Constructing Tolerance Limits On Order Statistics In Future Samples Coming From Location-Scale Distributions

Nicholas A. Nechval
•
Dept. of Mathematics, Baltic International Academy, Riga, Latvia
e-mail: nechval@junik.lv

Konstantin N. Nechval
•
Dept. of Applied Mathematics, Transport and Telecommunication Institute, Riga, Latvia
e-mail: konstan@tsi.lv

Vladimir F. Strelchonok
•
Dept. of Mathematics, Baltic International Academy, Riga, Latvia
e-mail: str@apollo.lv

Abstract

Although the concept of statistical tolerance limits has been well recognized for long time, surprisingly, it seems that their applications remain still limited. Analytic formulas for the tolerance limits are available in only simple cases. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations. In this paper, a new approach to constructing lower and upper tolerance limits on order statistics in future samples is proposed. Attention is restricted to location-scale distributions under parametric uncertainty. The approach used here emphasizes pivotal quantities relevant for obtaining tolerance factors and is applicable whenever the statistical problem is invariant under a group of transformations that acts transitively on the parameter space. It does not require the construction of any tables and is applicable whether the past data are complete or Type II censored. The proposed approach requires a quantile of the F distribution and is conceptually simple and easy to use. For illustration, the normal and log-normal distributions are considered. The discussion is restricted to one-sided tolerance limits. A practical example is given.

Keywords: *order statistics, F distribution, tolerance limits, location-scale distribution*

1. Introduction

Statistical tolerance limits are an important tool often utilized in areas such as engineering, manufacturing, and quality control for making statistical inference on an unknown population. As opposed to a confidence limit that provides information concerning an unknown population parameter, a tolerance limit provides information on the entire population; to be specific, one-sided tolerance limit is expected to capture a certain proportion or more of the population, with a given confidence level. For example, an upper tolerance limit for a univariate population is such that with a given confidence level, a specified proportion or more of the population will fall below the limit. A lower tolerance limit satisfies similar conditions. It is often desirable to have statistical tolerance limits available for the distributions used to describe time-to-failure data in reliability problems. For example, one might wish to know if at least a certain proportion, say β, of a manufactured product will operate at least T hours. This question can not usually be answered exactly, but it may be possible to determine a lower tolerance limit $L(X_1, \ldots, X_n)$, based on a preliminary random sample $(X_1, \ldots, X_n)$, such that one can say with a certain confidence γ that at least 100β % of the product will operate longer than $L(X_1, \ldots, X_n)$. Then reliability statements can be made based on $L(X_1, \ldots, X_n)$, or, decisions can be reached by comparing $L(X_1, \ldots, X_n)$ to T. Tolerance limits of the type mentioned above are considered in this paper. That is, if $f_\theta(x)$ denotes the density function of the parent population under consideration and if S is any statistic obtained from the preliminary random sample $(X_1, \ldots, X_n)$ of that population, then $L(S)$ is a lower γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{L(S)}^{\infty} f_\theta(x)dx \geq \beta\right) = \gamma, \tag{1}$$

and $U(S)$ is an upper γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{-\infty}^{U(S)} f_\theta(x)dx \geq \beta\right) = \gamma, \tag{2}$$

where θ is the parameter (in general, vector).

The common distributions used in life testing problems are the normal, log-normal, exponential, Weibull, and gamma distributions [1]. Tolerance limits for the normal distribution have been considered in [2], [3], [4], and others.

Tolerance limits enjoy a fairly rich history in the literature and have a very important role in engineering and manufacturing applications. Patel [5] provides a review (which was fairly comprehensive at the time of publication) of tolerance limits for many distributions as well as a discussion of their relation with confidence intervals for percentiles and prediction intervals. Dunsmore [6] and Guenther, Patil, and Uppuluri [7] both discuss 2-parameter exponential tolerance intervals and the estimation procedure in greater detail. Engelhardt and Bain [8] discuss how to modify the formulas when dealing with type II censored data. Guenther [9] and Hahn and Meeker [10] discuss how one-sided tolerance limits can be used to obtain approximate two-sided tolerance intervals by applying Bonferroni's inequality. Tolerance limits on order statistics in future samples coming from a two-parameter exponential distribution have been considered in [11].

In contrast to other statistical limits commonly used for statistical inference, the tolerance limits (especially on order statistics) are used relatively rarely. One reason is that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations.

In this paper, a new approach to constructing lower and upper tolerance limits on order statistics in future samples is proposed. For illustration, the normal and log-normal distributions that are commonly used in reliability and risk theory are considered. Although the concept of statistical tolerance limits has been well recognized for long time, surprisingly, it seems that their applications remain still limited.

2. Mathematical Preliminaries

2.1. Probability Distribution Function of Order Statistic

Theorem 1. If there is a random sample of m ordered observations $Y_1 \le \ldots \le Y_m$ from a known distribution (continuous or discrete) with density function $f_\theta(y)$, distribution function $F_\theta(y)$, then the probability distribution function of the kth order statistic Y_k, $k \in \{1, 2, \ldots, m\}$, is given by

$$P_\theta(Y_k \le y_k) = \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx, \tag{3}$$

where

$$f_{2(m-k+1),2k}(x) = \frac{1}{\mathrm{B}\left(\dfrac{2(m-k+1)}{2}, \dfrac{2k}{2}\right)} \left(\frac{2(m-k+1)}{2k}\right)\left(\frac{2(m-k+1)}{2k}x\right)^{2(m-k+1)/2-1}$$

$$\times \left(1 + \frac{2(m-k+1)}{2k}x\right)^{-\,[2(m-k+1)+2k]/2}, \quad x > 0, \tag{4}$$

is the probability density function of an F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. Suppose an event occurs with probability p per trial. It is well-known that the probability P of its occurring k or more times in m trials is termed a cumulative binomial probability, and is related to the incomplete beta function $I_x(a, b)$ as follows:

$$P \equiv \sum_{j=k}^{m} \binom{m}{j} p^j (1-p)^{m-j} = I_p(k, m-k+1). \tag{5}$$

