Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

Confidence interval, prediction interval and tolerance limits for a two-parameter Rayleigh distribution

K. Krishnamoorthy, Dustin Waguespack & Ngan Hoang-Nguyen-Thuy

To cite this article: K. Krishnamoorthy, Dustin Waguespack & Ngan Hoang-Nguyen-Thuy (2020) Confidence interval, prediction interval and tolerance limits for a two-parameter Rayleigh distribution, Journal of Applied Statistics, 47:1, 160-175, DOI: 10.1080/02664763.2019.1634681

To link to this article: https://doi.org/10.1080/02664763.2019.1634681

Published online: 25 Jun 2019.

JOURNAL OF APPLIED STATISTICS
2020, VOL. 47, NO. 1, 160–175
https://doi.org/10.1080/02664763.2019.1634681

Taylor & Francis
Taylor & Francis Group

Confidence interval, prediction interval and tolerance limits for a two-parameter Rayleigh distribution

K. Krishnamoorthy, Dustin Waguespack and Ngan Hoang-Nguyen-Thuy

Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA, USA

ABSTRACT

The problems of interval estimating the parameters and the mean of a two-parameter Rayleigh distribution are considered. We propose pivotal-based methods for constructing confidence intervals for the mean, quantiles, survival probability and for constructing prediction intervals for the mean of a future sample. Pivotal quantities based on the maximum likelihood estimates (MLEs), moment estimates (MEs) and the L-moments estimates (L-MEs) are proposed. Interval estimates based on them are compared via Monte Carlo simulation. Comparison studies indicate that the results based on the MEs and the L-MEs are very similar. The results based on the MLEs are slightly better than those based on the MEs and the L-MEs for small to moderate sample sizes. The methods are illustrated using an example involving lifetime data.

ARTICLE HISTORY
Received 24 October 2018
Accepted 16 June 2019

KEYWORDS
Confidence interval; moment estimates; prediction interval; precision; quantile estimation; tolerance interval

1. Introduction

The Rayleigh distribution with a single scale parameter has received much attention in the literature and it has been used in physics and engineering. Siddiqui [13] has noted that the one-parameter Rayleigh distribution arises as an asymptotic distribution in some two-dimensional random walk problems. Inference on the scale parameter is routinely obtained on the basis of square transformed data which can be regarded as a sample from an exponential distribution with a rate parameter [14]. For a Rayleigh distribution with a single scale parameter, Dey and Dey [2,3], Prakash [10] and Kotb and Raqab [7] have proposed inferential methods under different sampling designs such as ranked set sampling and progressively type II censored samples. Recently, there has been interest among researchers in developing inferential methods for Rayleigh distributions with an additional location or threshold parameter. Dey, Dey and Kundu [4] have noted the applicability of a two-parameter Rayleigh distribution to model failure time data. In particular, they noted that the hazard function is an increasing function of time, as a result, the distribution has attracted several researchers as such hazard functions commonly occur in failure time data analysis.

To describe some available estimates, we note that the probability density function (pdf) of a two-parameter Rayleigh distribution with location parameter a and the scale

CONTACT K. Krishnamoorthy krishna@louisiana.edu

parameter b is given by

$$f(x \mid a, b) = \frac{(x-a)}{b^2} \mathrm{e}^{-(1/2)((x-a)/b)^2}, \quad x > a, \quad b > 0. \tag{1}$$

The cumulative distribution function (cdf) is given by

$$F(x \mid a, b) = 1 - \mathrm{e}^{-(1/2)((x-a)/b)^2}, \quad x > a, \quad b > 0. \tag{2}$$

We shall denote the above distribution by Rayleigh (a, b).

Regarding inference on two-parameter Rayleigh distributions, Dey, Dey and Kundu [4] have derived the maximum likelihood estimates, moment estimates, and L-moment estimates, and evaluated their merits via Monte Carlo simulation. These authors have also proposed credible sets for a function of the location and scale parameters using a gamma prior and a non-proper uniform prior. It should be noted that the credible set is not in closed-form and it has to be obtained using the importance sampling method. Dey, Dey and Kundu [5] have extended these results when the data are progressively type II censored. Seo *et al.* [12] have developed exact confidence intervals (CIs) for the parameters a and b based on the upper record values. The exact CIs were obtained on the basis of some pivotal quantities (involving upper record values) whose distributions are known.

In failure time data analysis, it is of importance to find interval estimates for the mean, quantile, and survival probability, and prediction intervals (PIs) for the mean of a future sample. In this regard, we investigate the interval estimates based on the pivotal quantities involving equivariant estimators. The moment estimates (MEs), L-moments estimates (L-MEs) and the MLEs are all equivariant. Among these three equivariant estimates, the MLEs can not be expressed in closed-form and they can be obtained only by numerically while the MEs are some simple functions of the sample mean and variance, and the L-MEs are also some simple functions of order statistics. As the MEs and the L-MEs are easy to obtain, it is of interest to compare the results based on the pivotal quantities involving different equivariant estimates for various aforementioned problems. Furthermore, we noted that the derivation of the MLEs that is available in the literature is in error. Note that the pdf is defined only for $a < x$, and so the MLE of a should maximize the profile likelihood function subject to the constraint that a is less than the smallest order statistic. However, the proposed method in Dey, Dey and Kundu [4] could produce the MLE of the threshold parameter a that is greater than the smallest order statistic and/or is not the maxima of the profile likelihood function. Our derivation in this paper produce the MLE of a that is always less than the smallest order statistics and maximizes the profile likelihood function.

The rest of the article is organized as follows. In the following section, we provide the calculation of the MEs, L-MEs and the MLEs, and present pivotal quantities based on them. In Section 3, we consider interval estimation of the mean by proposing CIs based on the different pivotal quantities and compare them in terms of precision. As the distributions of the pivotal quantities are not in closed-form, the CIs can be obtained only by Monte Carlo simulation. In Section 4, we address the problems of constructing one-sided tolerance limits and constructing confidence limits for a survival probability. The problem of finding a PI for the mean of a future sample is considered in Section 5. For all the problems, we compare the results on the basis of different equivariant estimators. Our comparison studies indicate that the results based on the MEs and the L-MEs are very similar, and interval estimates based on the MLEs are slightly better than those based on the MEs. In Section 6, we briefly outline pivotal-based approach for type II censored case, and illustrate the method for finding confidence intervals for the mean. All the interval estimation methods are illustrated using a data set of lifetimes of drills in Section 7. Some concluding remarks are given in Section 8.

