

# Nonparametric Method to Estimate Tolerance Interval of Continuous Data of Unknown Distribution

Chang Chen, Yi Tsong<sup>\*</sup> and Meiyu Shen

Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Evaluation

<sup>\*</sup>**Corresponding Author:** Yi Tsong, Ph.D. Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, United States, E-mail: Yi.Tsong@fda.hhs.gov

**Received Date:** February 01, 2024 **Accepted Date:** March 01, 2024 **Published Date:** March 04, 2024

**Citation:** Chang Chen, Yi Tsong, Meiyu Shen (2024) Nonparametric Method to Estimate Tolerance Interval of Continuous Data of Unknown Distribution. J Biotechnol Biol 2: 1-12

## Abstract

[1] proposed a precision criterion for the sample size requirement for the estimation of two one-sided tolerance intervals for quality specification of normally distributed attribute data. However, often time the continuous quality data may be distributed with a skewed unknown distribution, multimodal distribution or truncated normal distribution. In this report, we adapt Wilks' proposal of using order statistics to determine the limit of the one-sided tolerance interval [2]. Based on Wilks' method, we can determine the minimum sample size requirements for one-sided and two one-sided tolerance intervals with a targeted order statistic as the limit(s). However, the limits determined with Wilks' method often lead to a tolerance interval with coverage less than the prespecified percentage  $p$  when the sample size is small and larger than  $p$  when the sample size is large. Therefore, we adapt the modification methods based on interpolation of order statistics to improve the precision of the estimated tolerance limits. Furthermore, for sample size cannot be determined using Wilks' method, we propose an interpolation of order statistics to determine the tolerance limits. A simulation study is conducted to illustrate the potential improvement.

**Keywords:** Nonparametric Method; Tolerance Interval; Attribute Data; Multimodal Distribution; Statistics; Tolerance Limits

## Introduction

Conventionally, the product quality specification and control chart limits are determined as the mean plus and minus 3 standard deviations with the assumption that the quality data is normally distributed. These limits are corresponding to the interval centered at mean with coverage of 97.3% of the distribution. Practically, it is determined by sample mean plus and minus 3 sample standard deviations. Such an interval is not a confidence interval of the statistical interval that covers 97.3% of the population. Statistically, we need to take consideration of the estimation error of such an interval. The statistical intervals that covers a fixed proportion of the population with a given confidence level is called tolerance interval. It has been proposed to use a two one-sided tolerance intervals approach for a drug product quality specification determination [3,4]. In order to avoid overestimating the tolerance interval when the coverage  $p$  is large, [1] proposed a precision criterion for the sample size requirement for the estimation of two one-sided tolerance intervals with 80% to 99% coverage as the quality specification. For a given confidence level,  $1 - \alpha$  (e.g. 95%) and coverage percentage  $p$ , the tolerance interval may lead to coverage much larger than  $p$  when the sample size is small [5]. In order to derive a precise tolerance interval, [6] proposed a "goodness" criterion for sample size determination.

However, often time the continuous quality data may be distributed with a skewed unknown distribution, a multimodal distribution or a truncated normal distribution. In this report, we adapt Wilks' proposal of using order statistics to determine the limit of the one-sided tolerance interval [2]. Based on Wilks' method, we can determine the minimum sample size requirements for one-sided and two one-

sided tolerance intervals with a targeted order statistic as the limit(s). The limits determined with Wilks' method often leads to a tolerance interval with coverage less than  $P$  when the sample size is small and larger than  $P$  when the sample size is large. In the literature, to overcome such weakness, method of interpolation of order statistics was proposed to improve the precision. In this paper, we discuss the application of Wilks' method and some improvement methods based on the interpolation of order statistics for the determination of product quality specification. We also propose an extension of the improved method to the tolerance interval estimation for sample sizes that are not derived directly using Wilk's method.

This paper is organized as follows. We state [2] approach and its extension to one-sided and two one-sided tolerance intervals in Section II. The sample size determination method and tables of minimum sample size requirement are presented in Section III. The method of using interpolation of order statistics to improve the precision will be discussed in Section IV. For sample size not listed in the tables, we propose an interpolation extension from the tables. It will also be presented in Section IV. A simulation study is conducted to demonstrate the improvement of interpolation methods to the Wilks' (1941) approach. The summary and conclusion will be given in Section V.

### WILKS (1941) APPROACH (From Chapter 8 of Statistical Tolerance Regions: Theory and Application by K. Krishmoorthy and Thomas Mathew)

Let  $X_1, X_2, \dots, X_n$  be a sample from a population with a continuous distribution  $F_X(x)$ . Let  $X_{(i)}$  denote the  $i$ th smallest of  $X_1, X_2, \dots, X_n$ . Then

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}$$

are the order statistics for the sample. Let  $Y = F_X(X)$ .

