1, η_0 and λ_2 are known hyperparameters, whereas α is a parameter whose distribution must be given at the third level of the hierarchy,

$$\alpha \sim p(\alpha).$$

Here, a conjugacy argument suggests the form

$$p(\alpha) = p(\alpha|c, d) \propto \frac{c^\alpha}{[\,\Gamma(\alpha + 1)\,]^d}, \qquad c, d > 0,$$

where the hyperparameters c and d could be chosen close to zero to represent vague prior knowledge.

It is worth noting that, under this specification, the assumed relationship among means and variances is supposed to hold only on the average. More precisely, we have

$$E(\sigma_1^2) = E(\tau_1^{-1}) = \frac{\mu^2}{\alpha}$$

$$E(\sigma_2^2) = E(\tau_2^{-1}) = \frac{\eta^2}{\alpha}.$$

Another interesting way in which this formulation can be interpreted follows from the observation that the likelihood function is given by

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{y}|\mu, \eta, \tau_1, \tau_2, \alpha) &= p(\boldsymbol{x}|\mu, \tau_1)\, p(\boldsymbol{y}|\eta, \tau_2) \\ &= N(\boldsymbol{x}|\boldsymbol{\mu}, \tau_1^{-1}\mathbf{I}_n)\, N(\boldsymbol{y}|\boldsymbol{\eta}, \tau_2^{-1}\mathbf{I}_m), \end{aligned}$$

where $\boldsymbol{\mu} = \mu\mathbf{1}_n$ and $\boldsymbol{\eta} = \eta\mathbf{1}_m$. Now, using the gamma distributions of the second level, we get the integrated likelihood

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{y}|\mu, \eta, \alpha) &= p(\boldsymbol{x}|\mu, \alpha)\, p(\boldsymbol{y}|\eta, \alpha) \\ &= \mathrm{St}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_1^{-1}, 2(\alpha+1))\, \mathrm{St}(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\Sigma}_2^{-1}, 2(\alpha+1)), \end{aligned}$$

where $\mathrm{St}(\cdot|\boldsymbol{a}, \boldsymbol{V}^{-1}, \beta)$ stands for a multivariate *Student's t* distribution with location vector $\boldsymbol{a}$, scale matrix $\boldsymbol{V}$ and β degrees of freedom, and

$$\boldsymbol{\Sigma}_1 = \mathrm{Diag}\left(\frac{\mu^2}{\alpha+1}, \dots, \frac{\mu^2}{\alpha+1}\right)$$

and

$$\boldsymbol{\Sigma}_2 = \mathrm{Diag}\left(\frac{\eta^2}{\alpha+1}, \dots, \frac{\eta^2}{\alpha+1}\right).$$

Consequently, under this integrated model the random variables $\{X_1, \dots, X_n\}$ are identically distributed according to a *Student's t* model. Clearly, provided that $\alpha > 0$, $E(X_i) = \mu$ and $Var(X_i) = \mu^2/\alpha; \quad i = 1, \dots, n$. Therefore, the normal distribution is replaced by a *Student's t* distribution where the mean and variance satisfy the original structural hypothesis. On the other hand, it should be noticed that, although the covariances are all equal to zero, the variables are no longer independent. A similar argument applies to $\{Y_1, \dots, Y_m\}$.

A Bayesian analysis of the model discussed in this section can be obtained from the joint posterior density

$$\begin{aligned} p(\mu, \eta, \tau_1, \tau_2, \alpha|\boldsymbol{x}, \boldsymbol{y}) &\propto N(\boldsymbol{x}|\boldsymbol{\mu}, \tau_1^{-1}\mathbf{I}_n)\, N(\boldsymbol{y}|\boldsymbol{\eta}, \tau_2^{-1}\mathbf{I}_m) \\ &\times \mathrm{Ga}(\tau_1|\alpha+1, \mu^2)\, N_T(\mu|\mu_0, \lambda_1^{-1}) \\ &\times \mathrm{Ga}(\tau_2|\alpha+1, \eta^2)\, N_T(\eta|\eta_0, \lambda_2^{-1}) \\ &\times p(\alpha) \end{aligned}$$

once a prior $p(\alpha)$ is given. Incidentally, it is interesting to note that α is closely related to the common coefficient of variation, ρ_*, of X and Y under the *Student's t* distributions discussed above (in fact $\rho_* = \alpha^{-1/2}$).

