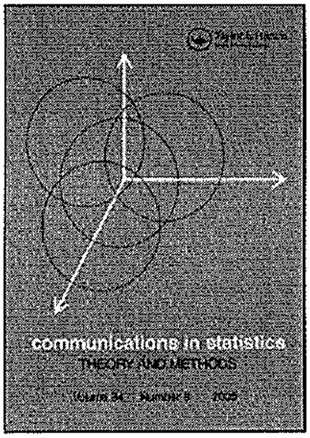

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Bayesian Inference of Scale Parameters in Exponential Family Using Conditionally Specified Priors

Panagis Moschopoulos [a]; Naijun Sha [a]
[a] Department of Mathematical Sciences, University of Texas at El Paso. El Paso, Texas. USA

Communications in Statistics—Theory and Methods, 34: 303–318, 2005

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1081/STA-200047409

Bayesian Inference of Scale Parameters in Exponential Family Using Conditionally Specified Priors

PANAGIS MOSCHOPOULOS AND NAIJUN SHA

Department of Mathematical Sciences, University of Texas at El Paso,
El Paso, Texas, USA

This article considers a unified Bayesian approach for comparing scale parameters in an exponential family, assuming other parameters such as location or shape, are known. Particular cases include the normal, gamma, Weibull, exponential, Poisson, and binomial distributions. Convenient conjugate priors are found in joint distribution conditionally specified. Hyperparameter elicitation techniques are discussed. The results are illustrated in the comparison of scales of two normals and two Weibull densities.

Keywords Bayesian analysis; Conjugate priors; Exponential family; Gibbs sampling; Normal distribution; Weibull distribution.

Mathematics Subject Classification 62E10; 62F15.

1. Introduction

Consider the following exponential class of densities

$$f(x, \theta) = r(x)g(\theta)\exp\{t(x)\theta\} \tag{1.1}$$

with a natural parameter θ. Particular cases in (1.1) are clearly the exponential with parameter θ, the Poisson(λ) with $\theta = log(\lambda)$, the Gamma(α, β) with known shape α and scale $\theta = \beta$, the Normal(μ, σ^2) with known mean μ and $\theta = 1/\sigma^2$, the binomial(n, p) with $\theta = log(p/(1-p))$, and the Weibull(α, β) with known shape α and scale $\theta = \beta$.

Consider two densities of the type (1.1), with scale parameters θ_1 and θ_2, respectively. Testing the ratio θ_1/θ_2 or the difference $\theta_1 - \theta_2$ based on two independent samples from the above densities is a well-known problem in statistics. For example, testing the ratios θ_1/θ_2 in two gamma, exponential, Weibull, or normal

Received April 30, 2003; Accepted July 7, 2004
Address correspondence to Panagis Moschopoulos, Department of Mathematical Sciences, University of Texas at El Paso, El Paso, TX 79968, USA; E-mail: peter@math.utep.edu

distributions is a well-known classical problem. In the case of Poisson, the ratio of the two parameters λ_1/λ_2 is equivalent to considering $\theta_1 - \theta_2$; in the case of two binomial densities one is also interested in the difference $\theta_1 - \theta_2$, which equals to the difference of log-odds ratio (or logit difference), $log\frac{p_1}{1-p_1} - log\frac{p_2}{1-p_2}$.

We present here a unified Bayesian approach to the problem of estimating θ_1 and θ_2 based on two independent samples from exponential family densities

$$f(x, \theta_i) = r_i(x)g_i(\theta_i)\exp\{t_i(x)\theta_i\}, \quad i = 1, 2. \tag{1.2}$$

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be the two independent samples and denote $x = (x_1, \ldots, x_{n_1})$, $y = (y_1, \ldots, y_{n_2})$. Then the joint density of x and y is

$$L(x, y|\theta_1, \theta_2) \propto [g_1(\theta_1)]^{n_1}[g_2(\theta_2)]^{n_2}\exp\left\{\sum_{i=1}^{n_1} t_1(x_i)\theta_1 + \sum_{j=1}^{n_2} t_2(y_j)\theta_2\right\} \tag{1.3}$$

2. Conditionally Specified Priors for θ_1 and θ_2

Recently, there has been considerable literature on bivariate distributions conditionally specified. The reader is referred to Arnold and Strauss (1988), Arnold et al. (1992), and Arnold et al. (1999). The last two references contain a host of bivariate distributions of (X, Y), specified by the conditional distributions of X given Y and Y given X. For the case where both conditionals are gamma, see Moschopoulos and Staniswalis (1994) and Arnold et al. (1999). Recently, Arnold et al. (1998) considered the use of conditionally specified priors in estimating the ratio of θ_1 and θ_2 in the case of two independent gamma densities.

Here we invoke a prior for (θ_1, θ_2) specified from the conjugate conditional distribution of $\theta_1|\theta_2$ and that of $\theta_2|\theta_1$. Let

$$f(\theta_1|\theta_2) \propto [g_1(\theta_1)]^{\eta_1}\exp\{s_1(\theta_2)\theta_1\} \tag{2.1}$$

$$f(\theta_2|\theta_1) \propto [g_2(\theta_2)]^{\eta_2}\exp\{s_2(\theta_1)\theta_2\}. \tag{2.2}$$

Note that for simplicity, we assumed that η_1 and η_1 are fixed parameters, and that only the "scale parameters" s_1 and s_2 are allowed to depend on conditioning variables and $s_1(\theta_2) < 0$ and $s_2(\theta_1) < 0$ for $\theta_1 > 0$ and $\theta_2 > 0$.

