

Tolerance Intervals for Exponentiated Scale Family of Distributions

D. T. SHIRKE,* R. R. KUMBHAR** & D. KUNDU†

*Department of Statistics, Shivaji University, Kolhapur, India; ** Department of Statistics, Kisan Veer Mahavidyalaya, Wai Satara, India; †Department of Mathematics, Indian Institute of Technology, Kanpur, India

ABSTRACT In this article we provide an asymptotic upper β -expectation and β -content γ -level tolerance intervals for a new family of distributions, namely the Exponentiated Scale family of distributions. Expected coverage of a proposed β -expectation Tolerance Interval is obtained. Bootstrap-based tolerance limits are obtained for data arising from an exponentiated exponential distribution.

KEY WORDS: β -expectation tolerance interval, β -content γ -level tolerance interval, expected coverage, exponentiated scale family, exponentiated exponential distribution.

Introduction

In general, the term Tolerance Interval (TI) is an interval determined from observed values of a random sample for the purpose of drawing inferences about the proportion of a distribution contained in that interval. Usually TI is designed to capture at least a given proportion of some distribution. Two types of TI have received considerable attention in the literature; β -expectation TI and β -content γ -level TI. In order to be more specific about the meaning of TI, let X be a measurable characteristic having a distribution function $F(x; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$. Let $L(\underline{X})$ and $U(\underline{X})$ be two functions of observations such that $L(\underline{X}) < U(\underline{X})$. Then (L, U) is called a β -content γ -level TI, if for given $\beta, \gamma \in (0, 1)$,

$$P \left\{ \int_L^U f(x; \theta) dx \geq \beta \right\} = \gamma, \quad \text{for every } \theta \in \Theta, \quad (1)$$

and if L and U are determined so that

$$E \left\{ \int_L^U f(x; \theta) dx \right\} b = \beta, \quad \text{for every } \theta \in \Theta \quad (2)$$

Correspondence Address: D.T. Shirke, Department of Statistics, Shivaji University, Kolhapur, India, 416004.
Email: dts_stats@unishivaji.ac.in

then (L, U) is called a β -expectation TI, where $f(x; \theta)$ is probability density function (pdf) of X . The quantity $\int_L^U f(x; \theta) dx$ is called the sample coverage and L and U are called lower and upper tolerance limits, respectively. If we set $L = -\infty$ and obtain U satisfying equations (1) and (2) then we get upper β -content γ -level TI and upper β -expectation TI respectively. Similarly, if we set $U = \infty$ and obtain L satisfying equations (1) and (2) then we get lower β -content γ -level TI and lower β -expectation TI respectively. It is easy to observe that the lower β -content γ -level tolerance limit is also an upper $(1-\beta)$ -content $(1-\gamma)$ -level tolerance limit and the lower β -expectation tolerance limit is the upper $(1-\beta)$ -expectation tolerance limit. Therefore, in the present study, we obtain only upper tolerance limits.

Wilks (1941) in his pioneering article first treated the problem of determining TIs. Since then, a large number of papers dealing with this and other aspects of tolerance limits have appeared in the literature. Jilek (1981) classified papers according to general results, distribution free results, normal and multivariate normal distributions, gamma, exponential, Weibull and other continuous and discrete distributions. Patel (1986) provided a review paper, which contains a large collection of known results on β -content γ -level TIs for different continuous and discrete univariate distributions. The main aim of this paper is to construct TIs for the following family of distributions,

$$\mathfrak{S} = \{F: F_X(x; \theta, \alpha) = [G(x/\theta)]^\alpha; \alpha > 0, \theta > 0\} \quad (3)$$

where $G(\cdot)$ is the known cumulative distribution function (cdf). We name this family as the *Exponentiated scale family* of distributions; which is analogous to Lehmann alternatives (see Lehmann, 1953). Under the assumption that all Cramer's regularity conditions are satisfied for F defined in equation (3), we obtain TI for this family.

