

Bayesian Prediction for Order Statistics from a General Class of Distributions Based on Left Type-II Censored Data

M.M. Mohie El-Din¹, Y. Abdel-Aty² and A.R. Shafay³

^{1,2} Department of Mathematics, Faculty of Science

Al-Azhar University, Cairo, Egypt

¹mmmmoheeldin@yahoo.com

²yahia1970@yahoo.com

³Department of Mathematics, Faculty of Science

Fayoum University, Fayoum, Egypt

a_elshafay2010@yahoo.com

ABSTRACT

In this paper, we consider a general form for the underlying distribution and a general conjugate prior. A general procedure is described to determine one-sample Bayesian prediction intervals for unknown lifetimes based on an observed left Type-II censored data. For the illustration of the developed results, the inverse exponential and inverse Rayleigh distributions are used as examples. Finally, some numerical results are presented to illustrate all the inferential results developed here.

Keywords: order statistics, left Type-II censored sample, Bayesian prediction, the inverse exponential distribution, the inverse Rayleigh distribution.

2010 Mathematics Subject Classification: 62G30, 62F15, 62N01.

1 Introduction

In reliability analysis, experiments often terminate before all units on test have failed due to cost and time considerations. In such cases, failure information is available only on part of the sample, and on all units that had not failed, only partial information will be available. Such data are said to be censored. One of the most common censoring schemes is left Type-II censoring scheme. it can be described as follows. Consider n identical units on a life-testing experiment and $r - 1$ of units had failed even before we started to make proper measurements on failure times. Hence, only the last $n - r + 1$ failures are observed and the first $r - 1$ failures are unobserved. There is a widespread application and use of left censoring or left censored data in survival analysis and reliability theory. For example, in medical studies patients are subject to regular examinations. Discovery of a condition only tells us that the onset of sickness fell in the period since the previous examination and nothing about the exact date of the attack. Thus the time elapsed since onset has been left censored. Different forms of censoring and inference based on such censored data have been discussed in the literature; see, for example, [1].

Let $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ be the order statistics from a random sample of size n from an absolutely continuous distribution function $F(x) \equiv F(x|\theta)$ with density function $f(x) \equiv f(x|\theta)$, where the parameter $\theta \in \Theta$ may be a real vector. These order statistics have been used in a wide range of problems, including robust statistical estimation, detection of outliers, characterization of probability distributions, goodness-of-fit tests, entropy estimation, analysis of censored samples, reliability analysis, quality control and strength of materials; for more details, see [2], [3], [4], [5], and the references contained therein. The joint density of left Type-II censored order statistics is given by

$$f_{\underline{X}}(\underline{x}) = c_{n-r} [F(x_r)]^{r-1} \prod_{i=r}^n f(x_i), \quad (1.1)$$

where $\underline{X} = (X_{r:n}, \dots, X_{n:n})$ is the vector of observable variables, $\underline{x} = (x_r, \dots, x_n)$ is a vector of realizations, and $c_{n-r} = \frac{n!}{(r-1)!}$.

Prediction of future events on the basis of the present knowledge is a fundamental problem of statistics, arising in many contexts and producing varied solutions. As in the case of estimation, a predictor can be either a point or an interval predictor. Several researchers have considered Bayesian prediction for future order statistics based on a right Type-II censored data; see [6], [7], [8], [9], [10], [11], and [12]. Recently, [13] have considered a general form for the underlying distribution, the general inverse exponential form, and a general conjugate prior and developed a general procedure for determining the two-sample Bayesian prediction intervals for future lifetimes based on a right Type-II censored sample. In this paper, we consider a general form for the underlying distribution, the general inverse exponential form, and a general conjugate prior and developed a general procedure for determining the one-sample Bayesian prediction intervals for ordered lifetimes based on a left Type-II censored data.

The rest of this paper is organized as follows. In Section 2, we present the general inverse exponential form for the underlying distribution and the structure of the prior and posterior distributions. In Section 3, we derive the one-sample Bayesian predictive cumulative function and the one-sample Bayesian prediction bounds for the s^{th} ($s < r$) ordered lifetime from the same sample when the observed data is a left Type-II censored sample from the general inverse exponential form for the underlying distribution. In Section 4, we present the results for the inverse exponential and inverse Rayleigh distributions as illustrative examples. Finally, in Section 5, we present some numerical results for illustrating all the inferential methods developed here.