It follows from (5) that

$$P_\theta\{Y_k \le y_k\} = \sum_{j=k}^{m} \binom{m}{j} [F_\theta(y_k)]^j [1 - F_\theta(y_k)]^{m-j} = I_{F_\theta(y_k)}(k, m-k+1)$$

$$= \frac{1}{\mathrm{B}(k, m-k+1)} \int_0^{F_\theta(y_k)} u^{k-1}(1-u)^{(m-k+1)-1} du = \frac{\left(\dfrac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\dfrac{2k}{2}, \dfrac{2(m-k+1)}{2}\right)} \int_0^{F_\theta(y_k)} u^{\frac{2(m-k+1)+2k}{2}}$$

$$\times \left(\frac{1-u}{u}\frac{2k}{2(m-k+1)}\right)^{2(m-k+1)/2-1} \frac{-2k}{2(m-k+1)}\left(-\frac{du}{u^2}\right)$$

$$= \frac{\left(\frac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\frac{2(m-k+1)}{2},\frac{2k}{2}\right)} \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}}^{\infty} x^{2(m-k+1)/2-1}\left(1+\frac{2(m-k+1)}{2k}x\right)^{-[2(m-k+1)+2k]/2} dx, \quad (6)$$

where

$$x = \frac{1-u}{u}\frac{2k}{2(m-k+1)}. \quad (7)$$

This ends the proof.

Corollary 1.1.

$$P_\theta(Y_k > y_k) = 1 - P_\theta\{Y_k \le y_k\} = \int_0^{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx. \quad (8)$$

Corollary 1.2. If $y_{k,m;\gamma}$ is the quantile of order γ for the distribution of Y_k, we have from (8) that $y_{k,m;\gamma}$ is the solution of

$$F_\theta(y_{k,m;\gamma}) = k\big/[k + (m-k+1)q_{2(m-k+1),2k;1-\gamma}], \quad (9)$$

where $q_{2(m-k+1),2k;1-\gamma}$ is the quantile of order $1-\gamma$ for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

2.2. Normal and Log-Normal Distributions

The normal and log-normal distributions are commonly used to model certain types of data that arise in several fields of engineering as, for example, different types of lifetime data (see, e.g., [12]). The goal of modeling certain types of data is to provide quantitative forecasts of various system performance measures such as service level, expected waiting time, agent's occupancy, schedule efficiency, cost etc. Evaluation of these performance measures is important to making optimal decisions about overall cost, system performance, which has to be within the allowable budget and other performance based constraints.

Particular properties of the log-normal random variable (as the non-negativeness and the skewness) and of the log-normal hazard function (which increases initially and then decreases) make log-normal distribution a suitable fit for some engineering data sets. The log-normal distribution is used to model the lives of units whose failure modes are of a fatigue-stress nature. Since this includes most, if not all, mechanical systems, the log-normal distribution can have widespread application. Consequently, the log-normal distribution is a good companion to the Weibull distribution when attempting to model these types of units. As may be surmised by the name, the log-normal distribution has certain similarities to the normal distribution. A random variable is log-normally distributed if the logarithm of the random variable is normally distributed. Because of this, there are many mathematical similarities between the two distributions. For example, the mathematical reasoning for the construction of the probability plotting scales and the bias of parameter estimators is very similar for these two distributions.

Nevertheless, the log-normal distribution differs from the normal distribution in several ways. A major difference is in its shape: where the normal distribution is symmetrical, a lognormal one is not. Because the values in a lognormal distribution are positive, they create a right skewed curve (Figure 1).

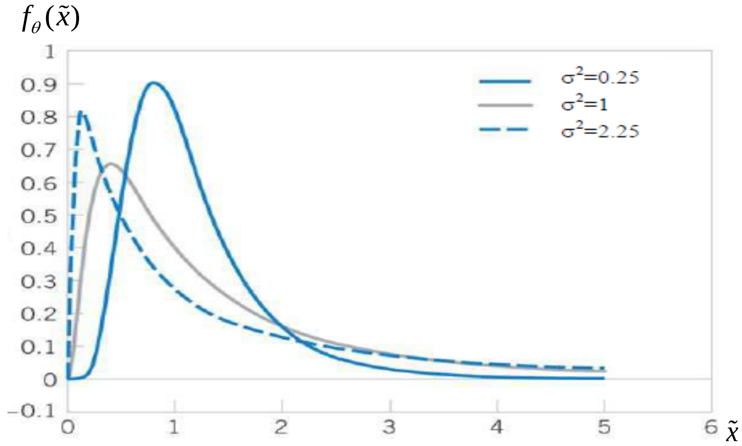

Figure 1. *Log-normal probability density functions with* μ*=0 for selected values of* σ^2.

The log-normal distribution has played major roles in diverse areas of science. Royston [13] modeled survival time in cancer with an emphasis on prognostic factors using the log-normal distribution. Log-normal distributions gave appropriate description of the overall service times and the service times of administrative, e-mail, miscellaneous and network jobs.

