J. Appl. Res. Ind. Eng. Vol. 4, No. 1 (2017) 67-74

**Journal of Applied Research on Industrial Engineering**
**www.journal-aprie.com**

# Optimal Number of Failures in Type II Censoring for Rayleigh Distribution

Elham Basiri*
*Department of statistics, Kosar University of Bojnord, Bojnord, Iran.*

**PAPER INFO**

**Chronicle:**
Received: 12 June 2017
Revised: 04 July 2017
Accepted: 09 July 2017
Available : 11 July 2017

**Keywords** :
Type II censoring.
Rayleigh distribution.
Bayesian point predictor.
Cost function.

**ABSTRACT**

Recently, Rayleigh distribution has received considerable attention in the statistical literature. This paper describes the Bayesian prediction of the one parameter Rayleigh distribution when the data are Type II censored data. For Type II censoring one question that arises is "How many failure is enough?". The aim of this paper is responding to this question by considering two criteria, total cost of experiment and mean squared prediction error in prediction problem. Towards this end, we find the Bayesian point predictor for the parameter of distribution. Then, the optimal value for number of failures is obtained when the mean squared prediction error and the total cost of experiment are bounded. Finally, to show the usefulness of the obtained results, a simulation study is presented.

## 1. Introduction

In reliability and life testing analysis, often the data are censored. Among the different censoring schemes, Type I and Type II censoring schemes are the most used ones in reliability and life testing experiments. In this paper, we mainly restrict our attention on Type II censoring. In Type II censoring it is assumed that $n$ items are put on a test. The integer $r$, $r < n$, is a pre-fixed value, and the experiment stops as soon as the $r$-th failure is observed. Then, the observed failure times $(X_{1:n}. \dots . X_{r:n})$ are called "Type II censored order statistics". For more details on Type II censoring we refer the reader to David and Nagaraja [1] and Arnold et al. [2].

Rayleigh distribution has been widely used in reliability theory and survival analysis, as its failure rate is a linear function of time. The origin and other aspects of this distribution can be found in Siddiqui [3] and Miller and Sackrowitz [4]. Several authors have carried out extensive studies as relate to the estimation, prediction and several other inferences with respect to Rayleigh distribution. For example, Kotb and Raqab [5] studied Bayesian estimation and prediction based on ordered ranked set sampling from Rayleigh distribution.

* Corresponding author
E-mail address: elham_basiri2000@yahoo.com
DOI: 10.22105/jarie.2017.48291

Exact confidence intervals and regions are proposed for the location and scale parameters of the Rayleigh distribution are obtained by Asgharzadeh et al. [6]. Seo et al. [7] provided the exact confidence intervals for unknown parameters and exact predictive intervals for the future upper record values by providing some pivotal quantities in the two-parameter Rayleigh distribution based on the upper record values.

Determination of optimal number of failures in type II censoring is an important issue in designing a life testing experiment. Suppose we are planning to collect a Type II censored sample from the one parameter Rayleigh distribution in order to find a point predictor for a future order statistic with smallest mean squared prediction error (MSPE) among other point predictors. Although, considering large values for failure numbers yields a point predictor with smaller MSPE, the average cost may increase considerably. One question arises here that "How many failure is enough?". In this paper we respond to the question by considering the prediction problem as well as cost function. Many authors have studied some similar problems by considering the cost of experiment. For example, Ng et al. [8] determined the optimal censoring plan in progressively Type II censoring based on some criteria such as the cost of experiment. Doostparast and Balakrishnan [9] discussed the optimal sample size for estimating the mean based on a criterion involving a cost function as well as the Fisher information based on records arising from a random sample. Ahmadi et al. [10] obtained optimal value for the sample size in the prediction problem by considering the total cost of experiment. Cordeiro and Pham [11] determined the optimum sample size on test which minimizes the expected total cost of performing the life testing subject to the unknown parameters of the Weibull distribution lifetime for a fixed number of failures. The optimization problem of sample size allocation when the competing risks data are from a progressive type-II censoring in a constant-stress accelerated life test with multiple levels, is studied by Huang and Wu [12].

The rest of this paper is organized as follows: In Section 2, we first obtain point Bayesian predictor for future order statistics from the one parameter Rayleigh distribution. Then, by considering two criteria as mean squared prediction error, MSPE, and total cost of the test optimal number of failures is determined. A simulation study is presented in Section 3.

