

## Inference for the Proportional Hazards Family under Progressive Type-II Censoring

A. Asgharzadeh<sup>1</sup>, R. Valiollahi<sup>2</sup>

<sup>1</sup>Department of Statistics, Faculty of Basic Science, University of Mazandaran, Iran. (a.asgharzadeh@umz.ac.ir)

<sup>2</sup>Department of Statistics, College of Sciences, Semnan University, Iran. (r.valiollahi@umz.ac.ir)

**Abstract.** In this paper, the well-known proportional hazards model which includes several well-known lifetime distributions such as exponential, Pareto, Lomax, Burr type XII, and so on is considered. With both Bayesian and non-Bayesian approaches, we consider the estimation of parameters of interest based on progressively Type-II right censored samples. The Bayes estimates are obtained based on symmetric and asymmetric loss functions. We also provide Bayes and empirical Bayes prediction intervals for the times to failure of units censored in multiple stages in a progressively censored sample. Finally, two numerical examples are given to illustrate the results.

---

*Key words and phrases:* Bayes and empirical Bayes estimation, maximum likelihood estimate, prediction, progressively Type-II right censoring, proportional hazard rate model, symmetric and asymmetric loss functions, uniformly minimum variance unbiased estimate.

## 1 Introduction

Censoring occurs when exact survival times are known only for a portion of the individuals or items under study. The complete survival times may not have been observed by the experimenter either intentionally or unintentionally. For example, individuals in a clinical trial may drop out of the study, or the study may have to be terminated early for lack of funds. In an industrial experiment, units may break accidentally. Data obtained from such experiments are called censored data.

In this paper, we consider a general scheme of progressively Type-II right censoring. The progressive Type-II right censoring, after starting the life-testing experiment with  $n$  units, arises as follows.  $n$  units are placed on a life-testing experiment and only  $m (< n)$  units are completely observed until failure. The censoring occurs progressively in  $m$  stages. These  $m$  stages offer failure times of the  $m$  completely observed units. At the time of the first failure (the first stage),  $R_1$  of the  $n - 1$  surviving units are randomly withdrawn (censored intentionally) from the experiment,  $R_2$  of the  $n - 2 - R_1$  surviving units are withdrawn at the time of the second failure (the second stage), and so on. Finally, at the time of the  $m$ th failure (the  $m$ th stage), all the remaining  $R_m = n - m - R_1 - \dots - R_{m-1}$  surviving units are withdrawn. We will refer to this as progressive Type-II right censoring scheme  $(R_1, R_2, \dots, R_m)$ . It is clear that this scheme includes the conventional Type-II right censoring scheme (when  $R_1 = R_2 = \dots = R_{m-1} = 0$  and  $R_m = n - m$ ) and complete sampling scheme (when  $n = m$  and  $R_1 = R_2 = \dots = R_m = 0$ ). For further details on progressively censoring, inferences and their applications, one may refer to Balakrishnan and Aggarwala (2000) and Balakrishnan (2007).

Let us consider the continuous random variable  $X$  with the cumulative distribution function (cdf)  $F(x; \theta)$ . In many situations  $F(x; \theta)$  can be written as

$$F(x; \theta) = 1 - [\bar{F}_0(x)]^\theta, \quad -\infty \leq c < x < d \leq \infty, \quad \theta > 0 \quad (1.1)$$

where  $\bar{F}_0(\cdot) = 1 - F_0(\cdot)$ , and  $F_0(\cdot)$  is an arbitrary continuous cdf with  $F_0(c) = 0$  and  $F_0(d) = 1$ . Here,  $\{F(x; \theta), \theta > 0\}$  is called a proportional hazards family with underlying distribution  $F_0$  (see Marshall and Olkin, 2007).

For the proportional hazards family (1.1), the two hazard rates corresponding to the distribution functions  $F$  and  $F_0$  are propor-

tional. The proportional hazards family has been extensively used in the literature to model failure time data. It is useful in estimating the survival function for right censored data. This family includes several well-known lifetime distributions such as exponential, Pareto, Lomax, Burr type XII, and so on. For more information on proportional hazards family see Marshal and Olkin (2007). See also Ahmadi et al. (2008, 2009) who have studied the problems of estimation and prediction for the proportional hazards family based  $k$ -record data.

From the model (1.1), the probability density function (pdf), the reliability function and hazard rate function (at some  $t$ ) are given, respectively, by

$$f(x; \theta) = \theta f_0(x) [\bar{F}_0(x)]^{\theta-1}, \quad -\infty \leq c < x < d \leq \infty, \quad (1.2)$$

$$R(t) = [\bar{F}_0(t)]^\theta, \quad (1.3)$$

and

$$H(t) = \theta \frac{f_0(t)}{\bar{F}_0(t)}, \quad (1.4)$$

where  $f_0(\cdot) = F'_0(\cdot)$  is the corresponding pdf.

In this paper, we consider the estimation problem with both Bayesian and non-Bayesian approaches for the proportional hazards family (1.1) based on progressively Type-II censored samples. In Section 2, the MLEs, the uniformly minimum variance unbiased estimates (UMVUEs) and Bayes estimates are derived for the unknown parameter, reliability function and hazard rate function based on progressively Type-II censored samples. The Bayes estimates are obtained based on square error, LINEX and general entropy loss functions. In Section 3, we provide Bayes and empirical Bayes prediction intervals for the times to failure of units censored in multiple stages in a progressively censored sample from the proportional hazards family. Finally, in Section 4, two numerical examples are given to illustrate the results .

## 2 Estimation

Let  $X_{1:m:n}, \dots, X_{m:m:n}$  denote a progressively Type-II censored sample from the proportional hazards family (1.1) obtained from a sample of size  $n$  with the censoring scheme  $(R_1, \dots, R_m)$ . To simplify the notation, we will use  $x_i$  in place of  $x_{i:m:n}$ . In this section, we consider

the problem of estimation with both Bayesian and non-Bayesian approaches for the unknown parameter  $\theta$ , the reliability function  $R(t)$ , and the hazard rate function  $H(t)$ .