Finally, log-normal distributions are self-replicating under multiplication and division, i.e., products and quotients of log-normal random variables are themselves log-normal distributions (Crow and Shimizu [14]; Aitchison and Brown [15]), a result often exploited in back-of-the-envelope calculations.

A positive random variable $\tilde{X}$ is said to be log-normally distributed with two parameters μ and σ^2 if $X = \ln \tilde{X}$ is normally distributed with mean μ and variance σ^2. The two-parameter log-normal distribution is denoted by $\Lambda(\mu,\sigma^2)$; the corresponding normal distribution is denoted by $N(\mu,\sigma^2)$. The probability density function (pdf) of $\tilde{X}$ having $\Lambda(\mu,\sigma^2)$ is

$$f_\theta(\tilde{x}) = \frac{1}{\tilde{x}\sigma\sqrt{2\pi}} \exp\left(-\frac{[\ln\tilde{x}-\mu]^2}{2\sigma^2}\right), \quad \tilde{x}>0, \quad -\infty<\mu<\infty, \quad \sigma>0, \tag{10}$$

where θ=(μ,σ^2). The cumulative distribution function (cdf)) of $\tilde{X}$ is given by

$$F_\theta(\tilde{x}) = \Pr(Z \le \tilde{x}) = \Phi\left(\frac{\ln\tilde{x}-\mu}{\sigma}\right). \tag{11}$$

It follows from (10) that

$$X \sim f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty, \tag{12}$$

that is, $X = \ln \tilde{X} \sim N(\mu,\sigma^2)$, where $\theta = (\mu,\sigma^2)$, $-\infty < \mu < \infty$ is the location parameter and $\sigma > 0$ is the scale parameter. The cdf of the normal distribution is given by

$$F_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) dx. \tag{13}$$

It is known (Nechval and Vasermanis [16]) that the complete sufficient statistic for the parametric vector θ, based on observations in a random sample (X_1, ..., X_n) of size n from the normal distribution (13) is given by

$$S = \left(\bar{X} = \sum_{i=1}^{n} X_i \,/\, n,\; S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \,/\, (n-1) \right). \tag{14}$$

Here the following theorem takes place.

Theorem 2. Let (X_1, ..., X_n) be a preliminary random sample from the normal distribution (13) , where it is assumed that the parametric vector $\theta = (\mu, \sigma^2)$ is unknown. Then the joint probability density function of the pivotal quantities,

$$V_1 = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}, \quad V_2 = \frac{(n-1)S_1^2}{\sigma^2}, \tag{15}$$

is given by

$$f(v) = f_1(v_1) f_2(v_2), \tag{16}$$

where

$$V = (V_1, V_2), \tag{17}$$

$$f_1(v_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v_1^2}{2} \right), \quad -\infty < v_1 < \infty, \tag{18}$$

$$f_2(v_2) = \frac{1}{2^{(n-1)/2}\Gamma\big((n-1)/2\big)} v_2^{(n-1)/2-1} \exp(-v_2/2), \quad v_2 \ge 0. \tag{19}$$

Proof. The joint density of X_1, ..., X_n is given by

$$f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{r} f_\theta(x_i) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2 \right)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 \right). \tag{20}$$

Using the invariant embedding technique (Nechval et al. [17], [18], [19]), we transform (20) to

$$f_\theta(x_1, \ldots, x_n) d\bar{x} ds_1^2 = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu)^2 \right) d\bar{x} ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} [(x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2] \right) d\bar{x} ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i-\bar{x})^2 + 2(\bar{x}-\mu) \sum_{i=1}^{n} (x_i-\bar{x}) + n(\bar{x}-\mu)^2 \right] \right) d\bar{x} ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2\right]\right) d\bar{x} ds_1^2$$

$$= n^{-1/2}(2\pi)^{-1/2} \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right) d\left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right)$$

$$\times (\pi)^{-(n-1)/2}(n-1)^{-(n-1)/2}(s_1^2)^{-(n-1)/2}\left(\frac{(n-1)s_1^2}{2\sigma^2}\right)^{(n-1)/2-1} \exp\left(-\frac{(n-1)s_1^2}{2\sigma^2}\right) d\left(\frac{(n-1)s_1^2}{2\sigma^2}\right)$$

$$\propto (2\pi)^{-1/2} \exp\left(-\frac{v_1^2}{2}\right) dv_1 \left(\frac{v_2}{2}\right)^{(n-1)/2-1} \exp\left(-\frac{v_2}{2}\right) d\left(\frac{v_2}{2}\right). \tag{21}$$

Normalizing (21), we obtain (16). This ends the proof.

Thus,

$$V_1 \sim N(0,1), \quad V_2 \sim \chi^2_{n-1}, \tag{22}$$

where V_2 is statistically independent of V_1.

Theorem 3. If V_1 is a normally distributed random variable with unit variance and zero mean, and V_2 is a chi-squared distributed random variable with n-1 degrees of freedom that is statistically independent of V_1, then

$$T = \frac{V_1+\Delta}{\sqrt{V_2/(n-1)}} = \frac{V_1+\Delta}{\sqrt{W}} \sim f_{n-1,\Delta}(t), \quad -\infty < t < \infty, \tag{23}$$

is a non-central t-distributed random variable with n-1 degrees of freedom and non-centrality parameter Δ, where

$$W = \frac{V_2}{n-1} = \frac{S_1^2}{\sigma^2} \sim f_{n-1}(w) = \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2}\Gamma\left((n-1)/2\right)} w^{(n-1)/2-1} \exp(-(n-1)w/2), \quad w \ge 0, \tag{24}$$