Suppose X_1, X_2, \dots, X_n are independent and identically distributed (i.i.d) with cdf $F_X(x; \theta, \alpha)$; members of \mathfrak{S} . The pdf of X_i is

$$f_{X_i}(x; \theta, \alpha) = \begin{cases} \alpha[G(x/\theta)]^{\alpha-1} G'_x(x/\theta) & x \in S, \theta, \alpha > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where S is the support of X , which is independent of parameters and $G'_x(\cdot)$ is the derivative of $G(\cdot)$ with respect to x , and is the pdf of X .

The rest of the paper is organized as follows. In the next section we construct an upper β -expectation TI and obtain its approximate expected coverage for the *exponentiated scale family* of distributions. The problem of construction of asymptotic β -content γ -level TI is presented in the section after. In the fourth section we obtain both the types of TIs using the results from the previous two sections for Exponentiated Exponential (EE) distribution introduced by Gupta *et al.* (1998). We study the performance of the proposed β -expectation TI for EE distribution through simulation and illustrate a procedure of construction of asymptotic β -content γ -level TI using the empirical distribution approach and bootstrap technique for real life data due to Lawless (1982) in the fifth section.

β -expectation Tolerance Interval

Let $\underline{\mu} = (\theta, \alpha)^T$ and $x_\beta(\underline{\mu}) \in S$ be the lower β th percentile of $F_X(x; \theta, \alpha)$. Therefore, we have $F_X(x_\beta(\underline{\mu}); \theta, \alpha) = \beta$ for $\beta \in (0, 1)$. This implies $x_\beta(\underline{\mu}) = \theta G^{-1}(\beta^{1/\alpha})$, where $G^{-1}(\cdot)$ is the inverse function of $G(\cdot)$. Since $\underline{\mu} = (\theta, \alpha)^T$ is unknown, we replace it by its

Maximum Likelihood Estimator (MLE). If $\hat{\underline{\mu}} = (\hat{\theta}, \hat{\alpha})^T$ is the MLE of $\underline{\mu}$, then by the invariance property of MLE we have MLE of $x_\beta(\underline{\mu})$ as $x_\beta(\hat{\underline{\mu}}) = \hat{\theta}G^{-1}(\beta^{1/\hat{\alpha}})$. Therefore we propose an approximate upper β -expectation TI as

$$I_1(\underline{X}) = (0, \hat{\theta}G^{-1}(\beta^{1/\hat{\alpha}})). \tag{5}$$

Expected coverage of equation (5) is given in the following theorem.

Theorem 1

The Approximate expected coverage of $I_1(\underline{X})$ is given by

$$E[F_X\{X_\beta(\hat{\underline{\mu}}); \underline{\mu}\}] = \beta + A(\underline{\mu})\frac{\sigma_1^2}{n} + B(\underline{\mu})\frac{\sigma_2^2}{n} + C(\underline{\mu})\frac{\sigma_{12}}{n} \tag{6}$$

where

$$A(\underline{\mu}) = 0.5\alpha(1 - \alpha)[G(t)]^{\alpha-2}[G'_\theta(t)]^2 + \alpha[G(t)]^{\alpha-1}\left[\frac{G''_{x\theta}(t)G'_\theta(t)}{G'_x(t)} - 0.5G''_{\theta\theta}(t)\right]$$

$$B(\underline{\mu}) = \alpha^{-1}[G(t)]^\alpha \log G(t)[1 + 0.5\alpha \log G(t)] \text{ and}$$

$$C(\underline{\mu}) = [G(t)]^\alpha \log G(t)\left[\frac{(\alpha - 1)G'_\theta(t)}{G(t)} + \frac{G''_{x\theta}(t)}{G'_x(t)}\right]$$

with $t = x/\theta$, $G'_x(t) = \partial G(t)/\partial x$, $G'_\theta(t) = \partial G(t)/\partial \theta$, $G''_{x\theta}(t) = \partial^2 G(t)/\partial x \partial \theta$, $G''_{\theta\theta}(t) = \partial^2 G(t)/\partial \theta^2$ while σ_1^2/n , σ_2^2/n are asymptotic variances of $\hat{\theta}$ and $\hat{\alpha}$ respectively and $\sigma_{12}/n = \text{Cov}(\hat{\theta}, \hat{\alpha})$ is an asymptotic covariance of $\hat{\theta}, \hat{\alpha}$. These quantities can be obtained from Fisher information matrix I.