## 2. Main results

Throughout this paper, let $Y_{s:m}$ be the $s$-th order statistic from a sample of size $m$ of independent and identically distributed (i.i.d.) continuous random variables from the one-parameter Rayleigh distribution, denoted by $Ray(\theta)$, with probability density function (pdf) and cumulative distribution function (cdf)

$$f_\theta(x) = 2\theta x e^{-\theta x^2}. \qquad x > 0.\ \ \theta > 0.$$

and

$$F_\theta(x) = 1 - e^{-\theta x^2}. \qquad x > 0\ .\ \theta > 0.$$

respectively. Also, let $\tilde{x} = (x_{1:n}. \dots . x_{r:n})$ be the first $r$ observed order statistics from the same distribution. We want to find the optimal value for the number of failures in the observed sample such that the point predictor of $Y_{s:m}$, say $\hat{Y}_{s:m}$, has small mean squared prediction error (MSPE) when the budget of the life test is restricted. To do this, first we obtain the point predictor $\hat{Y}_{s:m}$. The likelihood function takes the form (see for example, Arnold et al. [2] )

$$\begin{aligned} L(\theta|\tilde{x}) &= \frac{n!}{(n-r)!}\left(1 - F_\theta(x_r)\right)\prod_{i=1}^{r} f_\theta(x_i) \\ &= \frac{n!}{(n-r)!}\,\theta^r 2^r \left(\prod_{i=1}^{r} x_i\right) e^{-\theta T}. \end{aligned} \quad (1)$$

Where,

$$T = \sum_{i=1}^{r} x_i^2 + (n-r)x_r^2.$$

The conjugate prior distribution for $\theta$ is considered as:

$$\Pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}. \quad \theta > 0. a > 0. b > 0. \quad (2)$$

where $\Gamma$ is the complete Gamma function. Then, from (1) and (2) the posterior distribution will be easily obtained as :

$$\Pi(\theta|\tilde{x}) = \frac{(b+T)^{a+r}}{\Gamma(a+r)}\theta^{a+r-1}e^{-\theta(b+T)}. \quad \theta > 0. a > 0. b > 0. \quad (3)$$

On the other hand, the pdf of $Y_{s:m}$, is ( see for example, Arnold et al. [2])

$$f_{Y_{s:m}}(y) = \frac{2\theta y}{\beta(s.m-s+1)} e^{-\theta\,(m-s+1)y^2}(1 - e^{-\theta\, y^2})^{s-1}. \qquad y > 0. \ \theta > 0. \quad (4)$$

where $\beta$ is the complete beta function. So, from (3) and (4) the predictive density function for $Y_{s:m}$ can be obtained as

$$\begin{aligned} f^*_{Y_{s:m}}(y|\tilde{x}) &= \frac{2y(b+T)^{r+a}}{\beta(s.m-s+1)\Gamma(r+a)} \sum_{j=0}^{s-1} \binom{s-1}{j} (-1)^j \int_0^\infty \theta^{r+a} e^{-\theta[\,(m-s+1)y^2+b+T]}\, d\theta \\ &= \frac{2y(b+T)^{r+a}(r+a)}{\beta(s.m-s+1)} \sum_{j=0}^{s-1} \frac{\binom{s-1}{j}(-1)^j}{[\,(m-s+1)y^2+b+T]^{r+a+1}}. \end{aligned}$$

The point predictor for $Y_{s:m}$under squared error loss (SEL) function is:

$$\begin{aligned}\hat{Y}_{s:m} &= \frac{(b+T)^{r+a}}{\Gamma(r+a)}\sum_{j=0}^{s-1}\frac{\binom{s-1}{j}(-1)^j}{\beta(s.m-s+1)}\int_0^\infty \theta^{r+a}e^{-\theta[b+T]}\int_0^\infty 2y^2e^{-\theta\,(m-s+1)y^2}dy\ \ d\theta\\ &= \frac{\Gamma\left(\frac{3}{2}\right)(b+T)^{r+a}}{\Gamma(r+a)}\sum_{j=0}^{s-1}\frac{\binom{s-1}{j}(-1)^j}{\beta(s.m-s+1)(m-s+j+1)^{\frac{3}{2}}}\int_0^\infty \theta^{r+a-\frac{3}{2}}e^{-\theta[b+T]}\,d\theta\\ &= K(s.m)(b+T)^{\frac{1}{2}}\frac{\Gamma\left(r+a-\frac{1}{2}\right)}{\Gamma(r+a)}.\end{aligned}$$

Where,

$$K(s.m)=\Gamma\left(\frac{3}{2}\right)\sum_{j=0}^{s-1}\frac{\binom{s-1}{j}(-1)^j}{\beta(s.m-s+1)(m-s+j+1)^{\frac{3}{2}}}. \qquad (5)$$