$$f_{n-1,\Delta}(t) = \frac{(n-1)^{(n-1)/2}}{\sqrt{\pi}\Gamma\left((n-1)/2\right)2^{n/2}} \frac{\exp\left(-\frac{(n-1)\Delta^2}{2(t^2+n-1)}\right)}{(t^2+n-1)^{n/2}}$$

$$\times \int_0^\infty w_\bullet^{n/2-1} \exp\left(-\frac{1}{2}\left[w_\bullet^{1/2} - \frac{t\Delta}{\sqrt{t^2+n-1}}\right]^2\right) dw_\bullet, \quad -\infty < t < \infty, \tag{25}$$

is the probability density function of T,

$$W_\bullet = W(t^2+n-1), \tag{26}$$

$$F_{n-1,\Delta}(t) = \Pr(T \le t) = \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2}\Gamma\left((n-1)/2\right)} \int_0^\infty w^{(n-1)/2-1} \exp(-(n-1)w/2)\Phi(t\sqrt{w}-\Delta) dw \tag{27}$$

is the cumulative distribution function of T. $\Phi(x)$ is the standard normal distribution function. Note that the non-centrality parameter Δ may be negative.

Proof. It follows from (23) that

$$\Pr(T \le t \mid W = w) = \Pr\left(\frac{V_1 + \Delta}{\sqrt{w}} \le t \,\middle|\, w \right) = \Pr\left(V_1 \le t\sqrt{w} - \Delta\right) = \Phi(t\sqrt{w} - \Delta). \tag{28}$$

Since it follows from (24) and (28) that

$$\Pr(T \le t) = E\left\{\Pr(T \le t \mid W)\right\} = \int_0^\infty \Phi(t\sqrt{w} - \Delta)\, f_{n-1}(w)dw, \tag{29}$$

we get the cumulative distribution function $F_{n-1,\Delta}(t)$ of the non-central t-distribution given in (27). It is easy to show that the probability density function of T defined in (25) is given by

$$f_{n-1,\Delta}(t) = F'_{n-1,\Delta}(t). \tag{30}$$

This completes the proof.

3. Tolerance Limits on Order Statistic

3.1. Lower Tolerance Limit

Theorem 4. Let $X_1, \ldots, X_n$ be observations from a preliminary sample of size n from a normal distribution defined by the probability density function (12). Then a lower one-sided β-content tolerance limit at a confidence level γ, $L_k \equiv L_k(S)$ (on the kth order statistic Y_k, $k \in \{1, \ldots, m\}$, from a set of m future ordered observations $Y_1 \le \ldots \le Y_m$ also from the distribution (12)), which satisfies

$$\Pr\left(P_\theta(Y_k > L_k) \ge \beta \right) = \gamma, \tag{31}$$

is given by

$$L_k = \bar{X} + \eta_L S_1, \tag{32}$$

where

$$\eta_L = -t_{r,\Delta;\gamma} \big/ \sqrt{n}, \tag{33}$$

is the lower tolerance factor, $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central t-distribution with $r=n-1$ degrees of freedom and non-centrality parameter $\Delta = -z_{1-\delta_\beta}\sqrt{n}$, $z_{1-\delta_\beta}$ denotes the $1-\delta_\beta$ quantile of the standard normal distribution,

$$\delta_\beta = (m-k+1)q_{2(m-k+1),2k;\beta} \big/ [(m-k+1)q_{2(m-k+1),2k;\beta} + k], \tag{34}$$

$q_{2(m-k+1),2k;\beta}$ is the quantile of order β for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. It follows from (8), (13) and (31) that

$$\Pr\left(P_\theta(Y_k > L_k) \ge \beta \right) = \Pr\left(\int_0^{\frac{1-F_\theta(L_k)}{F_\theta(L_k)} \frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \ge \beta \right)$$

$$= \Pr\left(\frac{1-F_\theta(L_k)}{F_\theta(L_k)}\,\frac{2k}{2(m-k+1)} \geq q_{2(m-k+1),2k;\beta}\right) = \Pr\left(F_\theta(L_k) \leq \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;\beta}}\right)$$

$$= \Pr\left(\frac{1}{\sigma\sqrt{2\pi}}\int_{L_k}^{\infty}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy \geq \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k}\right)$$

$$= \Pr\left(\frac{1}{\sqrt{2\pi}}\int_{\frac{L_k-\mu}{\sigma}}^{\infty}\exp\left(-\frac{z^2}{2}\right)dz \geq \delta_\beta\right) = \Pr\left(\frac{1}{\sqrt{2\pi}}\int_{\infty}^{\frac{L_k-\mu}{\sigma}}\exp\left(-\frac{z^2}{2}\right)dz \leq 1-\delta_\beta\right)$$

$$= \Pr\left(\frac{L_k-\mu}{\sigma} \leq z_{1-\delta_\beta}\right) = \Pr\left(\frac{L_k-\bar{X}+\bar{X}-\mu}{\sigma} \leq z_{1-\delta_\beta}\right)$$

$$= \Pr\left(\frac{L_k-\bar{X}}{S_1}\sqrt{n}\,\frac{S_1}{\sigma} + \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq z_{1-\delta_\beta}\sqrt{n}\right) = \Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n}\right)$$

$$= \Pr\left(\frac{V_1 - z_{1-\delta_\beta}\sqrt{n}}{\sqrt{W}} \leq -\eta_L\sqrt{n}\right) = \Pr\left(\frac{V_1+\Delta}{\sqrt{W}} \leq -\eta_L\sqrt{n}\right) = \Pr\left(T \leq -\eta_L\sqrt{n}\right) = F_{r,\Delta}(t), \tag{35}$$

where

$$\eta_L = (L_k - \bar{X})\big/S_1, \tag{36}$$

is the lower tolerance factor,

$$\Delta = -z_{1-\delta_\beta}\sqrt{n}, \quad r = n-1, \quad t = -\eta_L\sqrt{n}. \tag{37}$$