2. Point estimates and pivotal quantities

Let $X_1, \ldots, X_n$ be a sample from a two-parameter Rayleigh distribution. Let $\bar{X}$ denote the sample mean and define the sample variance as $S^2 = (1/(n-1)) \sum_{i=1}^n (X_i - \bar{X})^2$.

2.1. Moment estimates

To obtain moment estimates (MEs) for a and b, we note that

$$\mu'_k = E(X^k) \quad \text{and} \quad \mu_{k,a} = E(X-a)^k = 2^{k/2} b^k \Gamma(k/2+1), \quad k = 1, 2, ..$$

Using the above results, it is easy to check that

$$E(X) = a + \sqrt{\frac{\pi}{2}} b \quad \text{and} \quad \mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{4-\pi}{2} b^2. \tag{3}$$

Replacing $E(X)$ by $\bar{X}$ and $\mathrm{Var}(X)$ by the sample variance S^2 in the above equation and then solving the equations for a and b, Dey *et al.* [4] have obtained moment estimates as

$$\hat{a} = \bar{X} - \sqrt{\frac{\pi}{4-\pi}} S \quad \text{and} \quad \hat{b} = \sqrt{\frac{2}{4-\pi}} S. \tag{4}$$

Dey *et al.* have also noted that the MEs are consistent to the corresponding estimators, and they are asymptotically bivariate normally distributed.

2.2. L-Moments estimates

Dey *et al.* [4] have also derived L-moments estimates using the method of Hosking [6]. To write these estimates, let $X_{i:n}$ denotes the ith order statistic for a sample of size n. Let

$$l_1 = \frac{1}{n} \sum_{i=1}^n X_{i:n} \quad \text{and} \quad l_2 = \frac{2}{n(n-1)} \sum_{i=1}^n (i-1) X_{i:n} - l_1.$$

In terms of l_1 and l_2 the L-moments estimates can be written as

$$\hat{a}_l = l_1 - \frac{\sqrt{2}}{\sqrt{2}-1} l_2 \quad \text{and} \quad \hat{b}_l = \frac{l_2}{\Gamma(3/2)(\sqrt{2}-1)}.$$

In the sequel, we shall denote the L-moments estimates by L-MEs.

2.3. Maximum likelihood estimates

For a given data $X_1, \ldots, X_n$, the likelihood function can be expressed as

$$L(a, b \mid X) = \frac{1}{b^{2n}} \prod_{i=1}^{n} (X_i - a) \exp\left[-\frac{1}{2} \left(\frac{X_i - a}{b} \right)^2 \right] I[X_{(1)} > a], \tag{5}$$

where $I[x]$ is the indicator function and $X_{(1)} = \min\{X_1, \ldots, X_n\}$. Letting $\theta = b^2$, the log-likelihood function, without the indicator function, can be expressed as

$$l(a, \theta \mid X) = -n \ln \theta + \sum_{i=1}^{n} \ln(X_i - a) - \frac{1}{2\theta} \sum_{i=1}^{n} (X_i - a)^2.$$

The partial derivative $\partial l(a, \theta \mid X)/\partial \theta = 0$ yields $\theta = \sum_{i=1}^{n} (X_i - a)^2/(2n)$. After replacing θ in $l(a, \theta)$ with this expression, we obtain the profile log-likelihood function, after omitting the constant term, as

$$l(a \mid X) = \sum_{i=1}^{n} \ln \left[\frac{(X_i - a)}{\sum_{j=1}^{n} (X_j - a)^2} \right]. \tag{6}$$

Using $\partial l(a \mid X)/\partial a = 0$, we see that the MLE of a is the root of the equation

$$h(a \mid X) = \frac{2n^2(\bar{X} - a)}{\sum_{i=1}^{n} (X_i - a)^2} - \sum_{i=1}^{n} (X_i - a)^{-1} = 0. \tag{7}$$

To solve the above equation numerically, Dey, Dey and Kundu [4] have proposed a fixed point scheme. Our investigation, however, indicates that the fixed point scheme does not converge. Alternatively, we find the MLE of a by maximizing the profile log-likelihood function $l(a)$ in (6) over $a < X_{(1)}$. As shown in the Appendix 1,

$$P\left(X_{(1)} - 12\hat{b}/\sqrt{n} \le a \le X_{(1)} \right) \approx 1, \quad \text{for } n \ge 3, \tag{8}$$

where $\hat{b}$ is the moment estimate of b given in (4). So the MLE of a lies in the interval $(X_{(1)} - 12\hat{b}/\sqrt{n},\ X_{(1)})$, and by optimizing $l(a \mid X)$ in (6) over $a \in (X_{(1)} - 12\hat{b}/\sqrt{n},\ X_{(1)})$, the MLE of a can be obtained (e.g. R function optimize()). Equivalently, the MLE of a can also be obtained as the root of the equation $h(a \mid X) = 0$ using the interval $(X_{(1)} - 12\hat{b}/\sqrt{n},\ X_{(1)})$ as a root bracketing interval (e.g. R function uniroot()). Let $\tilde{a}$ denote the MLE of a thus obtained. Then the MLE of θ is given by

$$\tilde{\theta} = \tilde{b}^2 = \frac{1}{2n} \sum_{i=1}^{n} (X_i - \tilde{a})^2.$$

2.4. Pivotal quantities

Let $\hat{a}_e$ and $\hat{b}_e$ be equivariant estimators of a and b, respectively. Then $(\hat{a}_e - a)/b$ and $\hat{b}_e/b$ are pivotal quantities (see [9], Theorem E2). As a consequence,

$$\frac{\hat{a}_e - a}{b} \sim \hat{a}_e^* \quad \text{and} \quad \frac{\hat{b}_e}{b} \sim \hat{b}_e^*, \tag{9}$$

where the notation '$\sim$' means 'distributed as' and $\hat{a}_e^*$ and $\hat{b}_e^*$ are the equivariant estimates based on a sample of size n from the Rayleigh distribution with $a = 0$ and $b = 1$.

It can be easily verified that the MEs, L-MEs and the MLEs are the location-scale equivariant, and the pivotal quantities based on them can be used to find confidence intervals for the parameters, mean and population quantiles. Let us refer to the pivotal quantities based on the MEs, L-MEs and the MLEs as the 'ME-pivot,' 'L-ME-pivot,' and 'MLE-pivot,' respectively.

