Bayesian analysis of selection models

M. J. BAYARRI & M. H. DeGROOT

1Departamento de Estadistica e I.O., Facultad de Matematicas, Av. Dr. Moliner 50, 46100 Burjasot, Valencia, Spain and 2Carnegie-Mellon University and Ohio State University

Abstract. In many situations, experimenters are not able to draw a random sample from the population in which they are interested, and statistical models that incorporate the nonrandomness or bias in their observations must be developed. We consider problems in which observations can be obtained only from certain selected portions of the population. Bayesian and related methods for the analysis of such selection models are discussed. Consideration is restricted to univariate problems in which the selection set is the upper tail of the population. Applications that are presented include the problem of appropriately analysing statistical significance as it is usually reported in the scientific literature.

1 Introduction: Weighted distributions and selection models

Consider a random variable $X$ that is distributed over a certain population according to the density $g(x|\theta)$ and it is desired to make inferences about the unknown value of the parameter $\theta$ ($\theta \in \Omega$). The usual statistical analysis assumes that a random sample from $g(x|\theta)$ can be obtained. In many situations, however, the probability or density that a particular observation $x$ will actually be obtained is proportional to $w(x)g(x|\theta)$, where $w(x)$ is a non-negative weight function that may itself depend on some parameters whose values are unknown. Thus, the observed sample is in fact a random sample from the weighted density

$$f(x|\theta) = \frac{w(x)g(x|\theta)}{E[w(X)]}.$$  

(1.1)

The concept of a weighted distribution such as (1.1) can be traced to Fisher (1934), although these models were first formulated in a unified way by Rao (1965). Good surveys on this topic are Patil (1984) and Rao (1985).

In this paper we are concerned with a special class of weighted distributions called selection models or truncation models. These are models in which

$$w(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{otherwise}, \end{cases}$$  

(1.2)

where $S$ is a subset of the sample space of $X$ and is called the selection set. In many problems the set $S$ is unknown and belongs to a class $S$. In such problems we write $w(x|S)$ to emphasise the dependence of $w(\cdot)$ on the unknown set $S$. We will let $f(x|\theta, S)$ denote the density (1.1) when the weight function $w(\cdot)$ is given by (1.2).

The name ‘selection models’ in this context is due to Fraser (1952, 1966), although the term ‘selection’ was used in a more general setting by Tukey (1949). We shall refer to a random sample from the selection model $f(x|\theta, S)$ as a selection sample. For a given selection sample $x=(x_1, \ldots, x_n)$, let

$$T(x) = \bigcap_{i \in S} \{S|x_i \in S \text{ for } i=1, \ldots, n\}.$$  

(1.3)
It is well-known (see, e.g. Fraser, 1952; Smith, 1957) that if $M(\theta)$ is a minimal sufficient statistic for the family $g(x; \theta); \theta \in \Omega$ then the pair $(M, T)$ is a jointly minimal sufficient statistic for the selection model $\{f(x; \theta, S); \theta \in \Omega, S \in S\}$.

Selection samples are common in scientific work and in many aspects of our daily lives. For example, a random sample from a uniform distribution over an unknown subset of the real line or the plane can be regarded as a selection sample. This problem and related problems are treated by DeGroot & Eddy (1983). In a medical context, we may be able to observe only patients who have a certain disease that is manifested only when a certain vector of relevant variables lies in some set $S$, and we are interested in the distribution of the vector in the entire population. Selection samples arise also in industrial settings when detailed measurements are made only on items which are within certain specified limits. The problem of sampling from a truncated binomial or Poisson distribution in which the zero class is missing has been widely treated in the literature (see, e.g. David & Johnson 1952; Irwin, 1959; Cohen, 1960; Dahiya & Gross, 1973; Sanathanan, 1977; Blumenthal & Sanathanan, 1980; and the survey paper by Blumenthal, 1981). Selection samples also occur in much more general settings: because of space and time restrictions, the news that is reported in newspapers or on television is a selection sample from the events of the day.

