CHAPTER I

BAYESIAN INFERENCE FOR SOME INCOMPLETELY SPECIFIED

PRIOR DISTRIBUTIONS

1. Introduction

In Bayesian approach the prior knowledge about the parameters is combined with the information in the sample to obtain posteriori inferences. Thus given sample observations, this approach typically depends on the prior distribution of the parameters θ, say, involved in the likelihood or the probability distribution of the sample observations. There exist situations where we have knowledge that the prior distribution of θ belongs to a certain class, but the particular member of the class that corresponds to the sample is sometimes not known; e.g. we may know the form of the prior density of θ, but some of the parameters or functions involved in this density may not be known. In this chapter, we consider problems concerning such incomplete specifications of the prior density.

In applied decision problems, the Bayes procedures usually depend on expectations of functions obtained
from posterior distributions; when these expectations depend on the unknown parameters or functions involved in the prior density \( g(\theta) \) of \( \theta \), these procedures cannot be applied directly. We suggest that, instead of the incompletely specified prior density \( g(\theta) \) of \( \theta \), a completely specified 'reduced prior' may be used. We give methods of obtaining the reduced prior distributions by using the uniform distribution for location and log scale parameters. The methods suggested here are useful in many applied Bayesian inference problems. This is illustrated in applications to samples from normal and exponential populations.

2. The Reduced Prior Distribution

We first define the dependence of a density function on the parameters.

The density function of \( y \) is said to depend on a location parameter \( \alpha \) and a scale parameter \( \beta \) if it is of the form

\[
h \left( \frac{y - \alpha}{\beta} \right) \left( \frac{1}{\beta} \right)
\]

where \( h \) integrates with respect to \( y \) (see Fisher(1936), Jeffreys(1961), Pitman(1933)). On similar lines the density
function of several variables \( y_i \) is said to depend on location parameters \( \alpha_i \) and scale parameters \( \beta_i \), if it is expressible in the form

\[
h\left( \frac{y_1 - \alpha_1}{\beta_1}, \frac{y_2 - \alpha_2}{\beta_2}, \ldots \right) \left( \frac{1}{\beta_1} \right) \left( \frac{1}{\beta_2} \right) \ldots
\]

**Definition of the Reduced Prior Distribution**

Let \( l(x | \theta) \) denote the likelihood corresponding to the sample \( x \) and parameters \( \theta \). It is given that the prior density of \( \theta \) is \( g(\theta) \), where some parameters \( \gamma \) involved in \( g(\theta) \) are not known.

For obtaining the posteriori inferences, we may then adopt a new prior measure \( \tilde{g}(\theta) \) called the reduced prior measure of \( g(\theta) \), and this is obtained by the following method:

**Method (a)**

Let \( T \) be a transformation of \( \theta \) to \( \lambda \),

\[ \theta = (\theta_1, \theta_2) \quad \text{and} \quad \lambda = (\lambda_1, \lambda_2) \quad \text{where} \quad \theta_2 \quad \text{has} \quad (r+s) \quad \text{components} \quad (r+s) \leq p, \quad \theta_2 = (\alpha_1, \alpha_2, \ldots, \alpha_r, \beta_1, \beta_2, \ldots, \beta_s) \quad \text{and} \quad \lambda_1 \quad \text{does not involve} \quad \gamma, \quad \text{such that} \]

(1) The prior marginal density \( g_1(\lambda_1) \) of \( \lambda_1 \) is independent of \( \gamma \),
(ii) The prior marginal density \( g_1(\theta_2) \) of \( \theta_2 \) involves \( \beta \), and

(iii) The components \((\alpha_1, \alpha_2, \ldots, \alpha_r)\) and \((\beta_1, \beta_2, \ldots, \beta_s)\)
of \( \theta_2 \) are respectively the location and the scale parameters in the function

\[
p(x, \theta_1, \theta_2) = \ell(x \mid \theta_2) g'(\theta_1, \theta_2) \quad \ldots \quad (2.1)
\]

where

\[
g'(\theta_1, \theta_2) = g_1(\lambda_1) \frac{\beta \lambda_1}{\lambda_1} \quad \ldots \quad (2.2)
\]

Then the function \( \bar{g}_T(\theta) \) defined by

\[
\bar{g}_T(\theta) = g'(\theta_1, \theta_2) p(\theta_2) \quad \ldots \quad (2.3)
\]

where

\[
p(\theta_2) = \prod_{i=1}^{s} (\beta_1)^{-1} \quad \ldots \quad (2.4)
\]

is called the reduced prior measure of \( g(\theta) \) with respect to the transformation \( T \).

The optimum transformations \( T \) correspond to the minimum \( (r+s) \), the component size of \( \theta_2 \).
Discussion

This method of reduction is justified on the following basis. We argue that, since we are not in a position to use the prior density \( g(\theta) \) in the given form, we may use a prior measure which loses as little knowledge as possible of the prior density \( g(\theta) \) in eliminating the unknown parameters \( \psi \). In the method suggested, this elimination involves reduction in the first instance to a known function \( g_1(\lambda_1) \) obtainable from \( g(\theta) \). And since

\[
g(\lambda) = g_1(\lambda_1) g(\theta_2 | \lambda_1), \quad (2.5)
\]