The conditionals (2.1) and (2.2) need to be compatible, i.e., define a joint density for θ_1 and θ_2. To find an expression for the joint density $f(\theta_1, \theta_2)$, we express it as a product of marginal and conditional densities both ways, i.e.,

$$f(\theta_1|\theta_2)f(\theta_2) = f(\theta_2|\theta_1)f(\theta_1), \tag{2.3}$$

substitute expressions (2.1) and (2.2), take natural logarithms, and differentiate with respect to θ_1 and then with respect to θ_2. This leads to

$$s_1'(\theta_2) = s_2'(\theta_1) \tag{2.4}$$

which implies

$$s_1(\theta_2) = \alpha\theta_2 + \beta, \quad s_2(\theta_1) = \alpha\theta_1 + \gamma \tag{2.5}$$

for some hyperparameters $\alpha < 0, \beta < 0, \gamma < 0$ to guarantee integrability of the densities above. Note that we have assumed $\theta_1 > 0$ and $\theta_2 > 0$ and that $\alpha = 0$ corresponds to the independence case. On substituting to (2.3), this gives the following marginal and joint densities, except of course for the specification of the normalizing constants

$$f(\theta_1) \propto [g_1(\theta_1)]^{\eta_1} \exp\{\beta\theta_1\} \tag{2.6}$$

$$f(\theta_2) \propto [g_2(\theta_2)]^{\eta_2} \exp\{\gamma\theta_2\} \tag{2.7}$$

$$f(\theta_1, \theta_2) \propto [g_1(\theta_1)]^{\eta_1} [g_2(\theta_2)]^{\eta_2} \exp\{\beta\theta_1 + \gamma\theta_2 + \alpha\theta_1\theta_2\}. \tag{2.8}$$

Now, from (1.3) and (2.8) we obtain the following posterior of θ_1 and θ_2 given x and y

$$\begin{aligned} f(\theta_1, \theta_2 | x, y) &\propto [g_1(\theta_1)]^{\eta_1+n_1} [g_2(\theta_2)]^{\eta_2+n_2} \\ &\times \exp\left\{ \left(\sum_{i=1}^{n_1} t_1(x_i) + \beta \right) \theta_1 + \left(\sum_{j=1}^{n_2} t_2(y_j) + \gamma \right) \theta_2 + \alpha\theta_1\theta_2 \right\} \end{aligned} \tag{2.9}$$

with the following conditional posteriors

$$f(\theta_1 | \theta_2, x, y) \propto [g_1(\theta_1)]^{\eta_1+n_1} \exp\left\{ \left(\sum_{i=1}^{n_1} t_1(x_i) + \beta + \alpha\theta_2 \right) \theta_1 \right\} \tag{2.10}$$

$$f(\theta_2 | \theta_1, x, y) \propto [g_2(\theta_2)]^{\eta_2+n_2} \exp\left\{ \left(\sum_{j=1}^{n_2} t_2(y_j) + \gamma + \alpha\theta_1 \right) \theta_2 \right\}. \tag{2.11}$$

Observe that both prior and posterior distributions belong to the same exponential family. Thus, the family of densities of the form (2.8) is a conjugate family of priors. Two of the hyperparameters in (2.8) are changed by the data, namely β and γ. The usual independent prior distributions for θ_1, θ_2 is included in the family (2.8) with all parameters being non zero except α. The standard non informative prior (see e.g., Berger, 1985, p. 85) is also included in (2.8). It corresponds to the case in which $g_1(\theta) = g_2(\theta) = \theta, \eta_1 = \eta_2 = -1$ and the others are zero. Thus, (2.8) is a rich family including the usual choices but is more flexible in that it includes a broader spectrum of prior beliefs than usual prior families.

3. Hyperparameter Assessment

Now we turn to the elicitation of hyperparameters for the prior (2.8). The approach used here arises from the characteristic of moments of exponential family densities. We propose eliciting values for the first two conditional moments for θ_1 and θ_2. These are obtained from the conditional priors in (2.1) and (2.2) that belong to the exponential family. On substituting $s_1(\theta_2)$ and $s_2(\theta_1)$ from (2.5), these become

$$f(\theta_1 | \theta_2) = [g_1(\theta_1)]^{\eta_1} e^{\beta\theta_1} h_1(\theta_2) \exp\{\alpha\theta_1\theta_2\} \tag{3.1}$$

$$f(\theta_2 | \theta_1) = [g_2(\theta_2)]^{\eta_2} e^{\gamma\theta_2} h_2(\theta_1) \exp\{\alpha\theta_1\theta_2\} \tag{3.2}$$

with normalizing factors $h_1(\theta_2)$ and $h_2(\theta_1)$ containing the parameters η_1, η_2, α, β, and γ. The dependence on η_1, η_2, α, β, and γ is suppressed from h_1 and h_2 for notational convenience. The conditional expectations and variances are expressed via h_1 and h_2 using well-known relations for exponential families

$$E(\theta_1|\theta_2) = -\frac{1}{\alpha}\frac{\partial log(h_1(\theta_2))}{\partial\theta_2}, \quad Var(\theta_1|\theta_2) = -\frac{1}{\alpha^2}\frac{\partial^2 log(h_1(\theta_2))}{\partial\theta_2^2}$$

$$E(\theta_2|\theta_1) = -\frac{1}{\alpha}\frac{\partial log(h_2(\theta_1))}{\partial\theta_1}, \quad Var(\theta_2|\theta_1) = -\frac{1}{\alpha^2}\frac{\partial^2 log(h_2(\theta_1))}{\partial\theta_1^2}. \tag{3.3}$$

To get appropriate values of these parameters, one can specify a series of values $\theta_2^{(1)}, \ldots, \theta_2^{(m_1)}$ and $\theta_1^{(1)}, \ldots, \theta_1^{(m_2)}$ according to previous knowledge or best guess from the expert at the value of $E(\theta_1|\theta_2 = \theta_2^{(i)})$ and $Var(\theta_1|\theta_2 = \theta_2^{(i)})$, and $E(\theta_2|\theta_1 = \theta_1^{(i)})$ and $Var(\theta_2|\theta_1 = \theta_1^{(i)})$. Denote these elicited values by

$$\zeta_i^{(1)} = E(\theta_1|\theta_2 = \theta_2^{(i)}), \quad \zeta_i^{(2)} = Var(\theta_1|\theta_2 = \theta_2^{(i)}), \quad i = 1, 2, \ldots, m_1 \tag{3.4}$$

$$\eta_j^{(1)} = E(\theta_2|\theta_1 = \theta_1^{(j)}), \quad \eta_j^{(2)} = Var(\theta_2|\theta_1 = \theta_1^{(j)}), \quad j = 1, 2, \ldots, m_2. \tag{3.5}$$

Solve these equations for the parameters $\eta_1, \eta_2, \alpha, \beta, \gamma$. If m_1, m_2 are large so that more equations need to be satisfied than the number of parameters, one can use an approximation method to estimate them, i.e., the least squares method. This will be illustrated with the examples later.