Proof

See the Appendix for the proof.

Lemma 1

$I_1(\underline{X})$ is consistent TI.

Proof

The proof follows from consistency of $\hat{\theta}$ and $\hat{\alpha}$ and by Cramer (1946).

β -Content γ -level Tolerance Interval

Let $I_2(\underline{X}) = (0, \delta\hat{X}_\beta)$ be an upper β -Content γ -level TI for the distribution having cdf (3). The factor $\delta > 0$ is to be determined such that $I_2(\underline{X})$ is a β -Content γ -level TI for $\beta \in (0, 1)$ and $\gamma \in (0, 1)$.

That is

$$P[F(\delta\hat{X}_\beta; \theta, \alpha) \geq \beta] = \gamma$$

Equivalently, we get

$$P\left[\hat{X}_\beta \geq \frac{\theta G^{-1}(\beta^{1/\alpha})}{\delta}\right] = \gamma \tag{7}$$

We note that $\hat{X}_\beta(\underline{\mu}) \rightarrow AN(X_\beta(\underline{\mu}), \sigma^2(\underline{\mu})/n)$, where $\sigma^2(\underline{\mu}) = HI^{-1}H^T$, I is the Fisher information matrix and $H = [\partial X_\beta(\underline{\mu})/\partial\theta, \partial X_\beta(\underline{\mu})/\partial\alpha]$.

Using the asymptotic distribution of $\hat{\underline{\mu}}$ we have

$$P\left[Z \leq \left\{\frac{\theta G^{-1}(\beta^{1/\alpha})}{\delta} - X_\beta(\underline{\mu})\right\} \left(\frac{\sqrt{n}}{\sigma(\underline{\mu})}\right)\right] = 1 - \gamma \tag{8}$$

where Z is a standard normal variate. Therefore, we choose δ such that

$$\left\{\frac{\theta G^{-1}(\beta^{1/\alpha})}{\delta} - X_\beta(\underline{\mu})\right\} \left(\frac{\sqrt{n}}{\sigma(\underline{\mu})}\right) = Z_{1-\gamma} \tag{9}$$

where $Z_{1-\gamma}$ is the $100(1 - \gamma)$ th percentile of the standard normal variate. From equation (9) we have,

$$\delta = \left[1 + \frac{Z_{1-\gamma}\sigma(\underline{\mu})/\sqrt{n}}{\theta G^{-1}(\beta^{1/\alpha})}\right]^{-1} \tag{10}$$

We note that δ depends on both the parameters θ and α . Replacing θ and α by their respective MLEs, an asymptotic upper β -content γ -level tolerance limit of $I_2(\underline{X})$ is proposed as

$$U(\underline{X}) = \left[1 + \frac{Z_{1-\gamma}\hat{\sigma}(\hat{\underline{\mu}})/\sqrt{n}}{\hat{\theta} G^{-1}(\beta^{1/\hat{\alpha}})}\right]^{-1} X_\beta(\hat{\underline{\mu}}) \tag{11}$$

In the following section we obtain both types of TI for an EE distribution as a particular case to illustrate the results in the previous sections.

Tolerance Intervals for EE Distribution

Several authors have considered the problem of constructing TIs for Gamma and Weibull distributions. Some of these are Bain (1978), Bain & Weeks (1965), Guenther (1972) etc. Gupta *et al.* (1998) introduced the EE distribution as a generalization of the standard exponential distribution. Gupta & Kundu (1999, 2001) in their series of papers have discussed the properties of EE distribution, including that the EE distribution is an alternative to Gamma and Weibull distributions. The comparison of the EE distribution with Gamma and Weibull distributions is done through real-life situations. In this paper, we obtain construction of TIs for EE distribution as an illustration of the general theory reported earlier.

