Bayesian Prediction of the Total Time on Test Using Doubly Censored Rayleigh Data

Mohamed Z. Raqab
Dept of Mathematics
University of Jordan
Amman 11942, JORDAN.
M Raqab@accessme.com.jo

Mohamed T. Madi
Department of Statistics
UAE University
P.O.Box 17555, Al-Ain, UAE
MMadi@uaeu.ac.ae

Abstract

In a Bayesian setting, and on the basis of a doubly censored random sample of failure times drawn from a Rayleigh distribution, Fernandez (2000, Statist. Probab. Lett., 48, 393-399) considered the problem of predicting an independent future sample from the same distribution. In this article, we extend his work to include the estimation of the predictive distribution of the total time on test up to a certain failure in a future sample, as well as that of the remaining testing time until all the items in the original sample have failed. An example is used to illustrate the prediction procedure.

1 Introduction

A group of n components are put on test, their lifetimes $X_1, X_2, ..., X_n$ are assumed to follow a Rayleigh distribution with parameter $\sigma$. This model has a wide range of applications including life testing; Polovko (1968) and communication engineering; Dyer and Whisenand (1973). We consider the situation in which some observations are initially censored and where the life test is terminated before all items on test have failed. The resulting sample is said to be type II doubly censored. A number of authors have considered Bayesian prediction problems for the Rayleigh distribution. Sinha and Howlader (1985) derived HPD prediction intervals for a future observation and for the $k^{th}$ order statistic in a future sample. Howlader and Hossain (1995) considered Bayesian estimation and prediction of the Rayleigh parameter and reliability using a type II censored sample. Recently, Fernández (2000) presented a Bayesian approach to predicting a future sample from the Rayleigh distribution based on type II doubly censored data.

In this paper, we use Gibbs sampling to obtain estimates for the predictive distribution of the total time on test up to a certain failure. We also consider the one-sample prediction scenario and estimate the predictive density function of the total remaining testing time until all the items in the original sample have failed.

In Section 2, we derive the predictive distributions of the failure times of the unfailed items as well as that of the first $t$ failure times in a future sample of size $m$ items. In Section 3, the derivations needed in the prediction procedure are presented. Section 4 includes an illustrative example using a real life data set, and a discussion of the results.

2 Predictive Distributions

In this section we present the Bayesian predictive distributions for the future failure times in the one-sample and two-sample prediction scenarios based on the observed doubly type II censored sample $X = (X_{(r)}, X_{(2)}, ..., X_{(s)}), 1 \leq r \leq s \leq n$. Here we consider the natural conjugate family of prior densities for $\sigma$ used in Fernandez (2000), namely:

$$ \pi(\sigma) = \sigma^{a-2b-1} \exp\left(-\frac{a}{2\sigma^2}\right), \sigma > 0, $$

where $a$ and $b$ are specified positive constants.
In his paper, Fernández (2000) discussed the Bayes and highest posterior density (HPD) estimates of \( \sigma \). We extend his work to include the Bayesian prediction of the failure times of the unfailed items \( Y_1 = (X_{(s+1)}, \ldots, X_{(n)}) \).

It can be shown that the joint predictive density function of \( Y_1 \) given \( X = x \) is given by:

\[
p(y_1|x) \propto \frac{\prod_{j=s+1}^n x_{(j)} (V_1 + a)^{S_r + b} \Psi_r(V_0 + a, n_r + b)}{(V_0 + a)^{n_r + b} \Psi_r(V_1 + a, S_r + b)}.
\]  

(2)

where

\[
V_1 = \sum_{i=r}^s x_{(i)}^2 + (n - s)x_{(n)}^2, \quad V_0 = \sum_{j=r}^n x_{(j)}^2, \quad j_i = j - i + 1, \quad \text{for any two integers } i \text{ and } j,
\]

and where

\[
\Psi_r(u, v) = \sum_{j=0}^{r-1} c_j (1 + \frac{jx_2}{u})^{-v}, \quad u, \quad v > 0
\]