4. Sampling Technique

With parameters $\eta_1, \eta_2, \alpha, \beta, \gamma$ assessed earlier, the inference can be made from the joint posterior distribution (2.9) through a simulation approach. Note that the conditional posteriors (2.10) and (2.11) are members of the same exponential family as the conditional priors. Posterior simulation of the parameters θ_1 and θ_2 is pretty straightforward through a Gibbs sampler technique (see e.g., Gelfand and Smith, 1990). Specifically, starting with an arbitrary initial pair of values $(\theta_1^{(0)}, \theta_2^{(0)})$, at the kth iteration we draw a sample $(\theta_1^{(k)}, \theta_2^{(k)})$ based on the values at the $(k-1)$th iteration through the conditional posterior distributions (2.10) and (2.11):

$$\theta_1^{(k)} \sim f(\theta_1|\theta_2^{(k-1)}, x, y) \tag{4.1}$$

$$\theta_2^{(k)} \sim f(\theta_2|\theta_1^{(k)}, x, y). \tag{4.2}$$

for $k = 1, 2, \ldots$. As k approaches infinity, the joint distribution of θ_1 and θ_2 can be shown to approach the posterior $f(\theta_1, \theta_2|x, y)$. So for sufficiently large k, after discarding the first k samples, the rest of $(\theta_1^{(k^*)}, \theta_2^{(k^*)})(k^* > k)$ can be regarded as approximately simulated values from the posterior distribution $f(\theta_1, \theta_2|x, y)$. Furthermore, these could also be treated as a simulated sample from the density of any variable transformation $w(\theta_1, \theta_2)$ with the function $w: R^2 \to R$. In particular, here interest will focus on the choice $w(\theta_1, \theta_2) = \theta_1/\theta_2$. We will thus be able to approximate the posterior distribution of θ_1/θ_2 by providing the sample mean, variance, and $100(1-\alpha)\%$ credible interval for θ_1/θ_2. Analogously, one could also consider the difference $\theta_1 - \theta_2$.

5. A Reparameterization

Just as in Arnold et al. (1998), we can consider inference on the ratio θ_1/θ_2 through the following transformation of parameters before analysis

$$\lambda_1 = \theta_1/\theta_2, \quad \lambda_2 = \theta_2.$$

The reparameterized likelihood, see (1.3), becomes

$$L(\lambda_1, \lambda_2) \propto [g_1(\lambda_1\lambda_2)]^{n_1}[g_2(\lambda_2)]^{n_2}\lambda_2 \exp\left\{\sum_{i=1}^{n_1} t_1(x_i)\lambda_1\lambda_2 + \sum_{j=1}^{n_2} t_2(y_j)\lambda_2\right\}. \tag{5.1}$$

Assuming that $g_1(\lambda_1\lambda_2)$ could be factorized as $\tilde{g}_1(\lambda_1)\tilde{g}_2(\lambda_2)$, one can still obtain an exponential family for the likelihood of (λ_1, λ_2)

$$L(\lambda_1, \lambda_2) \propto [\tilde{h}_1(\lambda_1)]^{n_1}[\tilde{h}_2(\lambda_2)]^{n_2}\exp\left\{\sum_{j=1}^{n_2} t_2(y_j)\lambda_2 + \sum_{i=1}^{n_1} t_1(x_i)\lambda_1\lambda_2\right\}$$

with $\tilde{h}_1(\lambda_1) = \tilde{g}_1(\lambda_1), \tilde{h}_2(\lambda_2) = \lambda_2^{\frac{1}{n_2}} g_2(\lambda_2)[\tilde{g}_2(\lambda_2)]^{\frac{n_1}{n_2}}$. The joint conjugate prior of (λ_1, λ_2) has the same form as (2.8) previously

$$f(\lambda_1, \lambda_2) \propto [\tilde{h}_1(\lambda_1)]^{\eta_1}[\tilde{h}_2(\lambda_2)]^{\eta_2}\exp\{\beta\lambda_1 + \gamma\lambda_2 + \alpha\lambda_1\lambda_2\} \tag{5.2}$$

which leads to the joint posterior

$$f(\lambda_1, \lambda_2 | x, y) \propto [\tilde{h}_1(\lambda_1)]^{\eta_1+n_1}[\tilde{h}_2(\lambda_2)]^{\eta_2+n_2} \times \exp\left\{\beta\lambda_1 + \left(\gamma + \sum_{j=1}^{n_2} t_2(y_j)\right)\lambda_2 + \left(\alpha + \sum_{i=1}^{n_1} t_1(x_i)\right)\lambda_1\lambda_2\right\} \tag{5.3}$$

with two conditional conjugate posteriors being

$$f(\lambda_1 | \lambda_2, x, y) \propto [\tilde{h}_1(\lambda_1)]^{\eta_1+n_1}\exp\left\{\left[\beta + \left(\alpha + \sum_{i=1}^{n_1} t_1(x_i)\right)\lambda_2\right]\lambda_1\right\} \tag{5.4}$$

$$f(\lambda_2 | \lambda_1, x, y) \propto [\tilde{h}_2(\lambda_2)]^{\eta_2+n_2}\exp\left\{\left[\left(\gamma + \sum_{j=1}^{n_2} t_2(y_j)\right) + \left(\alpha + \sum_{i=1}^{n_1} t_1(x_i)\right)\lambda_1\right]\lambda_2\right\}. \tag{5.5}$$

Gibbs sampler provides a natural way for posterior simulation of λ_1 and λ_2 as before.

6. Examples

We present two examples to illustrate the application of our methods.

6.1. *Testing the Ratio of Variances σ_1^2/σ_2^2 in Normal Distributions*

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be two independent samples from normal distributions with known means μ_1, μ_2 and unknown variances σ_1^2, σ_2^2.