### 2.1 Maximum Likelihood Estimation

Based on the progressively Type-II censored sample  $\underline{X} = (X_1, \dots, X_m)$ , the likelihood function is given (see Balakrishnan and Aggarwala, 2000) by

$$L(\theta|\underline{x}) = A \prod_{i=1}^m [f(x_i; \theta)[1 - F(x_i; \theta)]^{R_i}], \tag{2.1}$$

where  $A = n(n-1-R_1)(n-2-R_1-R_2) \cdots (n-m+1-R_1 \cdots -R_{m-1})$ . It follows from (1.1), (1.2) and (2.1), that

$$L(\theta|\underline{x}) = A \left[ \prod_{i=1}^m \frac{f_0(x_i)}{F_0(x_i)} \right] \theta^m e^{-\theta S(\mathbf{x})} \tag{2.2}$$

where  $S(\underline{x}) = -\sum_{i=1}^m (R_i+1) \ln \bar{F}_0(x_i)$ . Note that  $S(\mathbf{x})$  can be written as

$$S(\mathbf{x}) = \sum_{i=1}^m (R_i + 1)T_0(x_i),$$

where  $T_0(\cdot) = -\ln \bar{F}_0(\cdot)$  is the baseline cumulative hazard function.

The log-likelihood function can be written as

$$\ln L(\theta|\underline{x}) = \ln A + \sum_{i=1}^m \ln \left[ \frac{f_0(x_i)}{F_0(x_i)} \right] + m \ln(\theta) - \theta S(\underline{x}). \tag{2.3}$$

By using (2.3), the MLE of  $\theta$  is

$$\hat{\theta}_{ML} = \frac{m}{S(\underline{X})}.$$

**Example 2.1.1.** (i) (Exponential distribution): Taking  $\bar{F}_0(x) = e^{-x}$ ,  $x > 0$ , in (1.1),  $X$  has exponential distribution, and we obtain the MLE of  $\theta$  as

$$\hat{\theta}_{ML} = \frac{m}{\sum_{i=1}^m (R_i + 1)x_i}.$$

(ii) (Burr type XII distribution): Taking  $\bar{F}_0(x) = (1 + x^c)^{-1}$ ,  $x > 0$ ,  $c > 0$ , with known  $c$ , in (1.1),  $X$  has Burr type XII distribution, and we obtain the MLE of  $\theta$  as

$$\hat{\theta}_{ML} = \frac{m}{\sum_{i=1}^m (R_i + 1) \ln(1 + x_i^c)}.$$

The corresponding MLE of the reliability function  $R(t)$ , and hazard rate function  $H(t)$ , after replacing  $\theta$  by its MLE  $\hat{\theta}_{ML}$ , are given by

$$\hat{R}_{ML}(t) = [\bar{F}_0(t)]^{\hat{\theta}_{ML}} \quad , \quad \hat{H}_{ML}(t) = \frac{f_0(t)}{\bar{F}_0(t)} \hat{\theta}_{ML}. \quad (2.4)$$

## 2.2 UMVUEs

To obtain the UMVUEs of  $\theta$ ,  $R(t)$  and  $H(t)$ , we first consider the distribution of  $S(\underline{X}) = -\sum_{i=1}^m (R_i + 1) \ln \bar{F}_0(X_i)$ . We know that if  $X \sim F(x; \theta)$  in (1.2), then

$$U = F(X; \theta) = 1 - [\bar{F}_0(X)]^\theta,$$

is distributed as the Uniform(0,1) distribution. Suppose that  $U_1, U_2, \dots, U_m$  be a progressively Type-II censored sample from the  $U(0, 1)$  distribution. Then, from a known result about the progressively Type-II censored sample from the Uniform(0,1) distribution (see Balakrishnan and Aggarwala, 2000, P. 20), the random variables

$$\begin{aligned} V_1 &= \frac{1 - U_m}{1 - U_{m-1}}, \\ V_2 &= \frac{1 - U_{m-1}}{1 - U_{m-2}}, \\ &\vdots \\ V_{m-1} &= \frac{1 - U_2}{1 - U_1}, \\ V_m &= 1 - U_1 \end{aligned}$$

are all mutually independent random variables with

$$V_i \stackrel{d}{=} \text{Beta} \left( i + \sum_{j=m-i+1}^m R_j, 1 \right), \quad i = 1, 2, \dots, m.$$

Now, since

$$-2 \left( i + \sum_{j=m-i+1}^m R_j \right) \ln V_i \stackrel{d}{=} \chi_2^2, \quad i = 1, 2, \dots, m,$$

we conclude that

$$\begin{aligned}
 -2 \sum_{i=1}^m \left( i + \sum_{j=m-i+1}^m R_j \right) \ln V_i &= -2 \sum_{i=1}^m (R_i + 1) \ln[1 - U_i] \\
 &= -2\theta \sum_{i=1}^m (R_i + 1) \ln[\bar{F}_0(X_i)] \\
 &= 2\theta S(\underline{X}) \stackrel{d}{=} \chi_{2m}^2, \tag{2.5}
 \end{aligned}$$

where  $\chi_q^2$  denotes the  $\chi^2$ -distribution with  $q$  degrees of freedom. From (2.5), we find that  $S(\underline{X}) \sim \Gamma(m, \theta)$ , i.e.,  $S(\underline{X})$  is a gamma distributed random variable with parameters  $m$  and  $\theta$ . Now, it is easy to see that  $\hat{\theta}_{ML} = m/S(\underline{X})$  is a biased estimator of  $\theta$  and the unbiased estimator is  $(m - 1)/S(\underline{X})$ . Furthermore, from (2.2) it is clear that  $S(\underline{X})$  is a complete sufficient statistics and hence  $(m - 1)/S(\underline{X})$  is also the UMVUE of  $\theta$ .