\( g(\theta_2 | \lambda_1) \) involves \( \psi \). Also by condition (ii), the marginal density \( g_1(\theta_2) \) of \( \theta_2 \) involves \( \psi \). These functions represent the factors concerning the knowledge of \( \hat{\eta} \). Since these are not known completely, a reasonable prior measure is obtained to replace \( g(\theta) \). For this, we put

\[
p(x, \lambda_1 | \theta_2) = 1(x, \theta) g_1(\lambda_1) \quad (2.6)
\]

as representing the conditional joint distribution of \( x, \lambda_1 \) given \( \theta_2 \). Further on transforming \( \lambda_1 \) to \( \theta_1 \), we obtain the conditional joint distribution of \( x, \theta_1 \) given \( \theta_2 \).
Now by condition (iii), the components of $\theta_C$ are the location and scale parameters in the density function $p(x, \theta_1, \theta_2)$. We adopt for $\theta_2$ the prior distribution (2.4), i.e., the uniform prior distribution for location and log scale parameters advocated by Jeffreys (1961) to express initial ignorance of these parameters. See Lindley et al. (1960), Box and Tiao (1962, 1964), Cattman and Tiao (1964), Tiao and Sellner (1964). The effect of using this prior distribution is one of expressing that, the only information about these parameters is that contained in $p(x, \theta_1, \theta_2)$, the function defined in (2.7). Also, since we wish to utilise as much knowledge of $\theta$ as is possible to get by means of this method, it is natural to seek transformations $T$ which correspond to minimum $\nu + \sigma$.

This method of reducing the prior distribution depends on the particular transformation $T$ that has been adopted. Since the prior distribution is used for making posterior inferences about certain parametric functions, the transformation is often suggested by the problem at hand.

When the sample observations are given, the Bayesian is mainly concerned with posterior distributions. It is therefore interesting to note that in samples from normal
and exponential populations, the posterior distributions with respect to the reduced prior distributions are slight modifications of the corresponding distributions with respect to the original prior distribution.

It is sometimes possible to obtain the reduced prior $g(\theta)$ without using a transformation $T$. In such cases it is obtained by means of the following method.

**Method (b)**

If the conditional prior density $g(\theta_1 | \theta_2)$ of $\theta_1$ given $\theta_2$ is independent of the parameter $\gamma$ and if the function $p(x, \theta_1 | \theta_2)$ defined by

$$p(x, \theta_1 | \theta_2) = l(x | \theta) g(\theta_1 | \theta_2)$$

has the components of $\theta_2$ as location and scale parameters, then a reduced prior measure of $\theta$ is

$$\tilde{g}(\theta) = g(\theta_1 | \theta_2) p(\theta_2)$$

where $p(\theta_2)$ is given by (2.4).

This method is applicable when $\theta_2$ is sufficient for the distribution $g(\theta)$, i.e., when the density
function $g(\theta)$ factorises in the form

$$g(\theta) = g(\theta_1, \theta_2) g_1(\theta_2), \quad \ldots \quad (2.10)$$

and the conditional density $g(\theta_1 | \theta_2)$ is independent of $\theta_2$. However, comparing the optimum size of $\theta_2$, this method is only as good as the method (a). Since it can be noted that when (2.10) holds, there always exists a transformation $T$ which yields (2.9) as the reduced prior measure. And in some cases by means of the method (a), it may be possible to get $\theta_2$ of smaller component size. Since it is sometimes possible to obtain from $g_1(\theta_2)$ of (2.10) a prior measure $\overline{p}_T(\theta_2)$ different from $p(\theta_2)$, a reduced prior measure for $\theta$ is given by

$$g(\theta) = g(\theta_1, \theta_2) \overline{p}_T(\theta_2), \quad \ldots \quad (2.11)$$

The component-size $(r + \rho_2)$ in (2.11) is smaller than that in (2.9). Therefore $\overline{g}(\theta)$ in (2.11) is more informative than $\overline{g}(\theta)$ in (2.9).

We can distinguish between the two different procedures of obtaining the reduced prior measures both derivable from the method (a). We shall call these Method I and Method II.
Method I - This is same as the method (b) described above and is applicable when we can transform \( \theta \) to \( \lambda = (\lambda_1, \theta_2) \) such that

\[
g(\lambda) = g_1(\lambda_1) g_2(\theta_2)
\]  

(2.12)

and it gives the prior measure

\[
g(\theta) = g(\theta_1, \theta_2) p(\theta_2).
\]  

(2.13)

Method II - The reduced prior \( \bar{g}(\theta) \) in this case is obtained exclusively by using the method (a) and it refers to the case where method (b) is not applicable.

Using \( \bar{g}_T(\theta) \) for the prior distribution of \( \theta \), we obtain the posterior density \( p(\theta|x) \) of \( \theta \) given \( x \) as

\[
p(\theta|x) = \frac{1(x|\theta) \bar{g}_T(\theta)}{\int 1(x|\theta) \bar{g}_T(\theta) d\theta}
\]  

(2.14)

If \( p(\theta|x) = p(\theta|x) \) where \( x = x(\theta) \), the set of statistics \( x \) is called \( P-I \) (prior-incomplete) sufficient set of statistics for \( \theta \) relative to the prior density \( g(\theta) \) and the transformation \( T \), or the sufficient set \( \xi \).
statistics for $\theta$ relative to the reduced prior distribution $\pi_T(\theta)$.

The posterior marginal density of $\theta_{(1)}$, a subset of $\theta = (\theta_{(1)}, \theta_{(2)})$, is given by

$$p(\theta_{(1)} | x) = \int p(\theta_{(1)}, \theta_{(2)} | x) \, d(\theta_{(2)})$$  \hspace{1cm} (2.15)

If $p(\theta_{(1)} | x) = p(\theta_{(1)} | x_{(1)})$, the set of statistics $S_{(1)}$ is called P-I marginally sufficient set of statistics for $\theta_{(1)}$ relative to $g(\theta)$ and the transformation $T$, or the marginally sufficient set of statistics for $\theta_{(1)}$ relative to the reduced prior distribution $\pi_T(\theta)$.