The normal density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad x > 0 \tag{6.1}$$

which is one parameter exponential family density with the natural parameter $\theta = 1/\sigma^2$ and $r(x) = 1/\sqrt{2\pi}$, $g(\theta) = \theta^{1/2}$, $t(x) = -(x-\mu)^2/2$. From (2.8), one obtains a joint prior density of θ_1 and θ_2, conveniently reparameterized $\frac{\eta_1}{2} = \gamma_1 - 1$, $\frac{\eta_2}{2} = \gamma_2 - 1$, and positive constants $c_1 = -\beta$, $c_2 = -\gamma$, $c = -\alpha$

$$f(\theta_1, \theta_2) \propto \theta_1^{\gamma_1 - 1}\theta_2^{\gamma_2 - 1}\exp\{-(c_1\theta_1 + c_2\theta_2 + c\theta_1\theta_2)\} \tag{6.2}$$

Both conditionals are Gamma distributions with fixed shape parameters γ_1 and γ_2, i.e., $\theta_1|\theta_2 \sim \text{Gamma}(\gamma_1, c_1 + c\theta_2)$ and $\theta_2|\theta_1 \sim \text{Gamma}(\gamma_2, c_2 + c\theta_1)$. This model is discussed in Moschopoulos and Staniswalis (1994). Equivalently, both $\sigma_1^2|\sigma_2^2$ and $\sigma_2^2|\sigma_1^2$ have inverse Gamma distributions as usual conjugate priors for the scale of normal distributions. Then the joint posterior becomes

$$\begin{aligned} f(\theta_1, \theta_2|x, y) \propto\ & \theta_1^{\gamma_1 + n_1/2 - 1}\theta_2^{\gamma_2 + n_2/2 - 1} \\ & \times \exp\left\{ -\left[\left(c_1 + \frac{1}{2}\sum_{i=1}^{n_1}(x_i - \mu_1)^2\right)\theta_1 \right.\right. \\ & \left.\left. + \left(c_2 + \frac{1}{2}\sum_{j=1}^{n_2}(y_j - \mu_2)^2\right)\theta_2 + c\theta_1\theta_2\right]\right\} \end{aligned} \tag{6.3}$$

and it has the same form as the prior distribution with the updated hyperparameters shown in Table 1. The two conditional gamma posteriors are

$$\theta_1|\theta_2, x, y \sim \text{Gamma}\left(\gamma_1 + n_1/2, \left(\frac{1}{2}\sum_{i=1}^{n_1}(x_i - \mu_1)^2 + c_1 + c\theta_2\right)\theta_1\right) \tag{6.4}$$

$$\theta_2|\theta_1, x, y \sim \text{Gamma}\left(\gamma_2 + n_2/2, \left(\frac{1}{2}\sum_{j=1}^{n_2}(y_j - \mu_2)^2 + c_2 + c\theta_1\right)\theta_2\right). \tag{6.5}$$

6.1.1. Hyperparameters Settings. In this situation of Gamma prior conditionals, we have $h_1(\theta_2) = (c_1 + c\theta_2)^{\gamma_1}$, $h_2(\theta_1) = (c_2 + c\theta_1)^{\gamma_2}$ in (3.1) and (3.2), and the following

Table 1
Normal case: adjustments in the parameters in the prior family combined with likelihood

Parameters	Prior value	Posterior value
γ_1	γ_1^*	$\gamma_1^* + \frac{n_1}{2}$
γ_2	γ_2^*	$\gamma_2^* + \frac{n_2}{2}$
c_1	c_1^*	$c_1^* + \frac{1}{2}\sum_{i=1}^{n_1}(x_i - \mu_1)^2$
c_2	c_2^*	$c_2^* + \frac{1}{2}\sum_{j=1}^{n_2}(y_j - \mu_2)^2$
c	c^*	c^*

conditional means and variances from (3.3)

$$E(\theta_1|\theta_2) = \frac{\gamma_1}{c_1 + c\theta_2}, \quad Var(\theta_1|\theta_2) = \frac{\gamma_1}{(c_1 + c\theta_2)^2}$$
$$E(\theta_2|\theta_1) = \frac{\gamma_2}{c_2 + c\theta_1}, \quad Var(\theta_2|\theta_1) = \frac{\gamma_2}{(c_2 + c\theta_1)^2}.$$

As proposed in Sec. 3 for eliciting values for conditional means and variances of the parameters θ_1 and θ_2 with $\theta_1 = \sigma_1^{-2}$ and $\theta_2 = \sigma_2^{-2}$ in this normal case, we choose a series of values for $\theta_1, \theta_2, \zeta^{(1)}, \zeta^{(2)}, \eta^{(1)}, \eta^{(2)}$ in the conditional moments (3.4) and (3.5) leading to the following equations:

$$\frac{[\zeta_i^{(1)}]^2}{\zeta_i^{(2)}} = \gamma_1, \ \frac{\zeta_i^{(1)}}{\zeta_i^{(2)}} = c_1 + c\theta_2^{(i)}, \quad i = 1, 2, \ldots, m_1 \tag{6.6}$$

$$\frac{[\eta_j^{(1)}]^2}{\eta_j^{(2)}} = \gamma_2, \ \frac{\eta_j^{(1)}}{\eta_j^{(2)}} = c_2 + c\theta_1^{(j)}, \quad j = 1, 2, \ldots, m_2 \tag{6.7}$$

Five equations are needed to solve for the unknown parameters $\gamma_1, \gamma_2, c_1, c_2$, and c. In the case of more equations elicited, there will be no choice of these parameters to make all the above equations hold exactly. We propose obtaining the least squares estimates of γ's and c's from these linear equations. Specifically, we use the linear model $\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon$ where the unknown vector $\mathbf{b} = (\gamma_1, \gamma_2, c_1, c_2, c)'$, the design matrix $\mathbf{X}$, and the response vector $\mathbf{Y}$ are

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{m_1} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1}_{m_2} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1}_{m_1} & 0 & \theta_2 \\ 0 & 0 & 0 & \mathbf{1}_{m_2} & \theta_1 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \mathbf{Y}_4 \end{pmatrix}$$

with $\mathbf{1}_p$ being a p-column vector of all elements 1, $\theta_1 = (\theta_1^{(1)}, \ldots, \theta_1^{(m_2)})'$, $\theta_2 = (\theta_2^{(1)}, \ldots, \theta_2^{(m_1)})'$, $\mathbf{Y}_1 = \left(\frac{[\zeta_i^{(1)}]^2}{\zeta_i^{(2)}}\right)_{i=1,\ldots,m_1}$, $\mathbf{Y}_2 = \left(\frac{[\eta_j^{(1)}]^2}{\eta_j^{(2)}}\right)_{j=1,\ldots,m_2}$, $\mathbf{Y}_3 = \left(\frac{\zeta_i^{(1)}}{\zeta_i^{(2)}}\right)_{i=1,\ldots,m_1}$, $\mathbf{Y}_4 = \left(\frac{\eta_j^{(1)}}{\eta_j^{(2)}}\right)_{j=1,\ldots,m_2}$ and $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2\mathbf{I}$. The least squares solution is $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