In the two-sample prediction plan, one might be interested in predicting the first \( t \) ordered failure times of a future sample of size \( m \), \( Y_2 = (Y_{(1)}, \ldots, Y_{(t)}) \), \( 1 \leq t \leq m \) drawn from the Rayleigh distribution. The predictive density function of \( Y_2 \) given \( X = x \) is given by

\[
p(y_2|x) \propto \frac{\prod_{j=1}^t y_{(j)} (V_1 + V_2 + a)^{S_r + b + t} \Psi_r(V_1 + V_2 + a, S_r + b + t)}{(V_1 + V_2 + a)^{S_r + b + t} \Psi_r(V_1 + a, S_r + b)}
\]  

(3)

where

\[
V_2 = \sum_{i=1}^t y_{(i)}^2 + (m - t)y_{(n)}^2.
\]

Prediction of order statistics or a function of order statistics from the same or a future sample is of interest. Suppose all the items in a batch are put to test and the voltage is gradually increased until \( (s - r + 1) \) of the items fail. In the one-sample prediction problem, our interest is that of finding the future behavior for the total amount of the remaining testing time, \( Z_1 = \sum_{i=s+1}^n X_{(i)} - (n - s)X_{(n)} = \sum_{i=s+1}^n (X_{(i)} - X_{(n)}) \). In another situation, a manufacturer of a certain equipment might be interested in setting up a warranty for the equipment in a lot being sent out to the market. Using the information based on a censored sample, the objective is to predict the total time on test up to the \( t^{th} \) failure time in a future sample, \( Z_2 = \sum_{i=1}^t Y_{(i)} + (m - t)Y_{(t)} \).

It is important to point out here that the expressions in (2) and (3) do not allow us to obtain explicit forms for the predictive densities of \( Z_1 \) and \( Z_2 \). For this, we use the Gibbs sampler (Gelfand 2000) to generate samples from the predictive distributions of \( Z_1 \) and \( Z_2 \).

3 Prediction of the Total Time on Test

First, we consider the problem of predicting \( Y_1 = (X_{(s+1)}, \ldots, X_{(n)}) \) \( (r \leq s \leq n - 1) \) and then estimate the predictive density function of \( Z_1 \). The full conditional distributions of \( X_{(k)} \) \( (s + 1 \leq k \leq n) \), and \( \sigma \) are needed to implement the Gibbs sampling algorithm. By setting \( Y_{1(k)} = (x_{(s+1)}, \ldots, x_{(k-1)}, x_{(k+1)}, \ldots, x_{(n)}) \), we immediately obtain the conditional probability density function of \( X_{(k)} \) as

\[
\pi(x_{(k)}|X, Y_{1(k)}, \sigma) = \begin{cases} 
\frac{x_{(k)}^{(s+1) - 1} \exp(-x_{(k)}^2/2\sigma^2)}{\exp(-x_{(k+1)}^2/2\sigma^2) - \exp(-x_{(k-1)}^2/2\sigma^2)} \cdot I[x_{(k)} < x_{(k+1)}], & k = s + 1, \ldots, n - 1 \\
\frac{x_{(n)}^{(s+1) - 1} \exp \left[-(x_{(n)}^2 - x_{(n-1)}^2)/2\sigma^2 \right]}{\exp(-x_{(n-1)}^2/2\sigma^2) - \exp(-x_{(n)}^2/2\sigma^2)} \cdot I[x_{(n)} > x_{(n-1)}], & k = n.
\end{cases}
\]  

(4)
The full conditional posterior of \( \sigma \) is given by

\[
\pi(\sigma|x, y_1) = \frac{\sigma^{-(n_0 + b + 1)} \alpha_{0 \sigma}^{(n_0 + b)} \sum_{j=0}^{r-1} c_j \exp(-\alpha_j/2\sigma^2)}{\Gamma(n_r + b) 2^{n_r + b - 1} \Psi_r(\alpha_0, n_r + b)} .
\]

where

\[
\alpha_j = V_0 + a + jx_{(r)}^2, \quad j = 0, 1, ..., r - 1.
\]

Generation of values from the model in (5) can be carried out based on its cdfs. Generation from (4) uses the inverse cdf transformation method.