Substituting $G(x/\theta) = (1 - \exp(-x/\theta))$ in equation (3) we get

$$F_X(x; \theta, \alpha) = \begin{cases} (1 - \exp(-x/\theta))^\alpha & x > 0, \theta > 0, \alpha > 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

and

$$f_X(x; \theta, \alpha) = \begin{cases} \frac{\alpha}{\theta} \exp(-x/\theta)(1 - \exp(-x/\theta))^{\alpha-1} & x > 0, \theta > 0, \alpha > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

as the cdf and pdf of an EE distribution. In particular when $\alpha = 1$, it represents the exponential distribution. If $\alpha = k$ (non-negative known integer) then it gives the pdf of a lifetime distribution of the largest of k i.i.d. exponentials with mean θ . TIs for the same are discussed by Kumbhar & Shirke (2004). But when $\alpha > 0$ and is unknown, the problem of obtaining TIs becomes a problem of constructing TIs for an EE family of distribution. As parameters θ and α are unknown we obtain tolerance limits based on MLEs. Details regarding MLEs of θ and α have been given in Gupta & Kundu (2001).

The β th percentile of equation (12) is $X_\beta(\underline{\mu}) = -\theta \log(1 - \beta^{1/\alpha})$. Therefore, an upper β -expectation TI for EE distribution is obtained from equation (5) as

$$I_1(\underline{X}) = (0, -\hat{\theta} \log(1 - \beta^{1/\hat{\alpha}})) \tag{14}$$

and its expected coverage is given by Theorem 1, with

$$\begin{aligned} A(\underline{\mu}) &= 0.5\alpha^2 f(\alpha f - 1)(1 - f)^{\alpha-2} \theta^{-2} \\ B(\underline{\mu}) &= (1 - f)^\alpha \log(1 - f)[1 + 0.5\alpha \log(1 - f)]\alpha^{-1} \\ C(\underline{\mu}) &= (1 - f)^{\alpha-1} \log(1 - f)\{(1 - \alpha)f + (t - 1)\} \theta^{-1} \end{aligned}$$

where $f = \exp(-t)$. An asymptotic upper β -content γ -level TI for EE distribution is

$$I_2(\underline{X}) = (0, -\hat{\delta} \hat{\theta} \log(1 - \beta^{1/\hat{\alpha}})) \tag{15}$$

where factor $\hat{\delta}$ is given by

$$\hat{\delta} = \left[1 - \frac{Z_{1-\gamma} \sigma(\hat{\underline{\mu}}) / \sqrt{n}}{\hat{\theta} \log(1 - \beta^{1/\hat{\alpha}})} \right]^{-1} \tag{16}$$

In order to use $I_2(\underline{X})$, one needs to know expressions for asymptotic variances and covariance. However we avoid use of such expressions and adopt a re-sampling approach to estimate the variance and covariance of MLEs. Readers interested in detailed expressions may refer to Gupta & Kundu (2001). In the following section, we study the performance of $I_1(\underline{X})$ using simulation and illustrate the use of $I_2(\underline{X})$ by real life data using an empirical distribution approach and using the bootstrap technique.

Simulation Study

Note that expression (6) gives an approximate value of the actual coverage of the interval (14) for the EE distribution. Hence we use simulation to study the performance of the proposed β -expectation TI using MLEs of θ and α , namely, $\hat{\theta}$ and $\hat{\alpha}$ respectively. In the simulation study we generate n ($= 10, 25, 50$) observations from the EE distribution with $\theta = 1$ and $\alpha = 2$. We obtain $\hat{\theta}$ and $\hat{\alpha}$ by solving likelihood equations simultaneously. These estimates are then used to compute $X_{\beta}(\hat{\mu}) = -\hat{\theta} \log(1 - \beta^{1/\hat{\alpha}})$. Repeating the process 5000 times we obtain that many estimates of $X_{\beta}(\hat{\mu})$ and compute expected coverage of the interval $I_1(\underline{X})$. Corresponding values are tabulated in Table 1 for $\beta = 0.90, 0.95, 0.975$ and 0.99 .

From Table 1, we observe that for small n ; $I_1(\underline{X})$ underestimates coverage, while as n increases coverage converges to the desired value.