Since in many real applications, no prior knowledge is available about θwe may take $a=b=0$, i.e. the Jeffrey's non-informative prior for θ. Therefore

$$\hat{Y}_{s:m}=K(s.m)T^{\frac{1}{2}}\frac{\Gamma\left(r-\frac{1}{2}\right)}{\Gamma(r)}. \qquad (6)$$

Moreover, we can write

$$\begin{aligned}E(Y_{s:m}) &= \frac{\theta}{\beta(s.m-s+1)}\int_0^\infty 2y^2e^{-\theta\,(m-s+1)y^2}(1-e^{-\theta y^2})^{s-1}dy\\ &= \frac{\theta}{\beta(s.m-s+1)}\sum_{j=0}^{s-1}\binom{s-1}{j}(-1)^j\int_0^\infty 2y^2e^{-\theta\,(m-s+j+1)y^2}dy\\ &= \frac{1}{\theta^{\frac{1}{2}}}K(s.m).\end{aligned}$$

Also, if $Z_1.\cdots.Z_n$ are i.i.d random variables from the exponential distribution with parameter θ, then (see for example, Arnold et al. [2])

$$Y_{s:m}^2\overset{d}{=}\sum_{l=1}^{s}\frac{Z_l}{m-l+1},$$

where $\overset{d}{=}$ stands for identical in distribution. So, we have,

$$E(Y_{s:m}^2)=\frac{1}{\theta}g(s.m).$$

where

$$g(s.m)=\sum_{l=1}^{s}\frac{1}{m-l+1}$$

Then, we find that,

$$Var(Y_{s:m})=\frac{1}{\theta}\{g(s.m)-K^2(s.m)\}$$

Also, T has Gamma distribution with parameters r and θ. So, from (6) we can write

$$E(\hat{Y}_{s:m}) = \frac{1}{\theta^{\frac{1}{2}}} K(s,m) \frac{\Gamma(r-\frac{1}{2})\Gamma(r+\frac{1}{2})}{\Gamma^2(r)}$$

And,

$$E(\hat{Y}_{s:m}^2) = \frac{r}{\theta} K^2(s,m) \frac{\Gamma^2(r-\frac{1}{2})}{\Gamma^2(r)}.$$

So, we get,

$$Var(\hat{Y}_{s:m}) = \frac{1}{\theta} K^2(s,m) \frac{\Gamma^2(r-\frac{1}{2})}{\Gamma^2(r)} \{r - \frac{\Gamma^2(r+\frac{1}{2})}{\Gamma^2(r)}\}.$$

Thus, we can consider mean squared prediction error, MSPE, of $\hat{Y}_{s:m}$ as:

$$MSPE(\hat{Y}_{s:m}) = \frac{1}{\theta}\{K^2(s,m)Q(r) + g(s,m)\}, \tag{7}$$

Where,

$$Q(r) = \frac{\Gamma(r-\frac{1}{2})}{\Gamma^2(r)}\left[r\Gamma(r-\frac{1}{2}) - 2\Gamma(r+\frac{1}{2})\right]. \tag{8}$$

Another criterion considered in this paper is the total cost of test, which plays an important role in practices. The total cost associated with the information sample $\tilde{x} = (x_{1:n}, \cdots, x_{r:n})$ is given by

$$TC = c_0 + c_t X_{r:n} + nc_u,$$

are $c_u$ and $c_t$, $c_0$ wherethe sampling set-up cost or any other related cost involved in sampling, cost of total time on test and cost per unit, respectively. We consider the expected of cost function which is given by

$$\begin{aligned} E(TC) &= c_0 + c_t E(X_{r:n}) + nc_u \\ &= c_0 + \frac{c_t}{\theta^{\frac{1}{2}}} K(r,n) + nc_u, \end{aligned} \tag{9}$$

which depends on the unknown parameter $\theta$ and therefore it can be replaced by its preliminary estimator based on past experiments and pre-information.

In the sequel, we try to find optimal value for the number of failures, $r$, such that $E(TC) < c$ and $MSPE(\hat{Y}_{s:m}) < p$, where $c$ and $p$ are pre-fixed values. Based on (9), the first condition is equivalent to

$$K(r,n) < \frac{(c - c_0 - nc_u)}{c_t} \theta^{\frac{1}{2}}, \tag{10}$$

where $K(r,n)$ is defined as in (5). From (7), the second condition is equivalent to

$$Q(r) < \frac{(p\theta - g(s,m))}{K^2(s,m)}, \tag{11}$$

where $Q(r)$ is defined as in (8). Then, the optimal value for $r$, say $r_{opt}$, is satisfied in both conditions in (10) and (11).