3. Confidence interval for the mean

The mean of the Rayleigh(a, b) distribution (see (3)) is given by

$$a + cb, \quad \text{where } c = \sqrt{\frac{\pi}{2}}. \tag{10}$$

The quantity

$$u_n = \frac{a + cb - \hat{a}_e}{\hat{b}_e} = \frac{a - \hat{a}_e}{\hat{b}_e} + c\frac{b}{\hat{b}_e} \sim \frac{c - \hat{a}_e^*}{\hat{b}_e^*},$$

where $\hat{a}_e^*$ and $\hat{b}_e^*$ are as defined in (9), is a pivotal quantity. If $u_{n;\alpha}$ denotes the 100α percentile of u_n, then

$$\left(\hat{a}_e + u_{n;\alpha}\hat{b}_e,\ \hat{a}_e + u_{n;1-\alpha}\hat{b}_e\right) \tag{11}$$

is a $100(1 - 2\alpha)\%$ CI for the mean.

As all the proposed estimates are equivariant, the CIs for the mean based on the MLE-pivot, L-ME-pivot, and the ME-pivot are exact in the sense that the coverage probabilities are equal to the nominal level for all parameter values. However, the precisions of these three CIs could be different. We estimated the expected widths of the CIs based on all three pivots and reported them in Table 1. As the expected width of an equivariant CI does not depend on the location parameter a, without loss of generality, we can assume a to be zero when comparing CIs with respect to expected width. Examination of the expected widths clearly indicates that the ME-CIs and the L-ME CIs have practically the same expected width for all the cases considered in Table 1. Furthermore, the expected widths of MLE CIs are slightly shorter than the other two CIs.

Even though ME CIs and the L-ME CIs are simple to compute, the expected widths of the MLE CIs are very less than or equal to those of the other two CIs, and so we provide percentiles to compute 90, 95 and 99 percent MLE CIs for the mean of a Rayleigh distribution in Table 2. Details of calculation along with R code are given in Appendix 2.

Table 1. Expected widths of 95% CIs based on the ME-pivot, L-ME pivot and the MLE-pivot for the mean.

	$n = 5$			$n = 10$			$n = 20$			$n = 30$		
b	ME	L-ME	MLE	ME	L-ME	MLE	ME	L-ME	MLE	ME	L-ME	MLE
1	1.62	1.62	1.60	0.94	0.94	0.93	0.61	0.62	0.60	0.48	0.48	0.48
2	3.25	3.25	3.20	1.90	1.90	1.85	1.24	1.24	1.20	0.98	0.98	0.96
3	4.85	4.85	4.82	2.81	2.80	2.75	1.88	1.87	1.81	1.47	1.47	1.44
4	6.50	6.49	6.36	3.78	3.78	3.69	2.49	2.46	2.40	1.97	1.97	1.92
5	8.05	8.05	7.93	4.66	4.70	4.60	3.07	3.08	2.99	2.44	2.44	2.39

Table 2. Percentiles for computing 90%, 95% and 99% CIs for the mean based on the MLE-pivot.

n	5%	95%	2.5%	97.5%	.5%	99.5%
4	.472	2.43	.219	2.90	−.642	4.42
5	.635	2.14	.452	2.44	−.046	3.25
6	.728	1.98	.595	2.22	.246	2.86
7	.790	1.88	.679	2.07	.398	2.55
8	.832	1.81	.735	1.97	.506	2.37
9	.863	1.76	.778	1.90	.579	2.24
10	.892	1.72	.817	1.84	.648	2.13
11	.913	1.69	.840	1.80	.678	2.06
12	.931	1.65	.865	1.76	.720	2.00
13	.945	1.63	.882	1.73	.746	1.95
14	.962	1.61	.905	1.70	.784	1.91
15	.971	1.60	.916	1.68	.801	1.87
16	.982	1.58	.930	1.66	.818	1.83
17	.992	1.57	.941	1.65	.831	1.82
18	.999	1.55	.950	1.63	.849	1.79
19	1.01	1.55	.960	1.62	.865	1.77
20	1.02	1.53	.972	1.60	.878	1.75
25	1.04	1.50	1.00	1.55	.919	1.67
30	1.06	1.48	1.02	1.52	.950	1.63
35	1.07	1.45	1.04	1.49	.977	1.59
40	1.09	1.43	1.06	1.48	.998	1.56
45	1.10	1.43	1.07	1.47	1.01	1.55
50	1.11	1.42	1.08	1.45	1.03	1.52
60	1.12	1.40	1.09	1.43	1.05	1.50
70	1.13	1.39	1.11	1.42	1.06	1.48
80	1.14	1.38	1.12	1.41	1.08	1.46
90	1.14	1.37	1.12	1.40	1.09	1.45
100	1.15	1.36	1.13	1.39	1.09	1.43

4. One-sided tolerance limits and survival probability

One-sided tolerance limits (TLs) are the one-sided confidence intervals for a population quantile. For example, a p content – γ coverage one-sided upper TL is the $100\gamma\%$ upper confidence limit for the pth quantile of the population. From the cdf (2), it can be easily checked that the pth quantile of a two-parameter Rayleigh distribution is given by

$$q_p(a, b) = a + b\sqrt{-2\ln(1-p)}.$$

Since the above quantile is the same form as the mean in (10) with $c = q_p(0, 1)$, $(q_p(a, b) - \hat{a}_e)/\hat{b}_e \sim (q_p(0, 1) - \hat{a}_e^*)/\hat{b}_e^*$, where $\hat{a}^*$ and $\hat{b}^*$ are equivariant estimators based on a sample of size n from a Rayleigh$(0, 1)$ distribution. If $k_{p,\gamma}$ denotes the 100γ percentile of $(q_p(0, 1) - \hat{a}^*)/\hat{b}^*$ and $p > .5$, then

$$\hat{a}_e + k_{p,\gamma}\hat{b}_e \tag{12}$$

Table 3. Average values of (.90, .95) one-sided tolerance limits based on the MLE-pivot, ME-pivot and L-ME pivot.