6.1.2. Data Analysis. A data set in Mason et al. (1989) is used as an application of our method. The data are tensile-strength measurements on wire that is used in industrial equipment. The wire is manufactured by drawing the base metal through a die that has a specified diameter. The data displayed in Table 2 are tensile strengths of 18 samples taken from large spools of wire made with each of 2 dies. It is of interest to determine whether the use of different dies will affect the tensile strength of the wire. A larger sample indicates that the normal probability distribution is a reasonable assumption for these tensile-strength measurements and the mean measurement for the two are $\mu_1 = 84.7$ and $\mu_2 = 84.5$, respectively. We are only concerned with scale parameters σ_1 and σ_2. In this case, the maximum likelihood estimator (MLE) estimate for σ_1^2/σ_2^2 is 0.4697 and a 95% classical confidence interval for it based on the F distribution is (0.1810, 1.2191). For Bayesian analysis,

Table 2
Two normal samples $n_1 = n_2 = 18$

Die 1		Die 2	
85.769	86.725	79.424	82.912
86.725	84.292	81.628	83.185
87.168	84.513	82.692	86.725
84.513	86.725	86.946	84.070
84.513	84.513	86.725	86.460
83.628	82.692	85.619	83.628
82.912	83.407	84.070	84.513
84.734	84.070	85.398	84.292
82.964	85.337	86.946	85.619

first we need reasonable hyperparameter values. Suppose that our expert provided prior information in the form of the following subjective conditional means and variances:

$$E(\theta_1|\theta_2 = 0.03) = 0.2815, \quad Var(\theta_1|\theta_2 = 0.03) = 0.0184$$
$$E(\theta_1|\theta_2 = 0.09) = 0.2782, \quad Var(\theta_1|\theta_2 = 0.09) = 0.0180$$
$$E(\theta_1|\theta_2 = 0.012) = 0.2775, \quad Var(\theta_1|\theta_2 = 0.012) = 0.0180$$

$$E(\theta_2|\theta_1 = 0.03) = 0.252, \quad Var(\theta_2|\theta_1 = 0.03) = 0.0118$$
$$E(\theta_2|\theta_1 = 0.09) = 0.249, \quad Var(\theta_2|\theta_1 = 0.09) = 0.0115$$
$$E(\theta_2|\theta_1 = 0.012) = 0.248, \quad Var(\theta_2|\theta_1 = 0.012) = 0.01146$$

The least square estimates of these hyperparameters and their updated values in the posterior distribution are shown in Table 3.

6.1.3. Posterior Inference. A Gibbs sampler approach was used to generate a simulated sample from the posterior density of $(\sigma_1^{-2}, \sigma_2^{-2})$ through their posterior conditionals (6.4) and (6.5). 2500 runs were implemented and the first 500 discarded. The remaining 2000 realizations were used to obtain the approximate posterior distribution for $\lambda = \sigma_1^2/\sigma_2^2$. The choice of 500 warm-up iterations followed

Table 3
Normal data: hyperparameter values in the prior and posterior distributions

Parameters	Prior value	Posterior value
γ_1	4.2948	13.2948
γ_2	5.3800	14.3800
c_1	15.1940	32.0388
c_2	21.3531	57.2160
c	2.4549	2.4549

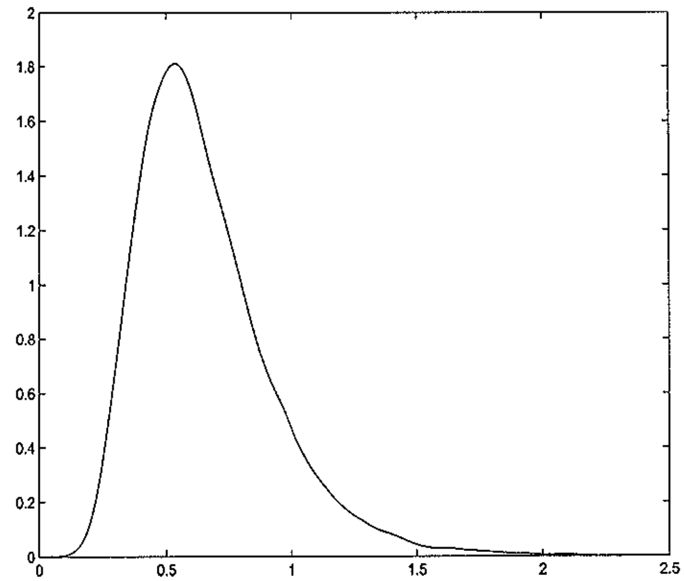

Figure 1. Informative priors: simulated density of σ_1^2/σ_2^2 using the Gibbs sampler with 2500 replications and 500 burn-in runs.

by 2000 utilized iterations was somewhat arbitrary. However, multiple simulations involving warm-ups of 500–3000 and run lengths of 2000–10000 yielded results that were graphically barely distinguishable from the 500–2000 results. The smoothed histogram shown in Fig. 1 was obtained by Gaussian kernel density estimation. The corresponding approximate posterior mean and variance are

$$E(\lambda) = 0.6524, \quad Var(\lambda) = 0.0719.$$

The 2.5th and 97.5th percentile for the simulated values λ yield approximate 95% credible interval for σ_1^2/σ_2^2, namely (0.2849, 1.3123). This Bayesian analysis, using an informative prior, confirms the classical statistician's conclusion that there is little difference between the two dies in the study.