2 Prior and posterior distributions

We consider here the general inverse exponential form for the underlying distribution, suggested by Mohie El-Din et al. (2011), that is described as follows; Motivated by the fact that the cumulative distribution function (*cdf*) can be written in the form

$$F(x|\theta) = \exp[-\lambda(x; \theta)], \quad (2.1)$$

where $\lambda(x; \theta) = -\ln F(x|\theta)$. Of course, some conditions need to be imposed so that $F(x|\theta)$ is a valid *cdf*. These conditions are: $\lambda(x; \theta)$ is continuous, monotone decreasing and differentiable

function, with $\lambda(x; \theta) \rightarrow \infty$ as $x \rightarrow -\infty$ and $\lambda(x; \theta) \rightarrow 0$ as $x \rightarrow \infty$. The probability density function (*pdf*) corresponding to (2.1) is given by

$$f(x|\theta) = \lambda'(x; \theta) \exp[-\lambda(x; \theta)], \quad (2.2)$$

where $\lambda'(x; \theta)$ is the derivative of $\lambda(x; \theta)$ with respect to x . With an appropriate choice of $\lambda(x; \theta)$, several distributions that are used in reliability studies can be obtained as special cases such as the inverse exponential, inverse Rayleigh, inverse Weibull, inverse Pareto, negative exponential, negative Weibull, negative Pareto, negative power, Gumbel, exponentiated-Weibull, loglogistic, Burr X, inverse Burr XII and inverse paralogistic distributions.

Upon using (2.1) and (2.2) in (1.1), we obtain the likelihood function of the left Type-II censored data $\underline{X} = (X_{r:n}, \dots, X_{n:n})$ as

$$L(\theta; \underline{x}) \propto A(\theta; \underline{x}) \exp[-B(\theta; \underline{x})], \quad (2.3)$$

where

$$A(\theta; \underline{x}) = \prod_{i=r}^n \{-\lambda'(x_{j_i}; \theta)\}$$

and

$$B(\theta; \underline{x}) = r\lambda(x_r; \theta) + \sum_{i=r+1}^n \lambda(x_i; \theta).$$

From the Bayesian viewpoint, the unknown parameter is regarded as a realization of a random variable, which has some prior distribution. We consider here a general conjugate prior, suggested by [11], that is given by

$$\pi(\theta; \delta) \propto C(\theta; \delta) \exp[-D(\theta; \delta)], \quad (2.4)$$

where $\theta \in \Theta$ is the vector of parameters of the distribution in (2.1) and δ is the vector of prior parameters. The prior family in (2.4) includes several priors used in the literature as special cases.

Then, from (2.3) and (2.4), the posterior density function of θ , given the left Type-II censored data, is readily obtained as

$$\pi^*(\theta|\underline{x}) = I^{-1} \eta(\theta; \underline{x}) \exp[-\zeta(\theta; \underline{x})], \quad (2.5)$$

where

$$\eta(\theta; \underline{x}) = C(\theta; \delta) A(\theta; \underline{x}), \quad \zeta(\theta; \underline{x}) = B(\theta; \underline{x}) + D(\theta; \delta),$$

and

$$I = \int_{\theta \in \Theta} \eta(\theta; \underline{x}) \exp[-\zeta(\theta; \underline{x})].$$

3 One-sample Bayesian predication

In this section, we present the one-sample Bayesian predictive cumulative function and the one-sample Bayesian prediction bounds for the s^{th} ($s < r$) ordered lifetime when the observed data is a left Type-II censored sample follows the distribution (2.1).