3 Application to normal samples

In this section, we illustrate the method of obtaining the reduced prior distributions by considering the case of samples from normal populations.

3.1 Sample from one normal population.

(1) Let $x$ denote the sample of size $n$ from a normal population $N(\mu, \sigma^2)$, where $\sigma^2$ is known. The likelihood of the sample is
\[ l(x|\theta) = \left(\sqrt{2\pi}\sigma\right)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right\}. \quad (3.1) \]

Suppose it is given that the prior density of \( \mu \) is

\[ g(\mu) \propto \exp \left\{ -\frac{n'}{2\sigma^2} (\mu - m)^2 \right\}. \quad (3.2) \]

Consider that the value of \( m \) is not known.

We make the transformation

\[ \theta_2 = \mu \quad \ldots \quad (3.3) \]

and obtain

\[ \bar{S}_1(\mu) \propto \theta. \quad \ldots \quad (3.4) \]

Using \( \bar{S}_1(\mu) \) for the prior distribution of \( \mu \), we get the posterior density of \( \mu \)

\[ p(\mu|x) \propto \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \mu)^2 \right\}. \quad (3.5) \]

(ii) Let \( \bar{x} \) denote the sample mean size \( n \) from a normal population \( \mathcal{N}(\mu, \sigma^2) \) where \( \mu \) is known.
Suppose it is given that the prior density of $\sigma$ is the conjugate prior density of Raiffa and Schaffer (1961) -

$$g(\sigma) \propto (w^2)^{v/2} (\sigma)^{-(v+1)} \exp \left\{ -\frac{1}{2} \frac{w^2}{\sigma^2} \right\}.$$  \hspace{1cm} (3.6)

Consider that $W$ is not known.

We make the obvious transformation

$$\theta_2 = \sigma \hspace{1cm} \cdots \hspace{1cm} (3.7)$$

and obtain

$$\tilde{g}(\sigma) \propto \sigma^{-1} \hspace{1cm} \cdots \hspace{1cm} (3.8)$$

Using this as the prior distribution, we obtain the posterior distribution of $\sigma$ -

$$p(\sigma | \mathcal{X}) \propto (\sigma)^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \right\}.$$  \hspace{1cm} (3.9)

(iii) Let $\mathcal{X}$ denote the sample from the normal population $\mathcal{N} (\mu, \sigma^2)$ where both $\mu$ and $\sigma$ are unknown.

Suppose it is given that the prior density of $(\mu, \sigma)$ is the conjugate prior density -
\[ g(\mu, \sigma) \propto \sigma^{-1} (\mu+1)^{-2} \exp \left[ - \frac{1}{2 \sigma^2} \left\{ \frac{X}{n} + \frac{1}{\sigma^2} \right\} \right] . \quad (3.10) \]

(a) Consider that the value of \( \mu \) is not known.

We make the transformation of the variables

\[ \lambda_1 = \sigma, \quad \theta_2 = \mu \quad \ldots \quad (3.11) \]

and obtain

\[ g_1(\sigma) = g^t(\sigma, \mu) \propto \sigma^{-(n^t+1)} \exp \left\{ - \frac{W^2}{2 \sigma^2} \right\} . \quad (3.12) \]

We have the reduced prior distribution \( E_\mu(\mu, \sigma) \)

\[ E_\mu(\mu, \sigma) \propto g^t(\sigma, \mu) . \quad \ldots \quad (3.13) \]

Using \( E_\mu(\mu, \sigma) \) for the prior distribution, we get posterior densities

\[ p(\mu, \sigma | x) \propto \sigma^{-(n+n^t+1)} \exp \left[ - \frac{1}{2 \sigma^2} \left\{ \frac{X}{n} + \frac{1}{\sigma^2} + \frac{(X-\mu)^2}{n} \right\} \right] . \quad (3.14) \]
and

\[ p(\sigma^2|\bar{x}) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} \left( \frac{1}{n} \sum (x_i - \bar{x})^2 \right) \right\} \quad \ldots \quad (3.16) \]

where

\[ s^2 = \sum (x_i - \bar{x})^2. \quad \ldots \quad (3.17) \]

(b) Consider that \( \Sigma \) is not known.

We make the transformation

\[ \lambda_1 = \frac{(\mu - \bar{x})}{\sigma}, \quad \theta_2 = \sigma \quad \ldots \quad (3.18) \]

and obtain

\[ g'(\mu, \sigma) \propto \sigma^{-1} \exp\left\{-\frac{n^*}{2\sigma^2} (\mu - \bar{m})^2 \right\} \quad \ldots \quad (3.19) \]

We have in this case the reduced prior measure

\[ \overline{g}(\mu, \sigma) \propto \sigma^{-1} g'(\mu, \sigma) \quad \ldots \quad (3.20) \]
The posterior densities with respect to the prior distribution (3.20) are

\[
p(\mu, \sigma | \bar{x}) \propto \sigma^{-(n+2)} \exp\left[ -\frac{1}{2\sigma^2} \left\{ s^2 + n(\bar{x} - \mu)^2 + n^1(\bar{y} - \mu)^2 \right\} \right] \]

.. (3.21)

\[
p(\mu, \sigma | \bar{x}) \propto \left\{ \frac{2}{n+n^1} \bar{x} - \frac{2}{n+n^1} \bar{y} \right\} \exp\left[ -\frac{1}{2\sigma^2} \left\{ s^2 + \frac{nn^1(\bar{x}-\bar{y})^2}{n+n^1} \right\} \right] \]

.. (3.22)

\[
p(\sigma | \bar{x}) \propto \sigma^{-(n+1)} \exp\left[ -\frac{1}{2\sigma^2} \left\{ s^2 + \frac{nn^1(\bar{x}-\bar{y})^2}{n+n^1} \right\} \right] . \] (3.23)

3.2 Samples from two normal populations

We now consider the problem of comparing two normal populations \(N(\mu_i, \sigma^2)\), \(i = 1, 2\). Let \(\bar{x}\) denote samples of size \(n\) each from the two populations. We are interested in inferences about \((\mu_1 - \mu_2)\).