In Table 1, we have presented some values of $r_{opt}$ by considering different values of s, c and p, when
$m = n = 10, c_0 = 1, c_t = 1, c_u = 0.2$ and $\theta = 1$.

In Table 1, dash (-) means that there is no $r_{opt}$ which satisfies both conditions (10) and (11). As one would expect, $r_{opt}$ increases as s increases when all other components are held fixed. By increasing the number of failures, total cost of test increases. Therefore, for fixed values of s and p, the maximum value of $r_{opt}$ is an increasing function of c. MSPE is a decreasing function of r. So, for fixed values of s and c, the minimum value of $r_{opt}$ is a decreasing function of $p$. Values of $r_{opt}$ for lower order statistics are smaller than those for upper order statistics. From Table 1, we can see that for upper order statistics there are some cases that $r_{opt}$ does not exist and for others cases $r_{opt}$ is not unique. In practice, the engineer may request only one value for r. So, this question arises for a practitioner that which $r_{opt}$ must be chosen in a practical situation. We respond that it depends on which criterion is more important for that user. If the cost of experiment is more important criterion than MSPE, the lower bound for $r_{opt}$ can be considered. On the other hand, if we consider the MSPE criterion as the most important one, we must prefer the upper bound for $r_{opt}$ , as we expected intuitively.

**Table 1.** Values of $r_{opt}$ for different values of s, c and p when
$m = n = 10, c_0 = 1, c_t = 1, c_u = 0.2 \quad and \quad \theta = 1$.

| | | P=0.05 | P=0.1 | P=0.5 |
|---|---|---|---|---|
| s=1 | c=3.5 | {2} | {1,2} | {1,2} |
| | c=4 | {2,…,6} | {1,…,6} | {1,…,6} |
| | c=4.5 | {2,…,9} | {1,…,9} | {1,…,9} |
| s=3 | c=3.5 | - | {2} | {1,2} |
| | c=4 | {5,6} | {2,…,6} | {1,…,6} |
| | c=4.5 | {5,…,9} | {2,…,9} | {1,…,9} |
| s=5 | c=3.5 | - | - | {2} |
| | c=4 | - | {4,…,6} | {2,…,6} |
| | c=4.5 | - | {4,…,9} | {2,…,9} |
| s=7 | c=3.5 | - | - | {2} |
| | c=4 | - | {6} | {2,…,6} |
| | c=4.5 | - | {6,…,9} | {2,…,9} |
| s=10 | c=3.5 | - | - | - |
| | c=4 | - | - | {3,…,6} |
| | c=4.5 | - | - | {3,…,9} |

(-) denotes there is not any value.

## 3. Simulation study

In this section, a simulation study is carried out in order to assess the performances of the results obtained in Section 2. To do this, the following algorithm has been applied.

**Algorithm 1.** Take θ = 1 and suppose values of s, m, n, c, p, $c_0$, $c_t$ and $c_u$ are all given. Then:

1- Choose $r_{opt}$ which satisfies both conditions (10) and (11).

2- Generate order statistics $\tilde{x} = (x_{1:n}, \cdots, x_{r_{opt}:n})$ and $Y_{s:m}$ from $Ray(1)$.

3- Obtained the point predictor for $Y_{s:m}$ by using (6).

4- Repeat the Steps 1–3 for K = 10000 times and let $\hat{Y}_{s:m}(i)$ obtained from Step 3 in the i-th iteration, i = 1, ..., K. Also, let $Y_{s:m}(i)$ be the s-th order statistic of a sample of size m and $\tilde{x}(i) = (x_{1:n}(i), \cdots, x_{r_{opt}:n}(i))$ be the observed Type II censored order statistics generated in Step 2 in the i-th iteration.

5- Then, calculate the mean point predictors (MPPs), the estimated MSPEs (EMSPEs) and the estimated expected cost functions (EECFs) by using the relations

$$\tilde{Y}_{s:m} = \frac{1}{K}\sum_{i=1}^{K} \hat{Y}_{s:m}(i),$$

$$EMSPE(\hat{Y}_{s:m}) = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{Y}_{s:m}(i) - Y_{s:m}(i)\right)^2,$$

and,

$$EECF(\hat{Y}_{s:m}) = c_0 + \frac{c_t}{K}\sum_{j=1}^{K}\sum_{i=1}^{r_{opt}} x_{i:n}(j) + nc_u$$

respectively.

