

Inferences on the Difference and Ratio of the Means of Two Independent Two-Parameter Exponential Distribution *

SHI JIANHONG LIN HONGMEI

(School of Mathematics and Computer Science, Shanxi Normal University, Linfen, 041004)

Abstract

Methods for interval estimation and hypothesis testing about the ratio of expected lifetimes of two independently distributed two-parameter exponential distribution based on the concept of generalized variable approach are proposed. As assessed by simulation, the coverage probabilities of the proposed approach are found to be very close to the nominal level even for small samples. The proposed new approaches are conceptually simple and are easy to use. Similar procedures are developed for constructing confidence intervals and hypothesis testing about the difference between means of two independent two-parameter exponential distribution.

Keywords: Two-parameter exponential distribution, ratio of means, generalized confidence intervals, generalized pivotal quantity.

AMS Subject Classification: 62F12.

§1. Introduction

It is well known that two-parameter exponential distribution, which occupies an important position in probability and statistical areas, has been widely used in practice, especially in the area of product lifetime. The density function of the two-parameter exponential distribution, $E(\mu, \theta)$, is defined as

$$f(x; \mu, \theta) = \frac{1}{\theta} e^{-(x-\mu)/\theta}, \quad x > \mu, \mu \geq 0, \theta > 0, \quad (1.1)$$

where μ is the gate parameter and θ is the scale parameter, denote $E(\mu, \theta)$. In lifetime data analysis, μ is referred to as a threshold or “guarantee time” parameter, and θ is the mean time to failure. The mean parameter for the $E(\mu, \theta)$ is defined as $m = \mu + \theta$. The two-parameter exponential distribution is useful to model lifetime distributions and wind energy distributions. Examples from many diverse fields such as cardiology, hydrology, demography and finance are given in Lawless (1982). See this book for more details about the two-parameter exponential distribution and its applications.

*The work was supported by the Youth Technology Research Foundation of Shanxi Province (2008021007) and the Natural Science Foundation of Shanxi Normal University (YZ06001).

Received April 24, 2012. Revised July 5, 2012.

Consider two-parameter exponential distribution, $E(\mu, \theta)$, Chiou (1997) proposed a method of finding a confidence interval for the scale parameter θ following a preliminary test. Roy and Mathew (2005) developed a new method based on the concept of generalized confidence intervals to find a generalized confidence limit for the reliability function $e^{-(x-\mu)/\theta}$. Li and Zhang (2010) considered the problem of estimation of the ratio of means of two two-parameter exponential distribution and established an asymptotic confidence interval for it. However, it is clear from the simulation results in Li and Zhang (2010) that the coverage probabilities of the asymptotic confidence interval were quite unsatisfactory for small samples. Therefore, it is desirable to develop an interval method when small samples.

The purpose of this paper is to develop a simple approach which can be used to obtain confidence intervals for the ratio of two two-parameter exponential distribution means. Toward this, we develop methods based on the concept of generalized variable. The generalized p -value was introduced by Tsui and Weerahandi (1989) and generalized confidence interval by Weerahandi (1993). The concepts of generalized confidence interval and generalized p -value have been widely applied to a variety of practical settings where standard methods are failed to produce satisfactory results. For example, see Weerahandi and Johnson (1992), Weerahandi (1991, 1995b), Weerahandi and Berger (1999), Zhou and Mathew (1994), Krishnamoorthy and Mathew (2003), Krishnamoorthy and Lu (2003) and Iyer et al. (2004). For a recipe of constructing generalized pivotal quantities (GPQs), see Iyer and Patterson (2002).

This article is organized as follows. In Section 2, the concepts of the generalized p -value and generalized confidence intervals are outlined. In Section 3, the GPQs for the ratio of two two-parameter exponential distribution means and for the difference between two two-parameter exponential distribution means are developed. Interval estimation and hypothesis testing procedures about the ratio of means and about the difference between two means are developed based on the GPQs. In Section 4, simulation studies are carried out to evaluate the coverage probabilities of the generalized confidence intervals and the asymptotic confidence intervals due to Li and Zhang (2010). Simulation studies indicate that the coverage probabilities of the generalized confidence intervals are very close to the nominal confidence level.