Then  $Y \sim \text{Uniform}(0, 1)$ . Consider the empirical distribution

$$\hat{F}(x) = \frac{\text{number of } X_i' \text{ s } \leq x}{n}, \quad (1)$$

the number of observations  $X'_{i \leq a}$  given value  $x$  is  $n \hat{F}_{n(x)} \sim \text{binomial}(n, F_X(x))$ . The probability density function (pdf) of  $X(i)$  is

$$f_{X(i)}(x) = \frac{n!}{(i-1)!(n-i)!} [F_X(x)]^{i-1} [1 - F_X(x)]^{n-i} f_X(x) \quad (2)$$

where  $f_X(x)$  is the pdf of  $X$ .

In order to construct a nonparametric  $(p, 1-\alpha)$  low-

er tolerance interval limit, we need to find the positive integer  $k$  so that

$$P_{X(k)} [P_X (X \geq X_{(k)} | X_{(k)}) \geq p] = 1 - \alpha. \quad (3)$$

The probability on the left hand side of (3) can be expressed as

$$P_{X(k)} [1 - F(X_{(k)}) \geq p] = P_{X(k)} [F(X_{(k)}) \leq 1 - p] = P(U_{(k)} \leq 1 - p),$$

with

$$U_{(k)} = F(X_{(k)}) \sim \text{beta}(k, n - k + 1),$$

where beta (a, b) is beta distribution with parameters a and b. [7] has shown that

$$P(U_{(k)} \leq 1 - p) = 1 - P(U_{(k)} \geq 1 - p) = 1 - P(Y \geq k - 1 | n, 1 - p)$$

$$P(Y \geq k | n, 1 - p) = P(n - Y \leq n - k | n, 1 - p) = P(W \leq n - k | n, p) \geq 1 - \alpha, \quad (4)$$

where  $y \sim \text{binomial}(n, 1-p)$ , and  $W = n - Y \sim \text{binomial}(n, p)$ . Thus,  $X(k)$  is the desired  $(p, 1-\alpha)$  lower tolerance limit if  $k$  is the largest integer that satisfies (4).

To construct a  $(p, 1-\alpha)$  upper tolerance limit, we need to find the largest positive integer  $m$  to satisfy

$$P_{X(m)} [P_X (X \leq X_{(m)} | X_{(m)}) \geq p] \geq 1 - \alpha. \quad (5)$$

If we construct the lower limit  $X_{(k)}$  and upper limit  $X_{(m)}$  of one-sided tolerance intervals derived from  $(p, 1-\alpha/2)$ , the two one-sided tolerance intervals is then defined as  $(X_{(k)}, X_{(m)})$ . With the same  $n, p$  and  $\alpha$ , it is easy to see that  $m = n - k + 1$ .

For any fixed sample size  $n$ , there may not exist order statistics that satisfy the requirements of one-sided tolerance limit. Let us consider sample size  $n$  requirement for estimation of  $(p, 1-\alpha)$  two-sided tolerance intervals in the form of  $(X_{(k)}, X_{(m)})$ , for any given  $k$ . That is, to solve for the smallest  $n$  for the following equation,

### Sample Size Determination for Estimation of One-Sided and Two One-Sided Tolerance Intervals

$$P_{X(k)}, \{P_X [X \geq X_{(k)} | X_{(k)}] \geq p\} \geq 1 - \alpha. \quad (6)$$

It can be simplified to

$$P(X \leq n - k) \geq 1 - \alpha$$

where  $X \sim \text{Binomial}(n, p)$ . It is held if and only if

$$(n - k) p^n - np^{n-k} + 1 \geq 1 - \alpha. \quad (7)$$

The sample size requirement for the two one-sided tolerance interval,  $(X_{(k)}, X_{(n-k+1)})$ , can be determined based

on the formula above. For  $\alpha=0.05$  and a given value of  $k$  and  $p$ , the sample size is the minimum  $N$  such that

$$P(X \leq N - k) \geq 1 - \alpha,$$

where  $X \sim \text{Binomial}(n, (1+p)/2)$ . The  $(p, 1-\alpha)$  two

one-sided non-parametric tolerance interval,  $(X_{(k)}, X_{(n-k+1)})$ , has the following property

$$P_{X_{(k)}} [P_X(X \leq X_{(k)} | X_{(k)}) \leq (1 - p) / 2] + P_{X_{(n-k+1)}} [P_X(X \geq X_{(n-k+1)} | X_{(n-k+1)}) \leq (1 - p) / 2] \geq 1 - \alpha.$$

We calculate the sample size requirement for  $(X_{(k)}, X_{(n-k+1)})$  to be the two one-sided tolerance interval ( $p$ ,

0.95) for  $p=0.8, 0.9, 0.95$  and  $0.99$ . The results are given in Table 1.