The likelihood of the sample is

\[
l(x; \theta) = (2\pi \sigma)^{-2n} \exp\left[ -\frac{1}{2\sigma^2} \sum_{i = 1}^{n} \left\{ \left( x_{i1} - \mu_1 \right)^2 + \left( x_{i2} - \mu_2 \right)^2 \right\} \right] . \]

.. (3.24)
Suppose it is given that the prior density of $\mu_1$, $\mu_2$ and $\sigma$ is
\[
g(\mu_1, \mu_2, \sigma) \propto \sigma^{-(\nu+3)} \exp\left[-\frac{1}{2\sigma^2} \left\{ \frac{\nu_1^2}{\sigma^2} + \frac{n'(\mu_1 - \mu)^2}{\sigma^4} + \frac{n'(\mu_2 - \mu)^2}{\sigma^4} \right\} \right]
\]
\[\text{(2.25)}\]

(a) Let us suppose that the value of $\mu$ is not known.

We use the transformation
\[
\lambda_1 = (\delta_1, \sigma), \delta_2 = \mu_2
\]
\[\text{(2.26)}\]
where
\[
\delta_1 = (\mu_1 - \mu_2) / \sqrt{2},
\]
\[\text{(2.27)}\]

We then obtain the reduced prior distribution
\[
\tilde{g}_T(\delta_1, \mu_2, \sigma) \propto \sigma^{-(\nu + 2)} \exp\left[-\frac{1}{2\sigma^2} \left\{ \frac{\nu_1^2}{\sigma^2} + \frac{n'(\mu_2 - \mu)^2}{\sigma^4} \right\} \right]
\]
\[\text{(2.28)}\]

With respect to the prior distribution (2.28), we get the posterior densities.
\[ p(s_1^1 | x) \propto \left\{ s^2 + \frac{w^2}{n+n'} d_1^2 + (n+n')(s - \frac{nd_1}{n+n'})^2 \right\}^{-(2n+n')/2} \]

and

\[ p(\sigma^2 | x) \propto \sigma^{-3(2n+n')} \exp \left\{ -\frac{1}{2\sigma^2} (s^2 + \frac{w^2}{n+n'} d_1^2) \right\} \quad (3.30) \]

where

\[ d_1 = (\bar{x}_1 - \bar{x}_2)/2, \quad s_1^2 = \frac{2}{n} \sum_{i=1}^{n} x_i^2 \quad \text{and} \quad s_1^2 = \frac{n}{n-1} (\bar{x} - \bar{x}_1)^2. \quad (3.31) \]

(b) Let us suppose that \( w \) and \( n \) are not known.

We make transformation

\[ \lambda_1 = \frac{s_1}{\sigma}, \ \theta_2 = (\mu_2, \sigma) \quad (3.32) \]

where \( s_1 = (\mu_1 - \mu_2)/\sqrt{2} \), and obtain the reduced prior distribution

\[ \beta_1(\theta_1, \mu_2, \sigma) \propto \sigma^{-2} \exp \left\{ -\frac{n' \delta_1^2}{2 \sigma^2} \right\} \quad (3.33) \]

The posterior marginal densities with respect to

\[ (3.33) \text{ are} \]

\[ p(s_1^1 | x) \propto \left\{ s^2 + \frac{nn'}{n+n'} d_1^2 + (n+n')(s - \frac{nd_1}{n+n'})^2 \right\}^{-n} \quad (3.34) \]
and

\[ p(\sigma^2 | x) \propto \sigma^{-2n} \exp\left\{ -\frac{1}{2\sigma^2} \left( s^2 + \frac{nn'}{n+n'} d_1^2 \right) \right\} \tag{3.35} \]

where \( d_1 \) is defined in (3.31).

### 3.3 Samples from Several Normal Populations

Let \( x \) denote samples of size \( n \) from each of \( p \) normal populations \( N(\mu_i, \sigma^2) \), \( i = 1, 2, \ldots, p \). Suppose that we wish to compare the \( p \) means \( \mu_1 \); the inferences about the contrasts in \( \mu_i \)'s will then be of special interest.