Note that the parameter of interest, $\phi = \mu/\eta$, does not appear explicitly in this hierarchical formulation of the problem. Thus, the required inferences about ϕ must be obtained from the joint posterior distribution of μ and η. Moreover, the notion of a 'vague' or 'noninformative' prior cannot be implemented in terms of ϕ. Rather, the relative weight of the information contained in the joint prior can be adjusted by a suitable selection of the hyperparameters μ_0, λ_1, η_0, λ_2, c and d .

A relatively simple analysis of the hierarchical model is provided by the Gibbs sampler (Gelfand and Smith, 1990), or, more generally, by the Metropolis-within-Gibbs algorithm (see, for example, Gilks, Richardson and Spiegelhalter, 1996), since in this case some of the required full conditionals turn out not to be standard densities. Such full conditionals are given as follows (dropping the dependence on $\boldsymbol{x}$ and $\boldsymbol{y}$ for notational convenience).

$$
\begin{aligned}
p(\mu|\eta,\tau_1,\tau_2,\alpha) &\propto \mu^{2(\alpha+1)}\, N_T(\mu|m_\mu,\sigma^2_\mu)\\
p(\eta|\mu,\tau_1,\tau_2,\alpha) &\propto \eta^{2(\alpha+1)}\, N_T(\eta|m_\eta,\sigma^2_\eta)\\
p(\tau_1|\mu,\eta,\tau_2,\alpha) &\propto \mathrm{Ga}(\tau_1|a_1,b_1)\\
p(\tau_2|\mu,\eta,\tau_1,\alpha) &\propto \mathrm{Ga}(\tau_2|a_2,b_2)\\
p(\alpha|\mu,\eta,\tau_1,\tau_2) &\propto \frac{[\, c\,\mu^2\eta^2\tau_1\tau_2\,]^\alpha}{\Gamma(\alpha+1)^{d+2}},
\end{aligned}
$$

where

$$
\begin{aligned}
m_\mu &= \frac{\lambda_1\mu_0 + n\tau_1\bar{x}}{\lambda_1+(n+2)\tau_1}, & \sigma^2_\mu &= [\,\lambda_1+(n+2)\tau_1\,]^{-1},\\
m_\eta &= \frac{\lambda_2\eta_0 + m\tau_2\bar{y}}{\lambda_2+(m+2)\tau_2}, & \sigma^2_\eta &= [\,\lambda_2+(m+2)\tau_2\,]^{-1},\\
a_1 &= \alpha+(n/2)+1, & b_1 &= \frac{ns_x^2+n(\mu-\bar{x})^2}{2}+\mu^2,\\
a_2 &= \alpha+(m/2)+1, & b_2 &= \frac{ms_y^2+m(\eta-\bar{y})^2}{2}+\eta^2.
\end{aligned}
$$

Example (continued). We now look again at the data of Table 1, concerning the percentage of fine gravel in two types of surface soil. We fit the hierarchical model described above using the Gibbs sampler. A Metropolis step was necessary to generate samples from the full conditionals of μ, η and $\log\alpha$. In all three cases we used a normal proposal distribution with mean at the mode and variance equal to minus 1.5 times the inverse Hessian evaluated at the mode.

A sample of size 2000 was drawn from the joint posterior distribution of μ,η,τ_1,τ_2 and α, from which a sample of the posterior distribution of ϕ was obtained. The resulting 95% HPD interval (Bayesian II) is presented in Table 2. Figure 2 shows the marginal posterior density of the parameter of interest, ϕ. As in the previous case, the Bayesian approach yields a narrower interval.