§2. The Generalized Inference

The generalized inference method is based on the concepts of generalized p -value and generalized confidence interval, introduced by Tsui and Weerahandi (1989) and Weerahandi (1993), respectively. These ideas have turned out be very satisfactory for obtaining tests and confidence intervals for many complex problems, especially involving nuisance pa-

rameters (Zhou and Mathew (1994); Weerahandi (1995a, 2004); Roy and Mathew (2005); Mathew and Webb (2005); Ho and Weerahandi (2007)).

Consider an observable random vector X with a probability distribution $P_\eta(\cdot)$, where $\eta = (\theta, \delta)$ is an unknown vector in parameter space Ω . $\theta = \theta(\eta)$ is the parameter of interest, and δ is the nuisance parameter. The problem of interest is to test the one-sided hypothesis

$$H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0.$$

Let x be the observed value of X and consider the generalized test variable $R = R(X; x, \eta)$, which depends on the observed value x and the parameters $\eta = (\theta, \delta)$, and satisfies the following properties:

- (a) The observed value $r = R(x; x, \eta)$ does not depend on unknown parameters.
- (b) When θ is specified, R has a probability distribution that is free of nuisance parameters.
- (c) For fixed x and η , $P\{R(X; x, \eta) \geq r | \theta\}$ is a monotonic function of θ for any given r .

Following the conditions in property (a)-(c), if $P\{R(X; x, \eta) \geq \theta\}$ is nondecreasing in θ , then the generalized p -value for testing

$$H_0 : \theta \leq \theta_0 \leftrightarrow H_1 : \theta > \theta_0$$

can be defined as

$$p = \sup_{\theta \leq \theta_0} P\{R(X; x, \eta) \geq r | \theta\} = P\{R(X; x, \eta) \geq r | \theta_0\}$$

and a small p -value indicates that the observed value does not support H_0 .

Under the same set-up, $T = T(X; x, \eta)$ is a generalized pivotal quantity if it satisfies the following properties:

- (a) The observed value $t = T(x; x, \eta)$ does not depend on nuisance parameters δ .
- (b) T has a probability distribution free of unknown parameters. (2.1)

Given a confidence coefficient γ , let T_γ be the 100γ th percentile of T . Then

$$\{\theta : T(X; x, \eta) \leq T_\gamma\}$$

is a one-sided confidence interval for θ .

§3. The Generalized Variable Approach

We shall first develop a GPQ for the mean of a two-parameter exponential distribution. Even though, one-sample case is not the primary interest of this article, it is

considered just to demonstrate that the generalized variable approach can produce exact results. Furthermore, details of the one-sample case is used to find the GPQ for the two-sample case.

3.1 A GPQ for the Mean of a Two-Parameter Exponential Distribution

Let $X_i \sim E(\mu, \theta)$, $i = 1, 2, \dots, n$ be a sample of size n from two-parameter exponential distribution populations. We shall present our results for the case of type II censored data. Thus let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$ denote the k smallest observations among X_1, X_2, \dots, X_n . Then

$$\hat{\mu} = X_{(1)} \quad \text{and} \quad \hat{\theta} = \frac{1}{k} \left(\sum_{i=1}^k X_{(i)} + (n-k)X_{(k)} - nX_{(1)} \right) \quad (3.1)$$

are jointly sufficient for μ and θ . Furthermore, $\hat{\mu}$ and $\hat{\theta}$ are the maximum likelihood estimators of μ and θ respectively. Also, $\hat{\mu}$ and $\hat{\theta}$ are independently distributed with

$$U = \frac{2n(\hat{\mu} - \mu)}{\theta} \sim \chi^2(2) \quad \text{and} \quad V = \frac{2k\hat{\theta}}{\theta} \sim \chi^2(2k-2), \quad (3.2)$$

where $\chi^2(m)$ denotes a central chisquare distribution with m df (see Lawless (1982), Section 3.5).