It follows from (31), (35) and (37) that the lower tolerance factor η_L should be chosen such that

$$F_{r,\Delta}(t) = F_{r,\Delta}(-\eta_L\sqrt{n}) = F_{r,\Delta}(t_{r,\Delta;\gamma}) = \gamma, \tag{38}$$

where $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central *t*-distribution with *r* degrees of freedom and non-centrality parameter Δ. It follows from (38) that

$$\eta_L = -t_{r,\Delta;\gamma}\big/\sqrt{n}. \tag{39}$$

It follows from (36) that $L_k = \bar{X} + \eta_L S_1$. This completes the proof.

Corollary 4.1. It follows from (35) that $\Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n}\right)$ can be transformed as follows:

$$\Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n}\right) = \Pr\left(V_1 \leq -\eta_L\sqrt{n}\sqrt{W} + z_{1-\delta_\beta}\sqrt{n}\right)$$

$$= \int_{-\infty}^{-\eta_L\sqrt{n}\sqrt{W}+z_{1-\delta_\beta}\sqrt{n}} f_1(v_1)dv_1 = \Phi(-\eta_L\sqrt{n}\sqrt{W} + z_{1-\delta_\beta}\sqrt{n}) = \Phi(t\sqrt{W} - \Delta), \tag{40}$$

where

$$t=-\eta_L\sqrt{n},\quad \Delta=-z_{1-\delta_\beta}\sqrt{n}. \tag{41}$$

Then it follows from (31) and (40) that t has to be found such that

$$t=\arg\left(E\left\{\Phi(t\sqrt{W}-\Delta)\right\}=\gamma\right)=\arg\left(\int_0^\infty \Phi(t\sqrt{w}-\Delta)\,f_r(w)dw=\gamma\right)$$

$$=\arg\left(\frac{r^{r/2}}{2^{r/2}\Gamma(r/2)}\int_0^\infty w^{r/2-1}\exp(-rw/2)\Phi(t\sqrt{w}-\Delta)\,dw=\gamma\right)=\arg\left(F_{r,\Delta}(t)=\gamma\right)=t_{r,\Delta;\gamma}, \tag{42}$$

where $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central t-distribution with r=n–1 degrees of freedom and non-centrality parameter Δ,

$$F_{r,\Delta}(t)=\Pr(T\le t)=\frac{r^{r/2}}{2^{r/2}\Gamma(r/2)}\int_0^\infty w^{r/2-1}\exp(-rw/2)\Phi(t\sqrt{w}-\Delta)dw. \tag{43}$$

is the cumulative distribution function of T,

$$f_{r,\Delta}(t)=F'_{r,\Delta}(t)=\frac{r^{r/2}\exp\left(-r\Delta^2/[2(t^2+r)]\right)}{\sqrt{\pi}\Gamma(r/2)2^{(r+1)/2}(t^2+r)^{(r+1)/2}}$$

$$\times\int_0^\infty w_\bullet^{(r-1)/2}\exp\left(-\frac{1}{2}\left[w_\bullet^{1/2}-\frac{t\Delta}{\sqrt{t^2+r}}\right]^2\right)dw_\bullet,\quad -\infty<t<\infty, \tag{44}$$

is the probability density function of T, where

$$W_\bullet=W(t^2+r). \tag{45}$$

Corollary 4.2. If

$$W_{\bullet\bullet}=W(t^2+r)/2, \tag{46}$$

then

$$f_{r,\Delta}(t)=F'_{r,\Delta}(t)=\frac{r^{r/2}\exp(-\Delta^2/2)}{\sqrt{\pi}\Gamma(r/2)(t^2+r)^{(r+1)/2}}\int_0^\infty w_{\bullet\bullet}^{(r-1)/2}\exp\left(-\left[w_{\bullet\bullet}-\frac{t\Delta\sqrt{2}}{\sqrt{t^2+r}}w_{\bullet\bullet}^{1/2}\right]\right)dw_{\bullet\bullet}$$

$$=\frac{r^{r/2}\exp(-\Delta^2/2)}{\sqrt{\pi}\Gamma(r/2)(t^2+r)^{(r+1)/2}}\sum_{j=0}^\infty\frac{\Gamma((r+j+1)/2)}{j!}\left(\frac{t\Delta\sqrt{2}}{\sqrt{t^2+r}}\right)^j,\quad -\infty<t<\infty. \tag{47}$$

This form of the density function is derived in Rao [20] and appears in Searle [21]. In both Rao and Searle, $\sqrt{\pi}$ is incorrectly omitted from the denominator. It should also be noted that the central t-distribution is just a special case of the non-central t with Δ = 0.