The likelihood of the sample is

\[ l(x | \theta) = (\sqrt{2\pi} \sigma)^{-np} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i,j} (x_{ij} - \mu_i)^2 \right\} \tag{3.36} \]

Suppose it is given that the prior density of \( \mu_1, i = 1, 2, \ldots, p \) and \( \sigma \) is

\[ g(\mu_1, \ldots, \mu_p, \sigma) \propto \sigma^{-(v'+p+1)} \exp\left[ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^{v+n'} \frac{\left( \mu_i - \bar{x}_i \right)^2}{\sigma^2} \right) \right] \]

\[ \ldots \tag{3.37} \]

Let us suppose that \( W \) and \( n \) are not known.
Let \( s_1, s_2, \ldots, s_{p-1} \) denote the \((p-1)\) orthogonal contrasts in \( \beta_1 : \beta_i = \sum_j l_{ij} \beta_j \) \((\sum_j l_{ij} = 0, \sum_j l_{ij} l_{1j} = 0, i \neq 1). \) We make the transformation

\[
\gamma_1 = \left( \frac{s_1}{\sigma}, \frac{s_2}{\sigma}, \ldots, \frac{s_{p-1}}{\sigma} \right), \quad \gamma_2 = \left( \frac{\mu}{\sigma} \right)
\]

and obtain the reduced prior measure

\[
\pi(\gamma_1; \gamma_2; \beta, \sigma) \propto \sigma^{-p} \exp \left\{ -\frac{n}{2\sigma^2} \sum_{i=1}^{p-1} \gamma_i^2 \right\}.
\]

The posterior joint density of \( s_i, i=1,2,\ldots, p-1 \)
is then given by

\[
p(s_1, s_2, \ldots, s_{p-1} | \bar{x}) \propto \left\{ s_i^2 + \frac{n}{n+n_i} s_b^2 + (n+n_i) \Omega \left( s_i - \frac{nd_i}{n+n_i} \right)^2 \right\}
\]

where \( s_i^2 = \sum s_i^2, \quad s_b^2 = n \sum (x_i - \bar{x})^2 \) is the sum of squares between means \((\bar{x} = \sum x_i / p)\) and \( d_i \) are obtained from \( s_i \) on replacing \( \mu_j \) by \( \bar{x}_j \).

The posterior distributions obtained in this section agree with the conventional usage as regards the sums
of squares and their degrees of freedom.

4. Application to samples from exponential populations

4.1 Sample from one exponential population

Let \( x \) denote the sample of size \( n \) from an exponential population with parameters \( \eta \) and \( \beta \). The likelihood of the sample is

\[
L(x|\theta) = \prod_{i=1}^{n} \left\{ \frac{1}{\beta} \exp \left( -\frac{x_i}{\beta} \right) \right\}, \quad x_i > \eta
\]  \hspace{1cm} (4.1)

(i) Suppose it is given that the prior density of \((\eta, \beta)\) is

\[
g(\eta, \beta) \propto \beta^{-(k+2)} \exp \left\{ -\frac{1}{\beta} \left( w + n \left( \eta - \delta \right) \right) \right\}, \quad \eta > 0
\]  \hspace{1cm} (4.2)

(a) Let us suppose that \( \delta \) is not specified.

We use the transformation \( \lambda_1 = \beta, \theta_2 = \eta \), and obtain the reduced prior distribution

\[
g(\eta, \beta) \propto \beta^{-(k+1)} \exp \left\{ -\frac{w}{\beta} \right\}
\]  \hspace{1cm} (4.3)
The posterior densities of \( \gamma \) and \( \beta \) with respect to the prior distribution (4.3) are

\[
p(\gamma | x) \propto \left\{ Z + W + n(x - \gamma) \right\}^{-(n+k)}, \quad \gamma < x^* \quad (4.4)
\]

and

\[
p(\beta | x) \propto \beta^{-(n+k)} \exp\left\{ -\frac{1}{\beta} (Z + W) \right\} \quad (4.5)
\]

respectively, where

\[
Z = \Sigma (x_i^* - x^*) \quad \ldots \quad (4.6)
\]

and \( x^* \) is the smallest member in the sample \( x \).

(b) Let us suppose that \( W \) is not specified.

We put \( \lambda_1 = (\gamma - \delta) / \beta \), \( \theta_2 = \beta \) and obtain the reduced prior measure

\[
\bar{\pi}_1(\eta, \beta) \propto \beta^{-2} \exp\left\{ -\frac{n'}{\beta} (\gamma - \delta) \right\}, \quad \gamma > \delta \quad (4.7)
\]

when \( n' \neq n \), the posterior densities of \( \gamma \) and \( \beta \) are given by

\[
p(\gamma | x) \propto \left\{ Z + (n'_n) (\gamma - \frac{n'_x - n_x^*}{n'_n - n}) \right\}^{-(n+1)} \quad (4.8)
\]

and
\[
p(\beta \mid x) \propto \beta^{-(n+1)} \exp\left(\frac{-x}{\beta}\right) \exp\left\{ - \frac{n}{\beta} (x^{*} - \delta) \right\} \exp\left\{ - \frac{n}{\beta} (x^{*} - \delta) \right\}.
\]

When \( n' = n \), the posterior distribution of \( \eta \) is

rectangular, \( 0 < \eta < x^{*} \), and the posterior density of \( \eta \) is

\[
p(\eta \mid x) \propto \beta^{-(n+2)} \exp\left[ - \frac{1}{\beta} \left( z + n (x^{*} - \delta) \right) \right]
\]

\[(4.10)\]

(ii) Suppose that the prior distribution of \((\eta, \beta)\) is

\[
g(\eta, \beta) \propto \beta^{-(k+2)} \exp\left[ - \frac{1}{\beta} \left( w + n' (s - \gamma) \right) \right], \ s > \gamma (4.11)
\]

(a) Consider that \( \delta \) is not known.

