Tolerance Intervals for Truncated Distributions

Florian Knicker, Christian E. Reuss, Christian Weckbacher, and Patric R. Stracke
Sanofi, MSAT Devices, QE Data Science
florian.knicker@sanofi.com, christian.reuss@sanofi.com, christian.weckbacher@sanofi.com, patric.stracke@sanofi.com

Abstract

Statistical tolerance intervals are widely used as an acceptance criterion in various applications. For example, for needle-based injection systems ISO 11608-1 specifies that a given proportion of the underlying base population must fall within specification limits at a certain confidence level. Accordingly, ISO 16269-6 describes procedures for the estimation and use of tolerance intervals. However, in practice the distributions of many features are subject to truncation due to some known physical or technical limits. When measuring metrical features the truncation threshold may become close to the observed values in testing. In such cases the resulting tolerance limits are prone to be imprecise or even mathematically be outside of the truncation limits . We submit that tolerance intervals may still be applied if the truncation of the underlying distribution can be appropriately modeled and accounted for. In this document we describe an approach for adequately handling tolerance intervals in the presence of truncation.

Introduction

In statistics, a truncated distribution is a conditional distribution that results from restricting the domain of some other probability distribution. Truncated distributions arise in practical statistics in cases, where for some reason, the range of values is limited to values which lie above or below a given threshold or within a specified range. Truncated distributions are not to be confused with censored distributions (see figure 1)

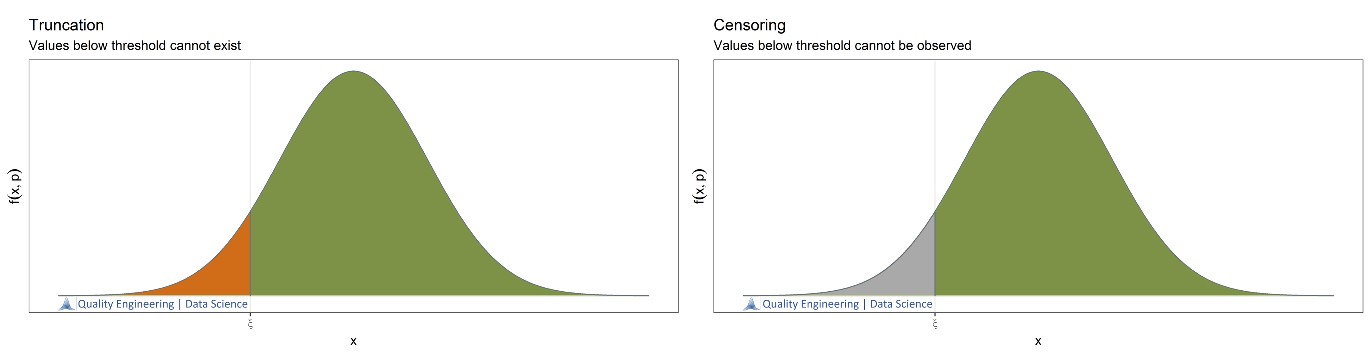

Figure 1: Truncated Density vs. Censored Density

Statistical tolerance intervals $I = [T_L, T_U]$ are defined to cover at least a probability content pc of a given random variable X with a confidence level of $1-\alpha$. The according mathematical expression is

$$P_p\left(P_X(T_L \le X \le T_U \mid p) \ge pc\right) = 1-\alpha,$$

where $p \in \mathbb{R}^n$ is the parameter vector of the distribution of X.

Technical Description

In testing of needle-based injection systems several device-related features are evaluated, which are subject to truncation. For example the expelled volume of a pen device is always non-negative. Thus, the dose accuracy distribution is left-truncated at threshold $\xi_l = 0$. In practice the truncation of features often can be neglected as actually considered values are sufficiently far away from the according truncation threshold, such that the influence on derived statistics is not significant. However in some cases a significance may arise.

ISO 11608-1 defines that in verification testing a statistical tolerance interval I shall be calculated according to [2] and compared to the specification range S, where truncation due to physical limits may be treated as one limit of the specification range. The acceptance criterion is that $I \subseteq S$ has to be fulfilled. A plot of the density of the normal distribution with T_L, T_U, L, and U is diagrammatically shown in figure 2. In that case we have $T_U \le U \Rightarrow I \subseteq S$. Thus the acceptance criterion is met.