We shall now develop a GPQ for the mean of a two-parameter exponential distribution, say $m = \mu + \theta$. The basic idea is to exhibit generalized pivot quantities for μ and θ say T_μ and T_θ , respectively, satisfying the properties in (2.1). That is, the distribution of (T_μ, T_θ) is free of any unknown parameters, and the observed value of (T_μ, T_θ) is (μ, θ) . Once this is done, a generalized pivot quantities, say T_m , for m is given by

$$T_m = T_\mu + T_\theta.$$

In order to derive T_μ and T_θ , let $\hat{\mu}_{\text{obs}}$ and $\hat{\theta}_{\text{obs}}$ denote the observed values of $\hat{\mu}$ and $\hat{\theta}$, respectively, where $\hat{\mu}$ and $\hat{\theta}$ are defined in (3.1). Using the distribution in (3.2), T_μ and T_θ are constructed by

$$T_\mu = \hat{\mu}_{\text{obs}} - \frac{k}{n} \frac{U}{V} \hat{\theta}_{\text{obs}} \quad \text{and} \quad T_\theta = \frac{2k\hat{\theta}_{\text{obs}}}{V}. \quad (3.3)$$

Hence a generalized pivot quantity T_m for m is given by

$$T_m = T_\mu + T_\theta = \hat{\mu}_{\text{obs}} - \frac{k}{n} \frac{U}{V} \hat{\theta}_{\text{obs}} + \frac{2k\hat{\theta}_{\text{obs}}}{V}. \quad (3.4)$$

It can be easily verified that the value of T_m at $(X_1, X_2, \dots, X_k) = (x_1, x_2, \dots, x_k)$ is m , and T_m has a probability distribution free of unknown parameters.

3.2 One-Sample Inference

Let X_1, X_2, \dots, X_n be a sample from a $E(\mu, \theta)$ distribution. We shall present our results for the case of type II censored data. Thus let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$ denote the k smallest observations among X_1, X_2, \dots, X_n .

The UMP-unbiased test for

$$H_0 : m = m_0 \leftrightarrow H_1 : m \neq m_0.$$

Due to the generalized pivot quantity T_m , the generalized p -value is given by $2 \min\{\mathbf{P}(T_m < m_0 | H_0), \mathbf{P}(T_m > m_0 | H_0)\}$. The $100(1 - \alpha)\%$ generalized confidence interval for m is obtained by $(T_m(\alpha/2), T_m(1 - \alpha/2))$, where $T_m(\alpha)$ is the 100α th percentile of T_m .

3.3 Two-Sample Inference based on Generalized Variable Method

Let $X_i \sim E(\mu_1, \theta_1)$, $Y_j \sim E(\mu_2, \theta_2)$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$ be two independent random samples of size n from two-parameter exponential distribution populations. We shall present our results for the case of type II censored data. Thus let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$ and $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(k)}$ denote the k smallest observations among X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_n , respectively. Then

$$\begin{aligned} \hat{\mu}_1 &= X_{(1)} & \text{and} & & \hat{\theta}_1 &= \frac{1}{k} \left(\sum_{i=1}^k X_{(i)} + (n-k)X_{(k)} - nX_{(1)} \right), \\ \hat{\mu}_2 &= Y_{(1)} & \text{and} & & \hat{\theta}_2 &= \frac{1}{k} \left(\sum_{i=1}^k Y_{(i)} + (n-k)Y_{(k)} - nY_{(1)} \right) \end{aligned} \quad (3.5)$$

are jointly sufficient for μ_i and θ_i , $i = 1, 2$. Furthermore, $\hat{\mu}_i$ and $\hat{\theta}_i$, $i = 1, 2$, are the maximum likelihood estimators of μ_i and θ_i , $i = 1, 2$, respectively. Also, $\hat{\mu}_i$ and $\hat{\theta}_i$, $i = 1, 2$, are independently distributed with

$$U_i = \frac{2n(\hat{\mu}_i - \mu_i)}{\theta_i} \sim \mathcal{X}^2(2) \quad \text{and} \quad V_i = \frac{2k\hat{\theta}_i}{\theta_i} \sim \mathcal{X}^2(2k-2), \quad i = 1, 2, \quad (3.6)$$

where $\mathcal{X}^2(m)$ denotes a central chisquare distribution with m df.