6.1.4. Noninformative Prior Analysis. If no prior information is available, a completely diffuse prior, $f(\sigma_1^2, \sigma_2^2) \propto 1$, or the standard independent non informative prior $f(\sigma_1^2, \sigma_2^2) \propto (\sigma_1^2\sigma_2^2)^{-1}$, could be used in the analysis. These are both improper priors that yield proper posterior distributions and use of the Gibbs sampler is legitimate. Under these two cases, using the same data, the posterior density curves are shown in Figs. 2 and 3 and approximate posterior means and variances are shown in Table 4. The results are not greatly at variance with the informative prior or the classical analysis.

For a more reliable analysis in lack of prior information, one could first begin with a diffuse prior and then "guess" a representative sample to be combined with the prior using Table 1. The resulting posterior hyperparameters then become the prior ones for subsequent analysis of the real data set.

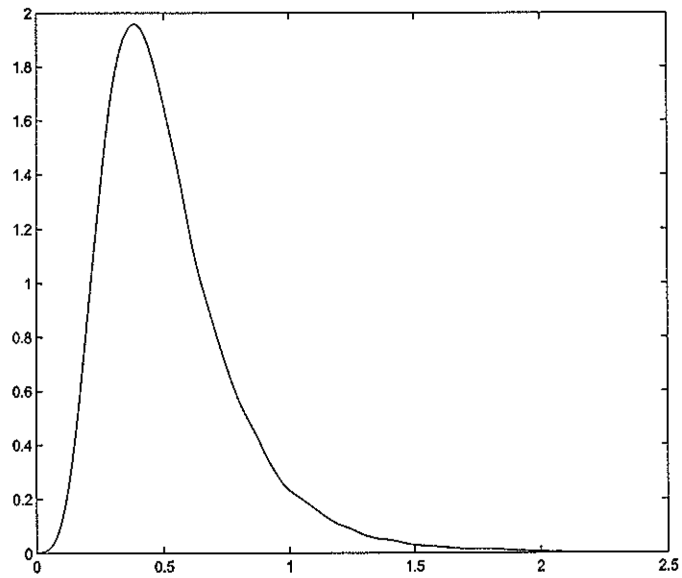

Figure 2. Diffuse prior: simulated density of σ_1^2/σ_2^2 using the Gibbs sampler with 2500 replications and 500 burn-in runs.

6.2. *Testing the Ratio of Scales β_1/β_2 in Weibull Distributions*

Suppose that two independent samples $x = (x_1, \ldots, x_{n_1})$ and $y = (y_1, \ldots, y_{n_2})$, are from Weibull distributions with known shape parameters α_1, α_2 and unknown scales β_1 and β_2 in which we are interested. The Weibull (α, β) probability density

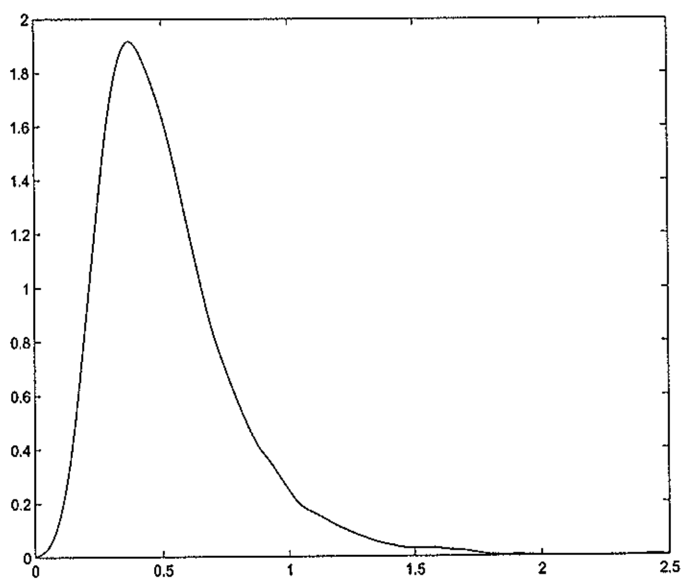

Figure 3. Noninformative independent prior: simulated density of σ_1^2/σ_2^2 using the Gibbs sampler with 2500 replications and 500 burn-in runs.

Table 4
Normal data: posterior inference of σ_1^2/σ_2^2 for different priors

Prior	Mean	Variance	95% CI
Informative	0.6524	0.0719	(0.2849, 1.3123)
Diffuse	0.5292	0.0764	(0.1763, 1.2327)
Independent	0.5294	0.0759	(0.1829, 1.2203)
Classical (F-stat)	0.4697 (MLE)		(0.1810, 1.2191)

function (pdf) is

$$f(x) = \alpha\beta x^{\alpha-1} e^{-x^{\alpha}\beta}, \quad x > 0, \tag{6.8}$$

a one-parameter exponential family density with scale parameter $\theta = \beta$ and $r(x) = \alpha x^{\alpha-1}$, $g(\theta) = \theta$, $t(x) = -x^{\alpha}$. The likelihood of the data is

$$L(\beta_1, \beta_2) \propto \beta_1^{n_1}\beta_2^{n_2}\exp\left\{-\left(\sum_{i=1}^{n_1} x_i^{\alpha_1}\beta_1 + \sum_{j=1}^{n_2} y_j^{\alpha_2}\beta_2\right)\right\}. \tag{6.9}$$

We choose the same form of the joint prior density as in the normal case with positive constants $\gamma_1, \gamma_2, c_1, c_2, c$

$$f(\beta_1, \beta_2) \propto \beta_1^{\gamma_1-1}\beta_2^{\gamma_2-1}\exp\{-(c_1\beta_1 + c_2\beta_2 + c\beta_1\beta_2)\}. \tag{6.10}$$

Again, both conditionals follow Gamma distributions with fixed shape parameters, i.e., $\beta_1|\beta_2 \sim \text{Gamma}(\gamma_1, c_1 + c\beta_2)$ and $\beta_2|\beta_1 \sim \text{Gamma}(\gamma_2, c_2 + c\beta_1)$. The posterior

$$\begin{aligned} f(\beta_1, \beta_2|x, y) &\propto \beta_1^{\gamma_1+n_1-1}\beta_2^{\gamma_2+n_2-1} \\ &\quad \times \exp\left\{-\left[\left(c_1 + \sum_{i=1}^{n_1} x_i^{\alpha_1}\right)\beta_1 + \left(c_2 + \sum_{j=1}^{n_2} y_j^{\alpha_2}\right)\beta_2 + c\beta_1\beta_2\right]\right\} \end{aligned} \tag{6.11}$$

has the same form as the prior with the updated hyperparameters shown in Table 5.