**Table 1:** Nonparametric Two One-sided Tolerance Intervals: Sample Size Requirements so that

$$P_{X_{(k)}, X_{(n-k+1)}} \left\{ \Pr(X \leq X_{(k)}) \leq \frac{1-p}{2} | X \right\} + P_{X_{(k)}, X_{(n-k+1)}} \left\{ \Pr(X \geq X_{(n-k+1)}) \leq \frac{1-p}{2} | X \right\} = 1 - \alpha$$

| $\alpha = 0.05$ |       |       |        |        |
|-----------------|-------|-------|--------|--------|
| Values of k     | p=0.8 | p=0.9 | p=0.95 | p=0.99 |
| 1               | 36    | 72    | 146    | 736    |
| 2               | 54    | 110   | 221    | 1113   |
| 3               | 70    | 142   | 287    | 1443   |
| 4               | 85    | 173   | 348    | 1751   |
| 5               | 100   | 202   | 407    | 2046   |
| 6               | 114   | 230   | 464    | 2331   |
| 7               | 127   | 258   | 519    | 2609   |
| 8               | 141   | 285   | 574    | 2881   |
| 9               | 154   | 312   | 627    | 3149   |
| 10              | 167   | 338   | 680    | 3413   |
| 11              | 180   | 364   | 732    | 3674   |
| 12              | 193   | 390   | 783    | 3933   |
| 13              | 206   | 415   | 834    | 4188   |
| 14              | 218   | 440   | 885    | 4442   |

|    |     |     |      |      |
|----|-----|-----|------|------|
| 15 | 231 | 465 | 935  | 4694 |
| 16 | 243 | 490 | 985  | 4944 |
| 17 | 255 | 515 | 1035 | 5192 |
| 18 | 267 | 540 | 1084 | 5439 |
| 19 | 280 | 564 | 1133 | 5685 |
| 20 | 292 | 589 | 1182 | 5929 |

For one-side tolerance interval  $(p, 0.95)$  for  $p=0.8, 0.9, 0.95$  and  $0.99$  can also be calculated with formula (7). The results are given in Table 2.

**Table 2:** Nonparametric One-sided Tolerance Intervals: Sample Size Requirements so that  $X_{(k)}$  is the Lower Tolerance Limit or  $X_{(n-k+1)}$  is the Upper Tolerance Limit

| $\alpha = 0.05$ |        |        |         |         |
|-----------------|--------|--------|---------|---------|
| Values of k     | p =0.8 | p =0.9 | p =0.95 | p =0.99 |
| 1               | 14     | 29     | 59      | 299     |
| 2               | 22     | 46     | 93      | 473     |
| 3               | 30     | 61     | 124     | 628     |
| 4               | 37     | 76     | 153     | 773     |
| 5               | 44     | 89     | 181     | 913     |
| 6               | 50     | 103    | 208     | 1049    |
| 7               | 57     | 116    | 234     | 1182    |
| 8               | 63     | 129    | 260     | 1312    |
| 9               | 69     | 142    | 286     | 1441    |
| 10              | 76     | 154    | 311     | 1568    |
| 11              | 82     | 167    | 336     | 1693    |
| 12              | 88     | 179    | 361     | 1818    |
| 13              | 94     | 191    | 386     | 1941    |
| 14              | 100    | 203    | 410     | 2064    |
| 15              | 106    | 215    | 434     | 2185    |
| 16              | 112    | 227    | 458     | 2306    |
| 17              | 118    | 239    | 482     | 2426    |
| 18              | 124    | 251    | 506     | 2546    |
| 19              | 129    | 263    | 530     | 2665    |
| 20              | 135    | 275    | 554     | 2784    |

Similarly, we calculate the coverage  $p$  for a given tolerance interval  $(X_{(k)}, X_{(n-k+1)})$  for a given  $k$  and  $n$ . The true

coverage  $p$  (with 95% confidence level) for selected  $k$  and  $n$  are given in Table 3.

**Table 3:** Nonparametric Two One-sided Tolerance Intervals: True Coverage Probability ( $p$ ) for a given Lower Tolerance Limit  $(X_{(k)}, X_{(n-k+1)})$