In order to derive the UMVUE of  $R(t)$ , define

$$g_t(u) = \begin{cases} \left( 1 + \frac{\ln \bar{F}_0(t)}{u} \right)^{m-1} & \text{if } u \geq -\ln \bar{F}_0(t) \\ 0 & \text{otherwise.} \end{cases}$$

Then, it is easy to verify that  $E[g_t(S)] = R(t)$ . Hence, the UMVUE of  $R(t)$  is

$$\hat{R}_{UMVU}(t) = \begin{cases} \left( 1 + \frac{\ln \bar{F}_0(t)}{S(\underline{X})} \right)^{m-1} & \text{if } S(\underline{X}) \geq -\ln \bar{F}_0(t) \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Also the UMVUE of  $H(t)$  is obtained as

$$\hat{H}_{UMVU}(t) = \frac{f_0(t)}{\bar{F}_0(t)} \frac{m - 1}{S(\underline{X})}. \tag{2.7}$$

It should be mentioned here that since  $2\theta S(\underline{X})$  has chi square distribution with  $2m$  degrees of freedom, therefore it can be used to construct confidence intervals or to conduct tests of hypotheses for the parameter  $\theta$ . For example, a two sided  $100(1 - \gamma)\%$  confidence interval for  $\theta$  is given by

$$\left[ \frac{\chi_{2m}^2(1 - \gamma/2)}{2S(\underline{X})}, \frac{\chi_{2m}^2(\gamma/2)}{2S(\underline{X})} \right],$$

where  $\chi_{2m}^2(\gamma/2)$  is the right-tailed  $\gamma/2$  percentile for chi-squared distribution with  $2m$  degrees of freedom.

### 2.3 Bayes Estimation

In the literature, most of the Bayesian inference procedures have been developed under the usual squared error loss (SEL) function. The symmetric nature of this function gives equal weight to overestimation as well as underestimation, while in the estimation of parameters of life time model, overestimation may be more serious than underestimation or vice-versa. For example, in the estimation of reliability and hazard rate functions, an overestimation is usually much more serious than an underestimation. In this case, the use of symmetrical loss function might be inappropriate as also emphasized by Basu and Ebrahimi (1991). In recent years, many authors have considered asymmetric loss functions in the Bayesian inference procedures, such as Basu and Ebrahimi (1991), Parsian and Nematollahi (1996), Moore and Papadopoulos (2000), Soliman (2005) and Ahmadi et al. (2005).

One of the most popular asymmetric loss function is the linear-exponential loss function (LINEX). This loss function was introduced by Varian (1975) and was extensively discussed by Zellner (1986). Under the assumption that the minimal loss occurs at  $\phi^* = \phi$ , the LINEX loss function for  $\phi = \phi(\theta)$  can be expressed as

$$L(\Delta) \propto \exp(a\Delta) - a\Delta - 1, \quad a \neq 0. \tag{2.8}$$

where  $\Delta = (\phi^* - \phi)$ ,  $\phi^*$  is an estimate of  $\phi$ . The sign and magnitude of the shape parameter  $a$  represents the direction and degree of symmetry, respectively. (If  $a > 0$ , the overestimation is more serious than underestimation, and vice-versa.) For  $a$  close to zero, the LINEX loss is approximately SEL and therefore almost symmetric.

The posterior expectation of the LINEX loss function (2.8) is

$$E_\phi[L(\phi^* - \phi)] \propto \exp(a\phi^*)E_\phi[\exp(-a\phi)] - a(\phi^* - E_\phi(\phi)) - 1, \tag{2.9}$$

where  $E_\phi(\cdot)$  denotes the posterior expectation with respect to the posterior density of  $\phi$ . The Bayes estimator of  $\phi$ , denote by  $\phi_{BL}^*$  under the LINEX loss function is the value  $\phi^*$  which minimizes (2.9).

It is

$$\phi_{BL}^* = -\frac{1}{a} \ln\{E_\phi[\exp(-a\phi)]\}, \tag{2.10}$$

provided that the expectation  $E_\phi[\exp(-a\phi)]$  exists and is finite.

Another useful asymmetric loss function is the general entropy loss (GEL) function

$$L(\phi^*, \phi) \propto \left(\frac{\phi^*}{\phi}\right)^q - q \ln\left(\frac{\phi^*}{\phi}\right) - 1, \quad q \neq 0, \tag{2.11}$$

whose minimum occurs at  $\phi^* = \phi$ . This loss function is a generalization of the entropy loss function used by Dey et al. (1987).

The Bayes estimator  $\phi_{BG}^*$  of  $\phi$  under the general entropy loss (2.11) is (see for example Soliman, 2005)

$$\phi_{BG}^* = (E_{\phi}(\phi^{-q}))^{-\frac{1}{q}}. \tag{2.12}$$

Under the assumption that the parameter  $\theta$  is unknown, we can use the conjugate gamma prior  $\Gamma(\alpha, \beta)$ , with pdf

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0, \quad (\beta > 0, \delta > 0). \tag{2.13}$$

The posterior density function of  $\theta$  given the data, denoted by  $\pi(\theta | \underline{x})$ , can be obtained using (2.2) and (2.13) as

$$\pi(\theta | \underline{x}) = \frac{[\beta + S(\underline{x})]^{m+\alpha}}{\Gamma(m + \alpha)} \theta^{m+\alpha-1} e^{-\theta[\beta+S(\underline{x})]}. \tag{2.14}$$

Under a squared error loss function, the usual estimate of a parameter is the posterior mean. Thus, Bayes estimators of the parameter, reliability function and failure rate function are obtained by using the posterior density (2.14). The Bayes estimator  $\hat{\theta}_{BS}$  of parameter  $\theta$  is

$$\hat{\theta}_{BS} = \frac{m + \alpha}{\beta + S(\underline{X})}. \tag{2.15}$$

The Bayes estimator,  $\hat{R}_{BS}$ , of the reliability function  $R(t)$  is

$$\begin{aligned} \hat{R}_{BS}(t) &= E[R(t)|\underline{X}] \\ &= E\left[(\bar{F}_0(t))^{\theta}|\underline{X}\right] \\ &= \left[1 - \frac{\ln \bar{F}_0(t)}{\beta + S(\underline{X})}\right]^{-(m+\alpha)}. \end{aligned} \tag{2.16}$$