	$a=3$											
	$n=10$						$n=20$					
	MLE		ME		L-ME		MLE		ME		L-ME	
b	LTL	UTL	LTL	UTL	LTL	UTL	LTL	UTL	LTL	UTL	LTL	UTL
1	3.02	5.98	2.97	6.02	2.98	6.01	3.20	5.66	3.15	5.68	3.17	5.67
1.5	3.02	7.46	2.95	7.51	2.96	7.51	3.31	6.98	3.24	7.02	3.25	7.00
2	3.03	8.95	2.91	9.04	2.95	9.03	3.41	8.32	3.32	8.34	3.33	8.36
2.5	3.04	10.5	2.95	10.5	2.94	10.5	3.51	9.64	3.39	9.69	3.41	9.68
3	3.04	11.9	2.92	12.0	2.92	12.0	3.61	10.9	3.48	11.0	3.48	11.0

is a (p, γ) one-sided upper TL for the Rayleigh distribution. The R code for computing CI for the mean (see Appendix 2), with a slight modification, can be used to estimate the percentiles of $(q_p(0,1) - \hat{a}^*)/\hat{b}^*$. Similarly, a (p, γ) lower TL can be expressed as

$$\hat{a}_e + k_{1-p,1-\gamma}\hat{b}_e, \tag{13}$$

where $k_{1-p,1-\gamma}$ is the $100(1-\gamma)$ percentile of $(q_{1-p}(0,1) - \hat{a}_e^*)/\hat{b}_e^*$ and $q_{1-p}(0,1) = \sqrt{-2\ln p}$.

To choose among the MEs, L-MEs and MLEs to compute the TLs, we estimated the average values of the (.90, .95) one-sided TLs obtained using these three estimates. These estimated expected values are reported in Table 3 for different sample sizes. We observe from the table that the lower TLs based on the MLEs are larger (larger is better) than those based on the MEs and L-MEs, and the upper TLs based on the MLEs are smaller (smaller is better) than those based on the MEs and L-MEs. The expected values in the table also indicate that TLs based on the MEs and L-MEs are practically the same. Thus, TLs based on the MLEs are preferable to those based on the MEs or L-MEs.

Since the TLs based on the MLEs are better than those based on the MEs and L-MEs, we provide factors for computing $(p, .95)$ one-sided TLs on the basis of MLEs in Table 4. The factors are given for sample sizes ranging from 5 to 100 and $p = .80, .90, .95$ and .99. As an example, the factor for computing one-sided (.90,.95) lower TL based on a sample of size 15 is .042 and the factor for computing one-sided upper TL is 2.87.

4.1. Survival probability

For a given $t > a$, the survival probability is given by

$$\tau = P(X > t \mid a, b) = \exp\left(-(t-a)^2/(2b^2)\right).$$

In applications, a lower confidence limit for τ is used to assess the reliability, and it can be found using a lower tolerance limit as follows. The lower tolerance limit in (13) is determined so that

$$P_{\hat{a}_e,\hat{b}_e}\left\{P_X\left(X \geq \hat{a}_e + k_{1-p,1-\gamma}\hat{b}_e \mid \hat{a}_e, \hat{b}_e\right) \geq p\right\} = \gamma. \tag{14}$$

Table 4. Factors for computing $(p, .95)$ one-sided tolerance limits based on the MLE-pivot.

	lower tolerance factor				upper tolerance factor			
	p				p			
n	.80	.90	.95	.99	.80	.90	.95	.99
5	−.425	−.896	−1.22	−1.67	3.43	4.35	5.18	6.73
6	−.197	−.599	−.878	−1.26	3.11	3.89	4.57	5.93
7	−.059	−.416	−.665	−1.01	2.90	3.60	4.23	5.43
8	.031	−.299	−.531	−.849	2.76	3.41	3.98	5.10
9	.115	−.209	−.430	−.729	2.66	3.27	3.82	4.86
10	.167	−.144	−.355	−.635	2.58	3.16	3.69	4.70
11	.208	−.091	−.293	−.565	2.52	3.09	3.58	4.54
12	.239	−.047	−.247	−.508	2.47	3.02	3.49	4.44
13	.266	−.012	−.207	−.459	2.42	2.96	3.43	4.33
14	.289	.018	−.168	−.423	2.39	2.92	3.37	4.26
15	.312	.042	−.134	−.386	2.35	2.87	3.31	4.19
16	.330	.067	−.115	−.355	2.33	2.84	3.28	4.14
17	.348	.089	−.095	−.331	2.31	2.81	3.23	4.09
18	.361	.100	−.069	−.306	2.29	2.79	3.21	4.03
19	.373	.119	−.057	−.286	2.27	2.75	3.17	3.99
20	.385	.133	−.044	−.269	2.25	2.73	3.14	3.96
21	.395	.145	−.025	−.252	2.23	2.71	3.12	3.93
22	.404	.158	−.014	−.240	2.22	2.69	3.10	3.91
23	.413	.167	−.001	−.227	2.21	2.67	3.07	3.87
24	.419	.177	.008	−.214	2.20	2.66	3.06	3.84
25	.428	.185	.016	−.203	2.19	2.64	3.04	3.82
30	.458	.220	.056	−.154	2.14	2.59	2.96	3.72
35	.478	.246	.085	−.126	2.11	2.54	2.92	3.66
40	.496	.265	.109	−.097	2.08	2.51	2.88	3.60
45	.506	.280	.124	−.079	2.06	2.49	2.84	3.55
50	.520	.293	.139	−.063	2.05	2.46	2.82	3.52
60	.535	.313	.160	−.038	2.02	2.43	2.78	3.46
70	.548	.327	.177	−.019	2.00	2.40	2.75	3.43
80	.557	.338	.190	−.005	1.98	2.38	2.73	3.40
90	.565	.347	.200	.007	1.97	2.37	2.71	3.37
100	.571	.355	.208	.015	1.96	2.35	2.69	3.35

Suppose p is determined so that $\hat{a} + k_{1-p,1-\gamma}\hat{b} = t$. Then (14) implies that

$$P_X(X > t) \geq p \quad \text{with probability } \gamma,$$

and so p is the $100\gamma\%$ lower confidence limit for $\tau = P(X > t)$. To find the value of p such that $\hat{a}_e + k_{1-p,1-\gamma}\hat{b}_e = t$, recall that $k_{1-p,1-\gamma}$ is the $100(1-\gamma)$ percentile of $(\sqrt{-2\ln p} - \hat{a}_e^*)/\hat{b}_e^*$, so p is to be determined so that

$$100(1-\gamma) \text{ percentile of } \frac{\sqrt{-2\ln p} - \hat{a}_e^*}{\hat{b}_e^*} = \frac{t - \hat{a}_e}{\hat{b}_e}. \tag{15}$$

The value of p that satisfies (15) can be estimated using Monte Carlo simulation and a bisection method as shown in the following algorithm.