| Sample Size | Values of K |       |       |       |       |       |       |       |       |       |
|-------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|             | 1           | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
| 20          | 0.663       | 0.502 | 0.366 | 0.242 | 0.126 | 0.017 | NA    | NA    | NA    | NA    |
| 25          | 0.725       | 0.592 | 0.479 | 0.375 | 0.278 | 0.185 | 0.097 | 0.012 | NA    | NA    |
| 30          | 0.768       | 0.655 | 0.558 | 0.469 | 0.385 | 0.305 | 0.228 | 0.154 | 0.082 | 0.012 |
| 35          | 0.799       | 0.701 | 0.616 | 0.538 | 0.465 | 0.394 | 0.327 | 0.261 | 0.197 | 0.134 |
| 40          | 0.823       | 0.736 | 0.661 | 0.592 | 0.526 | 0.463 | 0.403 | 0.344 | 0.287 | 0.230 |
| 45          | 0.842       | 0.764 | 0.697 | 0.634 | 0.575 | 0.518 | 0.464 | 0.41  | 0.358 | 0.308 |
| 50          | 0.857       | 0.787 | 0.725 | 0.669 | 0.615 | 0.563 | 0.513 | 0.465 | 0.417 | 0.371 |
| 50          | 0.857       | 0.787 | 0.725 | 0.669 | 0.615 | 0.563 | 0.513 | 0.465 | 0.417 | 0.371 |
| 60          | 0.880       | 0.821 | 0.769 | 0.721 | 0.676 | 0.632 | 0.589 | 0.548 | 0.508 | 0.468 |
| 70          | 0.897       | 0.845 | 0.801 | 0.759 | 0.720 | 0.682 | 0.645 | 0.609 | 0.574 | 0.539 |
| 80          | 0.909       | 0.864 | 0.825 | 0.788 | 0.753 | 0.720 | 0.687 | 0.655 | 0.624 | 0.594 |
| 90          | 0.919       | 0.879 | 0.844 | 0.811 | 0.780 | 0.750 | 0.721 | 0.692 | 0.664 | 0.637 |
| 100         | 0.927       | 0.891 | 0.859 | 0.829 | 0.801 | 0.774 | 0.747 | 0.722 | 0.696 | 0.672 |
| 110         | 0.934       | 0.900 | 0.871 | 0.844 | 0.819 | 0.794 | 0.770 | 0.746 | 0.723 | 0.700 |
| 120         | 0.939       | 0.908 | 0.882 | 0.857 | 0.833 | 0.810 | 0.788 | 0.767 | 0.745 | 0.724 |
| 130         | 0.944       | 0.915 | 0.891 | 0.868 | 0.846 | 0.825 | 0.804 | 0.784 | 0.764 | 0.745 |
| 140         | 0.947       | 0.921 | 0.898 | 0.877 | 0.856 | 0.837 | 0.818 | 0.799 | 0.781 | 0.762 |
| 150         | 0.951       | 0.926 | 0.905 | 0.885 | 0.866 | 0.847 | 0.829 | 0.812 | 0.795 | 0.778 |
| 160         | 0.954       | 0.931 | 0.911 | 0.892 | 0.874 | 0.857 | 0.84  | 0.823 | 0.807 | 0.791 |
| 170         | 0.957       | 0.935 | 0.916 | 0.898 | 0.881 | 0.865 | 0.849 | 0.834 | 0.818 | 0.803 |
| 180         | 0.959       | 0.938 | 0.920 | 0.904 | 0.888 | 0.872 | 0.857 | 0.843 | 0.828 | 0.814 |
| 190         | 0.961       | 0.942 | 0.924 | 0.909 | 0.893 | 0.879 | 0.865 | 0.851 | 0.837 | 0.823 |
| 200         | 0.963       | 0.944 | 0.928 | 0.913 | 0.899 | 0.885 | 0.871 | 0.858 | 0.845 | 0.832 |

The coverage  $p$  for a given tolerance interval  $(X_{(k)}, X_{(n)})$  or  $(X_{(1)}, X_{(n-k+1)})$  for a given  $k$  and  $n$  can also be calculated. The true coverage  $p$  (with 95% confidence level) for selected  $k$  and  $n$  are given in Table 4.

**Table 4:** Nonparametric One-sided Tolerance Intervals: True Coverage Probability ( $p$ ) for a given Lower Tolerance Limit  $X_{(k)}$  or Upper Tolerance Limit  $X_{(n-k+1)}$