The Bayes estimator,  $\hat{H}_{BS}$ , of the failure rate function  $H(t)$  is

$$\begin{aligned} \hat{H}_{BS}(t) &= E[H(t)|\underline{X}] \\ &= \frac{f_0(t)}{\bar{F}_0(t)} \hat{\theta}_{BS}. \end{aligned} \tag{2.17}$$

Under the LINEX loss function, the Bayes estimator  $\hat{\theta}_{BL}$  of  $\theta$  is obtained by using (2.10) and (2.14), and is given by

$$\hat{\theta}_{BL} = \frac{m + \alpha}{a} \ln \left[1 + \frac{a}{\beta + S(\underline{X})}\right], \quad a \neq 0. \tag{2.18}$$



Similarly, the Bayes estimator for the reliability function  $R(t)$  is given by

$$\begin{aligned} \widehat{R}_{BL}(t) &= -\frac{1}{a} \ln \left[ \int_0^\infty e^{-aR(t)} \pi(\theta | \underline{x}) d\theta \right] \\ &= -\frac{1}{a} \ln \left[ \int_0^\infty e^{-a[(\bar{F}_0(t))^\theta]} \frac{[\beta + S(\underline{x})]^{m+\alpha}}{\Gamma(m + \alpha)} \theta^{m+\alpha-1} e^{-\theta[\beta+S(\underline{x})]} d\theta \right]. \end{aligned}$$

By using the exponential series

$$\begin{aligned} e^{-a[(\bar{F}_0(t))^\theta]} &= \sum_{k=0}^\infty \frac{(-a)^k}{k!} [\bar{F}_0(t)]^{k\theta} \\ &= \sum_{k=0}^\infty \frac{(-a)^k}{k!} e^{\theta k \ln[\bar{F}_0(t)]} \end{aligned}$$

and after some simplification, we obtain

$$\widehat{R}_{BL}(t) = -\frac{1}{a} \ln \left[ \sum_{k=0}^\infty \frac{(-a)^k}{k!} \left[ 1 - \frac{k \ln \bar{F}_0(t)}{\beta + S(\underline{X})} \right]^{-(m+\alpha)} \right]. \quad (2.19)$$

Also, we obtain the Bayes estimator for the hazard rate function  $H(t)$  as

$$\begin{aligned} \widehat{H}_{BL}(t) &= -\frac{1}{a} \ln \left[ \int_0^\infty e^{-aH(t)} \pi(\theta | \underline{x}) d\theta \right] \\ &= -\frac{1}{a} \ln \left[ \int_0^\infty e^{-a\theta \frac{f_0(t)}{\bar{F}_0(t)}} \frac{[\beta + S(\underline{x})]^{m+\alpha}}{\Gamma(m + \alpha)} \theta^{m+\alpha-1} e^{-\theta[\beta+S(\underline{x})]} d\theta \right] \\ &= \frac{m + \alpha}{a} \ln \left[ 1 + \frac{a}{\beta + S(\underline{X})} \frac{f_0(t)}{\bar{F}_0(t)} \right]. \end{aligned} \quad (2.20)$$

Under the general entropy loss function, the Bayes estimator  $\widehat{\theta}_{BG}$  of  $\theta$  is obtained by using (2.12) and (2.14), and is given by

$$\widehat{\theta}_{BG} = [E(\theta^{-q} | \underline{X})]^{-\frac{1}{q}} = \left( \frac{\Gamma(m + \alpha)}{\Gamma(m + \alpha - q)} \right)^{\frac{1}{q}} [\beta + S(\underline{X})]^{-1}. \quad (2.21)$$

Similarly, the Bayes estimator for the reliability function  $R(t)$  is given by

$$\begin{aligned} \widehat{R}_{BG}(t) &= \left[ \int_0^\infty (R(t))^{-q} \pi(\theta | \underline{x}) d\theta \right]^{-\frac{1}{q}} \\ &= \left[ \int_0^\infty [(\bar{F}_0(t))^\theta]^{-q} \frac{[\beta + S(\underline{x})]^{m+\alpha}}{\Gamma(m + \alpha)} \theta^{m+\alpha-1} e^{-\theta[\beta+S(\underline{x})]} d\theta \right]^{-\frac{1}{q}}, \end{aligned}$$

and after some simplification, we obtain

$$\widehat{R}_{BG}(t) = \left[ 1 + \frac{q \ln \bar{F}_0(t)}{\beta + S(\underline{x})} \right]^{\frac{m+\alpha}{q}}. \tag{2.22}$$

For  $H(t)$ , we obtain

$$\begin{aligned} \widehat{H}_{BG}(t) &= [E([H(t)]^{-q}) | \underline{X}]^{-\frac{1}{q}} \\ &= \frac{f_0(t)}{\bar{F}_0(t)} \widehat{\theta}_{BG} \end{aligned} \tag{2.23}$$

### 2.4 Empirical Bayes Estimation

Assume that the conjugate family of prior distribution for  $\theta$  is the family of gamma distributions,  $\Gamma(\alpha, \beta)$ , with known  $\alpha$  and unknown  $\beta$ . The Bayes estimators obtained in the previous subsection are seen to depend on the parameter  $\beta$ . When the prior parameter  $\beta$  is unknown, we may use the empirical Bayes approach to get its estimate. From (2.2) and (2.13), we calculate the marginal pdf of  $\underline{x}$ , with density

$$\begin{aligned} m(\underline{x}|\beta) &= \int_0^\infty f(\underline{x} | \theta) \pi(\theta | \beta) d\theta \\ &= \int_0^\infty A \left[ \prod_{i=1}^m \frac{f_0(x_i)}{\bar{F}_0(x_i)} \right] \theta^m e^{-\theta S(\underline{x})} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\ &= A \left[ \prod_{i=1}^m \frac{f_0(x_i)}{\bar{F}_0(x_i)} \right] \frac{\Gamma(m + \alpha)}{\Gamma(\alpha)} \frac{\beta^\alpha}{[\beta + S(\underline{x})]^{m+\alpha}}. \end{aligned}$$