Algorithm 4.1:

(1) *For a given sample of size n, compute the MLEs $\tilde{a}$ and $\tilde{b}$ and compute the estimate of $\tau = P(X > t)$ as $p_0 = \exp(-.5(t - \tilde{a})^2/\tilde{b}^2)$ and the value of $t_0 = (t - \tilde{a})/\tilde{b}$.*

(2) *Generate, say,* 100, 000 *samples each of size n from the Rayleigh distribution with* $a = 0$ *and* $b = 1$.
(3) *Calculate the MLEs* $\tilde{a}_i^*$ *and* $\tilde{b}_i^*$ *based on the ith sample generated in the preceding step,* $i = 1, \ldots, 100, 000$
(4) *Denote the 100*$(1-\gamma)$ *percentile of* $(\sqrt{-2\ln p} - \tilde{a}^*)/\tilde{b}^*$ *by* Q_p *and set* $f(p) = Q_p - t_0$. *Note that for a given p,* Q_p *can be estimated using the simulated estimates* $\tilde{a}^*$ *and* $\tilde{b}^*$ *in Step* 3.
(5) *Using the* p_0 *in Step* 1 *and* $p_1 = .001$, *say, as the root bracketing values, the solution to the equation* $f(p) = 0$ *can be found using a bisection method. The root of the equation is a 100*γ *lower confidence limit for* $\tau = P(X > t)$.

Note that to compute $f(p)$ defined in the above algorithm at various values of p, we need to carry out the calculation in Step 4 only. The bisection scheme converges in a fewer steps with the aforementioned bracketing values in Step 5 of Algorithm 4.1.

5. Prediction intervals for the mean of a future sample

To find a prediction interval (PI) for the mean of a future sample of size m from a Rayleigh(a, b) distribution, let $\hat{a}_e$ and $\hat{b}_e$ denote the equivariant estimators based on a background sample of size n from a Rayleigh(a, b) distribution and let $\bar{Y}$ denote the mean of a future sample from the same Rayleigh(a, b) distribution. To find a PI for $\bar{Y}$, we note that

$$\frac{\bar{Y} - \hat{a}_e}{\hat{b}_e} = \frac{(\bar{Y} - a)/b - (\hat{a}_e - a)/b}{\hat{b}_e/b} \sim \frac{\bar{Y}^* - \hat{a}_e^*}{\hat{b}_e^*},$$

where, as defined earlier, $\hat{a}_e^*$ and $\hat{b}_e^*$ are equivariant estimators based on a sample of size n from the Rayleigh(0,1) distribution and $\bar{Y}^*$ is the mean of a sample of size m from the Rayleigh(0,1) distribution. If P_α denotes 100α percentile of $(\bar{Y}^* - \hat{a}_e^*)/\hat{b}_e^*$, then

$$\left(\hat{a}_e + P_\alpha \hat{b}_e,\ \hat{a}_e + P_{1-\alpha} \hat{b}_e\right)$$

is a $1 - 2\alpha$ PI for a future sample mean $\bar{Y}$. Note that the percentiles of $(\bar{Y}^* - \hat{a}_e^*)/\hat{b}_e^*$ can be estimated by Monte Carlo simulation.

To compare the PIs based on the MEs, L-MEs and MLEs, we estimated expected widths of all three PIs for some values of n and future sample size m, and reported them in Table 5. We see from this table that all three PIs have some desired properties; their expected widths decreasing with increasing sample sizes n and/or m. We once again see that the PIs based on the MEs and L-MEs are very similar. Furthermore, the expected widths of the PIs based on the MLEs are slightly smaller than those based on the MEs and L-MEs for all the cases considered. Thus, there is a slight improvement in the precision of the MLE prediction intervals over the other two prediction intervals.

As the PI based on the MLEs is better than other two PIs, we estimated percentiles (using the MLE-pivot) required to compute the 95% PIs for the mean of a future sample of size m based on a background sample of size n for a few values of n and m and reported them in Table 6. As an example, when $n = 15$ and $m = 10$, the 95% PI is given by $(\tilde{a} + 0.723\,\tilde{b},\ \tilde{a} + 1.93\,\tilde{b})$, where $\tilde{a}$ and $\tilde{b}$ are the MLEs based on a background sample of size 15.

Table 5. Expected widths of 95% PIs for the mean of a future sample of size m based on background sample of size n.

	$a=0$											
	$n=10,\ m=5$			$n=10,\ m=10$			$n=15,\ m=7$			$n=15,\ m=10$		
b	MLE	ME	L-ME	MLE	ME	L-ME	MLE	ME	L-ME	MLE	ME	L-ME
1	1.57	1.60	1.58	1.28	1.31	1.32	1.25	1.27	1.27	1.11	1.13	1.14
1.5	2.36	2.40	2.39	1.92	1.96	1.95	1.88	1.91	1.92	1.67	1.70	1.71
2	3.14	3.20	3.20	2.58	2.63	2.62	2.52	2.56	2.55	2.24	2.28	2.28
2.5	3.92	3.99	4.00	3.21	3.27	3.27	3.13	3.18	3.18	2.78	2.83	2.83
3	4.69	4.77	4.77	3.83	3.90	3.91	3.75	3.81	3.82	3.34	3.40	3.41

Table 6. Lower and upper percentiles for computing 95% PIs for the mean of a future sample of size m based on the background sample of size n using the MLEs.