We use the reduced prior distribution

\[
\tilde{g}(\eta, \beta) \propto \beta^{-(k+1)} \exp\left\{ - \frac{1}{\beta} w \right\}.
\]

\[(4.12)\]

The posterior densities of \( \eta \) and \( \beta \) with respect to
the prior density (4.12) are given by (4.4) and (4.5) re-
pectively.
(b) Consider that \( \theta \) is not known

We use the transformation \( \chi = \frac{(\gamma - \eta)}{\beta} \) and \( \theta = \beta \), and obtain the reduced prior distribution for \((\gamma, \beta)\)

\[
\pi_{T}(\gamma, \beta) \propto \beta^{-2} \exp \left\{ -\frac{1}{\beta} n'(\theta - \gamma) \right\}, \quad \theta > \gamma \tag{4.13}
\]

The posterior densities of \( \gamma \) and \( \beta \) with respect to the prior distribution (4.13) are

\[
p(\gamma | x) \propto \left\{ Z + n(x^e - \gamma) + n'(\theta - \gamma) \right\}^{-(n+1)} \tag{4.14}
\]

and

\[
p(\beta | x) \propto \beta^{-(n+1)} \exp \left\{ -\frac{1}{\beta} \left\{ Z + n(x^e - \gamma) \right\} \right\}
\begin{cases} 
& \text{if } \beta = \min(x^e, \gamma) \tag{4.15} \\
& \propto \beta^{-(n+1)} \exp \left\{ -\frac{1}{\beta} \left\{ Z + n'(\theta - x^e) \right\} \right\} \\
& \text{if } x^e = \min(x^e, \gamma).
\end{cases}
\]

4.2. **Samples from two exponential populations.**

We consider a problem of comparing two exponential populations having the same parameter \( \beta \). Let \( x \) denote the samples of size \( n \) from two such populations. The likelihood of the samples is
Suppose it is given that the prior density of \( (\eta_1, \eta_2, \beta) \) is

\[
g(\eta_1, \eta_2, \beta) \propto \beta^{-(k+3)} \exp \left[ -\frac{1}{\beta} \left( w + n' \left( \eta_1 - \xi \right) + n' \left( \eta_2 - \zeta \right) \right) \right] \exp \left\{ -\frac{n}{\beta} \left| \alpha_1 \right| \right\} \quad (4.17)
\]

Let us suppose that \( w \) and \( \beta \) are not known.

We make the transformation \( \lambda_1 = (\eta_1 - \eta_2) / \beta \), \( \lambda_2 = (\eta_2, \beta) \), and obtain the reduced prior measure for \( \alpha_1 = (\eta_1 - \eta_2) \), \( \eta_2 \) and \( \beta \)

\[
\pi_T (\alpha_1, \eta_2, \beta) \propto \beta^{-2} \exp \left\{ -\frac{n}{\beta} \left| \alpha_1 \right| \right\} \quad (4.18)
\]

Using (4.18) as the prior density, we get the posterior density of \( \alpha_1 \) as

\[
p(\alpha_1 | x) \propto \left\{ z_1 + z_2 + n \left| \alpha_1 - d_1 \right| + n' \left| \alpha_1 \right| \right\}^{-2n} \quad (4.19)
\]

where \( d_1 = \hat{x}_1 - \hat{x}_2 \).
5. The prior distribution using Jeffreys Invariance Principle.

When initially we are ignorant about the parameters, Jeffreys has suggested a method of obtaining the prior distribution which is invariant to nonsingular transformations. A justification of the method was given by Lindley (1960).

Using Jeffreys approach when \( \theta \) is a single parameter, we may proceed to obtain the prior distribution as follows:

We assume that \( g(\theta) = g(\theta | \phi) \) is the conditional prior density of \( \theta \) given \( \phi \) and that \( \phi \) has a density. (Here we consider that the range of the distribution \( g(\theta | \phi) \) does not depend on \( \phi \)). We then seek an objective prior density of \( \phi \) say \( k(\phi) \). The final prior density of \( \theta \) is then taken as

\[
\int g(\theta | \phi) k(\phi) \, d\phi \quad \cdots \quad (6.1)
\]

Using Jeffreys approach we take

\[
k(\phi) = \left[ -E \left\{ \frac{\delta^2 \log f(x | \phi)}{\delta \phi^2} \right\} \right]^{-\frac{1}{2}} \quad \cdots \quad (6.2)
\]
where
\[ f(x|\theta) = \int f(x|\theta_0) g(\theta|\theta_0) d\theta. \quad (5.3) \]

The prior density (5.1) is an average prior density of \( g(\theta_0) \).

Let us investigate how this prior density (5.1) compares with the reduced prior density in the normal case.

(i) Consider that we have a sample from \( N(\mu, \sigma^2) \) where \( \sigma^2 \) is known. Suppose it is known that the prior density of \( \mu \) is normal with mean \( \mu \) and variance \( \sigma^2/n \) and that \( \mu \) is not known.

In this case the average prior density (5.1) obtained by using (5.2) and (5.3) coincides with the reduced prior density
\[ \bar{g}(\mu) \propto c. \]

(ii) Consider the case where the mean of the normal population is known. Suppose it is known that, a priori, \( \sigma \) has the density
\[ g(\sigma) \propto (\bar{w}^2)^{v'/2} (\sigma^{-v'-1}) \exp \left\{ -\frac{1}{2} \frac{\bar{w}^2}{\sigma^2} \right\} \]
and that \( \bar{w} \) is not known.
In this case also the average prior density obtained by using (5.2) and (5.3) coincides with the reduced prior distribution

\[ p(\sigma) \propto \sigma^{-1}. \]

The prior distributions can be obtained by the above approach in many situations involving a single real parameter. When more than one parameter are involved, the extension based on this approach is less satisfactory (see Lindley (1960)). The prior distributions obtained by Jeffreys approach and by the reduced prior method do not in general coincide.