In this paper, we will restrict ourselves to problems in which $g(x; \theta)$ is a univariate distribution and the selection set $S$ is the upper tail of the distribution containing all the values $x \geq \tau$. Thus, we will assume that $X_1, \ldots, X_n$ is a selection sample from the density

$$f(x; \theta, \tau) = \frac{g(x; \theta)}{1 - G(\tau; \theta)} \quad \text{for } x \geq \tau,$$

(1.4)

where $f$ may represent either an absolutely continuous or a discrete distribution. In this problem, $T(x) = \min(x_1, \ldots, x_n)$. Therefore, if $\tau$ is unknown, the jointly minimal sufficient statistic for $(\theta, \tau)$ is the minimal sufficient statistic for $\theta$ together with $\min(x_1, \ldots, x_n)$.

In Section 2 we consider selection models in which the parameter $\theta$ is known and it is desired to make inferences about the parameter of selection $\tau$. In Section 3 we consider selection models in which the selection set is known and it is desired to make inferences about the unknown parameter $\theta$. The results are then applied in Section 4 to the important problem of appropriately analysing statistical significance as it is usually reported in the scientific literature. We conclude the paper with a discussion in Section 5 of selection models in which both $\theta$ and $\tau$ are unknown but the probability content of the selection set is known.

A related problem of interest in the design and analysis of selection models is that of comparing the information contained in a selection sample with the information in a random sample from the unrestricted density $g(x; \theta)$. This problem is studied in Bayarri & DeGroot (1986).

2 Selection from a known distribution

In this section, we will assume that the parameter $\theta$ is known, so the model (1.4) reduces to $f(x; \tau)$ and the joint density of the selection sample $x = (x_1, \ldots, x_n)$ is

$$f(x; \tau) = \prod_{i=1}^{n} \frac{g(x_i)}{[1 - G(\tau)]^n} \quad \text{for } \tau \leq \min(x_1, \ldots, x_n),$$

(2.1)
and \( f(x|\tau) = 0 \) otherwise, where \( g(x) \) is a known density and \( G(x) \) is the corresponding d.f. Since the likelihood function \( L(\tau) \) is of the form
\[
L(\tau) \propto [1 - G(\tau)]^{-a} \quad \text{for } \tau \leq \min(x_1, \ldots, x_n)
\]
and \( L(\tau) = 0 \) otherwise, the natural choice for a conjugate prior density for \( \tau \) is of the form
\[
\xi_1(\tau) \propto [1 - G(\tau)]^a \quad \text{for } \tau \leq b
\]
where the hyperparameters \( a \) and \( b \) are constants to be specified. A distribution of this form may be appropriate in problems in which the support of \( g(\cdot) \) is bounded on the left, so the possible values of \( \tau \) are also bounded on the left. In this case, the prior density \( \xi_1(\tau) \) can be normalised for any value of \( a \) so that it is a proper probability distribution. However, if the possible values of \( \tau \) are unbounded from the left then the density \( \xi_1(\tau) \) will be improper for every real number \( a \), regardless of whether \( a \) is positive, negative, or zero. Although improper prior distributions are widely used in Bayesian statistics, the density \( \xi_1(\tau) \) seems to be of no use whatsoever because it follows from the statement just given that the posterior distribution of \( \tau \) will also be improper for all values of \( n \) and all values of \( x_1, \ldots, x_n \). Thus, it seems desirable to use a different form for the conjugate prior distribution in this problem.

There are many different families of proper prior distributions for \( \tau \) that are closed under sampling. For example, any density of the form
\[
\xi(\tau) \propto h(\tau) \left[1 - G(\tau)\right]^a \quad \text{for } \tau \leq b
\]
will have this property, where \( h(\cdot) \) could be any function for which the right-hand side of (2.4) is integrable over \( \tau \). In particular, \( h(\cdot) \) could be taken to be of the form
\[
h(\tau) = \begin{cases} 
0 & \text{for } \tau < c, \\
1 & \text{for } \tau \geq c,
\end{cases}
\]
which has the effect of specifying a fixed lower bound \( c \) for the possible values of \( \tau \).

However, a more suitable choice for \( h(\cdot) \) is \( h(\cdot) = g(\cdot) \). It can be verified that the resulting prior density
\[
\xi_2(\tau) \propto g(\tau) \left[1 - G(\tau)\right]^a \quad \text{for } \tau \leq b
\]
is proper for every real number \( a \). The Bayesian analysis of this selection model is then straightforward.