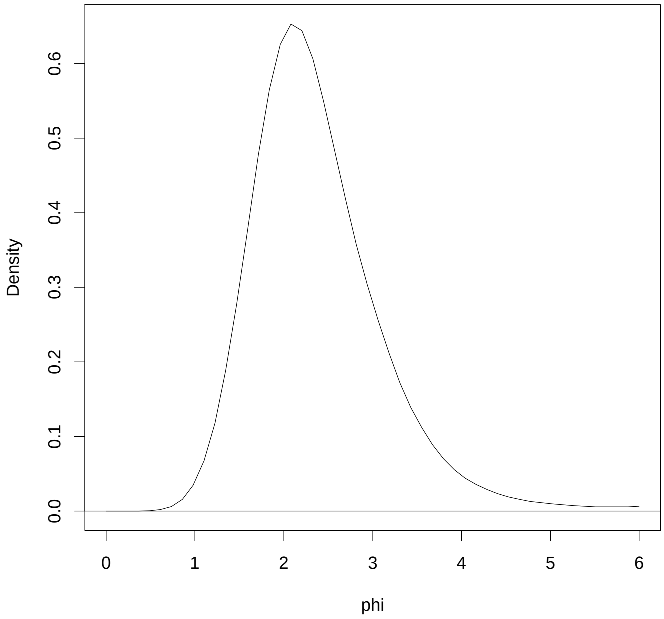

Figure 2: *Posterior density of ϕ for the hierarchical model*

4 The General Model

As discussed in the previous section, the model proposed by Cox assumes an arbitrary known value of k. For all practical purposes, however, he only deals with the case $k = 2$ where the confidence limits can be calculated as the roots of a quadratic equation. For $k > 2$, the roots of a higher order polynomial must be found and Cox provides no explicit answer for this situation. In any case the problem of assessing the value of k remains unsolved.

More generally, k might be considered as yet another unknown parameter in the model. However, the resulting inferential problem can become quite involved. It is worth noting that if k is unknown the validity of the mean-variance relationship is open to criticism. As an alternative, Cox's model can simply be seen as a particular case of the general situation where the variances are not equal. In fact, a statistical analysis of this general model might provide a more robust solution to the original problem in the sense that it does not involve any structural hypothesis.

Here we deal with the Bayesian analysis of the general model (2), where $\sigma_1 > 0$ and $\sigma_2 > 0$ are unknown and no relationship with the means is assumed. For this model we have the likelihood function

$$L(\phi, \eta, \sigma_1, \sigma_2) \quad \propto \quad \sigma_1^{-n}\, \sigma_2^{-m} \, \exp \left\{ -\tfrac{1}{2} \left(\tfrac{1}{\sigma_1^2} \textstyle\sum_{i=1}^{n} (x_i - \phi\eta)^2 + \tfrac{1}{\sigma_2^2} \sum_{j=1}^{m} (y_j - \eta)^2 \right) \right\}. \quad (8)$$

If we now adopt the ordered parametrisation $(\phi, \eta, \sigma_1, \sigma_2)$, the resulting reference prior is given by

$$\pi(\phi, \eta, \sigma_1, \sigma_2) \propto \phi^{-1/2}\, \sigma_1^{-1}\, \sigma_2^{-1}; \quad \phi > 0, \eta > 0, \sigma_1 > 0, \sigma_2 > 0, \tag{9}$$

which yields the posterior

$$\begin{aligned} \pi(\phi, \eta, \sigma_1, \sigma_2 | \boldsymbol{x}, \boldsymbol{y}) \;&\propto\; \phi^{-1/2} \sigma_1^{-(n+1)} \exp\left\{ -\tfrac{1}{2\sigma_1^2} \textstyle\sum_{i=1}^n (x_i - \phi\eta)^2 \right\} \\ &\times\; \sigma_2^{-(m+1)} \exp\left\{ -\tfrac{1}{2\sigma_2^2} \textstyle\sum_{j=1}^m (y_j - \eta)^2 \right\}. \end{aligned} \tag{10}$$

This posterior turns out to be proper, in contrast with the reference posterior for Cox's model. Here σ_1 and σ_2 can be integrated out analytically, leading to the marginal posterior $\pi(\phi, \eta | \boldsymbol{x}, \boldsymbol{y})$ from which a further (numerical) integration can produce the posterior distribution of the parameter of interest, $\pi(\phi | \boldsymbol{x}, \boldsymbol{y})$.