Corollary 4.3. If k=m=1, then

$$\delta_\beta=\beta,\quad \Delta=-z_{1-\beta}\sqrt{n}. \tag{48}$$

Corollary 4.4. Let $\tilde{X}_1$ ≤ … ≤ $\tilde{X}_n$ be ordered observations from a preliminary sample of size n from a log-normal distribution defined by the probability density function (10). Then a lower one-sided β-content tolerance limit at confidence level γ, $\tilde{L}_k\equiv\tilde{L}_k(S)$ (on the kth order statistic $\tilde{Y}_k$,

$k\in\{1, \ldots, m\}$, from a set of m future ordered observations $\tilde{Y}_1 \le \ldots \le \tilde{Y}_k$ also from the distribution (10)), which satisfies

$$\Pr\left\{P_\theta(\tilde{Y}_k > \tilde{L}_k) \ge \beta\right\} = \gamma, \tag{49}$$

is given by

$$\tilde{L}_k = \exp(L_k) = \exp\left(\bar{X} + \eta_L S_1\right), \tag{50}$$

where

$$X_i = \ln \tilde{X}_i,\ i \in \{1, \ldots, n\},\quad \bar{X} = \sum_{i=1}^{n} X_i / n,\ S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1),$$

$$\Delta = -z_{1-\delta_\beta}\sqrt{n},\quad \delta_\beta = (m-k+1)q_{2(m-k+1),2k;\beta} \big/ [(m-k+1)q_{2(m-k+1),2k;\beta} + k],$$

$$t_{r,\Delta;\gamma} = \arg\left[F_{r,\Delta}(t) = \gamma\right],\quad r = n-1,\quad \eta_L = -t_{r,\Delta;\gamma} \big/ \sqrt{n}. \tag{51}$$

3.2. Upper Tolerance Limit

Theorem 5. Let $X_1, \ldots, X_n$ be observations from a preliminary sample of size n from a normal distribution defined by the probability density function (12). Then an upper one-sided β-content tolerance limit at a confidence level γ, $U_k \equiv U_k(S)$ (on the kth order statistic Y_k from a set of m future ordered observations $Y_1 \le \ldots \le Y_m$ also from the distribution (12)), which satisfies

$$\Pr\left\{P_\theta(Y_k \le U_k) \ge \beta\right\} = \gamma, \tag{52}$$

is given by

$$U_k = \bar{X} + \eta_U S_1, \tag{53}$$

where

$$\eta_U = t_{r,\Delta;1-\gamma} \big/ \sqrt{n}, \tag{54}$$

is the upper tolerance factor, $t_{r,\Delta;1-\gamma}$ is the quantile of order $1-\gamma$ for the non-central t-distribution with $r=n-1$ degrees of freedom and non-centrality parameter $\Delta = -z_{1-\delta_{1-\beta}}\sqrt{n}$, $z_{1-\delta_{1-\beta}}$ denotes the $1-\delta_{1-\beta}$ quantile of the standard normal distribution,

$$\delta_{1-\beta} = (m-k+1)q_{2(m-k+1),2k;1-\beta} \big/ [(m-k+1)q_{2(m-k+1),2k;1-\beta} + k], \tag{55}$$

$q_{2(m-k+1),2k;1-\beta}$ is the quantile of order $1-\beta$ for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. It follows from (3), (13) and (52) that

$$\Pr\left\{P_\theta(Y_k \le U_k) \ge \beta\right\} = \Pr\left(\int_{\frac{1-F_\theta(U_k)}{F_\theta(U_k)} \frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx \ge \beta \right)$$

$$= \Pr\left(\int_0^{\frac{1-F_\theta(U_k)}{F_\theta(U_k)} \frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \le 1-\beta \right) = \Pr\left(\frac{1-F_\theta(U_k)}{F_\theta(U_k)} \frac{2k}{2(m-k+1)} \le q_{2(m-k+1),2k;1-\beta} \right)$$

$$= \Pr\left(F_\theta(U_k) \ge \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;1-\beta}} \right)$$

$$= \Pr\left(\frac{1}{\sigma\sqrt{2\pi}} \int_{U_k}^{\infty} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} \right) dy \le \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta}+k} \right)$$

$$= \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\frac{U_k-\mu}{\sigma}}^{\infty} \exp\left(-\frac{z^2}{2} \right) dz \le \delta_{1-\beta} \right) = \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\infty}^{\frac{U_k-\mu}{\sigma}} \exp\left(-\frac{z^2}{2} \right) dz \ge 1-\delta_{1-\beta} \right)$$

$$= \Pr\left(\frac{U_k-\mu}{\sigma} \ge z_{1-\delta_{1-\beta}} \right) = \Pr\left(\frac{U_k-\bar{X}+\bar{X}-\mu}{\sigma} \ge z_{1-\delta_{1-\beta}} \right)$$

$$= \Pr\left(\frac{U_k-\bar{X}}{S_1}\sqrt{n}\frac{S_1}{\sigma} + \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \ge z_{1-\delta_{1-\beta}}\sqrt{n} \right) = \Pr\left(\eta_U\sqrt{n}\sqrt{W} + V_1 \ge z_{1-\delta_{1-\beta}}\sqrt{n} \right)$$

$$= \Pr\left(\frac{V_1 - z_{1-\delta_{1-\beta}}\sqrt{n}}{\sqrt{W}} \ge -\eta_U\sqrt{n} \right) = \Pr\left(\frac{V_1+\Delta}{\sqrt{W}} \ge -\eta_U\sqrt{n} \right) = \Pr\left(T \ge -\eta_U\sqrt{n} \right) = 1-F_{r,\Delta}(t), \quad (56)$$

where

$$\eta_U = (U_k - \bar{X}) \big/ S_1, \quad (57)$$

is the upper tolerance factor,

$$\Delta = -z_{1-\delta_{1-\beta}}\sqrt{n}, \quad r = n-1, \quad t = -\eta_U\sqrt{n}. \quad (58)$$

It follows from (49), (56) and (58) that the upper tolerance factor η_U should be chosen such that

$$F_{r,\Delta}(t) = F_{r,\Delta}(-\eta_U\sqrt{n}) = F_{r,\Delta}(t_{r,\Delta;1-\gamma}) = 1-\gamma, \quad (59)$$

where $t_{r,\Delta;1-\gamma}$ is the quantile of order $1-\gamma$ for the non-central *t*-distribution with *r* degrees of freedom and non-centrality parameter Δ. It follows from (59) that

$$\eta_U = -t_{r,\Delta;1-\gamma} \big/ \sqrt{n}. \quad (60)$$

It follows from (57) that $U_k = \bar{X} + \eta_U S_1$. This completes the proof.