When \( G \) is absolutely continuous, so \( g(\cdot) \) is a p.d.f., the problem can be reduced to making inferences about the parameter of a uniform distribution in the following way: Let
\[
Y = 1 - G(X) \quad \text{and} \quad \beta = 1 - G(\tau).
\]
Then the transformed selection sample \( Y_1, \ldots, Y_n \) is a random sample from the uniform distribution on the interval \((0, \beta)\) where \( \beta < 1 \). In this formulation, the usual conjugate prior p.d.f. for \( \beta \) is of the form
\[
\xi_3(\beta) \propto \beta^a \quad \text{for } \beta \geq 1 - b,
\]
where the hyperparameters \( a \) and \( b \) are arbitrary constants \((0 < b < 1)\). The p.d.f. \( \xi_3(\beta) \) is a truncated Pareto distribution and it is proper for every real number \( a \).

It can be easily verified that \( \beta \) has the p.d.f. \( \xi_3(\cdot) \) given by (2.7) if and only if \( \tau \) has the p.d.f. \( \xi_2(\cdot) \) given by (2.5). This fact provides another justification for the choice of \( \xi_2(\cdot) \) as the conjugate prior p.d.f. for \( \tau \) in this problem.

It should be noted that whenever \( G \) is a known, absolutely continuous d.f., this reduction of a selection sample to a random sample from a uniform distribution can
be carried out for any type of selection set and not just for the upper tail. Thus, the
analysis of these selection samples can be accomplished by standard methods.

3 Known selection set

In this section we will assume that the parameter $\theta$ in the unrestricted model $g(x|\theta)$ is
unknown but the selection parameter $\tau$ is known. Thus, the selection model (1.4)
reduces to $f(x|\theta)$ and the joint density of the selection sample $x=(x_1, \ldots, x_n)$ is

$$f(x|\theta) = \prod_{i=1}^{n} g(x_i|\theta)/[1-G(\tau|\theta)]^a$$

for $\tau \leq \min(x_1, \ldots, x_n)$ (3.1)

and $f(x|\theta)=0$ otherwise. In this problem the minimal sufficient statistic for the family
$\{f(x|\theta); \theta \in \Omega\}$ is just the minimal sufficient statistic for the unrestricted family $\{g(x|\theta); \theta \in \Omega\}$.

An example of the selection problem in which $\tau$ is known is that of sampling from a
truncated binomial or Poisson distribution in which the zero class is missing. Some
references for this problem were given in Section 1. More generally, the lowest $\tau$
classes might be missing, where $\tau$ is a known non-negative integer.

A selection sample with known $\tau$ arises naturally in the sampling of historical
records. In such problems, observations from the entire historical population of
interest are usually not available and inferences have to be based on observations that
have been recorded just for some selected group. For example, estimation of the
distribution of the height of the population might have to be based on the historical
observations of the heights of members of the army for which there was a minimum
required height (Wachter & Trussell, 1982).

A particularly simple model is obtained for a selection sample from an exponential
distribution with parameter $\theta$ when the value of $\tau$ is known. In this case, the
distribution of each observation $X_i-\tau$ is again exponential with parameter $\theta$, so the
analysis is straightforward.

In general, for any given values of $x_1, \ldots, x_n$, the likelihood function $L(\theta)$ is
proportional to the right-hand side of (3.1). The Bayesian analysis of this model
cannot usually be carried out in closed form because of the appearance of $\theta$ through
$G(\tau|\theta)$ in the denominator. Furthermore, the choice of an appropriate prior distribution
for $\theta$ in this problem is not clear.