| Sample Size | Values of K |       |       |       |       |       |       |       |       |       |
|-------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|             | 1           | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
| 20          | 0.860       | 0.783 | 0.717 | 0.656 | 0.598 | 0.544 | 0.492 | 0.441 | 0.393 | 0.346 |
| 25          | 0.887       | 0.823 | 0.768 | 0.718 | 0.670 | 0.624 | 0.580 | 0.537 | 0.496 | 0.456 |
| 30          | 0.904       | 0.851 | 0.804 | 0.761 | 0.720 | 0.681 | 0.642 | 0.606 | 0.570 | 0.534 |
| 35          | 0.917       | 0.871 | 0.830 | 0.793 | 0.757 | 0.722 | 0.689 | 0.656 | 0.625 | 0.594 |
| 40          | 0.927       | 0.886 | 0.850 | 0.817 | 0.785 | 0.754 | 0.725 | 0.696 | 0.667 | 0.640 |
| 45          | 0.935       | 0.898 | 0.866 | 0.836 | 0.808 | 0.780 | 0.753 | 0.727 | 0.702 | 0.676 |
| 50          | 0.941       | 0.908 | 0.879 | 0.852 | 0.826 | 0.801 | 0.776 | 0.753 | 0.729 | 0.706 |
| 60          | 0.951       | 0.923 | 0.898 | 0.875 | 0.853 | 0.832 | 0.812 | 0.792 | 0.772 | 0.752 |
| 70          | 0.958       | 0.934 | 0.912 | 0.892 | 0.873 | 0.855 | 0.837 | 0.820 | 0.803 | 0.786 |
| 80          | 0.963       | 0.942 | 0.923 | 0.905 | 0.889 | 0.873 | 0.857 | 0.841 | 0.826 | 0.811 |
| 90          | 0.967       | 0.948 | 0.931 | 0.916 | 0.901 | 0.886 | 0.872 | 0.858 | 0.845 | 0.831 |
| 100         | 0.970       | 0.953 | 0.938 | 0.924 | 0.910 | 0.897 | 0.885 | 0.872 | 0.860 | 0.848 |
| 120         | 0.975       | 0.961 | 0.948 | 0.936 | 0.925 | 0.914 | 0.903 | 0.893 | 0.882 | 0.872 |
| 140         | 0.978       | 0.966 | 0.955 | 0.945 | 0.935 | 0.926 | 0.917 | 0.908 | 0.899 | 0.890 |
| 160         | 0.981       | 0.970 | 0.961 | 0.952 | 0.943 | 0.935 | 0.927 | 0.919 | 0.911 | 0.903 |
| 180         | 0.983       | 0.973 | 0.965 | 0.957 | 0.949 | 0.942 | 0.935 | 0.928 | 0.921 | 0.914 |
| 200         | 0.985       | 0.976 | 0.968 | 0.961 | 0.954 | 0.948 | 0.941 | 0.935 | 0.928 | 0.922 |

When an experimenter wants to determine the lower and upper limits of the specification that covers at least 90% of the product population, for a continuous attribute with unknown distribution, he needs to plan an experiment with at least sample size of 72. On the other hand, for  $k=1$ ,  $(X_{(k)}, X_{(n-k+1)})$  is the range of minimum to maximum (table1). With sample size  $n=35$ , the range determined by (min, max) is 80% of the population centered at median (table3).

### Modified Wilks' (1941) Approach

As shown in Section III, for tolerance interval  $(p, 1-\alpha)$  and any  $n$  satisfies sample size requirement,  $X_{(k)}$  is the lower limit of confidence interval of  $(p, 1-\alpha)$  if  $k$  is the largest integer satisfying the inequality [4]. Therefore, the confidence interval  $(X_{(k)}, X_{(n)})$  would have a confidence level  $1 - \alpha' \geq 1 - \alpha$  and  $(X_{(k-1)}, X_{(n)})$  would be a tolerance interval with confidence level  $1 - \alpha'' < 1 - \alpha$ . A naive modification would be use the average of  $X_{(k)}$  and  $X_{(k+1)}$  to replace  $X_{(k)}$  as the lower limit. It is represented as  $L(X)$ ,

$$L(X) = \frac{X_{(k)} + X_{(k+1)}}{2}.$$

[8-10] investigated interpolation of two order statistics to improve the precision of the estimate. The interpolation can be expressed as follow,

$$L(X) = X_{(k)} + \frac{(1 - \alpha') - (1 - \alpha)}{(1 - \alpha') - (1 - \alpha'')} * (X_{(k+1)} - X_{(k)})$$

[11] presented a detailed discussion of the interpolation methods and proposed an extrapolation for tolerance interval limit based on sample sizes smaller than the minimum required. However, for quality specification, the coverage is typically high and often the range of values are restricted, we don't recommend to use the extrapolation

$$L(X) = X_{(k)} + \frac{n - S_{(k)}}{S_{(k+1)} - S_{(k)}} * (X_{(k+1)} - X_{(k)}) .$$

For example, with sample size = 150, for a tolerance interval of 95% coverage and with 95% confidence lev-