It is interesting to note, however, that the posterior on $(\mu,\ \eta,\ \sigma_1,\ \sigma_2)$ induced by (10) is given by

$$\begin{aligned} \pi(\mu, \eta, \sigma_1, \sigma_2 | \boldsymbol{x}, \boldsymbol{y}) \;&\propto\; \mu^{-1/2} \sigma_1^{-(n+1)} \exp\left\{ -\tfrac{1}{2\sigma_1^2} \textstyle\sum_{i=1}^n (x_i - \mu)^2 \right\} \\ &\times\; \eta^{-1/2} \sigma_2^{-(m+1)} \exp\left\{ -\tfrac{1}{2\sigma_2^2} \textstyle\sum_{j=1}^m (y_j - \eta)^2 \right\}, \end{aligned} \tag{11}$$

where $\mu > 0$, $\eta > 0$, $\sigma_1 > 0$ and $\sigma_2 > 0$. It then becomes clear that μ and σ_1 are independent of η and σ_2 *a posteriori*. In fact, (11) suggests an alternative procedure to obtain the marginal posterior for the parameter of interest. This will be discussed in the following example.

Example (continued). We now re-analyse the fine gravel data using the general model discussed above. To this end, we use a sampling-based approach following Smith and Gelfand (1992). Our aim is to generate a sample from the marginal posterior distribution of ϕ. It is clear from (11) that we can generate samples separately from the posterior distributions of (μ, σ_1) and (η, σ_2), from which the sample of ϕ can then be obtained.

Recall that $\tau_1 = 1/\sigma_1^2$. Then the posterior for (μ, τ_1) can be written as

$$\pi(\mu, \tau_1 | \boldsymbol{x}) \propto \mu^{-1/2}\, N_T(\mu | \bar{x}, (n\tau_1)^{-1})\, \mathrm{Ga}(\tau_1 | (n-1)/2, ns_x^2/2) \tag{12}$$

where N_T denotes a normal density truncated at zero. We worked in terms of $\theta = \log \mu$ and $\psi = \log \tau_1$. Unlike $\pi(\mu, \tau_1 | \boldsymbol{x})$, the posterior density of (θ, ψ) is unimodal. Let $\hat{\pi}(\theta, \psi | \boldsymbol{x})$ denote a bivariate normal approximation to $\pi(\theta, \psi | \boldsymbol{x})$ with mean at the mode and covariance matrix equal to minus 1.5 times de inverse Hessian matrix evaluated at the mode. Thus we can first generate a sample $\{(\tilde{\theta}^i, \tilde{\psi}^i)\}_{i=1}^N$ from $\hat{\pi}(\theta, \psi | \boldsymbol{x})$ and then resample with weights proportional to

$$w_i = \frac{\pi(\tilde{\theta}_i, \tilde{\psi}_i | \boldsymbol{x})}{\hat{\pi}(\tilde{\theta}_i, \tilde{\psi}_i | \boldsymbol{x})}.$$

A completely analogous procedure can be employed to simulate samples from $\pi(\eta, \sigma_2|\boldsymbol{y})$.

Table 2: 95% interval limits for ϕ

Procedure	Interval limits		Interval width
Cox	1.27	6.68	5.41
Elston	1.13	7.85	6.72
Bayesian I	1.26	4.16	2.90
Bayesian II	1.15	3.76	2.61
Bayesian III	0.63	5.29	4.66

A final sample of size 2000 was drawn from the posterior distribution (11) using the method described above. Figure 3 shows the resulting marginal posterior density of ϕ. Table 2 presents the corresponding 95% HPD interval (Bayesian III) along with two confidence intervals reported by Cox. The first is based on the model discussed in Section 2.1, whereas the second is derived from a classical analysis of the general model, due to Elston (1969).