The difficulties can be illustrated for a selection sample from a normal distribution.
Suppose that $g(x|\theta)$ is the p.d.f. of a normal distribution with mean $\theta$ and variance 1,
and just a single observation $X$ is to be obtained from $f(x|\theta)$. Then the likelihood
function for $\theta$ is

$$L(\theta) \propto \frac{\phi(x-\theta)}{1-\Phi(\tau-\theta)},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the p.d.f. and d.f. of the standard normal distribution,
respectively. Thus, a conjugate prior density for $\theta$ based directly on the form of $L(\theta)$ is

$$\xi(\theta) \propto \frac{\exp[-h(\theta-m)^2/2]}{[1-\Phi(\tau-\theta)]^a},$$

where the hyperparameters $m$, $h$, and $a$ are specified constants. Although the form of
any conjugate prior distribution is always related to the experiment to be performed,
the form of (3.3) seems especially unsuitable because of its explicit dependence on the
value of $\tau$. In most problems, $\tau$ is fixed and conveys no information of $\theta$. Furthermore, the use of the conjugate prior (3.3) does not eliminate the difficulties that we have mentioned in carrying out an analysis of the posterior distribution of $\theta$.

It is more appropriate in this problem to use a prior distribution that is conjugate with respect to the unrestricted model $[g(x|\theta); \theta \in \Omega]$. Thus, in this example, we would choose the prior distribution of $\theta$ simply to be a normal distribution, which is equivalent to using the value $a=0$ in (3.3). With this choice we have eliminated the dependence of the prior distribution on $\tau$. However, we have not eliminated any of the difficulties of the posterior analysis, which remains essentially unchanged.

If the prior distribution of $\theta$ is normal with mean $m_0$ and precision $h_0$, then the posterior p.d.f. $\xi(\theta|x)$ will be of the form (3.3) with $a=1$ and

$$m = \frac{h_0 m_0 + x}{h_0 + 1}, \quad h = h_0 + 1.$$  \hspace{1cm} (3.4)

It should be noted that the expressions for $m$ and $h$ are the same as those that would be obtained for the posterior hyperparameters of the unrestricted model. For any function $\Psi(\theta)$, we shall let $E^*[\Psi(\theta)]$ denote the expectation with respect to the normal distribution with mean $m$ and precision $h$; that is, the expectation with respect to the posterior distribution that would be obtained if the value $x$ were observed in the unrestricted model. Furthermore, let

$$r(\theta) = [1 - \Phi(\tau - \theta)]^{-1}. \hspace{1cm} (3.5)$$

Then the posterior mean and variance of $\theta$ under the p.d.f. $\xi(\theta|x)$ obtained in the selection model are:

$$E(\theta|x) = E^*[\theta r(\theta)]/E^*[r(\theta)]$$

$$\text{Var}(\theta|x) = [E^*[\theta^2 r(\theta)]/E^*[r(\theta)] - (E(\theta|x))^2]. \hspace{1cm} (3.6)$$

The expectations in (3.6) have to be evaluated numerically or otherwise approximated in some fashion. Let

$$M(y) = \frac{1 - \Phi(y)}{\phi(y)} \hspace{1cm} (3.7)$$

denote Mills’ ratio at any point $y$; i.e. $M(y)$ is the inverse of the hazard rate function at $y$. Then the delta method applied to each of the factors of $E(\theta|x)$ in (3.6) yields:

$$\frac{E^*[r(\theta)]}{r(m)} \approx 1 + \frac{1}{2hM(\tau - m)} \left[ \frac{2}{M(\tau - m)} - (\tau - m) \right],$$

$$\frac{E^*[\theta r(\theta)]}{r(m)} \approx 1 + \frac{1}{2hM(\tau - m)} \left[ m \left( \frac{2}{M(\tau - m)} - (\tau - m) \right) - 2 \right].$$

4 Analysis of reported significant results

An important example of selection models with a known value of $\tau$ occurs when only experimental results that are found to be ‘statistically significant’ are published in the scientific literature. This situation is very common both because the editors of some journals will publish only articles in which statistical significance has been obtained, and because many experimenters themselves regard their results as being useless unless the results are statistically significant and will not even submit them for publication.

The unfortunate effects that these policies can have on scientific learning have been
discussed by several authors, including Sterling (1959), Dawid & Dickey (1977), Rosenthal (1979), Salsburg (1985), and Hedges & Olkin (1985, Chapter 14). The practice of publishing only statistically significant results is, in turn, a natural consequence of the misuse of statistical methods in many areas of application through the almost exclusive reliance on tests of hypotheses (Zellner, 1980; DeGroot & Mezzich, 1985).