Based on  $m(\underline{x} | \beta)$ , we obtain an estimate,  $\widehat{\beta}$ , of  $\beta$ . The MLE of  $\beta$  is

$$\widehat{\beta} = \frac{\alpha}{m} S(\underline{X}). \tag{2.24}$$

Now, by substituting  $\widehat{\beta}$  for  $\beta$  in the different Bayes estimators, we obtain the empirical Bayes estimators of  $\theta$ ,  $R(t)$  and  $H(t)$  as follows:

$$\widehat{\theta}_{EBS} = \frac{m}{S(\underline{X})}, \tag{2.25}$$

$$\widehat{R}_{EBS}(t) = \left[ 1 - \frac{m \ln \bar{F}_0(t)}{(m + \alpha) S(\underline{X})} \right]^{-(m+\alpha)}, \tag{2.26}$$

$$\widehat{H}_{EBS}(t) = \frac{f_0(t)}{\bar{F}_0(t)} \frac{m}{S(\underline{X})}, \tag{2.27}$$

$$\widehat{\theta}_{EBL} = \frac{m + \alpha}{a} \ln\left[1 + \frac{ma}{(m + \alpha)S(\underline{X})}\right], \quad a \neq 0, \tag{2.28}$$

$$\widehat{R}_{EBL}(t) = -\frac{1}{a} \ln \left[ \sum_{k=0}^{\infty} \frac{(-a)^k}{k!} \left[ 1 - \frac{mk \ln \bar{F}_0(t)}{(m + \alpha)S(\underline{X})} \right]^{-(m+\alpha)} \right], \tag{2.29}$$

$$\widehat{H}_{EBL}(t) = \frac{m + \alpha}{a} \ln \left[ 1 + \frac{ma}{(m + \alpha)S(\underline{X})} \right] \frac{f_0(t)}{\bar{F}_0(t)}, \tag{2.30}$$

$$\widehat{\theta}_{EBG} = \left( \frac{\Gamma(m + \alpha)}{\Gamma(m + \alpha - q)} \right)^{\frac{1}{q}} \frac{m}{(\alpha + m)S(\underline{x})}, \tag{2.31}$$

$$\widehat{R}_{EBG}(t) = \left[ 1 + \frac{mq \ln \bar{F}_0(t)}{(\alpha + m)S(\underline{x})} \right]^{\frac{m+\alpha}{q}}, \tag{2.32}$$

$$\widehat{H}_{EBG}(t) = \frac{f_0(t)}{\bar{F}_0(t)} \widehat{\theta}_{EBG}. \tag{2.33}$$

### 3 Prediction

Based on the progressively Type-II right censored sample  $\underline{X} = (X_1, \dots, X_m)$  from the proportional hazards family (1.1), our interest is to find prediction interval for the life-lengths  $X_{s:R_i}$  ( $s = 1, 2, \dots, R_i$ ;  $i = 1, 2, \dots, m$ ) of all censored units in all  $m$  stages of censoring. Here  $Y = X_{s:R_i}$  denotes the  $s$ -th order statistic out of  $R_i$  removed units at stage  $i$  ( $i = 1, 2, \dots, m$ ). Let  $\underline{x} = (x_1, \dots, x_m)$  and  $Y = y$  denote the observed value of  $\underline{X}$  and the unobserved value of  $Y$ , respectively. The conditional distribution of  $Y = X_{s:R_i}$  given  $\underline{X}$  is just the distribution of  $Y$  given  $X_i = x_i$  due to the well-known Markovian property of progressively Type-II censored ordered statistics. It follows (see Balakrishnan and Aggarwala, 2000), that

$$\begin{aligned} f(y|x_i; \theta) &= s \binom{R_i}{s} f(y; \theta) [F(y; \theta) - F(x_i; \theta)]^{s-1} [1 - F(y; \theta)]^{R_i-s} \\ &\times [1 - F(x_i; \theta)]^{-R_i}, \quad y \geq x_i. \end{aligned}$$

For the proportional hazards family, with cdf and pdf given by (1.1) and (1.2), the function  $f(y|x_i; \theta)$  is given by

$$f(y|x_i; \theta) = s \binom{R_i}{s} \theta \frac{f_0(y)}{\bar{F}_0(y)} \left[ (\bar{F}_0(y))^\theta \right]^{R_i-s+1} \left[ (\bar{F}_0(x_i))^\theta - (\bar{F}_0(y))^\theta \right]^{s-1} \times \left[ (\bar{F}_0(x_i))^\theta \right]^{-R_i}, \quad y \geq x_i. \tag{3.1}$$

The Bayes predictive density function of  $Y = X_{s:R_i}$  given  $X_i = x_i$  is given by

$$f^*(y|x_i) = \int_{\Theta} f(y|x_i, \theta) \pi(\theta|\underline{x}) d\theta. \tag{3.2}$$

By substituting (2.14) and (3.1) into (3.2), we get

$$f^*(y|x_i) = \int_0^\infty s \binom{R_i}{s} \theta \frac{f_0(y)}{\bar{F}_0(y)} \left[ (\bar{F}_0(y))^\theta \right]^{R_i-s+1} \left[ (\bar{F}_0(x_i))^\theta - (\bar{F}_0(y))^\theta \right]^{s-1} \times \left[ (\bar{F}_0(x_i))^\theta \right]^{-R_i} \frac{[\beta + S(\underline{x})]^{m+\alpha}}{\Gamma(m+\alpha)} \theta^{m+\alpha-1} e^{-\theta[S(\underline{x})+\beta]} d\theta, \quad y \geq x_i, \tag{3.3}$$