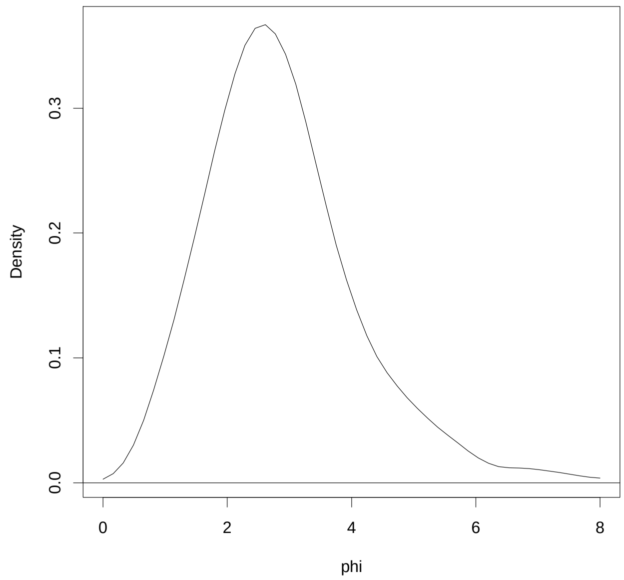

Figure 3: *Posterior density of ϕ for the general model*

As expected, for the general model both the frequentist and Bayesian intervals are wider than the corresponding intervals under Cox's model. On the other hand, for each of the models the Bayesian intervals are significantly narrower than the corresponding confidence intervals. Moreover, for the more structured model proposed by Cox, the confidence interval is about 15% wider than the Bayesian interval obtained for the general model.

In order to assess the adequacy of Cox's model we now turn our attention to the posterior distribution of $\delta = (\sigma_1/\mu) - (\sigma_2/\eta)$. This parameter is just the difference of the coefficients of variation of the two populations and must be zero if Cox's model is indeed correct. Figure 4 shows the marginal posterior distribution of δ for the prior (9). Similar results can be obtained if one uses, for example, Jeffreys' noninformative prior. No conclusive evidence against Cox's model seems to be provided by the data in this case.

Note that Cox's model (as described in Section 2) is a curved exponential family. Hence it is usually more difficult to analyse than the general model considered in this section (which is essentially a regular exponential family). Moreover, the general model is often more realistic, so in practice one would usually consider trying out the general model first.

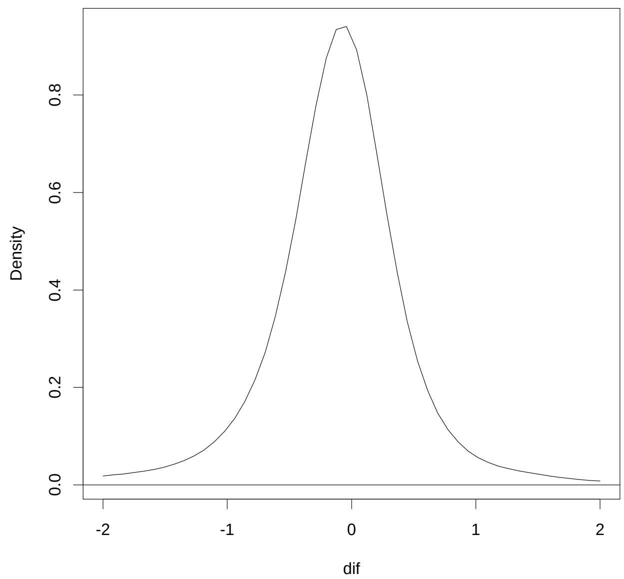

Figure 4: *Posterior density of δ*

5 Concluding Remarks

The problem of making inferences about the ratio of two normal means can be addressed within at least two different scenarios. First, if the corresponding variances are supposed to be equal then quite a simple solution can be obtained using the Bayesian approach. This is in contrast with the methodological difficulties arising in the calculation of the confidence intervals obtained from frequentist methods. Second, if the assumption of equal variances cannot be sustained, the frequentist approach can only offer approximate answers. On the other hand, within the Bayesian framework the corresponding posterior distribution provides an exact solution to any inferential problem.