$n=10$			$n=15$			$n=20$			$n=25$		
m	L	U	m	L	U	m	L	U	m	L	U
1	−.050	3.21	1	.054	3.01	1	.100	2.93	1	.130	2.89
3	.358	2.44	5	.583	2.11	3	.487	2.23	5	.642	2.00
5	.495	2.24	8	.683	1.97	6	.664	1.99	10	.781	1.82
7	.573	2.14	10	.723	1.93	8	.720	1.91	15	.837	1.75
9	.611	2.09	15	.779	1.85	15	.820	1.79	20	.874	1.71
11	.647	2.04	17	.790	1.84	20	.848	1.75	30	.909	1.66
15	.692	2.01	20	.802	1.82	30	.885	1.71	40	.931	1.64

6. Inference based on a type II censored sample

If the samples are type II censored, then the pivotal quantities based on equivariant estimators are still valid, and so the pivotal-based inference can be obtained in a straightforward manner. The properties of pivot-based procedures for type II censored samples should be similar to those of the complete sample case. In type I censoring, failures are recorded until a mission time, and so the number of failed items r in a random sample is a random variable. Because the distribution of r depends on the parameters, the pivotal quantities given earlier are no longer valid for the type I censored samples. However, we can use the pivotal quantities for type II censored samples as approximations for the type I censored samples [11].

In the following, we shall provide a method of obtaining the MLEs based on a type II censored sample. Let $X_{(1)} < \cdots < X_{(r)}$ be a set of failure times recorded from a sample of n test items. After omitting the indicator function $I[X_{(1)} > a]$, the log-likelihood function can be expressed as

$$\ln L = -r \ln \theta + \sum_{i=1}^{r} \ln(X_{(i)} - a) - \frac{1}{2\theta} \sum_{i=1}^{n} (X_i^* - a)^2,$$

where $\theta = b^2$, $X_i^* = X_{(i)}$ for $i = 1, \ldots, r$ and $X_i^* = X_{(r)}$ for $i = r+1, \ldots, n$.

The equation $\partial \ln L/\partial\theta = 0$ yields $\theta = \sum_{i=1}^{n}(X_i^* - a)^2/(2r)$. Using this expression for θ in $\partial \ln L/\partial a = 0$, we see that the MLE of a is the solution of the equation

$$g(a) = 2r \frac{\sum_{i=1}^{n}(X_i^* - a)}{\sum_{i=1}^{n}(X_i^* - a)^2} - \sum_{i=1}^{r} \frac{1}{X_i^* - a} = 0.$$

Table 7. Percentiles to compute 90% CIs for the mean based on censored samples.

	$n = 10$			$n = 15$			$n = 20$			$n = 30$	
r	5%	95%	r	5%	95%	r	5%	95%	r	5%	95%
3	.898	5.24	3	.936	5.88	4	.955	3.75	10	1.02	1.91
5	.894	2.50	5	.945	2.71	7	.981	2.18	13	1.03	1.74
7	.893	1.98	8	.962	1.95	10	.995	1.83	17	1.04	1.62
9	.892	1.78	10	.967	1.78	13	1.02	1.68	20	1.05	1.57
			13	.972	1.65	15	1.02	1.62	23	1.06	1.53
						17	1.01	1.58	27	1.06	1.49

As shown in the uncensored case, it can be shown that

$$P\left(X_{(1)} - 12\hat{b}_r/\sqrt{r} \le a \le X_{(1)}\right) \approx 1, \quad \text{for } r \ge 3, \tag{16}$$

where $\hat{b}_r$ is the moment estimate of b given in (4) based on the r uncensored data. The MLE of a can be found as the root of the equation $g(a) = 0$ using the interval $(X_{(1)} - 12\hat{b}_r/\sqrt{r},\ X_{(1)})$ as the root bracketing interval. The MLE of θ can be obtained as $\tilde{\theta} = \sum_{i=1}^{n}(X_i^* - \tilde{a})^2/(2r)$, where $\tilde{a}$ is the MLE of a. The R code given in Appendix 2 can used to find the MLEs based on a censored/uncensored sample.

Pivotal quantities for all the problems considered for the uncensored case can be readily obtained using the above MLEs for the censored case. For the sake of completeness, in the following, we shall describe the pivotal-based method for finding CI for the mean based on a type II censored sample.

6.1. Confidence interval for the mean based on a type II censored sample

As in the case of uncensored sample, the quantity

$$u_{r,n} = \frac{a + \sqrt{\pi/2}b - \tilde{a}_r}{\tilde{b}_r} \sim \frac{\sqrt{\pi/2} - \tilde{a}_r^*}{\tilde{b}_r^*}, \tag{17}$$

where $\tilde{a}_r^*$ and $\tilde{b}_r^*$ are the MLEs based on a censored sample of size n with r uncensored observations from a Rayleigh$(0, 1)$ distribution. If $u_{r,n;\alpha}$ denote the 100α percentile of the above pivotal quantity, then $(\tilde{a}_r + u_{r,n;\alpha}\tilde{b}_r, \tilde{a}_r + u_{r,n;1-\alpha}\tilde{b}_r)$ is an exact $1 - 2\alpha$ CI for the mean.

For easy verification by a reader, we estimated the 5th and 95th percentiles of the quantity $u_{r,n}$ in (17) for some values of r and n and presented them in Table 7. These percentiles were estimated using simulation consisting of 100,000 runs. As an example, to compute a 90% CI for the mean based on a sample of size 15 with 10 uncensored observations, the 5th percentile is 0.967 and the 95th percentile is 1.78, and $(\tilde{a}_r + 0.967\tilde{b}_r, \tilde{a}_r + 1.78\tilde{b}_r)$ is a 90% CI for the mean.

7. An example

In the production process of a manufacturing factory cutting machines are commonly used, and drill is one of the important components in cutting machines. Drills of different sizes are needed in the production process and the factory purchases the 1.88-mm

Table 8. Lifetime (in minutes) of a sample of 1.8-mm drills.

105	105	95	87	112	80	95	97	77	103	78	87	107	96	79
91	108	97	80	76	92	85	76	96	77	80	100	94	82	104
91	95	93	99	99	94	84	99	91	85	86	79	89	89	100

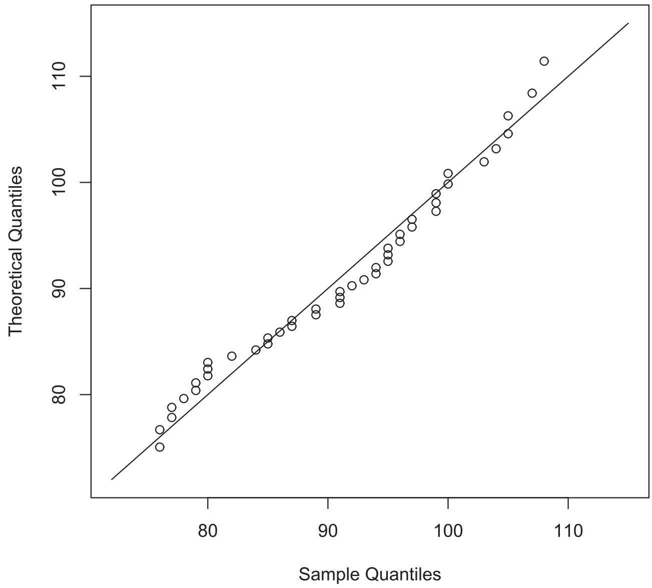

Figure 1. Q-Q plot of drill lifetime data.