By using bivariate expansion, we have

$$\left[ (\bar{F}_0(x_i))^\theta - (\bar{F}_0(y))^\theta \right]^{s-1} = \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^j \left[ \bar{F}_0(y) \right]^{\theta j} \left[ \bar{F}_0(x_i) \right]^{\theta(s-j-1)}. \tag{3.4}$$

From (3.4), the equation (3.3) can be rewritten as

$$f^*(y|x_i) = s \binom{R_i}{s} \frac{f_0(y)}{\bar{F}_0(y)} \frac{m+\alpha}{\beta + S(\underline{x})} \times \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^j \left[ 1 - \frac{(R_i - s + j + 1) \ln\left(\frac{\bar{F}_0(y)}{\bar{F}_0(x_i)}\right)}{\beta + S(\underline{x})} \right]^{-(m+\alpha+1)}. \tag{3.5}$$

Now, for constructing a Bayesian prediction interval for  $Y = X_{s:R_i}$ , we consider the predictive function  $P(Y \leq \lambda|x_i)$ , for some positive  $\lambda$ . It follows from (3.5), that

$$P(Y \leq \lambda|x_i) = \int_{x_i}^\lambda f^*(y|x_i) dy$$

$$\begin{aligned}
 &= s \binom{R_i}{s} \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^j \frac{1}{R_i - s + j + 1} \\
 &\times \left[ 1 - \left( 1 - \frac{(R_i - s + j + 1) \ln(\frac{\bar{F}_0(\lambda)}{\bar{F}_0(x_i)})}{\beta + S(\underline{x})} \right)^{-(m+\alpha)} \right]. \quad (3.6)
 \end{aligned}$$

Hence, the  $100(1 - \gamma)\%$  prediction interval for  $Y = X_{s:R_i}$  is given by  $(L(x_i), U(x_i))$ , where  $L(x_i)$  and  $U(x_i)$  are the lower and upper prediction bounds, respectively, satisfying

$$Pr[Y \leq L(x_i)|x_i] = \frac{\gamma}{2}, \quad \text{and} \quad P[Y \leq U(x_i)|x_i] = 1 - \frac{\gamma}{2}. \quad (3.7)$$

Iterative numerical methods are required to obtain the lower and upper  $100(1 - \gamma)\%$  prediction bounds for  $Y$  by finding  $\lambda$  from (3.6), using (3.7).

For the special case, when  $s = 1$ , the Eqs. (3.5) and (3.6) reduce to:

$$f^*(y|x_i) = R_i \frac{f_0(y)}{\bar{F}_0(y)} \frac{m + \alpha}{\beta + S(\underline{x})} \left[ 1 - R_i \frac{\ln(\frac{\bar{F}_0(y)}{\bar{F}_0(x_i)})}{\beta + S(\underline{x})} \right]^{-(m+\alpha+1)} \quad y \geq x_i,$$

and

$$P(Y \leq \lambda|x_i) = 1 - \left[ 1 - \frac{R_i}{\beta + S(\underline{x})} \ln\left(\frac{\bar{F}_0(\lambda)}{\bar{F}_0(x_i)}\right) \right]^{-(m+\alpha)}.$$

In this case, we can obtain the Bayes prediction bounds for  $Y = X_{s:R_i}$  as

$$L(x_i) = F_0^{-1} \left[ 1 - \bar{F}_0(x_i) \exp \left\{ \frac{\beta + S(\underline{x})}{R_i} \left( 1 - \left( 1 - \frac{\gamma}{2} \right)^{-\frac{1}{m+\alpha}} \right) \right\} \right],$$

and

$$U(x_i) = F_0^{-1} \left[ 1 - \bar{F}_0(x_i) \exp \left\{ \frac{\beta + S(\underline{x})}{R_i} \left( 1 - \left( \frac{\gamma}{2} \right)^{-\frac{1}{m+\alpha}} \right) \right\} \right]. \quad (3.8)$$

**Example 3.1.** For the case of exponential distribution, the equation (3.8) reduces to:

$$L(x_i) = -\ln \left[ 1 - \exp \left\{ -x_i + \frac{\beta + \sum_{i=1}^m (R_i + 1) \ln x_i}{R_i} \left( 1 - \left( 1 - \frac{\gamma}{2} \right)^{-\frac{1}{m+\alpha}} \right) \right\} \right],$$

and

$$U(x_i) = -\ln \left[ 1 - \exp \left\{ -x_i + \frac{\beta + \sum_{i=1}^m (R_i + 1) \ln x_i}{R_i} \left( 1 - \left( \frac{\gamma}{2} \right)^{-\frac{1}{m+\alpha}} \right) \right\} \right].$$

If the conjugate family of prior distribution for  $\theta$  is the family of Gamma distributions,  $\Gamma(\alpha, \beta)$  with known  $\alpha$  and unknown  $\beta$ . Then the parameter  $\beta$  in the Bayes prediction bounds has to be estimated. We may use the empirical Bayes approach to estimate  $\beta$ . By substituting  $\hat{\beta}$  for  $\beta$  in the Bayes prediction bounds, we can obtain the empirical Bayes prediction bounds for  $Y = X_{s:R_i}$ .