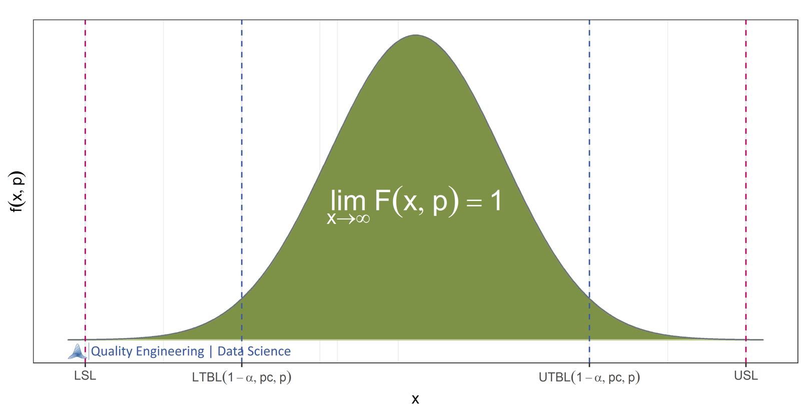

Figure 2: Plot of the density of a given distribution with upper/lower tolerance band limit (T_U, T_L) and upper/lower specification limit (U, L)

In certain cases it may occur that one tolerance band limit may exceed the according physical limit, though it is known that the physical limit can never be exceeded. In that case that violation is to be rated as being practically not existent. However, there may be an impact on the other side of the tolerance band limit and should be treated adequately.

Approach

The case of a left-truncation is considered, see figure 3. I.e., it exists a $\xi_l \in \mathbb{R}$ such that a probability content of $F(\xi_l, p)$, the proportion left to the truncation, does practically not exist.

The upper tolerance band limit indicates that a certain probability content lies within the according tolerance interval. In case the proportion $F(\xi_l, p)$ of practically not relevant content is significant, the upper tolerance band limit may need some transformation to reflect that truncation.

The required probability content pc may need to be transformed ($\widehat{pc} = T_L(pc)$), as diagrammatically shown in figure 3.

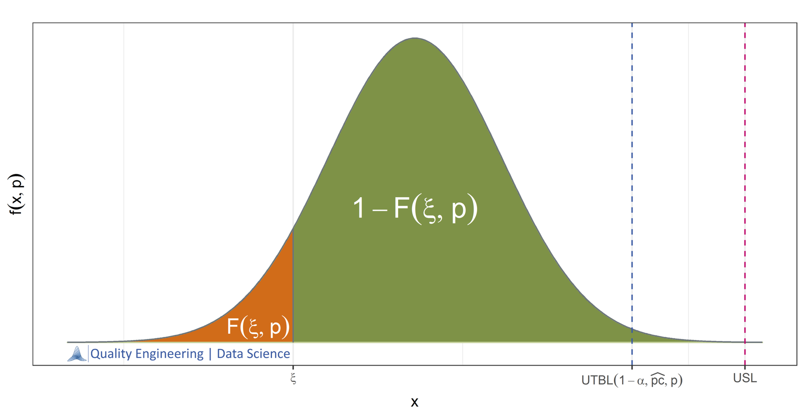

Figure 3: Plot of the density of a given distribution with left truncation at ξ. The upper tolerance band limit (T_U) is transformed according to transformed probability content $\widehat{pc}$ and upper specification limit (U)

In the first step we calculate the actual probability content which is covered by the tolerance interval, i.e. corrected by subtraction of the according proportion due to the truncation

$$pc_{\text{act}} = \frac{pc - F(\xi_l, p)}{1 - F(\xi_l, p)}.$$

In the second step the difference between the required and the actual probability content has to be added to the required probability content to derive the transformed probability content.

$$\widehat{pc} = T_L(pc) = pc + (pc - pc_{\text{act}}) = 2pc - \frac{pc - F(\xi_l, p)}{1 - F(\xi_l, p)} = \left(2 - \frac{1}{1 - F(\xi_l, p)}\right) pc + \frac{F(\xi_l, p)}{1 - F(\xi_l, p)} \quad (1)$$

Thus, a transformation has been derived, such that a tolerance interval of the truncated distribution (random variable X_t) can be calculated by evaluating the underlying distribution without truncation (random variable X):

$$P_p\left(P_X(\xi \le X_t \le T_U \mid p) \ge pc\right) = 1-\alpha \;\Leftrightarrow\; P_p\left(P_X(X \le T_U \mid p) \ge \widehat{pc}\right) = 1-\alpha.$$

Derivation of the formula for right truncation is a simple term conversion and directly analogous to the case of left truncation.