drills from a supplier. Lifetime data for the drills are collected during the production process and reported in Table 1 of Chen, Wang and Ye [1]. The data are reproduced here in Table 8. Rayleigh quantile-quantile (Q-Q) plot of the data in Figure 1 is approximately linear. The p-value of the Kolmogorov-Smirnov test for checking a Rayleigh distribution is 0.642. Thus, the data fit a Rayleigh distribution quite well.

In the following, we shall use the lifetime data in Table 8 to illustrate the construction of various statistical intervals described earlier.

Confidence Intervals for the Mean Lifetime: The 95% CIs for the mean life of 1.88-mm drills based on different pivotal quantities are given in the following table.

Methods	est of a and b	lower and upper percentiles	CIs
MLE	72.84 and 14.79	1.068 and 1.466	(88.64, 94.52)
ME	72.82 and 14.84	1.069 and 1.466	(88.68, 94.58)
L-ME	72.19 and 15.34	1.071 and 1.465	(88.62, 94.66)

As the sample size of 45 is considerably large, the CIs based on all three pivots are practically the same.

Tolerance Limits: In the following table, we provide (.90, .95) one-sided tolerance limits.

Methods	est of a and b	lower and upper factors	lower and upper TLs
MLE	72.84 and 14.79	.280 and 2.49	77.0 and 109.7
ME	72.82 and 14.84	.267 and 2.49	76.8 and 109.8
L-ME	72.19 and 15.34	.279 and 2.46	76.5 and 109.9

Notice that the one-sided upper TLs based on all three methods are essentially the same. The lower TL based on the MLEs is larger than the other two TLs. The MLE TL is 77.0 can be interpreted as follows: at least 90% of the drills will last at least 77 minutes with confidence 95%.

Confidence Bounds for $P(X > LSL)$: The lower specification limit (LSL) set by the factory is 80 minutes, and it is desired to estimate the percentage of drills whose lifetime exceeds the LSL. One could assess this percentage by finding a 95% lower confidence limit for $P(X > 80)$. Using Algorithm 1 with $N = 100{,}000$ and MLEs, we estimated the 95% lower confidence bound as 0.812. That is, about 81.2% of drills will last 80 minutes or more with confidence 95%. The lower confidence bound on the basis of MEs is 0.806 and the one on the basis of the L-moment estimates is 0.799. Notice that lower bounds based on the MEs and L-MEs are smaller than the lower bound based on the MLEs.

Prediction Intervals: To illustrate the method of finding a PI for the mean of a future sample, we found 95% PIs for the mean lifetime in a future sample of size $m = 15$ drills using different methods and presented them in the following table.

Methods	est of a and b	lower and upper factors	PIs
MLE	72.84 and 14.79	0.880 and 1.679	(85.86, 97.67)
ME	72.82 and 14.84	0.884 and 1.676	(85.94, 97.69)
L-ME	72.19 and 15.34	0.890 and 1.672	(85.84, 97.84)

The PIs based on the MLEs, MEs and L-MEs are very close to each other. For example, the MLE PI can be interpreted as 'the mean lifetime of a future sample of 15 drills is between 85.86 and 97.67 minutes with confidence 95%.'

It should be noted that the solutions to all of the above problems by different methods are practically the same because the sample size 45 is considerably large.

Censored Case: To illustrate the pivotal-based approach for the censored case, we assume that the largest 14 observations out of these 45 measurements were censored. That is, $n = 45$, $r = 31$ and $x_{(31)} = 96$. We computed the MLEs as $\hat{a}_r = 72.35$ and $\hat{b}_r = 15.74$. To construct a 95% CI for the mean, we estimated the 2.5th percentile as 1.06 and the 97.5th percentile as 1.54. The 95% CI for the mean is

$$(72.35 + 1.06 \times 15.74,\ 72.35 + 1.54 \times 15.74) = (89.03, 96.59).$$

The above CI is in slightly wider than the MLE-pivot CI (88.64, 94.52), which is based on all 45 measurements.

8. Conclusion

We have described the methods of obtaining CIs, PIs and one-sided tolerance limits on the basis of ME-pivotal, L-ME-pivotal and MLE-pivotal quantities. For all the problems considered, the results based on the ME- and L-ME-pivotal quantities are similar and are easy to obtain. The results based on the MLE-pivotal quantities are slightly better than those based on the ME- and L-ME-pivotal quantities. Although, proposed methods involve Monte Carlo simulation, the results are exact except for the simulation errors. A simulation consisting of 100,000 runs would suffice to obtain accurate estimates for practical applications. For example, the R code provided in Appendix 2 takes less than two seconds for computing a 95% CI for the mean (for the example in Section 7) using simulation with $N = 100{,}000$ runs.

Solutions to two-sample problems, such as finding a CI for the difference between means of two Rayleigh distributions or finding a lower confidence limit for a stress-strength reliability parameter, can be readily obtained using the fiducial approach along the lines of Krishnamoorthy and Xia [8] who obtained solutions to various two-sample problems involving two-parameter exponential distributions. We are currently working on such problems and plan to publish elsewhere.