Table 5
Weibull case: adjustments in the parameters in the prior family combined with likelihood

Parameters	Prior value	Posterior value
γ_1	γ_1^*	$\gamma_1^* + n_1$
γ_2	γ_2^*	$\gamma_2^* + n_2$
c_1	c_1^*	$c_1^* + \sum_{i=1}^{n_1} x_i^{\alpha_1}$
c_2	c_2^*	$c_2^* + \sum_{j=1}^{n_2} y_j^{\alpha_2}$
c	c^*	c^*

6.2.1. Hyperparameter Settings. Since the prior distributions are the same as the normal case before, we conclude that we should have the same relations for the first two moments

$$\frac{[\zeta_i^{(1)}]^2}{\zeta_i^{(2)}} = \gamma_1, \frac{\zeta_i^{(1)}}{\zeta_i^{(2)}} = c_1 + c\beta_2^{(i)}, \quad i = 1, 2, \ldots, m_1 \tag{6.12}$$

$$\frac{[\eta_j^{(1)}]^2}{\eta_j^{(2)}} = \gamma_2, \frac{\eta_j^{(1)}}{\eta_j^{(2)}} = c_2 + c\beta_1^{(j)}, \quad j = 1, 2, \ldots, m_2 \tag{6.13}$$

for the elicited values $\beta_1^{(i)}, \zeta_i^{(1)}, \zeta_i^{(2)}, \beta_2^{(j)}, \eta_j^{(1)}, \eta_j^{(2)}, i = 1, \ldots, m_1, j = 1, \ldots, m_2$. The same techniques in the previous example can be adopted to solve γ's and c's.

6.2.2. Data Analysis. Here we use the data sets in Harter and Moore (1965) for the application of our method. The data are simulated from two populations whose survival distributions are Weibull. Table 6 lists the numerical values for both samples. The shape parameters are $\alpha_1 = 1.94$ and $\alpha_2 = 2.75$, and the location parameters are 10 and 20, respectively. We are concerned with the scale parameters β_1 and β_2. In this case, the MLE estimate for β_1/β_2 is 40.7943 and a 95% classical confidence interval for it based on the F distribution is (26.2363, 63.4303). For Bayesian analysis, our guess values for prior information are as follows

$$E(\beta_1|\beta_2 = 0.01) = 0.276$$
$$Var(\beta_1|\beta_2 = 0.01) = 0.0181$$
$$E(\beta_2|\beta_1 = 0.08) = 0.205$$
$$E(\beta_2|\beta_1 = 0.015) = 0.206$$
$$Var(\beta_2|\beta_1 = 0.08) = 0.0085.$$

Based on these guessed values, we obtain the five hyperparameters from (6.12) and (6.13), and their updated values in the posterior are shown in Table 7.

Table 6
Two Weibull samples $n_1 = n_2 = 40$

Weibull 1				Weibull 2			
15	65	92	124	40.9	81.4	108.7	129.7
20	68	95	126	52.2	85.4	109.3	130.8
27	68	100	127	53.2	86.0	111.6	134.1
42	71	102	134	59.4	86.3	113.1	137.5
43	74	102	149	60.0	87.4	114.2	139.2
43	75	112	152	66.8	88.5	117.7	140.3
44	75	113	153	77.3	89.9	121.6	143.0
46	76	116	161	78.0	92.4	121.9	143.8
64	77	117	168	79.7	93.0	127.6	183.3
65	78	124	205	81.1	93.2	128.0	185.1

Table 7
Weibull data: hyperparameter values in the prior and posterior distributions

Parameters	Prior value	Posterior value
γ_1	4.2162	44.2162
γ_2	5.1873	45.1873
c_1	15.2301	2.62×10^5
c_2	24.7440	1.09×10^7
c	3.9037	3.9037

6.2.3. Posterior Inference. The Gibbs sampler approach is used to generate a simulated sample from the posterior density of (β_1, β_2). 3000 runs were implemented and the first 500 as burn-in period. The remaining 2500 realizations were used to obtain approximate posterior distribution for $\lambda = \beta_1/\beta_2$. Multiple simulations involving burn-in of 500–3000 and run lengths of 3000–10000 yielded similar results as the 500–2500. The smoothed histogram is shown in Fig. 4 using Gaussian kernel density estimation. The corresponding approximate posterior mean and standard error are

$$E(\lambda) = 41.6974, \quad SE(\lambda) = 8.9938.$$

The 2.5th and 97.5th percentile for the simulated values λ yield approximate 95% credible interval for β_1/β_2, namely (26.7831, 61.7613). This result shows that two scales are greatly different and the first is much larger than the second one,

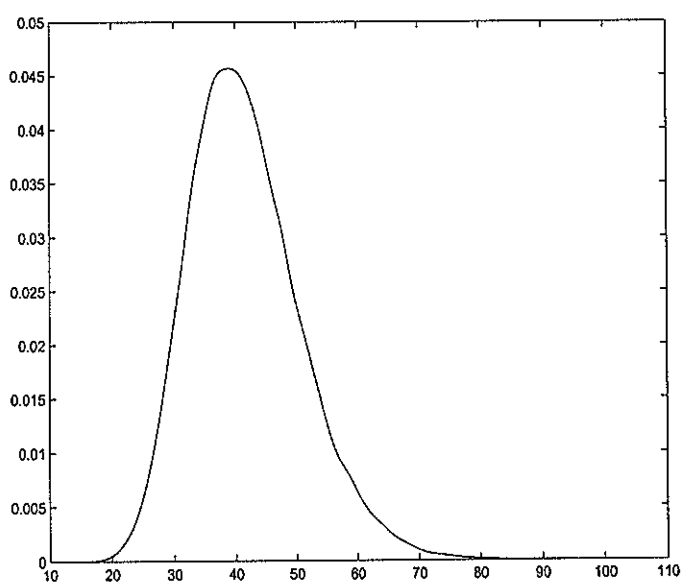

Figure 4. Informative priors: simulated density of β_1/β_2 using the Gibbs sampler with 3000 replications and 500 burn-in runs.