We shall construct confidence intervals and hypothesis testing about the ratio of expected lifetimes of two independently distributed two-parameter exponential distribution. The GPQ for m_i based on (3.4) is given by

$$T_{m_i} = \hat{\mu}_{i\text{obs}} - \frac{k}{n} \frac{U_i}{V_i} \hat{\theta}_{i\text{obs}} + \frac{2k\hat{\theta}_{i\text{obs}}}{V_i}, \quad i = 1, 2.$$

Then the GPQ for $R = m_2/m_1$ is given by

$$T_R = T_{m_2}/T_{m_1}. \quad (3.7)$$

For testing $H_0 : m_2/m_1 = 1 \leftrightarrow H_1 : m_2/m_1 \neq 1$, the generalized p -value is given by $2 \min\{\mathbf{P}(T_R < 1|H_0), \mathbf{P}(T_R > 1|H_0)\}$.

Similarly, the GPQ for $D = m_2 - m_1$ is given by

$$T_D = T_{m_2} - T_{m_1},$$

and the generalized p -value for testing $H_0 : m_1 = m_2 \leftrightarrow H_1 : m_1 \neq m_2$ is given by $2 \min\{\mathbf{P}(T_D < 0|H_0), \mathbf{P}(T_D > 0|H_0)\}$.

For given two independent two-parameter exponential samples, the appropriate percentiles of T_R form a $1 - \alpha$ confidence interval for R . Hence the $100(1 - \alpha)\%$ generalized confidence interval for R is obtained by

$$\text{GCI} = \left(T_R\left(\frac{\alpha}{2}\right), T_R\left(1 - \frac{\alpha}{2}\right)\right), \quad (3.8)$$

where $T_R(\alpha)$ is the 100α th percentile of T_R . Similarly, the percentiles of T_D can be used to construct a confidence interval for D .

3.4 Computing Algorithms

For a given data set $X_i \sim E(\mu_1, \theta_1)$, $Y_j \sim E(\mu_2, \theta_2)$, $i = 1, \dots, n$, $j = 1, \dots, n$, the generalized confidence intervals for R and the generalized p -values for testing can be computed using the following steps:

1. Compute $(x_{(1)}, y_{(1)}), (x_{(2)}, y_{(2)}), \dots, (x_{(k)}, y_{(k)})$.
2. Randomly generate $U_i \sim \mathcal{X}^2(2)$, $V_i \sim \mathcal{X}^2(2k - 2)$, $i = 1, 2$, and then calculate T_R from (3.7).
3. Repeat Step 2 a total M times and obtain an array of T_R 's.
4. Rank this array of T_R 's from small to large.

The 100α -th percentile of T_R 's, $T_R(\alpha)$, is an estimation of the lower bound of the one-sided $100(1 - \alpha)\%$ confidence interval and $(T_R(\alpha/2), T_R(1 - \alpha/2))$ is a two-sided $100(1 - \alpha)\%$ confidence interval. The generalized p -value for testing $R = r$ vs. $R \neq r$ is $2 \min\{\mathbf{P}(T_R \geq r), \mathbf{P}(T_R \leq r)\}$. The probability $\mathbf{P}(T_R \geq r)$ can be estimated by the proportion of the T_R 's in Step 3 that are greater than or equal to r . Similarly, $\mathbf{P}(T_R \leq r)$ can be also be estimated.

Remark 1 The asymptotic $100(1 - \alpha)\%$ confidence interval for R , derived by Li and Zhang (2010), is given by

$$\text{ACI} = \left(\frac{\hat{R}}{1 + Z_{1-\alpha/2} / \left[\frac{k-2}{\sqrt{k-1}} \left(\frac{\hat{\mu}_2}{\hat{\theta}_2} - \frac{1}{n(k-2)} \right) \right]}, \frac{\hat{R}}{1 + Z_{\alpha/2} / \left[\frac{k-2}{\sqrt{k-1}} \left(\frac{\hat{\mu}_2}{\hat{\theta}_2} - \frac{1}{n(k-2)} \right) \right]} \right), \quad (3.9)$$

where $\hat{R} = (\hat{\mu}_2 + \hat{\theta}_2)/(\hat{\mu}_1 + \hat{\theta}_1)$, and Z_m denotes the $100m$ th percentile of the standard normal distribution.