Corollary 5.1. Let $\tilde{X}_1, \ldots \le \tilde{X}_n$ be observations from a preliminary sample of size *n* from a log-normal distribution defined by the probability density function (10). Then an upper one-sided β-content tolerance limit at confidence level γ, $\tilde{U}_k \equiv \tilde{U}_k(S)$ (on the *k*th order statistic $\tilde{Y}_k$, $k\in\{1, \ldots, m\}$, from a set of *m* future ordered observations $\tilde{Y}_1 \le \ldots \le \tilde{Y}_k$ also from the distribution (10)), which satisfies

$$\Pr \; P_\theta(\tilde{Y}_k \le \tilde{U}_k) \ge \beta \; = \gamma, \tag{61}$$

is given by

$$\tilde{U}_k = \exp(U_k) = \exp(\bar{X} + \eta_U S_1), \tag{62}$$

where

$$X_i = \ln \tilde{X}_i, \; i \in \{1, \ldots, n\}, \quad \bar{X} = \sum_{i=1}^{n} X_i / n, \; S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1),$$

$$\Delta = -z_{1-\delta_{1-\beta}} \sqrt{n}, \quad \delta_{1-\beta} = (m-k+1) q_{2(m-k+1),2k;1-\beta} \Big/ [(m-k+1) q_{2(m-k+1),2k;1-\beta} + k],$$

$$t_{r,\Delta;1-\gamma} = \arg\left[F_{r,\Delta}(t) = 1-\gamma\right]. \quad r = n-1, \quad \eta_U = -t_{r,\Delta;1-\gamma} \Big/ \sqrt{n}. \tag{63}$$

Remark 1. It will be noted that an upper tolerance limit may be obtained from a lower tolerance limit by replacing β by $1-\beta$, γ by $1-\gamma$.

4. Practical Example

A manufacturer of semiconductor lasers has the data on lifetimes (in terms of hours) obtained from testing n=10 semiconductor lasers. These data are given in Table 1.

Table 1. *The data on lifetimes obtained from testing n=10 semiconductor lasers*

Observations (in terms of hours)									
$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$	$\tilde{X}_1$
18657	18960	19771	21015	21183	21960	22881	24642	25373	27373

A buyer tells the laser manufacturer that he wants to place two orders for the same type of semiconductor lasers to be shipped to two different destinations. The buyer wants to select a random sample of m=5 semiconductor lasers from each shipment to be tested. An order is accepted only if all of 5 semiconductor lasers in each selected sample meet the warranty lifetime (in terms of hours). What warranty lifetime (in terms of hours) should the manufacturer offer so that all of 5 semiconductor lasers in each selected sample meet the warranty with probability of 0.95?

In order to find this warranty lifetime, the manufacturer wishes to use a random sample of size n=10 given in Table 1 and to calculate the lower one-sided simultaneous tolerance limit $L_{k=1}(S)$ (warranty lifetime) which is expected to capture a certain proportion, say, β=0.95 or more of the population of selected items (m=5), with the given confidence level γ=0.95. This tolerance limit is such that one can say with a certain confidence γ that at least 100β % of the semiconductor lasers in each sample selected by the buyer for testing will operate longer than $L_1(S)$.

Goodness-of-fit testing. It is assumed that the data of Table 1 follow the log-normal probability distribution (10), where the parameters μ and σ are unknown. Thus, for the above example, we have that n =10, m =5, k = 1, β = 0.95, γ = 0.95,

$$S = \left(\bar{X} = \sum_{i=1}^{n} X_i / n = 10, \; S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1) = 0.016302 \right). \tag{64}$$

We assess the statistical significance of departures from the model (10) by performing the Anderson–Darling goodness-of-fit test. The Anderson–Darling test statistic value is determined by

$$A^2 = -\left[\sum_{i=1}^{n}(2i-1)\left(\ln F_\theta(x_i) + \ln\left(1 - F_\theta(x_{n+1-i})\right)\right)\right]\Big/ n - n, \quad (65)$$

where $F_\theta(\,)$ is the cumulative distribution function of $X = \ln \tilde{X}$,

$$\theta = (\mu = \bar{x}, \sigma = s_1), \quad (66)$$

n is the number of observations.

The result from (65) needs to be modified for small sampling values. For the normal distribution the modification of A^2 is

$$A^2_{\text{mod}} = A^2(1 + 0.75/n + 2.25/n^2). \quad (67)$$

The A^2_{mod} value must then be compared with critical values, A^2_α, which depend on the significance level α and the distribution type. As an example, for the normal distribution the determined A^2_{mod} value has to be less than the following critical values for acceptance of goodness-of-fit (see Table 2):

Table 2. *Critical values for* A^2_{mod}

α	0.1	0.05	0.025	0.01
A^2_α	0.631	0.752	0.873	1.035

For this example, α=0.05, $A^2_{\alpha=0.05} = 0.752$,

$$A^2 = -\left[\sum_{i=1}^{10}(2i-1)\left(\ln F_\theta(x_i) + \ln\left(1 - F_\theta(x_{n+1-i})\right)\right)\right]\Big/ 10 - 10 = 0.193174, \quad (68)$$

$$A^2_{\text{mod}} = A^2(1 + 0.75/10 + 2.25/10^2) = 0.212 < A^2_{\alpha=0.05} = 0.752. \quad (69)$$

Thus, there is not evidence to rule out the log-normal model (10).