$$L(X) = X_{(1)} + \frac{150 - 146}{188 - 146} * (X_{(2)} - X_{(1)}) ,$$

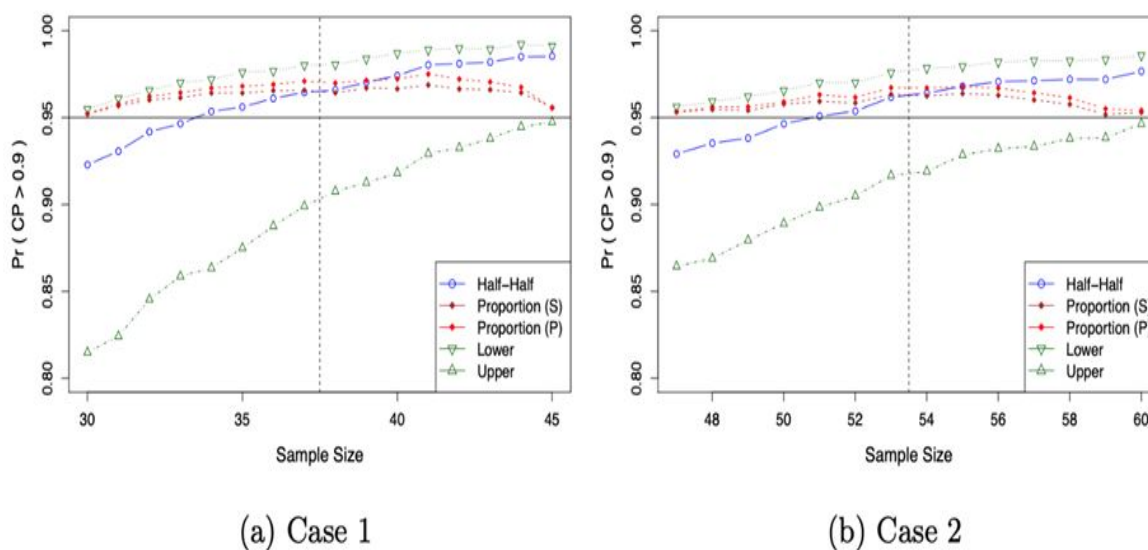
where  $X_{(1)}$  and  $X_{(2)}$  are the smallest and the second

method for quality specification.

We also propose a method to determine the lower limit for any given sample size  $n$  between  $S_{(k)}$ , the sample size required for the  $k$ th order statistics, and  $S_{(k+1)}$ , sample size required for the  $(k + 1)$ th order statistics.

el, the lower limit of the two one-sided tolerance interval should be

smallest observations respectively.



**Figure 1:** Determination of Nonparametric Lower Tolerance Limits when the Underlying Distribution is  $N(1,2)$ : True Probability of Coverage Probability greater than 0.9 depending on Sample Size

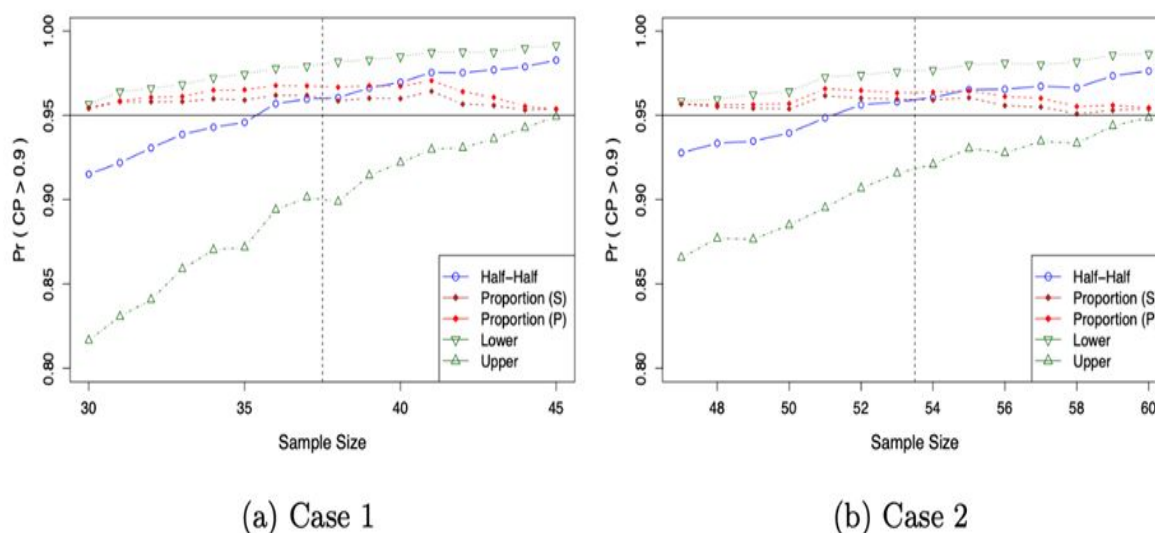
In order to illustrate the improvement of the confidence level of modified Wilks' approach. We compare the three modified approaches or tolerance interval (0.90, 0.95) with Wilks' approach and use the  $X_{(k+1)}$  as the lower limit assuming  $k=1$ . We generated 10,000 random samples from

$N(1,7)$ , normal distribution with mean = 0 and variance =7;  $\text{Exp}(7)$ , exponential distribution with rate parameter  $\lambda =7$ ; and chi-squared distribution  $\chi^2(1)$ , degree =1. We compare the discrepancy between the probability of coverage probability greater than the pre-specified 0.90. The results are giv-