Example

A force feature of a pen device is tested and the values are known to be normally distributed. Due to physical reasons, it is known that no negative forces can occur. I.e., there is a left truncation at $\xi_l = 0$ N. Specification limits are defined as $L = 0$ N and $U = 10$ N.

Table 1: Descriptive statistics and paramters for tolerance band calculation (Standard vs. Truncated Approach)

n	$\bar{x}$	s	pc	$\widehat{pc}$	k	$\widehat{k}$	T_U in N	$\widehat{T_U}$ in N	U in N
60	3.225 N	1.622 N	0.975	0.976	2.670	2.396	7.6	7.1	10

The acceptance criterion for the test is that at least a probability content $pc = 97.5\,\%$ of the values lies within the specification limits at a confidence level of $1-\alpha = 95\,\%$. Descriptive statistics and parameters of the test are given in table 1. The lower tolerance band limit is calculated as $T_L = \bar{x} - ks = -1.11$ N, when the truncation is not considered. In order to consider the truncation of the distribution correctly the approach derived above may be used. The given pc is transformed according to equation 1:

$$\widehat{pc} = \left(2 - \frac{1}{1 - 0.02337}\right) 0.975 + \frac{0.02337}{1 - 0.02337} = 0.97560$$

To consider the truncation a transformed k value (for one-sided limits) is calculated as $\widehat{k}(\widehat{pc}, \alpha, n) = 2.396$. This results in a transformed upper tolerance band limit $\widehat{T_U} = 7.11$ N $< U$. Thus, the acceptance criterion is still met.

Simulation

A simulation study has been carried out with random numbers drawn from normal distribution $\mathcal{N}(2, 2.25)$. A sample size of $n = 50,000$ observations has been used. Truncation threshold has been chosen as $\xi_l = 0$. A tolerance interval with a confidence level of $1-\alpha = 95\%$ and probability content of $pc = 97.5\%$. It is then checked how many observations are contained in the interval $[0, T_U(1-\alpha, \widehat{pc}, 2, 2.25)]$. The simulation has been repeated $1,000$ times with varying seed for the random number generator. An overview of simulation results can be found in table 2. It can be seen that the mean covered proportion is slightly above the desired probability content. This is due to the fact that the upper tolerance band limit is determined as an upper confidence limit of the pc-quantile in order to contain **at least** the desired proportion. A further feature of the tolerance interval simulation is that the desired probability content should be covered with a confidence of 95%. In the simulation this is confirmed as the 5% quantile of the simulated 97.5% quantiles has a relative error of $8 \cdot 10^{-7}$ compared to the desired value of 97.5%. Thus, the simulation shows the correctness of the presented approach.

Table 2: Overview of proportion covered in simulation. Requirement: At least 97.5% shall be covered at a confidence level of 95%

Sample Size	Repetitions	Min	Mean	Max	5% Quantile
50,000	1,000	0.9741	0.9760	0.9787	0.9750008

Conclusions

A transformation of the required probability content in case of truncated distribution was derived.

$$\widehat{pc} = \begin{cases} \left(2 - \frac{1}{1 - F(\xi_l, p)}\right) pc + \frac{F(\xi_l, p)}{1 - F(\xi_l, p)} & \text{in case of left truncation} \\ \left(2 - \frac{1}{F(\xi_r, p)}\right) pc + \frac{1 - F(\xi_r, p)}{F(\xi_r, p)} & \text{in case of right truncation} \end{cases}$$

This transformation enables precise calculation of tolerance intervals in case of truncation. In particular it solves the issue of tolerance limits lying outside of truncation limits.

References

[1] ISO 11608-1. *Needle-based injection systems for medical use - Requirements and test methods. Part 1: Needle-based injection systems, 2014.*
[2] ISO 16269-6. *Statistical Interpretation of data - Part 6: Determination of statistical tolerance intervals, 2014.*
[3] G.D. Faulkenberry and D.L. Weeks. "Sample Size Determination for Tolerance Limits". In: *Technometrics* 10 (2 1968).
[4] T. Mathew K. Krishnamoorthy. *Statistical Tolerance Regions.* 1st ed. Wiley, 2009. ISBN: 978-0-470-38026-0.
[5] D.S. Young. "tolerance: An R Package for Estimating Tolerance Intervals". In: *Journal of Statistical Software* 36 (5 2010).