It is well known that the joint density function of the s^{th} and r^{th} order statistics ($s < r$) from a sample of size m from a continuous distribution with a *cdf* $F(x)$ and a *pdf* $f(x)$ is given by

$$f_{X_{s:n}, X_{r:n}}(x_s, x_r) = \frac{n!}{(s-1)!(r-s-1)!(n-r)!} [F(x_s)]^{s-1} [F(x_r) - F(x_s)]^{r-s-1} \times [1 - F(x_r)]^{n-r} f(x_s) f(x_r) \quad (3.1)$$

and the marginal density function of r^{th} order statistics $X_{r:n}$ is given by

$$f_{X_{r:n}}(x_r) = \frac{n!}{(n-r)!(r-1)!} [F(x_r)]^{r-1} [1 - F(x_r)]^{n-r} f(x_r). \quad (3.2)$$

By using the Markov Chain property of the order statistics we have

$$\begin{aligned} f_{X_{s:n}}(x_s | \underline{x}) &= f(x_s | x_r) = \frac{f_{X_{s:n}, X_{r:n}}(x_s, x_r)}{f_{X_{r:n}}(x_r)} \\ &= \frac{(r-1)!}{(s-1)!(r-s-1)!} [F(x_s)]^{s-1} [F(x_r) - F(x_s)]^{r-s-1} \frac{f(x_s)}{[F(x_r)]^{r-1}}. \end{aligned} \quad (3.3)$$

Upon using (2.1) and (2.2) in (3.3), we obtain the conditional density of $X_{s:n}$, given the left Type-II censored data $\underline{X} = (X_{r:n}, \dots, X_{n:n})$, as

$$f_{X_{s:n}}(x_s | \underline{x}) = \frac{(r-1)!}{(s-1)!} \sum_{w=0}^{r-s-1} c_w (r-s-1) \{-\lambda'(x_s; \theta)\} \exp[-(s+w)\{\lambda(x_s; \theta) - \lambda(x_r; \theta)\}], \quad (3.4)$$

where $c_w (r-s-1) = \frac{(-1)^w}{w!(r-s-w-1)!}$ for $w = 0, \dots, r-s-1$.

From (2.5) and (3.4), we simply obtain the Bayesian predictive density function of $X_{s:n}$ as

$$\begin{aligned} f_{X_{s:n}}^*(x_s | \underline{x}) &= \int_{\theta \in \Theta} f_{X_{s:n}}(x_s | \underline{x}) \pi^*(\theta | \underline{x}) d\theta \\ &= \frac{(r-1)! I^{-1}}{(s-1)!} \sum_{w=0}^{r-s-1} c_w (r-s-1) \int_{\theta \in \Theta} \{-\lambda'(x_s; \theta)\} \eta(\theta; \underline{x}) \\ &\quad \times \exp[-\zeta(\theta; \underline{x}) - (s+w)\{\lambda(x_s; \theta) - \lambda(x_r; \theta)\}] d\theta. \end{aligned} \quad (3.5)$$

From (3.5), we simply obtain the predictive cumulative distribution function $F_{X_{s:n}}^*(t | \underline{x})$ as

$$\begin{aligned} F_{X_{s:n}}^*(t | \underline{x}) &= \int_0^t f_{X_{s:n}}^*(x_s | \underline{x}) dx_s \\ &= \frac{(r-1)! I^{-1}}{(s-1)!} \sum_{w=0}^{r-s-1} \frac{c_w (r-s-1)}{s+w} \int_{\theta \in \Theta} \eta(\theta; \underline{x}) \exp[-\zeta(\theta; \underline{x}) - (s+w)\{\lambda(t; \theta) - \lambda(x_r; \theta)\}] d\theta. \end{aligned} \quad (3.6)$$

Consequently, the Bayesian predictive bounds of a two-sided equi-tailed $100(1-\gamma)\%$ interval for $X_{s:n}$, $1 \leq s \leq r-1$, can be obtained by solving the following two equations:

$$F_{X_{s:n}}^*(L_{X_{s:n}} | \underline{x}) = \frac{\gamma}{2} \quad \text{and} \quad F_{X_{s:n}}^*(U_{X_{s:n}} | \underline{x}) = 1 - \frac{\gamma}{2},$$

where $L_{X_{s:n}}$ and $U_{X_{s:n}}$ denote the lower and upper bounds, respectively.

4 Illustrative examples

In this section, we discuss the Bayesian prediction problems for the inverse exponential and inverse Rayleigh distributions as illustrative examples.