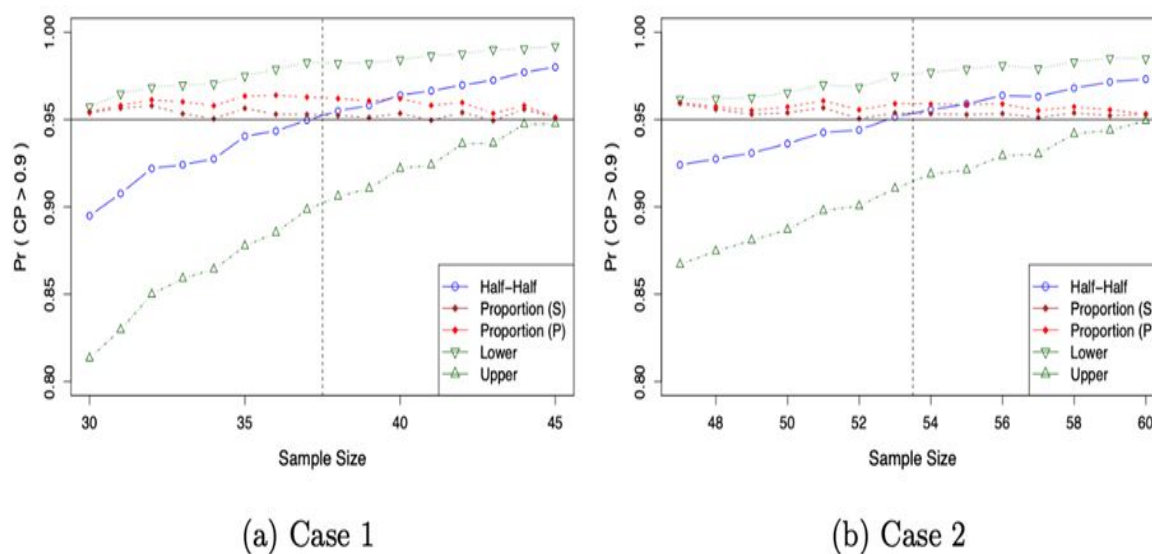


en in Figures 1, 2 and 3. In Figures 1 to 3, Half-Half stands for the average of the two order statistics, Proportion ( $p$ ) stands for the interpolation of two order statistics, Proportion ( $s$ ) stands for the interpolation of order statistics based on the sample sizes. In addition, the determination of lower and upper limits without interpolation are also presented in

those figures, which are labeled by Lower and Upper, which stands for the smaller order statistics and the larger order statistics, respectively. For example, with sample size = 150, for a tolerance interval of 95% coverage and with 95% confidence level, then  $X_{(1)}$  would be chosen as the lower limit under the approach of Lower;  $X_{(2)}$  would be chosen as the lower limit under the approach of Upper.



**Figure 2:** Determination of Nonparametric Lower Tolerance Limits when the Underlying Distribution is  $Exp(0.7)$ : True Probability of Coverage Probability greater than 0.9 depending on Sample Size



**Figure 3:** Determination of Nonparametric Lower Tolerance Limits when the Underlying Distribution is chi-square (1): True Probability of Coverage Probability greater than 0.9 depending on Sample Size

## Discussion and Conclusion

[1] Propose to use the “goodness” criterion proposed by [6] to determine the sample size for two one-sided tolerance intervals for product quality specifications. More specifically, it is proposed to use  $\delta=(1-p)/4$  for the “goodness” criterion. When the data are continuous but following unknown distribution, typically the tolerance interval is estimated using order statistics. We adapt Wilks’ proposal of using order statistics to determine the limit of the one-sided tolerance interval [2]. Based on Wilks’ method, we can determine the minimum sample size requirements for one-sided and two one-sided tolerance intervals with a targeted order statistic as the limit(s). Furthermore, we adapt the interpolation of order statistics approach proposed by [8-11] to improve the precision of the estimate. We also propose to use interpolation method to determine the limits of tolerance interval for sample sizes not presented in the minimum sample size tables. Through a simulation study, we have demonstrated that both methods can improve the pre-

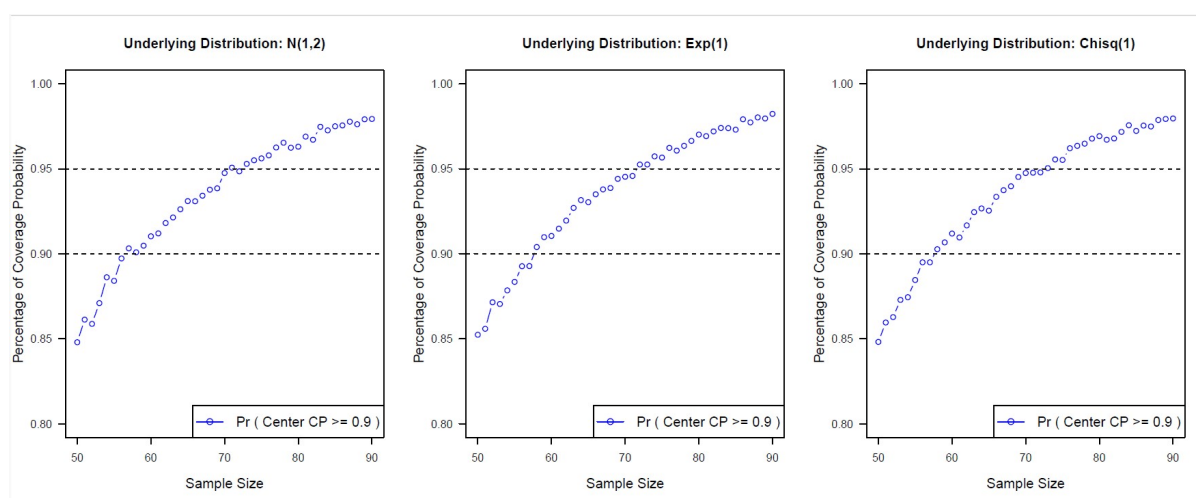
cision of the estimated nonparametric tolerance intervals with more accurate coverages given certain confidence levels.