We shall present our discussion of this problem in the context of one-sided tests of hypotheses. Consider an experiment in which a random sample $Y_1, \ldots, Y_n$ is obtained from a density $h(y|\theta)$, where the value of the real-valued parameter $\theta$ is unknown. Suppose that it is desired to test the following hypotheses at a certain level of significance $\alpha (0 < \alpha < 1)$:

$$
H_0: \theta \leq \theta_0,
H_1: \theta > \theta_0,
$$

(4.1)

where $\theta_0$ is a specified number. Suppose also that the test calls for rejecting $H_0$ whenever some test statistic $U$ is greater than or equal to a specified critical value $\tau$.

We shall let $g(u|\theta)$ denote the density of $U$. In many common problems, the distribution of $U$ is stochastically increasing in $\theta$, and the critical value $\tau$ is chosen so that $\Pr(U \geq \tau|\theta = \theta_0) = \alpha$. Suppose now that as readers of some scientific journal, we will learn the results of this experiment only if they lead to the rejection of $H_0$; that is, only if $U \geq \tau$. In this situation, the density of any value of $U$ that we will actually get to observe is not simply $g(u|\theta)$ but is given by the selection model

$$
f(u|\theta) = \frac{g(u|\theta)}{1 - G(\tau|\theta)} \quad \text{for } u \geq \tau.
$$

(4.2)

When the appropriate model (4.2) is used, it is quite possible that an analysis of the 'significant' values of $U$ will actually provide strong evidence in favor of the null hypothesis $H_0$.

As an illustration, suppose that $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with unknown mean $\mu$ and known variance $\sigma^2$, and it is desired to test the hypotheses

$$
H_0: \mu \leq 0,
H_1: \mu > 0.
$$

(4.3)

The uniformly most powerful test for any specified value of $\alpha$ is to reject $H_0$ whenever

$$
U = \frac{n^{1/2}}{\sigma} \bar{Y} - \Phi^{-1}(1 - \alpha).
$$

(4.4)

Here, the distribution of the test statistics $U$ is normal with mean $\theta = n^{1/2} \mu/\sigma$ and variance 1.

If we can only observe a 'significant' value of $U$, then the p.d.f. of our observation will be

$$
f(u|\theta) = \frac{\phi(u - \theta)}{1 - \Phi(\tau - \theta)} \quad \text{for } u \geq \tau,
$$

(4.5)

where $\tau = \Phi^{-1}(1 - \alpha)$. The likelihood function for $\theta$ based on (4.5) is the same as that given by (3.2), and the Bayesian analysis discussed in Section 3 applies. To see how a 'significant' value of $U$ when properly analysed can support the null hypothesis $H_0$, it is enlightening to calculate the maximum likelihood estimate of $\theta$ for an observed value $U = u$. 

It can be shown from (4.5) that the MLE \( \hat{\theta} \) is the unique solution of the following equation:

\[(u - \hat{\theta})M(\tau - \hat{\theta}) = 1\]  \hspace{1cm} (4.6)

where Mills’ ratio \( M(\cdot) \) is defined by (3.7). The value of \( \hat{\theta} \) for \( \alpha = 0.01, 0.05, \) and \( 0.10 \) and various observed values of \( u \) are given in Table 1. Of course, the most widely used criterion for statistical significance is \( \alpha = 0.05 \). It is unlikely that an editor or an experimenter would publish only results that were significant at the level \( \alpha = 0.01 \). In Table 1 we also give the \( p \)-values corresponding to the observed values of \( u \) as they would be reported in the literature, that is, calculated from the standard normal distribution. Since only ‘significant’ values of \( U \) can be observed, all the \( p \)-values must be less than \( \alpha \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( u )</th>
<th>( p )</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>2.792</td>
<td>0.0026</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.665</td>
<td>0.0038</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2.588</td>
<td>0.0048</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>2.538</td>
<td>0.0056</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>2.503</td>
<td>0.0062</td>
<td>-3</td>
</tr>
<tr>
<td>0.05</td>
<td>2.249</td>
<td>0.0124</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.063</td>
<td>0.0195</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.955</td>
<td>0.0253</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>1.888</td>
<td>0.0305</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>1.844</td>
<td>0.0326</td>
<td>-3</td>
</tr>
<tr>
<td>0.10</td>
<td>1.985</td>
<td>0.0236</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.755</td>
<td>0.0397</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.625</td>
<td>0.0526</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>1.546</td>
<td>0.0611</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>1.495</td>
<td>0.0675</td>
<td>-3</td>
</tr>
</tbody>
</table>