§4. Simulation and Conclusion

In this section, we report the results of using Monte Carlo simulations to investigate the estimated coverage probabilities of the confidence intervals (3.8) and (3.9) and their expected lengths. For this, we chose the sample sizes $n = 500$, the type II censored data $k = 25, 50, \dots, 500$. Furthermore, the values of μ and θ were chosen to be $\mu_1 = \mu_2 = 2$, $\theta_1 = 10$, $\theta_2 = 15$. The coverages and expected lengths of the confidence intervals were obtained using 10000 simulations. For $1 - \alpha = 0.95$, Table 1 show estimated coverage probabilities of the confidence intervals (3.8) and (3.9), ACI and GCI, and their expected lengths.

Table 1 The coverage probabilities and expected lengths of 95% confidence intervals

k	Coverage probabilities		Expected lengths	
	GCI	ACI	GCI	ACI
25	0.9531	0	1.5278	*
50	0.9482	0	1.0065	*
75	0.9463	0	0.8055	*
100	0.9550	0.0003	0.6923	*
125	0.9498	0.0020	0.6120	*
150	0.9514	0.0145	0.5624	*
175	0.9491	0.0840	0.5174	5.8926
200	0.9537	0.2718	0.4825	5.3963
225	0.9480	0.5777	0.4530	5.1424
250	0.9558	0.8400	0.4286	5.4283
275	0.9465	0.9686	0.4100	4.9365
300	0.9475	0.9968	0.3912	4.2684
325	0.9492	0.9998	0.3752	4.0440
350	0.9495	1	0.3631	4.3455
375	0.9530	1	0.3495	5.4294
400	0.9512	1	0.3383	4.7700
425	0.9498	1	0.3277	4.2557
450	0.9479	1	0.3189	3.9145
475	0.9458	1	0.3094	3.5895
500	0.9500	1	0.3031	3.3594

The absence of the confidence intervals indicated by *.

As can be seen from Table 1, the generalized confidence interval, GCI, provides very satisfactory coverage probabilities for various choices of sample sizes and the coverage probabilities of GCI are close to the nominal confidence level of 95%. Furthermore, the expected lengths of GCI are shorter than that of ACI in all conditions. However, through the numerical results, the coverages and expected lengths of the asymptotic confidence intervals by Li and Zhang (2010) are quite unsatisfactory, especially for small samples. The asymptotic confidence intervals are not existent when $k < 150$. The ACI performs well in coverage probability when k are close to n , but its expected lengths are much larger than those of GCI.

For overall comparisons, our approach based on generalized confidence interval is far better than the method proposed by Li and Zhang (2010).

Table 2 Monte Carlo estimates of the sizes

k	μ_1	μ_2	θ_1	θ_2	GP				
25					0.021	0.021	0.022	0.021	0.013
50	2	4	12	10	0.019	0.023	0.021	0.022	0.027
75					0.026	0.024	0.015	0.018	0.021
100					0.031	0.029	0.034	0.030	0.022
125	2	4	12	10	0.031	0.034	0.027	0.032	0.029
150					0.035	0.026	0.024	0.027	0.030
175					0.027	0.027	0.037	0.049	0.039
200	2	4	12	10	0.034	0.045	0.047	0.026	0.036
225					0.031	0.028	0.029	0.032	0.034
250					0.044	0.042	0.032	0.038	0.046
275	2	4	12	10	0.046	0.038	0.046	0.038	0.037
300					0.038	0.039	0.046	0.042	0.036
325					0.058	0.042	0.035	0.037	0.053
350	2	4	12	10	0.039	0.042	0.043	0.047	0.054
375					0.047	0.036	0.047	0.049	0.042
400					0.058	0.042	0.035	0.037	0.053
425	2	4	12	10	0.039	0.042	0.043	0.047	0.054
450					0.047	0.036	0.047	0.049	0.042
475					0.058	0.042	0.035	0.037	0.053
500	2	4	12	10	0.039	0.042	0.043	0.047	0.054

An important benchmark for a given statistical test is the size value of the test. Table 2 gives the size value of the test at the nominal level of $\alpha = 0.05$ with $M = 1000$. From the numerical results in Table 2, it appears that the test based on the generalized p -value has a type I error probability very close to the nominal level for all cases.