To obtain an upper β -content γ -level TI, using equations (15) and (16) we require asymptotic variances and covariance of the MLEs $\hat{\theta}$ and $\hat{\alpha}$. There are two types of problems in setting the TI here. The first one is that expressions for asymptotic variance and covariance of the MLE involve integrals to be solved numerically (see Gupta & Kundu, 2001). The second and the most serious is that asymptotic variance σ_1^2 for $\hat{\theta}$ takes two different functional forms depending on $\alpha > 2$ or $\alpha \leq 2$. Therefore, it is difficult to choose either of these, as expression for σ_1^2 , since α is unknown. We resolve both these problems by using an empirical distribution approach and bootstrap technique. In the following we provide an illustrative example in which both the approaches to construct β -content γ -level TI have been used.

Illustrative Example

Gupta & Kundu (2001) reported that the EE distribution can be used as an alternative to Gamma and Weibull distributions. They have given two data sets, where it is observed that, in the first data set, the EE distribution has a better fit compared to the Weibull and Gamma distributions, and for the other data set Weibull has a better fit than EE and Gamma distributions. We consider the first data set and obtain an upper limits of $I_1(\underline{X})$ and $I_2(\underline{X})$ for different values of β . The data set is as follows.

Data set (Lawless, 1982, p. 228): The data given here arose in tests on endurance of deep groove ball bearings. The data are the number of million revolutions before failure for each of the 23 ball bearings in the life test and they are: 17.23, 28.92, 33.00, 41.52, 42.12, 45.60, 48.80, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04 and 173.40.

Now we obtain $U(\underline{X})$ of $I_2(\underline{X})$ using the following three methods.

Table 1. Expected coverage of $I_1(\underline{X})$ TI when $\theta = 1$ and $\alpha = 2$

β	N			
	10	25	50	100
0.90	0.8664	0.8896	0.8922	0.9000
0.95	0.9224	0.9392	0.9424	0.9452
0.975	0.9496	0.9682	0.9730	0.9720
0.99	0.9684	0.9818	0.9850	0.9878

- M1. Using MLEs of θ, α based on original data and estimates of $\sigma(\underline{\mu})$ using the empirical distribution approach.
- M2. Using MLEs of θ, α based on the original data and bootstrap estimates of $\sigma(\underline{\mu})$.
- M3. Using bootstrap estimates of θ, α and $\sigma(\underline{\mu})$.

For these data, MLEs of θ and α are $\hat{\theta} = 30.963509$ and $\hat{\alpha} = 5.283209$. Using the empirical distribution approach, estimates of variances and covariance of MLEs $\hat{\theta}$ and $\hat{\alpha}$ are given by, respectively

$$(\hat{\sigma}_{e1})^2 = 41.0468, \quad (\hat{\sigma}_{e2})^2 = 4.53834 \quad \text{and} \quad \hat{\sigma}_{e12} = -11.68207.$$

Using the bootstrap technique we generate 25,000 random samples (with replacement) each of size 23 from the original sample. For each of the bootstrap sample we compute a MLE of θ and α . Based on such 25,000 MLEs we obtain $\hat{\theta}_b, \hat{\alpha}_b, \hat{\sigma}_{b1}, \hat{\sigma}_{b2}$ and $\hat{\sigma}_{b12}$ by taking averages, variances and covariances of MLEs respectively.

For the given data set we get $\hat{\theta}_b = 29.952213$ and $\hat{\alpha}_b = 6.820379$ with

$$(\hat{\sigma}_{b1})^2 = 34.028919, \quad (\hat{\sigma}_{b2})^2 = 16.049763 \quad \text{and} \quad \hat{\sigma}_{b12} = -16.247841.$$

Based on above estimates, upper limits for different values of β when $\gamma = 0.90$ and 0.95 are tabulated in Table 2.

It is observed that all three approaches that we have used lead to somewhat the same TI. The bootstrap-based approach has a wider length of TIs as against the one based on the partial bootstrap and empirical distribution approach. However, for higher values of β , a partial bootstrap approach gives shorter length TIs. From the practical viewpoint it is recommended that an empirical distribution approach is better, which involves few computational complexities as compared with the other two.

Upper β -expectation tolerance limits using the above data set for different values of β are obtained as follows.