In the future research, we will use formula (4) to construct of the lower limit of the lower confidence limit of percentile estimation. It leads to the determination of the cut point in immunogenicity studies. It can further be used for testing the difference of percentiles of two distributions and assessing the comparability of biologic products.

## Acknowledgement

The authors want to acknowledge the contributions of Drs. Atiar Rahman, Chao Wang, Yu-Ting Weng, Xiaoyu Cai, Shaobo Liu, Xutong Zhao and Tengfei Li of FDA in participating in the discussion during the development of the procedure.

## Supplementary Material



Simulation results for the  $(P, 1-\alpha)$  nonparametric two one-sided tolerance interval  $(X_{(k)}, X_{(N-k+1)})$  when  $k = 1$ ,  $\alpha = 0.05$ . The figure below presented the center coverage of the tolerance interval for different sample size  $N$  under three types of underlying distributions. The minimum sample size requirement to obtain a center coverage greater than 0.9 is consistent with the results in Table 1 with  $k = 1$  and  $P = 0.9$ .

## References

1. Chen C, Zhao X, Shen M, Tsong Y, (2024) Statistical considerations for using tolerance interval to determine product specification. Submitted to Journal of Biopharmaceutical Statistics.
2. Wilks SS (1941) Determination of sample sizes for setting tolerance limits. *Annals of Mathematics Statistics*, 12: 91-6.
3. Dong X, Tsong Y, Shen M, Zhong J (2015) Using tolerance intervals for assessment of pharmaceutical quality. *J. of Biopharm. Statistics*.
4. Tsong Y, Wang T, Hu X (2019) Statistical considerations in setting quality specification limits using quality data. In *Pharmaceutical Statistics*, Liu R. and Tsong Y (eds.), Springer Proceedings in Mathematics & Statistics 218.
5. Chakraborti S, Li J, (2012) Confidence interval estimation of a normal percentile. *The American Statistician*, 61: 331-6.
6. Faulkenberry GD, Daly JC (1970) Sample size for tolerance limits on a normal distribution. *Tenometrics*, 12: 813-21.
7. Krishnamoorthy K, Mathew T (2009) *Statistical tolerance regions: Theory, Applications and Computation*. John Wiley & Sons, Inc.
8. Hettmansperger TP, Sheather SJ (1986) Confidence intervals based on interpolated order statistics. *Statist. And Probability Letters*, 4: 75-9.
9. Beran R, Hall P, (1993) Interpolated nonparametric prediction intervals and confidence intervals. *J. of Royal Statist. Society. Series B (Methodological)* 55: 643-52.
10. Hutson AD (1999) Calculating nonparametric confidence intervals for quantiles using fractional order statistics. *J. of Applied Statistics* 26: 343-53.
11. Young DS, Mathew T (2014) Improved nonparametric tolerance intervals based on interpolated and extrapolated order statistics. *J. of Nonparametric Statistics* 26: 415-32.
12. Dong X, Tsong Y, Shen M (2015) Statistical considerations in setting product specifications. *J. of Biopharmaceutical Statistics*, 25: 280-94.
13. Frey J (2010) Data-driven nonparametric tolerance sets. *J. of Nonparametric Statistics* 22: 169-80.
14. Nyblom J (1992) Note on interpolated order statistics. *Statist. & Probability Letters* 14: 129-31.
15. Odeh RE, Chou YM, Owen DB, (1987) The precision for coverage and sample size requirements for normal tolerance intervals. *Commun. Statist. – Simula*, 16: 969-85.
16. Stiegler SM (1977) Fractional order statistics , with applications. *J. of American Statistical Association* 72: 544-50.

**Submit your manuscript to a JScholar journal and benefit from:**

- ¶ Convenient online submission
- ¶ Rigorous peer review
- ¶ Immediate publication on acceptance
- ¶ Open access: articles freely available online
- ¶ High visibility within the field
- ¶ Better discount for your subsequent articles

Submit your manuscript at  
<http://www.jscholaronline.org/submit-manuscript.php>