Since \( \mu = \sigma \hat{\theta}/n^{1/2} \), it follows from (4.3) that negative values of \( \hat{\theta} \) support the null hypothesis \( H_0 \). The basic conclusion to be drawn from the discussion in this section is that even observed values of \( u \) that appear to be highly significant can yield MLE’s that are very negative and, therefore, give strong support to \( H_0 \). For example it can be seen from Table 1 that when \( \alpha = 0.05 \), a \( p \)-value as small as 0.033 yields \( \hat{\theta} = -3 \), an estimate that is at least 3 standard deviations away from the parameter values in \( H_1 \). Similar behavior is found for \( \alpha = 0.01 \) and \( \alpha = 0.10 \).

These results emphasise the fact that under current statistical practice, a published \( p \)-value only slightly smaller than the effective \( \alpha \) can only be regarded as supporting \( H_0 \) rather than rejecting \( H_0 \). In fact, as the \( p \)-value approaches \( \alpha \) in this selection model, \( \theta \rightarrow -\infty \).

5 Unknown selection set with a fixed probability

In this section, we will consider selection models in which the selection set is the upper tail of the density \( g(x|\theta) \) and both \( \theta \) and \( \tau \) are unknown. However, we will assume that the selection set has a fixed, known probability \( \alpha \). Thus, there is a known relationship between \( \tau \) and \( \theta \). An illustration of this type of problem is provided by the medical example mentioned in the introduction. Suppose that it is known that the proportion \( \alpha \) of individuals in some population have a certain disease, and that an individual has
the disease whenever the value of an associated latent random variable \( X \) exceeds some critical threshold \( \tau \). On the basis of the values of \( X \) observed in a random sample of patients having the disease it is desired to make inferences about the distribution of \( X \) in the whole population.

The density for this selection model is

\[
f(x|\theta) = \frac{1}{\alpha} g(x|\theta) \quad \text{for} \quad x \geq G^{-1}(1 - \alpha|\theta) \tag{5.1}
\]

and \( f(x|\theta) = 0 \) otherwise. As before, the jointly minimal sufficient statistic for \( \theta \) based on a selection sample \( x_1, \ldots, x_n \) from \( f(x|\theta) \) is the minimal sufficient statistic for the unrestricted model \( g(x|\theta) \) together with the statistic \( T = \min(x_1, \ldots, x_n) \). The likelihood function for \( \theta \) based on \( x_1, \ldots, x_n \) is

\[
L(\theta) \propto \prod_{i=1}^{n} g(x_i|\theta) \quad \text{for} \quad \theta \in \Omega(\alpha, t), \tag{5.2}
\]

where \( \Omega(\alpha, t) \) is the set of all values of \( \theta \) such that \( G(t|\theta) \geq 1 - \alpha \) and \( t \) is the observed value of \( T \). A conjugate prior density based directly on the form of \( L(\theta) \) given in (5.2) would require that the range \( \Omega(\alpha, a) \) of possible values of \( \theta \) be dependent not only on a hyperparameter \( a \) whose value can be specified arbitrarily, but also on \( \alpha \). For reasons similar to those presented in Section 3, it seems desirable to eliminate the dependence of the prior distribution on \( \alpha \). Thus, in problems where there is a suitable family of conjugate prior distributions for the unrestricted model \( g(x|\theta) \), it is usually appropriate to use this family for the selection model as well. In effect, this type of prior is equivalent to choosing the value \( a = \infty \) for the hyperparameter in \( \Omega(\alpha, a) \).