### 4 Numerical Examples

In this section, two numerical examples are presented to illustrate all the estimation and prediction methods described in the preceding sections. We consider the exponential distribution  $E(1/\theta)$  with cdf

$$F(x, \theta) = 1 - e^{-\theta x}, \quad x > 0, \quad \theta > 0,$$

as a special case from the model (1.1). Here, we have

$$\bar{F}_0(x) = e^{-x} \quad \text{and} \quad S(\underline{x}) = \sum_{i=1}^m (R_i + 1) \ln x_i.$$

**Example 4.1. (Simulated Data):** The MLEs, the UMVUEs and Bayes estimates for  $\theta$ ,  $R(t)$  and  $H(t)$  (at  $t = 3$ ), and Bayes and empirical Bayes prediction intervals are computed as described in Sections 2 and 3 according to the following steps:

- (i) For given values of  $\alpha = 1.5$  and  $\beta = 2$ , we generate  $\theta = 0.744$  from the prior pdf (2.13).
- (ii) Using the value  $\theta = 0.744$  from step (i), we generate a progressively Type-II censored sample of size  $m = 10$  with the censoring scheme

$$R = (1, 0, 1, 2, 0, 0, 3, 0, 1, 2)$$

from the exponential distribution according to the algorithm presented in Balakrishnan and Sandhu (1995). The sample generated is

0.0089 0.0199 0.0768 0.2186 0.2693  
 0.3253 0.7980 0.9508 1.0819 2.3185

- (iii) Using this sample, we obtain the MLEs, the UMVUEs and Bayes estimates of  $\theta$ ,  $R(t)$  and  $H(t)$  (at  $t = 3$ ). These estimates are summarized in Table 1.
- (iv) The MLE  $\hat{\beta} = 2.206$  is computed using (2.24).
- (v) Applying the estimate  $\hat{\beta} = 2.206$  in the Bayes estimates, we also computed and reported the empirical Bayes estimates in Table 1.
- (vi) We also computed and reported the 95% Bayes and empirical Bayes prediction intervals for  $Y = X_{s:R_i}$  ( $s = 1, 2, \dots, R_i$ ;  $i = 1, 2, \dots, 10$ ) in Tables 2.

**Table 1.** The MLEs, UMVUEs and Bayes and Empirical Bayes estimates of the parameter, reliability and hazard rate functions in Example 4.1.

	ML	UMVU	BS	BL				BG			
				a				q			
				-2	-0.5	0.5	5	-5	-3	-1	1
$\hat{\theta}$	0.680	0.612	0.688	0.733	0.699	0.678	0.602	0.804	0.747	0.688	0.629
$\hat{R}(t = 3)$	0.130	0.128	0.150	0.157	0.151	0.148	0.134	0.229	0.192	0.150	0.103
$\hat{H}(t = 3)$	0.680	0.612	0.688	0.733	0.699	0.678	0.602	0.804	0.747	0.688	0.629
			EBS	EBL				EBG			
$\hat{\theta}$			0.680	0.724	0.690	0.670	0.596	0.794	0.738	0.680	0.621
$\hat{R}(t = 3)$			0.153	0.160	0.155	0.151	0.137	0.232	0.195	0.153	0.106
$\hat{H}(t = 3)$			0.680	0.724	0.690	0.670	0.596	0.794	0.738	0.680	0.621

**Table 2.** 95% Bayesian and Empirical Bayesian prediction intervals for  $X_{s:R_i}$  in Example 4.1

Bayesian prediction interval	$X_{1:R_1}$	$X_{1:R_3}$	$X_{1:R_4}$	$X_{2:R_4}$	$X_{1:R_7}$
	(0.046,6.326)   (0.114,6.394)   (0.237,3.377)   (0.016,7.920)   (0.810,2.904)				
	$X_{2:R_7}$	$X_{3:R_7}$	$X_{1:R_9}$	$X_{1:R_{10}}$	$X_{2:R_{10}}$
	(0.937,4.979)   (1.274,9.347)   (1.119,7.399)   (2.337,5.447)   (2.116,10.020)				
Empirical Bayesian prediction interval	$X_{1:R_1}$	$X_{1:R_3}$	$X_{1:R_4}$	$X_{2:R_4}$	$X_{1:R_7}$
	(0.046,6.404)   (0.114,6.472)   (0.237,3.416)   (0.013,8.015)   (0.810,2.930)				
	$X_{2:R_7}$	$X_{3:R_7}$	$X_{1:R_9}$	$X_{1:R_{10}}$	$X_{2:R_{10}}$
	(0.939,5.031)   (1.280,9.452)   (1.119,7.477)   (2.337,5.516)   (2.113,10.115)				

**Example 4.2. (Real Data):** We consider the following set of data reported in Nelson (1982, Table 1.1). Nelson presents the results of a life-test experiment in which specimens of a type of electrical insulating fluid were subject to a constant voltage stress(34 KV/minutes). The 19 times to breakdown are:

0.19 0.78 0.96 1.31 2.78 3.16 4.15 4.67 4.85 6.50  
7.35 8.01 8.27 12.06 31.75 32.52 33.91 36.71 72.89

We checked the validity of the exponential model based on the parameter  $\theta = 0.07$ , using the Kolmogorov-Smirnov (K-S) test. It is observed that the K-S distance is  $K - S = 0.2464$  with a corresponding  $p - vale = 0.1678$ . This indicates that the exponential model is adequate for these data. Let us consider the following progressively Type-II censored sample of size  $m = 8$  generated randomly from the  $n = 19$  observations. The observations and the censoring scheme applied, are reported in Table 3. These data has been used earlier by Viveros and Balakrishnan (1994) and Basak et al. (2006).

**Table 3.** Progressively censored data given in Example 4.2.

$i$	1	2	3	4	5	6	7	8
$X_i$	0.19	0.78	0.96	1.31	2.78	4.85	6.50	7.35
$R_i$	0	0	3	0	3	0	0	5

To compute the Bayes and empirical Bayes estimates and Bayes and empirical Bayes prediction intervals , since we do not have any prior information, we assume that  $\alpha = \beta = 0$ . Although it implies an improper prior on  $\theta$ , but the corresponding posterior is proper. The MLEs, the UMVUEs and Bayes estimates for  $\theta$ ,  $R(t)$  and  $H(t)$  (at  $t = 3$ ), and Bayes prediction intervals are given in Tables 4 and 5. Note that in this case  $\hat{\beta} = \beta$ , and hence the empirical Bayes estimates and empirical Bayes prediction intervals correspond with the Bayes estimates and Bayes prediction intervals.