6. Prior distributions for certain other incomplete specifications.

Let \( l(x; \theta) \) denote the likelihood of the sample observations. Suppose that the prior knowledge of the parameters is as follows: The prior density of \( \theta_1 \), a subset of \( \theta = (\theta_1, \theta_2, \theta_3) \), is \( g_2(\theta_1) \) where the function \( g_2(\theta_1) = 0 \) known except for the value of \( \theta_2 \); (in other words, \( \theta_2 \) appears in the prior density of \( \theta_1 \) and no other knowledge of \( (\theta_2, \theta_3) \) is available initially).
If the components of \((\theta_2, \theta_3)\) are the \(r\) location parameters \(\alpha\) and the \(s\) scale parameters \(\beta\) in the density function \(p(x_1, \theta_1 | \theta_2, \theta_3)\)

\[
p(x_1, \theta_1 | \theta_2, \theta_3) = 1(\theta_1 | \theta_2) \ g_{\theta_2} (\theta_1) \quad \ldots \quad (6.1)
\]
a completed prior distribution of \(\theta\) may be taken as

\[
\bar{g}(\theta) = g_{\theta_2} (\theta_1) \ p(\theta_2, \theta_3) \quad \ldots \quad (6.2)
\]
where

\[
p(\theta_2, \theta_3) = \prod_{i=1}^{s} (\theta_2^{-1})^i \ldots \quad (6.3)
\]

If \(g_{\theta_2} (\theta_1)\) involves unknown parameters \(\varphi\), we can sometimes obtain the reduced prior distribution \(\bar{g}_{\theta_2} (\theta_1)\) by using the method (a). Here \(\bar{g}_{\theta_2} (\theta_1)\) is the reduced prior distribution of \(\theta_1\) when \(\theta_2\) is assumed as known. The final prior distribution for \(\theta\) is then given by

\[
\bar{g}(\theta) = \bar{g}_{\theta_2} (\theta_1) \ p(\theta_2, \theta_3) \quad \ldots \quad (6.4)
\]
where \(p(\theta_2, \theta_3)\) is defined in (6.3)

**Illustrations**

1. Given a sample of size \(n\) from a normal population \(N(\mu, \sigma^2)\) where \(\mu\) and \(\sigma^2\) are unknown, suppose it is
known that, a priori, $\mu$ has a normal density with mean $\mu$ and variance $\sigma^2$, where $\sigma^2$ is not known but the ratio $(\sigma^2/\sigma^2) = 1/n'$ is known.

Thus a priori we have

$$g_\sigma(\mu) \propto \sigma^{-1} \exp \left\{ -\frac{n'}{2\sigma^2} (\mu-m)^2 \right\}. \tag{6.5}$$

The completed prior distribution of $(\mu, \sigma)$ may be taken as

$$g(\mu, \sigma) \propto \sigma^{-1} g_\sigma(\mu)$$

$$\propto \sigma^{-2} \exp \left\{ -\frac{n'}{2\sigma^2} (\mu-m)^2 \right\}. \tag{6.6}$$

2. Consider the problem where we wish to compare two normal populations $N(\mu_1, \sigma^2), i=1,2$, and we are given samples of size $n$ each from the two populations. Suppose it is known that, a priori, $\mu_1, \mu_2$ have independent normal densities with equal means $m$ and equal variances $\sigma^2$, and that $m$ and $\sigma^2$ are unknown, but the ratio $\sigma^2/\sigma^2 = 1/n'$ is known. See Duncan (1961, 1965).

We define $s_1 = (\mu_1 - \mu_2)/\sqrt{2}$ and $s_2 = (\mu_1 - \mu_2)/(2\sqrt{2})$. We have the prior joint density of $s_1, s_2$.
From this, we obtain the reduced prior distribution

\[ g_\sigma(\delta_1, \delta_2) \propto \sigma^{-2} \exp \left[ -\frac{n'}{2\sigma^2} \left( \delta_1^2 + (\delta_2 - \sqrt{2n})^2 \right) \right]. \quad (6.7) \]

The final prior distribution of \((\delta_1, \delta_2, \sigma)\) may be taken as

\[ \bar{g}(\delta_1, \delta_2, \sigma) \propto \sigma^{-1} \exp \left\{ -\frac{n'}{2\sigma^2} \delta_1^2 \right\}. \quad (6.8) \]

7. The nature of posterior distributions.

In this section, we discuss the nature of the posterior distributions in general when the reduced priors are used. The differences between the methods I and II given in Section 2 are clearly brought out by these posterior distributions.

**Method I**

This method is used when the prior density

\[ g(\lambda_1, \theta_2) \] satisfies
\[ g(\lambda_1, \theta_2) = g_1(\lambda_1) g_1(\theta_2) \quad \ldots \quad (7.1) \]

and \( g_1(\lambda_1) \) is independent of \( \theta \). The reduced prior for \( (\lambda_1, \theta_2) \) is then given by

\[ g(\lambda_1, \theta_2) = g_1(\lambda_1) p(\theta_2) \quad \ldots \quad (7.2) \]

Given the sample observations \( \mathbf{x} \), the posterior joint density of \( (\lambda_1, \theta_2) \) with respect to the prior distribution \( (7.2) \) is given by

\[ p(\lambda_1, \theta_2 | \mathbf{x}) \propto l(\mathbf{x} | \theta) g_1(\lambda_1) p(\theta_2) \quad \ldots \quad (7.3) \]

On integrating \( (7.3) \) over \( \theta_2 \), we obtain the posterior marginal density of \( \lambda_1 \)

\[ p(\lambda_1 | \mathbf{x}) \propto g_1(\lambda_1) \int l(\mathbf{x} | \theta) p(\theta_2) d\theta_2 \quad \ldots \quad (7.4) \]