References

- [1] Weerahandi, S., Generalized Confidence Intervals, *Journal of the American Statistical Association*, **88**(1993), 899–905.
- [2] Li, J.B. and Zhang, R.Q., Inference of parameters ratio in two-parameter exponential distribution, *Chinese Journal of Applied Probability and Statistics*, **26**(2010), 81–88.
- [3] Lawless, J.F., *Statistical Models and Methods for Lifetime Data*, Wiley, New York, 1982.
- [4] Chiou, P., Interval estimation of scale parameters following a pre-test for two exponential distributions, *Computational Statistics & Data Analysis*, **23**(1997), 477–489.
- [5] Roy, A. and Mathew, T., A generalized confidence limit for the reliability function of a two-parameter exponential distribution, *Journal of Statistical Planning and Inference*, **128**(2005), 509–517.
- [6] Tsui, K.W. and Weerahandi, S., Generalized p -values in significance testing of hypotheses in the presence of nuisance parameters, *Journal of the American Statistical Association*, **84**(1989), 602–607.
- [7] Weerahandi, S., Testing reliability in a stress-strength model when X and Y are normally distributed, *Technometrics*, **34**(1992), 83–91.
- [8] Weerahandi, S., Testing variance components in mixed models with generalized p values, *Journal of the American Statistical Association*, **86**(1991), 151–153.
- [9] Weerahandi, S., ANOVA under unequal error variances, *Biometrics*, **51**(1995b), 589–599.
- [10] Weerahandi, S. and Berger, V.W., Exact inference for growth curves with intraclass correlation structure, *Biometrics*, **55**(1999), 921–924.
- [11] Zhou, L. and Mathew, T., Some tests for variance components using generalized p values, *Technometrics*, **36**(1994), 394–402.
- [12] Krishnamoorthy, K. and Mathew, T., Inferences on the means of lognormal distributions using generalized p -values and generalized confidence intervals, *Journal of Statistical Planning and Inference*, **115**(2003), 103–121.
- [13] Krishnamoorthy, K. and Lu, Y., Inferences on the common mean of several normal populations based on the generalized variable method, *Biometrics*, **59**(2003), 237–247.
- [14] Iyer, H.K., Wang, C.M.J. and Mathew, T., Models and confidence intervals for true values in inter-laboratory trials, *Journal of the American Statistical Association*, **99**(2004), 1060–1071.
- [15] Iyer, H.K. and Patterson, P.D., A recipe for constructing generalized pivotal quantities and generalized confidence intervals, *Technical Report 2002-10*, Department of Statistics, Colorado State University, Fort Collins, 2002, Available at <http://www.stat.colostate.edu/research/2002.pdf>.
- [16] Weerahandi, S., *Exact Statistical Methods for Data Analysis*, Springer-Verlag, New York, 1995a.
- [17] Weerahandi, S., *Generalized Inference in Repeatde Measures*, Wiley, New York, 2004.

- [18] Mathew, T. and Webb, D.W., Generalized p values and confidence intervals for variance components: applications to army test and evaluation, *Technometrics*, **47**(2005), 312–322.
- [19] Ho, Y.Y. and Weerahandi, S., Analysis of repeated measures under unequal variances, *Journal of Multivariate Analysis*, **98**(2007), 493–504.

两个独立服从双参数指数分布产品平均寿命比率的统计推断

史建红 林红梅

(山西师范大学数学与计算机科学学院, 临汾, 041004)

本文利用广义 p 值和广义置信区间理论, 研究了两独立服从双参数指数分布产品平均寿命比率的统计推断问题. 给出了平均寿命比率的广义置信区间, 并对该区间的覆盖率和区间长度进行了数据模拟, 模拟结果与已有文献中的近似置信区间进行了比较, 结果显示本文给出的广义置信区间的区间覆盖率和区间长度都要优于近似置信区间, 特别是在小样本的情况下.

关键词: 双参数指数分布, 平均寿命比率, 广义置信区间, 广义枢轴量.

学科分类号: O211.3.