Acknowledgements

The authors are grateful to two reviewers for providing valuable comments and suggestions which enhanced the first version of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] P. Chen, B.X. Wang, and Z-S. Ye, *Yield-based process capability indices for nonnormal continuous data*, J. Qual. Technol. 51 (2019), pp. 171–180.
[2] S. Dey and T. Dey, *Rayleigh distribution revisited via extension of Jeffreys prior information and new loss function*, REVSTAT 9 (2011), pp. 213–226.
[3] S. Dey and T. Dey, *Statistical inference for the Rayleigh distribution under progressively type-II censoring with binomial removal*, Appl. Math. Model. 38 (2014), pp. 974–982.
[4] S. Dey, T. Dey, and D. Kundu, *Two-parameter Rayleigh distribution: Different methods of estimation*, Amer. J. Math. Management Sci. 33 (2014), pp. 55–74.
[5] T. Dey, S. Dey, and D. Kundu, *On progressively type II censored two-parameter Rayleigh distribution*, Comm. Statist. Simulation Comput. 45 (2016), pp. 438–455.
[6] J.R.M. Hosking, *L-moment: Analysis and estimation of distributions using linear combinations of order statistics*, J. R. Stat. Soc. Ser. B 52 (1990), pp. 105–124.
[7] M.S. Kotb and M.Z. Raqab, *Bayesian inference and prediction of the Rayleigh distribution based on ordered ranked set sampling*, Comm. Statist. Simulation Comput. 47 (2018), pp. 905–923.
[8] K. Krishnamoorthy and Y. Xia, *Confidence intervals for a two-parameter exponential distribution: One- and two-sample problems*, Comm. Statist. Theory Methods 47 (2018), pp. 935–952.
[9] J.F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Hoboken, NJ, 2003.
[10] G. Prakash, *Progressively censored Rayleigh data under Bayesian estimation*, Int. J. Intell. Technol. Appl. Statist. 8 (2015), pp. 257–273.

[11] J. Schmee, D. Gladstein, and W. Nelson, *Confidence limits of a normal distribution from singly censored samples using maximum likelihood*, Technometrics 27 (1985), pp. 119–128.
[12] J.-I. Seo, J.-W. Jeon, and S.-B. Kang, *Exact interval inference for the two-parameter Rayleigh distribution based on the upper record values*, J. Probab. Statist. (2016). Available at http://dx.doi.org/10.1155/2016/8246390.
[13] M.M. Siddiqui, *Some problems connected with Rayleigh distributions*, J. Res. Natl. Bur. Stand. D Radio Propagation 66D (1962), pp. 167–174.
[14] M.M. Siddiqui, *Statistical inference for Rayleigh distributions*, Radio Sci. J. Res. NBS/USNC-URSI 68D (1964), pp. 1005–1010.

Appendix 1

To prove (8), we first note that the pdf of $X_{(1)} = \min\{X_1, \ldots, X_n\}$ is given by

$$f_{X_{(1)}}(x) = \frac{n(x-a)}{b^2} \exp\left[-\frac{n}{2}\left(\frac{x-a}{b}\right)^2\right], \quad x > a.$$

For a $t > 0$,

$$\begin{aligned} P\left(X_{(1)} - t \le a \le X_{(1)}\right) &= P\left(a \le X_{(1)} \le a + t\right) \\ &= \int_a^{a+t} f_{X_{(1)}}(x)\,\mathrm{d}x \\ &= \int_0^{nt^2/2b^2} \mathrm{e}^{-y}\,\mathrm{d}y \\ &= 1 - \exp\left(-\frac{1}{2}nt^2/b^2\right). \end{aligned}$$

As $1 - \mathrm{e}^{-u}$ is practically close to unity for $u \ge 20$, we see that

$$P\left(X_{(1)} - t \le a \le X_{(1)}\right) \approx 1 \quad \text{for } t = b\sqrt{\frac{40}{n}}.$$

Since b is usually unknown, we can find an interval of the form $(X_{(1)} - c^*\hat{b}/\sqrt{n} \le a \le X_{(1)})$, where $\hat{b}$ is the moment estimate in (4) and c^* is a positive constant, so that

$$P\left(X_{(1)} - c^*\hat{b}/\sqrt{n} \le a \le X_{(1)}\right) \approx 1.$$

It is difficult to calculate the above probability numerically, but it can estimated as follows. Noting that $\hat{b}$ is a function of $\sum_{i=1}^n (X_i - \bar{X})^2$, it can be readily verified that the above probability, when n is fixed, remains the same for all a and $b > 0$. Using simulation, we find the value $c^* = 12$ for which the probability is close to one for all $n \ge 3$. For example, when $c^* = 12$, the probability is .976 for $n = 3$, .997 when $n = 5$ and .9995 when $n = 7$. Thus, we conclude that

$$P\left(X_{(1)} - 12\hat{b}/\sqrt{n} \le a \le X_{(1)}\right) \approx 1, \quad \text{for } n \ge 3.$$

Appendix 2

To compute a CI for the mean on the basis of the MLEs, we shall briefly outline the computational details on estimating the percentiles of $(\sqrt{\pi/2} - \tilde{a}^*)/\tilde{b}^*$, where $\tilde{a}^*$ and $\tilde{b}^*$ are the MLEs based on a sample of size n from a Rayleigh(0,1) distribution, using R. Applying the inverse probability integral transform to the cdf in (2), it can be easily verified that if $U \sim$ uniform(0, 1) distribution, then $\sqrt{-2\ln U} \sim$ Rayleigh$(0, 1)$ distribution. The following R code can be used to estimate the percentiles of $(\sqrt{\pi/2} - \tilde{a}^*)/\tilde{b}^*$ based on simulation with N runs.

```
****************************************************
# N = number of simulation runs
u = runif(N*n)
xm = matrix(sqrt(-2*log(u)), N, n)
mle = apply(xm, 1, function(x) MLES(x))
pivot = (sqrt(pi/2)-mle[1,])/mle[2,]
perc = quantile(pivot, c(alpha,1-alpha))
*****************************************************
```

The interval ($\tilde{a}$ + perc[1] $\tilde{b}$, $\tilde{a}$ + perc[2] $\tilde{b}$), where $\tilde{a}$ and $\tilde{b}$ are the MLEs based on a sample of size n from a Rayleigh(a, b) distribution, is a $1 - 2\alpha$ CI for the mean $a + \sqrt{\pi/2}b$. The MLEs can be computed using the following R function.

```
****************************************************
MLES = function(x){
n = length(x)
xb = mean(x); s = sd(x)
b.hat = sqrt(2/(4-pi))*s
fn = function(a){
ssq = sum((x-a)^2)
y = 2*n*sum(x-a)/ssq-sum(1/(x-a))
return(y)
}
a0 = x[1]-12*b.hat/sqrt(n); a1 = x[1]
a.mle = uniroot(fn, interval = c(a0,a1), tol = 10^-5,
   maxiter = 20)$root
b.mle = sqrt(.5*sum((x-a.mle)^2)/n)
return(c(a.mle,b.mle))
}
****************************************************
```