**Table 2.** Values of MPPs, EMSPEs and EECFs for different values of $r_{opt}$ and different values of s, c and p when $m = n = 10,\ m = n = 10, c_0 = 1, c_t = 1, c_u = 0.2$ and $\theta = 1$.

| $r_{opt}$ | p | c | s | MPP | EMSPE | EECF |
|---|---|---|---|---|---|---|
| 2 | 0.05, 0.1, 0.5 | 3.5, 4, 4.5 | 1 | 0.286 | 0.029 | 3.383 |
| | 0.1, 0.5 | 3.5, 4, 4.5 | 3 | 0.477 | 0.014 | 3.329 |
| | 0.5 | 3.5, 4, 4.5 | 5 | 0.960 | 0.120 | 3.498 |
| | 0.5 | 3.5, 4, 4.5 | 7 | 1.331 | 0.158 | 3.410 |
| 6 | 0.05, 0.1, 0.5 | 4, 4.5 | 1 | 0.254 | 0.017 | 3.488 |
| | 0.05, 0.1, 0.5 | 4, 4.5 | 3 | 0.713 | 0.007 | 3.777 |
| | 0.1, 0.5 | 4, 4.5 | 5 | 1.170 | 0.004 | 3.923 |
| | 0.1, 0.5 | 4, 4.5 | 7 | 0.760 | 0.011 | 3.409 |
| | 0.5 | 4, 4.5 | 10 | 1.563 | 0.081 | 3.493 |
| 9 | 0.05, 0.1, 0.5 | 4.5 | 1 | 0.361 | 0.046 | 3.997 |
| | 0.05, 0.1, 0.5 | 4.5 | 3 | 0.706 | 0.020 | 4.057 |
| | 0.1, 0.5 | 4.5 | 5 | 0.634 | 0.074 | 3.648 |
| | 0.1, 0.5 | 4.5 | 7 | 1.062 | 0.001 | 3.744 |
| | 0.5 | 4.5 | 10 | 2.140 | 0.499 | 3.968 |

Based on Algorithm 1 and the results in Table 1, we have computed the values of MPPs, EMSPEs and EECFs for different values of s when $c_0 = 1, c_t = 1, c_u = 0.2$ and $m = n = 10$. The results are tabulated in Table 2. From Table 2 we can see that in all cases both conditions $EECF < c$ and $EMSPE < p$ are satisfied.

## 4. Acknowledgment

The author would like to thank the referees for their valuable suggestions that improved on the original version of the manuscript.

## References

[1] David, H. A. & Nagaraja, H. N. (2003). *Order Statistics*. Third edition, John Wiley & Sons, Hoboken, New Jersey, (2003).

[2] Arnold, B. C, Balakrishnan, N. & Nagaraja, H. N. (2008). *A first course in order statistics.* Philadelphia: SIAM.

[3] Siddiqui, M. M. (1962). Some problems connected with Rayleigh distributions. *Journal of Research of the National Bureau of Standards, 60,* 167-174.

[4] Miller, K. S. & Sackrowitz, H. (1967). Relationships between biased and unbiased Rayleigh distributions. SIAM, *journal on applied Mathematics, 15,* 1490-1495.

[5] Kotb, M. S. &. Raqab, M. Z. (2017). Bayesian inference and prediction of the Rayleigh distribution based on ordered ranked set sampling. *Communications in Statistics-Simulation and Computation*, (just-accepted).

[6] Asgharzadeh, A. Fernández, A. J. & Abdi, M. (2017). Confidence sets for the two-parameter Rayleigh distribution under progressive censoring. *Applied Mathematical Modelling, 47,* 656-667.

[7] Seo, J. I. Jeon, J. W. & Kang, S. B. (2016). Exact Interval Inference for the Two-Parameter Rayleigh Distribution Based on the Upper Record Values. *Journal of Probability and Statistics*, 2016.

[8] Ng, H. K. T. Chan, P. S. & Balakrishnan, N. (2004). Optimal progressive censoring plans for the Weibull distribution. *Technometrics, 46,* 470-481.

[9] Doostparast, M. & Balakrishnan, N. (2010). Optimal sample size for record data and associated cost analysis for exponential distribution. *Journal of Statistical Computation and Simulation, 80(12*), 1389-1401.

[10] Ahmadi, J. Basiri, E. & MirMostafaee, S.M.T.K. (2016). Optimal random sample size based on Bayesian prediction problem of exponential lifetime and application to real data. *Journal of the Korean Statistical Society, 45(2),* 221-237.

[11] Cordeiro, J. B. &Pham, H. (2017). Optimal design of life testing cost model for Type-II censoring Weibull distribution lifetime units with respect to unknown parameters. *International Journal of System Assurance Engineering and Management, 8(1),* 28-32.

[12] Huang, S. R. & Wu, S. J. (2017). Optimal sample size allocation for accelerated life test under progressive type-II censoring with competing risks. *Journal of Statistical Computation and Simulation, 87(1),* 1-16.