For the 2-sample prediction problem, we set \( y_2(k) = (y(1), ..., y(k-1), y(k+1), ..., y(t)) \). The conditional probability density function of \( y(k) \) (\( k = 1, ..., t \)) is then found to be:

\[
\pi(y(k)|x, y_2, \sigma) = \begin{cases} \\
\frac{y(k) \exp(-y(k)^2/2\sigma^2) I[y(k-1) < y(k) < y(k+1)]}{\exp(-y_{(t-1)}^2/2\sigma^2) - \exp(-y_{(k+1)}^2/2\sigma^2)}, & k = 1, 2, ..., t - 1 \\
m_t \Psi_t(y_k^2 - y_{(t-1)}^2)/2\sigma^2 I[y(t) > y(t-1)], & k = t.
\end{cases}
\]

and the full conditional distribution of \( \sigma \) is immediately obtained as

\[
\pi(\sigma|x, y_2) = \frac{\sigma^{-2(S_r + b + t + 1)} \beta_0^{S_r + b + t} \sum_{j=0}^{r-1} c_j \exp(-\beta_j/2\sigma^2)}{\Gamma(S_r + b + t) 2^{S_r + b + t - 1} \Psi_r(\beta_0, S_r + b + t)} ,
\]

where

\[
\beta_j = V_1 + V_2 + a + jx_{(r)}^2, \quad j = 0, 1, ..., r - 1.
\]

4 Illustrative Example and Discussion

In this section, we show, through an example, how the Gibbs sampler is used in predicting the future failure times and then make some predictive inferences about \( Z_1 \). After setting the initial values for \( \sigma \) and \( y \), a Gibbs sampler single chain of 600 iterations is run and used as input in Raftery and Lewis Fortran program to determine the required number of iterations that should be run to attain convergence. Subsequent to convergence, 1000 draws of equally spaced variates were then collected for the parameter \( \sigma \) as well as for the future failure times \( y \). The iterations for \( \sigma \) were drawn using its cdf by means of the IMSL subroutines DRNGS and DRNGCT.

Example: Consider the following Type II doubly censored data which represent the number of revolutions to failure (in hundreds of millions) for each of 23 ball bearings; Leblein and Zelen (1956).

\[
-6.9, -5.8, -0.2, 0.24, 0.34, 0.46, 0.53, 0.54, 0.56, 0.58, 0.60, 0.67, 0.68, 0.69, 0.80, 0.84, 0.93, 0.94, -6.9, -5.8, -0.2,
\]

These data are examined and a one parameter Rayleigh model provides a satisfactory fit. Here, we have censored the first three and the last five observations. We then predict the total remaining testing time, \( Z_1 = \sum_{j=19}^{23} (X_{(j)} - X_{(18)}) \) and present its simulated predictive distribution using the prior parameters values \( a = b = 2 \). In this case \( n = 23, r = 4 \) and \( s = 18 \).

| Table 1: Simulated Percentiles of the Estimated Distributions of \( \sigma \) and \( Z_1 \) |
|---|---|---|---|---|---|---|---|
| \( p \) | 0.005 | 0.025 | 0.05 | 0.5 | 0.95 | 0.975 | 0.995 |
| \( \sigma \) | 0.4990 | 0.5000 | 0.5020 | 0.5840 | 0.7110 | 0.7420 | 0.8410 |
| \( Z_1 \) | 0.2360 | 0.3930 | 0.4870 | 1.3260 | 2.9470 | 3.3700 | 4.2950 |
From Table 1, one can establish prediction intervals of $\sigma$ and $Z_1$. For example, 95% prediction intervals for $\sigma$ and $Z_1$ are respectively, (0.5, 0.742), and (0.393, 3.370). Figure 1 illustrates the predictive density estimate of $Z_1$. The Bayesian predictive estimate of the remaining total time on test $Z_1$ is equal to 1.47 and turns out to exactly equal to the true observed remaining testing time. Experimentation with different starting values as well as with different values of the prior parameters led to the same results, which confirms convergence of the sampler and shows very good stability with respect to the prior settings.

This paper was aimed at extending the work of Fernández (2000) who considered the Rayleigh prediction problem based on doubly censored data. In his article, the author derived the predictive density and the Bayes predictive estimator of the $k^{th}$ order statistic in a future sample. Both the density and the estimator are obtained in closed forms. We have tackled the same problem but for the remaining and the future total times on test $Z_1$ and $Z_2$. It was seen that while the analytical approach to the problem is very complex, Gibbs sampling proved to be an effective alternative way for estimating the predictive distributions of these future quantities.

References