4.1 The inverse exponential distribution

The distribution function in this case is

$$F(x|\theta) = 1 - \exp[-\theta/x], \quad x > 0, \quad (4.1)$$

where $\theta > 0$, and so we have

$$\lambda(x; \theta) = \frac{\theta}{x} \quad \text{and} \quad \lambda'(x; \theta) = -\frac{\theta}{x^2}. \quad (4.2)$$

For the case when θ is unknown, we use the conjugate gamma prior for θ with density

$$\pi(\theta; \delta) = \frac{d^c}{\Gamma(c)} \theta^{c-1} \exp[-\theta d], \quad \theta > 0, \quad (4.3)$$

where c and d are positive constants, and so we have

$$C(\theta; \delta) = \theta^{c-1} \quad \text{and} \quad D(\theta; \delta) = \theta d, \quad (4.4)$$

where $\delta = (c, d)$.

Then, the predictive cumulative function of $X_{s:n}$ in this case is given by

$$\begin{aligned} F_{X_{s:n}}^*(t|\underline{x}) &= \frac{(r-1)! I_*^{-1}}{(s-1)!} \sum_{w=0}^{r-s-1} \frac{c_w(r-s-1)}{s+w} \int_0^\infty \theta^{n-r+c} \exp[-\theta \{ \frac{r-s-w}{x_r} + \sum_{i=s+1}^n \frac{1}{x_i} + \frac{s+w}{t} + d \}] d\theta \\ &= \frac{(r-1)! I_*^{-1}}{(s-1)!} \sum_{w=0}^{r-s-1} \frac{c_w(r-s-1)}{s+w} \left[\frac{r-s-w}{x_r} + \sum_{i=s+1}^n \frac{1}{x_i} + \frac{s+w}{t} + d \right]^{-(n-r+c+1)}, \quad (4.5) \end{aligned}$$

where

$$I_* = \frac{I}{\Gamma(n-r+c+1)} = \left[\frac{r}{x_r} + \sum_{i=r+1}^n \frac{1}{x_i} + d \right]^{-(n-r+c+1)}.$$

4.2 The inverse Rayleigh distribution

The distribution function in this case is

$$F(x|\theta) = 1 - \exp[-\theta/x^2], \quad x > 0, \quad (4.6)$$

where $\theta > 0$, and so we have

$$\lambda(x; \theta) = \frac{\theta}{x^2} \quad \text{and} \quad \lambda'(x; \theta) = -\frac{2\theta}{x^3}. \quad (4.7)$$

For the case when θ is unknown, we use the conjugate gamma prior for θ with density

$$\pi(\theta; \delta) = \frac{d^c}{\Gamma(c)} \theta^{c-1} \exp[-\theta d], \quad \theta > 0, \quad (4.8)$$

where c and d are positive constants, and so we have

$$C(\theta; \delta) = \theta^{c-1} \quad \text{and} \quad D(\theta; \delta) = \theta d, \quad (4.9)$$

where $\delta = (c, d)$.

Then, the predictive cumulative function of $X_{s:n}$ in this case is given by

$$\begin{aligned} F_{X_{s:n}}^*(t|\underline{x}) &= \frac{(r-1)!}{(s-1)!} \sum_{w=0}^{r-s-1} \frac{c_w(r-s-1)}{s+w} \int_0^\infty \theta^{n-r+c} \exp[-\theta \{ \frac{r-s-w}{x_r^2} + \sum_{i=s+1}^n \frac{1}{x_i^2} + \frac{s+w}{t^2} + d \}] d\theta \\ &= \frac{(r-1)!}{(s-1)!} \sum_{w=0}^{r-s-1} \frac{c_w(r-s-1)}{s+w} \left[\frac{r-s-w}{x_r^2} + \sum_{i=s+1}^n \frac{1}{x_i^2} + \frac{s+w}{t^2} + d \right]^{-(n-r+c+1)}, \quad (4.10) \end{aligned}$$

where

$$I_* = \frac{I}{\Gamma(n-r+c+1)} = \left[\frac{r}{x_r^2} + \sum_{i=r+1}^n \frac{1}{x_i^2} + d \right]^{-(n-r+c+1)}.$$

5 Numerical results

To illustrate the inferential procedures developed in the preceding sections, we present here a numerical study for the inverse exponential and inverse Rayleigh distributions.