Finding lower tolerance limit (warranty lifetime for semiconductor laser). Now the lower one-sided simultaneous β-content tolerance limit at the confidence level γ, $L_1 \equiv L_1(S)$ (on the order statistic Y_1 from a set of m = 5 future ordered observations $Y_1 \le \ldots \le Y_m$) can be obtained from (50).

Since m=5, k=1, β=0.95, it follows from (51) that:

$$\delta_\beta = (m-k+1)q_{2(m-k+1),2k;\beta}\Big/[(m-k+1)q_{2(m-k+1),2k;\beta} + k] = 0.989796, \quad (70)$$

$$r = n - 1 = 9, \quad \Delta = -z_{1-\delta_\beta}\sqrt{n} = 7.3325, \quad \gamma = 0.95, \quad (71)$$

the quantile of order γ for the non-central t-distribution with r degrees of freedom and non-centrality parameter Δ is given by

$$t_{r,\Delta;\gamma} = \arg\left(F_{r,\Delta}(t) = \gamma\right) = 12.5512, \quad (72)$$

the lower tolerance factor is given by

$$\eta_L = -t_{r,\Delta;\gamma}\Big/\sqrt{n} = -3.969. \quad (73)$$

Now it follows from (50), (64) and (73) that

$$L_{k=1} = \exp\left(\bar{X} + \eta_L S_1\right) = 13270. \tag{74}$$

Statistical inference. Thus, the manufacturer has 95% assurance that at least 100β % of the semiconductor lasers in each sample (m=5) selected by the buyer for testing will operate (in terms of hours) no less than L_1=13270 hours.

5. Conclusion

This paper introduces a methodology to construct the one-sided tolerance limits on order statistics in future samples coming from location-scale distributions under parametric uncertainty. For illustration, the normal and log-normal distributions are considered. These distributions play a vital role in many applied problems of biology, economics, engineering, financial risk management, genetics, hydrology, mechanics, medicine, number theory, statistics, physics, psychology, reliability, etc., and have been extensively studied, both from theoretical and applications point of view, by many researchers, since its inception.

It will be noted that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations. The concept proposed in this paper can be extended to two-sided tolerance limits too.

References

[1] Mendenhall, V. (1958). A bibliography on life testing and related topics. *Biometrika*, XLV: 521-543.

[2] Guttman, I. (1957). On the power of optimum tolerance regions when sampling from normal distributions. *Annals of Mathematical Statistics*, XXVIII: 773-778.

[3] Wald, A. and Wolfowitz, J. (1946). Tolerance limits for a normal distribution. *Annals of Mathematical Statistics*, XVII: 208-215.

[4] Wallis, W. A. (1951). Tolerance intervals for linear regression. In: *Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, pp. 43-51.

[5] Patel, J. K. (1986). Tolerance limits: a review. *Communications in Statistics: Theory and Methodology*, 15: 2719-2762.

[6] Dunsmore, I. R. (1978). Some approximations for tolerance factors for the two parameter exponential distribution. *Technometrics*, 20: 317-318.

[7] Guenther, W. C., Patil, S. A., and Uppuluri, V. R. R. (1976). One-sided β-content tolerance factors for the two parameter exponential distribution. *Technometrics*, 18: 333-340.

[8] Engelhardt, M. and Bain, L. J. (1978). Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution. *Technometrics*, 20: 37-39.

[9] Guenther, W. C. (1972). Tolerance intervals for univariate distributions. *Naval Research Logistics Quarterly*, 19: 309-333.

[10] Hahn, G. J. and Meeker, W. Q. Statistical Intervals: A Guide for Practitioners, John Wiley, 1991.

[11] Nechval, N. A. and Nechval, K. N. (2016). Tolerance limits on order statistics in future samples coming from the two-parameter exponential distribution. *American Journal of Theoretical and Applied Statistics*, 5: 1-6.

[12] Meeker, W.Q. and Escobar, L. A. Statistical Methods for Reliability Data, John Wiley,

1998.

[13] Royston, P. (2001). The log-normal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Statistica Neerlandica*, 55: 89-104.

[14] Crow, E.L. and Shimizu, K. Lognormal Distributions, Theory and Applications, Marcel Dekker, 1998.

[15] Aitchison, J. and Brown, J.A.C. The Lognormal Distribution, Cambridge University Press, 1957.

[16] Nechval, N. A. and Vasermanis, E. K. Improved Decisions in Statistics, Izglitibas Soli, 2004.

[17] Nechval, N. A., Nechval, K. N., and Vasermanis, E. K. (2003). Effective state estimation of stochastic systems. *Kybernetes (An International Journal of Systems & Cybernetics)*, 32: 666-678.

[18] Nechval, N. A., Berzins, G., Purgailis, M., and Nechval, K. N (2008). Improved estimation of state of stochastic systems via invariant embedding technique. *WSEAS Transactions on Mathematics*, 7: 141-159.

[19] Nechval, N. A., Purgailis, M., Nechval, K. N., and Strelchonok, V. F. (2012). Optimal predictive inferences for future order statistics via a specific loss function. *IAENG International Journal of Applied Mathematics*, 42: 40-51.

[20] Rao, C. R. Linear Statistical Inference and its Applications, Wiley, 1965.

[21] Searle, S. R. Linear Models, Wiley, 1971.