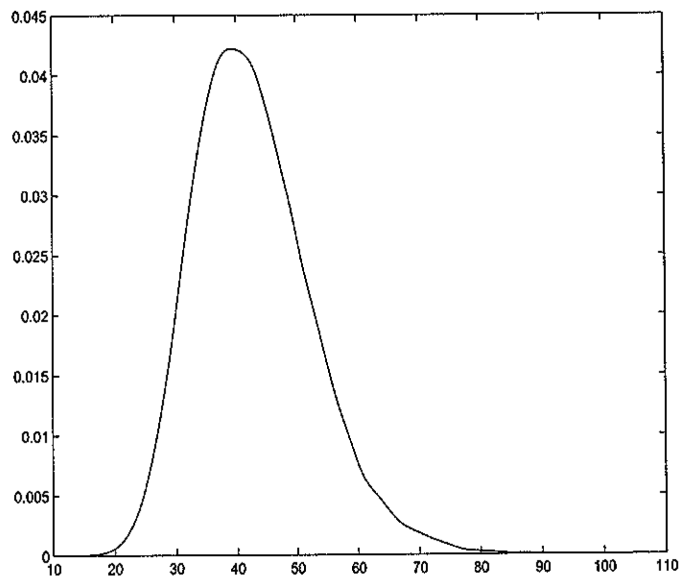

Figure 5. Diffuse prior: simulated density of β_1/β_2 using the Gibbs sampler with 3000 replications and 500 burn-in runs.

which is consistent with the conclusion in the classical hypothesis test (Harter and Moore, 1965, p. 235).

6.2.4. Noninformative Prior Analysis. A vague prior, $f(\beta_1, \beta_2) \propto 1$ and the standard independent noninformative prior $f(\beta_1, \beta_2) \propto (\beta_1\beta_2)^{-1}$, could also be used for the

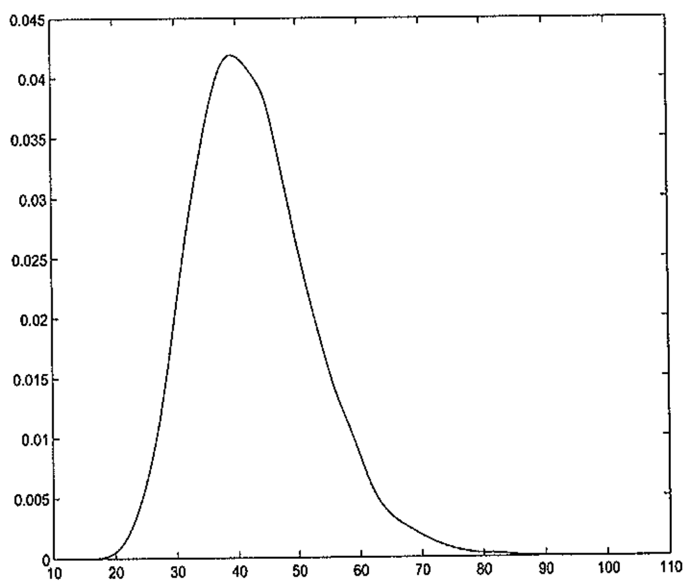

Figure 6. Noninformative independent prior: simulated density of β_1/β_2 using the Gibbs sampler with 3000 replications and 500 burn-in runs.

Table 8
Weibull data: posterior inference of β_1/β_2 for different priors

Prior	Mean	Standard error	95% CI
Informative	41.6974	8.9938	(26.7831, 61.7613)
Diffuse	42.7571	9.5832	(26.9973, 64.1665)
Independent	42.7861	9.7536	(26.8471, 64.5607)
Classical (F-stat)	40.7943 (MLE)		(26.2363, 63.4303)

analysis. The generated samples through Gibbs sampling produce the estimates of posterior density curves shown in Figs. 5 and 6 and approximate posterior means and variances are shown in Table 8. These results are comparable with those of the informative prior or the classical analysis.

7. Conclusion

We proposed a unified Bayesian method for comparing scale parameters of two distributions in an exponential family, assuming other parameters such as location or shape are known. A conjugate joint prior distribution is determined conveniently via fully specified conditionals. Vague and independent prior analysis are recognizable as special cases in our general prior settings. Several methods for the assessment of prior hyperparameter values have been given. Two analyses of real data are provided to illustrate our methods for applications. Matlab codes for the procedures presented in this article are available from the authors.

Acknowledgments

This research is partially supported by BBRC/RCMI NIH grant 2G12RR08124 and University Research Institute (UTEP) grant. The authors thank the Editor and referees for their helpful comments.

References

Arnold, B. C., Castillo, E., Sarabia, J. M. (1992). *Conditionally Specified Distributions. Lecture Notes in Statistics.* Eds. Berger, J., Fienberg, S., Gani, J., Krickeberg, K., Olkin, I., Singer, B. New York: Springer-Verlag.

Arnold, B. C., Castillo, E., Sarabia, J. M. (1998a). Bayesian analysis for classical distributions using conditionally specified priors. *Sankhyā Series B, Pt.* 2, 228–245.

Arnold, B. C., Castillo, E., Sarabia, J. M. (1998b). The use of conditionally conjugate priors in the study of ratios of Gamma scale parameters. *Comput. Statist. Data Anal.* 27: 125–139.

Arnold, B. C., Castillo, E., Sarabia, J. M. (1999). *Conditional Specification of Statistical Models.* New York: Springer.

Arnold, B. C., Strauss, D. J. (1988). Bivariate distributions with exponential conditionals. *J. Amer. Statist. Assoc.* 83:522–527.

Berger, J. O. (1985). *Statistical Decision Theory and Data Analysis, 2nd ed.* New York: Springer.

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

Harter, H. L., Moore, A. H. (1965). Maximum likelihood estimation of the parameters of Gamma and Weibull populations from complete and censored samples. *Technometrics* 7:639–243.

Mason, R. L., Gunst, R. F., Hess, J. L. (1989). *Statistical Design and Analysis of Experiments.* New York: John Wiley & Sons.

Moschopoulos, P., Staniswalis, J. G. (1994). Estimation given conditionals from an exponential family. *Amer. Statist. Assoc.* 48(4):271–275.