Example 1: To illustrate the prediction results for the the inverse exponential(θ) distribution when the parameter θ is unknown, we generated order statistics from a sample of size $n = 20$ from the inverse exponential distribution. The generated order statistics from the inverse exponential distribution (with $\theta = 5$) are as follows:

1.1223, 1.4670, 2.4733, 2.4801, 2.6632, 2.7619, 3.5070, 3.7613, 3.9203, 5.2278, 6.09525, 6.1555, 7.2720, 7.7767, 7.8503, 8.8911, 10.0302, 11.0626, 19.7074 and 61.0742.

We assume these data to have come from the inverse exponential(θ) distribution, where the parameter θ is unknown. Based on left Type-II censoring schemes with $r = 10$, we then used the results presented earlier in Subsection 4.1 to construct 95% one-sample Bayesian prediction intervals for the order statistics $X_{s:n}$, $s = 1, \dots, 9$, from the same sample. To examine the sensitivity of the Bayesian prediction intervals with respect to the hyper-parameters (c, d) , we used four different choices of the hyper-parameters (c, d) : $(5, 1)$, $(4.8, 1)$, $(5, 1.2)$, $(0, 0)$. The corresponding results for the one-sample prediction, for these four choices of the hyper-parameters, are presented in Table 1.

Example 2: To illustrate the prediction results for the the inverse Rayleigh(θ) distribution when the parameter θ is unknown, we generated order statistics from a sample of size $n = 20$ from the inverse Rayleigh distribution. The generated order statistics from the inverse Rayleigh distribution (with $\theta = 10$) are as follows:

1.3325, 1.8573, 2.3255, 2.4919, 2.5969, 2.6855, 2.7378, 2.7612, 2.9035, 4.06054, 4.2340, 4.3375, 4.45054, 4.4556, 4.5248, 4.9985, 6.4970, 6.8262, 9.4232 and 20.9983.

We assume these data to have come from the inverse Rayleigh(θ) distribution, where the parameter θ is unknown. Based on left Type-II censoring schemes with $r = 10$, we then used

Table 1: 95% Bayesian prediction bounds for $X_{s:n}$, $s = 1, \dots, 9$, from the inverse exponential distribution.

s	(c, d)							
	$(5, 1)$		$(4.8, 1)$		$(5, 1.2)$		$(0, 0)$	
	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$
1	0.7097	3.2973	0.7001	3.2745	0.6680	3.1675	0.6847	3.4538
2	1.0903	4.1147	1.0762	4.0909	1.0293	3.9786	1.0488	4.2765
3	1.4281	4.8292	1.4104	4.8060	1.3518	4.6963	1.3758	4.9818
4	1.7749	5.5089	1.7541	5.4875	1.6846	5.3869	1.7153	5.6427
5	2.1570	6.1709	2.1333	6.1525	2.0534	6.0667	2.0931	6.2785
6	2.5997	6.8118	2.5734	6.7976	2.4835	6.7319	2.5354	6.8883
7	3.1387	7.4093	3.1103	7.4002	3.0113	7.3584	3.0788	7.4536
8	3.8384	8.5494	3.8089	8.5533	3.7032	8.5711	3.7906	8.5312
9	4.8626	9.2091	4.8345	9.2087	4.7301	9.2068	4.8393	9.2106

the results presented earlier in Subsection 4.2 to construct 95% one-sample Bayesian prediction intervals for the order statistics $X_{s:n}$, $s = 1, \dots, 9$, from the same sample. To examine the sensitivity of the Bayesian prediction intervals with respect to the hyper-parameters (c, d) , we used four different choices of the hyper-parameters (c, d) : $(10, 1)$, $(9, 1)$, $(10, 1.5)$, $(0, 0)$. The corresponding results for the one-sample prediction, for these four choices of the hyper-parameters, are presented in Table 2.