**Table 4.** The MLEs, UMVUEs and Bayes and Empirical Bayes estimates of the parameter, reliability and hazard rate functions in Example 4.2.

	ML	UMVU	BS	BL				BG			
				a				q			
				-2	-0.5	0.5	5	-5	-3	-1	1
$\hat{\theta}$	0.110	0.096	0.110	0.112	0.110	0.110	0.106	0.136	0.123	0.110	0.096
$\hat{R}(t=3)$	0.719	0.745	0.724	0.730	0.725	0.722	0.706	0.741	0.733	0.724	0.714
$\hat{H}(t=3)$	0.110	0.096	0.110	0.112	0.110	0.110	0.106	0.136	0.123	0.110	0.096
			EBS	EBL				EBG			
$\hat{\theta}$			0.110	0.112	0.110	0.110	0.106	0.136	0.123	0.110	0.096
$\hat{R}(t=3)$			0.724	0.730	0.725	0.722	0.706	0.741	0.733	0.724	0.714
$\hat{H}(t=3)$			0.110	0.112	0.110	0.110	0.106	0.136	0.123	0.110	0.096

**Table 5.** 95% Bayesian and Empirical Bayesian prediction intervals for  $X_{s:R_i}$  in Example 4.2.

Bayesian prediction interval	$X_{1:R_3}$	$X_{2:R_3}$	$X_{3:R_3}$	$X_{1:R_5}$	$X_{2:R_5}$	$X_{3:R_5}$
	(1.037,15.154)	(1.821,29.541)	(3.874,59.666)	(2.857,16.975)	(2.061,31.361)	(5.695,61.486)
	$X_{1:R_8}$	$X_{2:R_8}$	$X_{3:R_8}$	$X_{4:R_8}$	$X_{5:R_8}$	
	(7.396,15.867)	(6.956,22.646)	(8.683,31.404)	(5.681,44.830)	(12.604,74.182)	
Empirical Bayesian prediction interval	$X_{1:R_3}$	$X_{2:R_3}$	$X_{3:R_3}$	$X_{1:R_5}$	$X_{2:R_5}$	$X_{3:R_5}$
	(1.037,15.154)	(1.821,29.541)	(3.874,59.666)	(2.857,16.975)	(2.061,31.361)	(5.695,61.486)
	$X_{1:R_8}$	$X_{2:R_8}$	$X_{3:R_8}$	$X_{4:R_8}$	$X_{5:R_8}$	
	(7.396,15.867)	(6.956,22.646)	(8.683,31.404)	(5.681,44.830)	(12.604,74.182)	

From Table 1 and for the sample generated in Example 4.1, we observe most of estimates of parameter  $\theta$  is underestimated, and when  $q = -3$  and  $q = -5$ , the Bayes estimates relative to GEL function are overestimated. From Tables 1 and 4, as anticipated, we note that for for  $a$  close to 0, and  $q = -1$ , all Bayes and empirical Bayes estimates relative to both LINEX loss and GEL function are very close to the corresponding estimates under SEL function. From Tables , we observe that the empirical Bayes estimators and predictors are very close to the Bayes estimators and predictors.

### Acknowledgements

The authors would like to thank the associate editor and a anonymous referee for their thorough review of the paper and their valuable suggestions that improved on the original version of the manuscript.

## References

- Ahmadi, J., Jafari Jozani, M., Marchand, E., and Parsian, A. (2008), Prediction of  $k$ -records from a general class of distributions under balanced type loss functions. *Metrika*, in Press.
- Ahmadi, J., Jafari Jozani, M., Marchand, E., and Parsian, A. (2009), Bayesian estimation based on  $k$ -record data from a general class of distributions under balanced type loss functions. *J. Statist. Plann. Inference*, **139**, 1180- 1189.
- Ahmadi, J., Doostparast, M., and Parsian, A. (2005), Estimation and prediction in a two parameter exponential distribution based on  $k$ -record values under LINEX loss function. *Commun. Statist. Theory. Meth.*, **34**, 795-805.
- Balakrishnan, N. (2007), Progressive censoring methodology: an appraisal. *Test*, **16**, 211-259.
- Balakrishnan, N. and Aggarwala, R. (2000), *Progressive Censoring: Theory, Methods and Applications*. Boston: Birkhäuser.
- Balakrishnan, N. and Sandhu, R. A. (1995), A simple simulational algorithm for generating progressive Type-II censored samples. *Amer. Statist.*, **49**, 229-230.
- Basak, I., Basak, P., and Balakrishnan, N. (2006), On some predictors of times to failure of censored items in progressively censored samples. *Comput. Statist. Data Analys.*, **50**, 1313-1337.
- Basu, A. P. and Ebrahimi, N. (1991), Bayesian approach to life testing and reliability estimation using asymmetric loss function. *J. Statist. Plann. Inference.*, **29**, 21-31.
- Dey, D. K., Ghosh, M., and Srinivasan, C. (1987), Simultaneous estimation of parameters under entropy loss. *J. Statist. Plann. Inference.*, **25**, 347-363.
- Moore, D. and Papadopoulos, A. S. (2000), The Burr type XII distribution as a failure model under various loss functions. *Microelectron. Reliab.*, **40**, 2117-2122.
- Marshall, A. W. and Olkin, O. (2007), *Life Distributions*. New York: Springer.

- Nelson, W. (1982), *Applied life data analysis*. New York: Wiley.
- Parsian, A. and Nematollahi, N. (1996), Estimation of scale parameter under entropy loss function. *J. Statist. Plann. Inference.*, **52**, 77-391.
- Soliman, A. A. (2005), Estimation of parameters of life from progressively censored data using Burr-XII model. *IEE Trans. Reliab.*, **54**(1), 34-42.
- Varrian, H. R. (1975), A Bayesian approach to real estate assessment. In: *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Sarage.*, Amsterdam: North Holland, 195-208.
- Zellner, A. (1986), Bayesian estimation and prediction using asymmetric loss function. *J. Amer. Statist. Assoc.*, **81**, 446-451.