We conclude this section by deriving the Bayes estimator \( \hat{E}(\theta|x) \) for selection samples from the exponential and normal distributions. Suppose first that

\[
g(x|\theta) = \theta e^{-\alpha x} \quad \text{for} \quad x > 0, \tag{5.3}
\]

so that it follows from (5.1) that

\[
f(x|\theta) = \frac{\theta}{a} e^{-\alpha x} \quad \text{for} \quad x > -\frac{\log \alpha}{\theta}. \tag{5.4}
\]

The likelihood function based on a selection sample \( x = (x_1, \ldots, x_n) \) from (5.4) is

\[
L(\theta) \propto \theta^r e^{-\alpha r} \quad \text{for} \quad \theta > \theta_t = -\frac{\log \alpha}{t}, \tag{5.5}
\]

where

\[
r = \sum_{i=1}^{n} x_i \quad \text{and} \quad t = \min(x_1, \ldots, x_n).
\]

It follows from (5.5) that the MLE \( \hat{\theta} \) is

\[
\hat{\theta} = \begin{cases} 
n/r & \text{if} \ n/r \geq \theta_t \\
\theta_t & \text{otherwise} \end{cases}. \tag{5.6}
\]

It should be noted that if \( \alpha < 1/e \), then \( \hat{\theta} \) must be \( \theta_t \).

If the prior distribution of \( \theta \) is a gamma distribution with parameters \( a_0 \) and \( b_0 \), then the posterior p.d.f. of \( \theta \) will be

\[
\tilde{\xi}(\theta|x) = c(a_1, b_1, \theta_t) \Gamma(\theta | a_1, b_1) \quad \text{for} \quad \theta > \theta_t, \tag{5.7}
\]
where \( a_i = a_0 + n, \ b_i = b_0 + r, \Gamma(\theta|a_i, \ b_i) \) is the p.d.f. of the gamma distribution with parameters \( a_i \) and \( b_i \), and
\[
[c(a, b, \alpha)]^{-1} = \int_0^{\infty} \Gamma(\theta|a, b) d\theta.
\] (5.8)

The mean of this distribution is
\[
E(\theta|x) = [a_i c(a_i, b_i, \theta)]/[b_i c(a_i + 1, b_i, \theta)].
\] (5.9)

Next suppose that \( g(x|\theta) \) is the p.d.f. of a normal distribution with unknown mean \( \theta \) and variance 1. In this case, the likelihood function based on a selection sample is
\[
L(\theta) \propto \exp \left[ -\frac{n}{2} (2x\theta - \theta^2) \right] \quad \text{for} \ \theta < \theta_t = t - \Phi^{-1}(1 - \alpha).
\] (5.10)

It follows from (5.10) that the MLE \( \hat{\theta} \) is
\[
\hat{\theta} = \begin{cases} \bar{x} & \text{if } \bar{x} \leq \theta_t, \\ \theta_t & \text{otherwise}. \end{cases}
\] (5.11)

It should be noted that if \( \alpha < 1/2 \), then \( \hat{\theta} \) must be \( \theta_t \).

If the prior distribution of \( \theta \) is a normal distribution with mean \( m_0 \) and precision \( h_0 \), then the posterior p.d.f. of \( \theta \) will be
\[
\xi(\theta|x) = h_1^{1/2} \phi\left[h_1^{1/2} (\theta - m_1)/\sqrt{h_1}\right]/\Phi[h_1^{1/2} (\theta - m_1)] \quad \text{for} \ \theta < \theta_t,
\] (5.12)

where
\[
m_1 = \frac{h_0 m_0 + n \bar{x}}{h_0 + n} \quad \text{and} \quad h_1 = h_0 + n.
\] (5.13)

It can be shown that the mean of this distribution is
\[
E(\theta|x) = m_1 - [h_1^{1/2} M(h_1^{1/2}(m_1 - \theta))]^{-1}
\] (5.14)

where, as before, \( M(\cdot) \) is Mills' ratio.

More general selection models, with both an unknown value of \( \theta \) and an unknown selection set will be treated in future papers.

**Acknowledgement**

This research was supported in part by the Spanish Ministry of Education and Science and the Fulbright Association under grant 85–07399 and by the National Science Foundation under grant DMS–8320618.

**References**