Table 2: 95% Bayesian prediction bounds for $X_{s:n}$, $s = 1, \dots, 9$, from the inverse Rayleigh distribution.

s	(c, d)							
	$(10, 1)$		$(9, 1)$		$(10, 1.5)$		$(0, 0)$	
	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$	$L_{X_{s:n}}$	$U_{X_{s:n}}$
1	0.5174	1.9997	0.4928	1.9535	0.4777	1.9076	0.3124	1.6194
2	0.7782	2.4036	0.7434	2.3587	0.7226	2.3137	0.4802	2.0218
3	0.9983	2.7340	0.9565	2.6928	0.9315	2.6515	0.6319	2.3722
4	1.2141	3.0299	1.1668	2.9942	1.1381	2.9585	0.7905	2.7051
5	1.4407	3.3021	1.3892	3.2731	1.3574	3.2444	0.9682	3.0297
6	1.6903	3.5513	1.6358	3.5301	1.6014	3.5094	1.1779	3.3450
7	1.9771	3.7717	1.9214	3.7587	1.8850	3.7463	1.4382	3.6409
8	2.3247	3.9480	2.2705	3.9427	2.2339	3.9376	1.7834	3.8917
9	2.7903	4.0490	2.7434	4.0484	2.7097	4.0479	2.3010	4.0478

Concluding remarks

1. It is evident from Tables 1 and 2 that, in the cases of the inverse exponential and inverse Rayleigh distributions, the lower as well as upper bounds are relatively insensitive to the specification of the hyper-parameters (c, d) .
2. If the vector of prior parameters δ is unknown, the empirical Bayes approach could be used in estimating such prior parameters based on past samples; see, for example, [14]. Alternatively, one could use the hierarchical Bayesian method in which some suitable prior for δ could be proposed; see, for example, [15] and [16]. Work in these directions are

currently under progress and we hope to report these findings in a future paper.

References

- [1] A. C. Cohen, *Truncated and Censored Samples*, Marcel Dekker, Inc., New York, 1991.
- [2] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, *A First Course in Order Statistics*, Wiley, New York, 1992.
- [3] H. A. David and H. N. Nagaraja, *Order Statistics*, Third edition, John Wiley & Sons, Hoboken, New Jersey, 2003.
- [4] N. Balakrishnan and C. R. Rao, *Order Statistics: Theory and Methods*, Handbook of statistics - vol. 16, North-Holland, Amsterdam, 1998.
- [5] N. Balakrishnan and C. R. Rao, *Order Statistics: Theory and Methods*, Handbook of statistics - vol. 17, North-Holland, Amsterdam, 1998.
- [6] I. R. Dunsmore, The Bayesian predictive distribution in life testing models, *Technometrics*, **16** (1974), 455–460.
- [7] A. M. Nigm and H. I. Hamdy, Bayesian prediction bounds for the Pareto lifetime model, *Communications in Statistics – Theory and Methods* **16** (1987), 1761–1772.
- [8] A. M. Nigm and H. I. Hamdy, Prediction bounds for the Burr model, *Communications in Statistics – Theory and Methods* **17** (1988), 289–297.
- [9] A. M. Nigm, An informative Bayesian prediction for the Weibull lifetime distribution, *Communications in Statistics – Theory and Methods* **18** (1989), 897–911.
- [10] E. K. AL-Hussaini and Z. F. Jaheen, Bayesian prediction bounds for the Burr type XII model, *Communications in Statistics – Theory and Methods* **24** (1995), 1829–1842.
- [11] E. K. AL-Hussaini, Predicting observables from a general class of distributions, *Journal of Statistical Planning and Inference* **79** (1999), 79–91.
- [12] M. Z. Raqab and M. T. Madi, Bayesian inference for the generalized exponential distribution, *Journal of Statistical Computation and Simulation* **75** (2005), 841–852.
- [13] M. M. Mohie El-Din, Y. Abdel-Aty, and A. R. Shafay, Two sample Bayesian prediction for order statistics based on the inverse exponential type distributions using right censored sample, *Journal of the Egyptian Mathematical Society* (2011) (to appear).
- [14] J. S. Maritz and T. Lwin, *Empirical Bayes Methods*, Second edition, Chapman & Hall, London, England, 1989.
- [15] S. Geisser, On hierarchical Bayes procedures for predicting simple exponential survival, *Biometrics* **46** (1990), 225–230.
- [16] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, John Wiley & Sons, New York, 1994.