Thus the posterior marginal density of \( \lambda_1 \) is proportional to the product of the marginal prior density of \( \lambda_1 \) and the average of the likelihood with respect to the uniform distribution of location and logscale parameters of \( \theta_2 \). This averaging of the likelihood over the uniform distribution in many cases effects a reduction of the sample space, i.e., it reduces the size of the minimal sufficient set.
The posterior marginal density of $\theta_2$ is given by

$$p(\theta_2 | x) \propto p(\theta_2) \int l(x | \theta) g_1(\lambda_1) \, d\lambda_1 \ldots$$  \hspace{2cm} (7.5)

The posterior distributions w.r.t. the reduced prior distribution (7.2) afford certain direct comparisons with those w.r.t. the prior distribution (7.1). Let

$$p_o(\lambda_1, \theta_2 | x), \ p_o(\theta_2 | x) \text{ etc. denote the posterior densities of } (\lambda_1, \theta_2), \ \theta_2 \text{ etc. w.r.t. the prior distribution (7.1). We have}$$

$$\frac{p(\lambda_1, \theta_2 | x)}{p_o(\lambda_1, \theta_2 | x)} \propto \frac{p(\theta_2)}{g_1(\theta_2)} \ldots$$  \hspace{2cm} (7.6)

where the proportionality factor depends on $x$ and $\theta$. Also we have

$$\frac{p(\theta_2 | x)}{p_o(\theta_2 | x)} \propto \frac{p(\theta_2)}{g_1(\theta_2)}$$  \hspace{2cm} (7.7)

the proportionality factor being the same as in (7.6). Therefore from (7.6) and (7.7), we obtain
\[ p(\lambda_1, \theta_2 | x) = p_0(\lambda_1, \theta_2 | x) \]  \hspace{1cm} (7.8)

and this leads to the relationship

\[ p(\lambda_1, \theta_2 | x) = p_0(\lambda_1, \theta_2 | x) \cdot p(\theta_2 | x). \]  \hspace{1cm} (7.9)

The relationship (7.9) shows that the effect of reduction on the posterior joint density of \((\lambda_1, \theta_2)\) is that of changing only the posterior marginal density of \(\theta_2\).

We may also consider the posterior distribution of \(\theta = (\theta_1, \theta_2)\) w.r.t. the prior distribution (7.2). The posterior joint density of \(\theta\) is given by

\[ p(\theta_1, \theta_2 | x) \propto 1(x | \theta) g(\theta_1 | \theta_2) p(\theta_2) \]  \hspace{1cm} (7.10)

where the conditional density function \(g(\theta_1 | \theta_2)\) satisfies

\[ g(\theta_1 | \theta_2) = g_1(\lambda_1) \left| \frac{\partial \lambda_1}{\partial \theta_1} \right| \]  \hspace{1cm} (7.11)

Similar to relations (7.6), (7.8) and (7.9) we have
Method IX

This method is used when

\[ g(\lambda_1, \theta_2) = g_1(\lambda_1) g(\theta_2 | \lambda_1) \]  \quad (7.15)

where \( g_1(\lambda_1) \) is independent of \( \lambda \) and

\[ g(\theta_2 | \lambda_1) \neq g_1(\theta_2) \]  \quad (7.16)

The reduced prior distribution of \( (\lambda_1, \theta_2) \) is in this case given by

\[ f(\lambda_1, \theta_2) = g_1(\lambda_1) p(\theta_2), \]  \quad (7.17)

and this is similar in form to (7.2).
The posterior marginal densities of $\lambda_1$ and $\theta_2$ are given by (7.4) and (7.5) respectively. However the posterior distributions obtained by this method do not allow direct comparisons with those w.r.t. $g(\lambda_1, \theta_2)$ defined by (7.15).

In applications we are concerned with certain loss functions or utility functions which depend only on some parameters involved in the likelihood function. In adopting method II, suitable transformations should include in $\theta_1$ the parameters involved in the loss or the utility functions. It may include some nuisance parameters, (parameters not involved in the loss or utility functions), in order that they provide some additional information.

8. Some remarks

8.1. Consider a situation where all the conditions stated in Section 2 for the application of method (a) except condition (ii) hold. And suppose that the prior density of $\theta$ is $g(\theta_1, \theta_2)$ which includes an unknown parameter $\theta_2$, and that the marginal prior density $g_1(\theta_1)$ is independent of $\theta_2$ and the marginal prior density $g_2(\theta_2)$ is also independent of $\theta_2$. In such situations if posterior inferences of $\theta_1$ only are required, we may use a prior distribution

$$g(\theta_1, \theta_2) \propto g_1(\theta_1) p(\theta_2).$$
Similarly if posterior inferences about $\theta_2$ only are required, we may use the prior distribution

$$g(\theta_1, \theta_2) \propto g_1(\theta_2) p(\theta_2).$$

Satisfactory inferences about $(\theta_1, \theta_2)$ jointly cannot be obtained by using either of these prior distributions.

### 3.2.

The prior measures obtained here will be useful in situations where variation in the values of the unknown parameters in $g(\theta)$ are such that they lead to inferences which differ to a large extent. For example, consider a multiple action Bayes rule. In situations where the posterior utility of every action in the set $A$ of $a$'s is practically the same for every permissible value of the unknown parameter, then the question of using the reduced prior will not arise. On the other hand in situations where the posterior utility of actions differ considerably for variation in the values of $\gamma$, the reduced prior distribution obtained from known density functions provides an adaptable solution.

### References