β :	0.90	0.95	0.975	0.99
$U(\underline{X})$:	121.5273	143.6577	165.4435	194.0060

Since the family (3) can accommodate many more distributions, the theory developed here can be extended for other situations also, such as exponentiated Weibull distribution introduced by Mudholkar *et al.* (1995).

Table 2. Upper β -content γ -level tolerance limits of $I_2(\underline{X})$

β	$\gamma = 0.90$			$\gamma = 0.95$		
	Using M1	Using M2	Using M3	Using M1	Using M2	Using M3
0.90	125.9264	126.5334	129.9994	127.2216	128.0166	131.4347
0.95	149.2426	149.1953	152.3091	150.8925	150.8306	154.0013
0.975	172.2330	171.6668	174.3413	174.2442	173.5024	176.3134
0.99	202.4053	201.3016	203.2974	204.9002	203.4533	205.6622

References

- Atwood, C. L. (1984) Approximate tolerance intervals based on maximum likelihood estimator, *Journal of American Statistical Association*, 79, pp. 459–465.
- Bain, L. J. (1978) *Statistical Analysis of Reliability and Life Testing Model* (New York: Marcel Dekker).
- Bain, L. J. & Weeks, D. L. (1965) Tolerance limits for generalised gamma distribution, *Journal of American Statistical Association*, 60, pp. 1142–1152.
- Cramer, H. (1946) *Mathematical Methods of Statistics* (Princeton, NJ: Princeton University Press).
- Guenther, W. C. (1972) Tolerance intervals for univariate distributions, *Naval Research Logistic Quarterly*, 19, pp. 309–333.
- Gupta, R. D. & Kundu, D. (1999) Generalized exponential distribution, *Australian and New Zealand Journal of Statistics*, 41(2), pp. 173–188.
- Gupta, R. D. & Kundu, D. (2001) Exponentiated Exponential family: an alternative to Gamma and Weibull Distributions, *Biometrical Journal*, 423(1), pp. 117–130.
- Gupta, R. C. et al. (1998) Modeling failure time data by Lehmann alternatives, *Communication in Statistics, Theory and Methods*, 27, pp. 887–904.
- Jilek, M. (1981) A bibliography of statistical tolerance regions, *Mathematische Operationsforschung und Statistik – Series Statistics*, 12, pp. 441–456.
- Kumbhar, R. R. & Shirke, D. T. (2004) Tolerance limits for lifetime distribution of k -unit parallel system, *Journal of Statistical Computations and Simulations*, 74(3), pp. 201–213.
- Lawless, J. F. (1982) *Statistical Models and Methods for Lifetime Data* (New York: Wiley).
- Lehmann, E. L. (1953) The power of Rank tests, *Annals of Mathematical Statistics*, 24, pp. 23–43.
- Mudholkar, G. S. et al. (1995) The exponentiated Weibull family: a reanalysis of the bus-motor failure data, *Technometrics*, 37, pp. 436–445.
- Patel, J. K. (1986) Tolerance interval – a review, *Communications in Statistics, Theory and Method*, 15(19), 2719–2762.
- Wilks, S. S. (1941) Determination of sample sizes for setting tolerance limits, *Annals of Mathematical Statistics*, 12, pp. 91–96.

Appendix

Proof of Theorem 1

We obtain approximate expected coverage of $I_1(\underline{X})$ using Atwood (1984). As $\hat{\underline{\mu}} = (\hat{\theta}, \hat{\alpha})^T$ is asymptotically unbiased for $\underline{\mu} = (\theta, \alpha)^T$ we have

$$E\{F[X_{\beta}(\hat{\underline{\mu}}); \underline{\mu}]\} \cong \beta - 0.5\text{tr}(F_{02}I^{-1}) + F_{01}^T I^{-1} F_{11} / F_{01}$$

where $F_{10} = \partial F / \partial x$, $F_{01} = \partial F / \partial \underline{\mu}$, $F_{11} = \partial^2 F / \partial x \partial \underline{\mu}^T$, $F_{02} = \partial^2 F / \partial \underline{\mu} \partial \underline{\mu}^T$ and all the derivatives are evaluated at $x = X_{\beta}$. By setting $t = x/\theta$ and ignoring the terms of higher order, after simplification we get the required result as stated in Theorem 1.