Further, between these two extreme scenarios one may consider several structural hypotheses. In particular, Cox's model assumes a specific form relating the mean and the variance of each of the two populations. It so happens, however, that the statistical analysis of this model is rather involved unless $k = 2$. In this case Cox's model has an interesting interpretation since the two populations will then share the same coefficient of variation. The frequentist solution proposed by Cox is based upon Fieller's theorem and thus suffers from the same problems arising in the analysis of the simpler, homoscedastic model. The hierarchical formulation of Cox's model discussed in Section 3 provides a robust alternative while maintaining the assumption that the variances are related to the means. For this, and for the original formulation of Cox, the Bayesian approach leads to a proper posterior distribution which is the basis for any inference concerning the parameter of interest.

For the fine gravel data set analysed in the example, the HPD intervals for the ratio of the means are more precise under Cox's model (with $k = 2$) than they are under the general model. More interestingly, the Bayesian interval for the general model is narrower than the corresponding confidence interval obtained by Cox for his model. This is not surprising since the intervals provided by Fieller's theorem are not optimal in any sense whereas the Bayesian HPD regions are constructed so as to have minimum length. It could therefore be argued that a frequentist analysis of Cox's model is of little worth in this case.

Acknowledgements. This work was partially supported by the Sistema Nacional de Investigadores, Mexico.

References

Bernardo, J.M. (1977). Inferences about the Ratio of Normal Means: A Bayesian Approach to the Fieller-Creasy Problem. In *Recent Developments in Statistics.* (J.R. Barra et al. eds.) Amsterdam: North-Holland, pp. 345-350

Berger, J.O., and Bernardo, J.M. (1992), On the Development of Reference Priors (with discussion). In *Bayesian Statistics 4.* (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds.) Oxford: Clarendon Press, pp. 35-60.

Bliss, C.I. (1970). *Statistics in Biology*, Vol. 2. New York: McGraw-Hill.

Cox, C.P. (1985). Interval Estimates for the Ratio of the Means of two Normal Populations with Variances Related to the Means. *Biometrics,* **41**, 261-265.

Creasy, M.A. (1954). Limits for the Ratio of Means. *Journal of the Royal Statistical Society B*, **16**, 186-194.

Darby, S.C. (1980). A Bayesian Approach to Parallel Line Bioassay. *Biometrika,* **67**, 607-612.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

Elston, R.C. (1969). An Analogue to Fieller's Theorem Using Scheffe's Solution to the Fisher-Behrens Problem. *The American Statistician,* **23**, 26-28.

Fieller, E.C. (1954). Some Problems in Interval Estimation. *Journal of the Royal Statistical Society B*, **16**, 175-185.

Finney, D.J. (1978). *Statistical Method in Biological Assay*, 3rd. ed. London: Griffin.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association,* **85**, 398-409.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice.* (W.R. Gilks, S. Richarson, and D.J. Spiegelhalter, eds.) London: Chapman and Hall, pp. 1-19.

Kappenman, R.F., Geisser, S., and Antle, C.E. (1970). Bayesian and Fiducial Solutions for the Fieller-Creasy Problem. *Sankhya B,* **32**, 331-340.

Mendoza, M. (1988). Inferences About the Ratio of Linear Combinations of the Coefficients in a Multiple Regression Model. In *Bayesian Statistics 3.* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds.) Oxford: Oxford University Press, pp. 705-711.

Press, S.J. (1969). The t-Ratio Distribution. *Journal of the American Statistical Association,* **64**, 242-252.

Raftery, A.E. and Schweder, T. (1993). Inference About the Ratio of Two Parameters, with Application to Whale Censusing. *The American Statistician*, **47**, 259-264.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian Statistics without Tears: a Sampling-resampling Perspective. *American Statistician,* **46**, 84-88.

Srivastava, M.S. (1986). Multivariate Bioassay, Combination of Bioassays and Fieller's Theorem. *Biometrics,* **42**, 131-141.

Zellner, A. (1978). Estimation of Functions of Population Means and Regression Coefficients Including Structural Coefficients: A Minimum Expected Loss (MELO) Approach. *Journal of Econometrics,* **8**, 127-158.