Check for updates

Received: 15 September 2022 | Revised: 31 January 2023 | Accepted: 16 February 2023

DOI: 10.1002/wics.1611

FOCUS ARTICLE

WIREs COMPUTATIONAL STATISTICS WILEY

Tolerance limits for mixture-of-normal distributions with application to COVID-19 data

Hsiuying Wang

Institute of Statistics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Correspondence
Hsiuying Wang, Institute of Statistics, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan.
Email: wang@stat.nycu.edu.tw

Funding information
Ministry of Science and Technology, Grant/Award Number: 109-2118-M-009-005-MY2

Edited by: James Gentle, Commissioning Editor and Co-Editor-in-Chief and David Scott, Review Editor and Co-Editor-in-Chief

Abstract

Tolerance intervals (TIs) are widely used in various applications including manufacturing engineers, clinical research, and pharmaceutical industries. TIs can be used to construct limits of control charts for monitoring quality characteristics. For manufacturing processes where multiple factors may contribute to defects or multiple-stream processes, a mixture distribution of several suitable probabilistic models may be a better choice than a simple distribution for modeling the data. TIs for the normal mixture distribution have been studied in the literature. This article reviews the TIs of the normal mixture distribution, the applications of the mixture distribution, and the control charts of the mixture distribution. A rule for constructing modified two-sided TIs of the normal mixture distribution is summarized, and this rule may be extended to construct modified two-sided TIs for general mixture distributions. The feasibility of using TIs to build control charts for mixture distributions is also discussed. A real data example of coronavirus disease 2019 is used to illustrate the method by linking the TI to control charts.

This article is categorized under:

Statistical Learning and Exploratory Methods of the Data Sciences > Clustering and Classification

KEYWORDS

control chart, mixture distribution, tolerance interval

1 | INTRODUCTION

A tolerance interval (TI) is a method to cover a fixed proportion of the population with a specified confidence level (Meeker et al., 2017). Compared with other types of interval estimation, TIs are not as well-known as confidence intervals, which are widely used in statistical inference to cover population parameters with a specified confidence level. Confidence intervals are a very useful tool for parameter estimation of parametric distributions (Brown et al., 2001; Short, 2021; Wang, 2008) when the estimated target is a point. However, sometimes the characteristic of interest is an enclosure interval rather than a point, in which case TI can be a very useful tool for estimating this specific enclosure interval.

TIs have various applications in pharmaceutical industries and quality control. In pharmaceutical industries, three tests including the dose content uniformity, delivered dose uniformity, and dissolution tests are important tools for the assessment of pharmaceutical quality (Rahman et al., 2021; Tsong et al., 2015; Wang, 2007). The TI approach can be

 © 2023 Wiley Periodicals LLC.

applied to set the decision rules of these tests (Dong et al., 2015). The endpoints of a TI are called tolerance limits. An application of TIs in manufacturing involves comparing customer-specified specification limits with tolerance limits (Little, 2016). The proportion of nonconforming parts of the population is a quantity of interest in manufacturing. A TI-based calculation approach was proposed to investigate the proportion of nonconforming parts (Pusztai & Kemény, 2020).

There are different types of TI under the frequentist and Bayesian frameworks. The methods for constructing TIs and sample size determination of TIs for the normal distribution date back to the 1940s (Wald, 1943; Wald & Wolfowitz, 1946; Wilks, 1941, 1942). The sample size determination for the two-sided TIs controlling both tails of the normal distribution was investigated (Chou & Mee, 1984). The earlier studies on constructing tolerance limits for the parametric continuous and discrete univariate distributions have been reviewed including the normal, exponential, Weibull, generalized gamma, Cauchy, logistic, Poisson, binomial, and negative-binomial (Patel, 1986). The construction of TIs for discrete variables has been critical in industrial applications. The calculation of the exact minimum coverage probabilities of TIs for Poisson and binomial variables was studied (Wang & Tsung, 2009), and the Edgeworth expansion was used to construct desirable TIs for Poisson and binomial variables (Cai & Wang, 2009). The widely-used TIs for common continuous and discrete distributions were introduced in books (Krishnamoorthy & Mathew, 2009; Meeker et al., 2017). The tolerance limits for common distributions can be calculated using R or Matlab software (Young, 2010). This paper reviews TIs for the mixture distribution under the frequentist framework.

The rest of this paper is organized as follows. In Section 2, different types of TI are introduced, and the related studies are reviewed. Section 3 reviews the mixture distributions and their applications. Section 4 reviews the methods for constructing TIs of the normal mixture distribution and summarizes a rule for constructing modified two-sided TIs that may be extended to general mixture distributions. In Section 5, control charts for the mixture distribution are reviewed. The potential of using TIs to build control limits for general mixture distributions is discussed. A real data example using the coronavirus disease 2019 (COVID-19) data to illustrate the procedure of constructing TI for the normal mixture distribution, and the connection of TI to the control chart is provided in Section 6. Finally, Section 7 summarizes the main points of this paper.

2 | TYPE OF TOLERANCE INTERVAL

There are different types of TI. Under the frequentist framework, there are usually two typical types of TIs: β-content TI and β-expectation TI, where β denotes a proportion between 0 and 1 (Krishnamoorthy & Mathew, 2009; Mee, 1984).

2.1 | β-Content **TI**

An interval $(L^*(X), U^*(X))$ is said to be a two-sided β-content, $1-\alpha$ confidence TI, denoted by $(\beta, 1-\alpha)$ TI, for F if

$$P[[F(U^*(X)) - F(L^*(X))] \geq \beta] = 1-\alpha,$$

where X is a random variable with a distribution function F. One-sided tolerance limits can be defined in a similar way. A tolerance limit $L(X)$ is said to be a lower $(\beta, 1-\alpha)$ tolerance limit for F if $P\{[1-F(L(X))] \geq \beta\} = 1-\alpha$, and a tolerance limit $U(X)$ is said to be an upper $(\beta, 1-\alpha)$ tolerance limit for F if $P\{[F(U(X))] \geq \beta\} = 1-\alpha$.

2.2 | β-Expectation **TI**

An interval $(L_E(X), U_E(X))$ is said to be a two-sided β-expectation TI for F if

$$\mathrm{Exp}[F(U_E(\mathrm{X})) - F(L_E(X))] = \beta$$

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In addition to the TIs constructed under the frequentist framework, TIs under the Bayesian framework have been proposed. Let $\pi(\theta)$ denote a prior distribution for θ and $\pi(\theta|\mathrm{x})$ be the posterior distribution of θ. An interval $(L_B(\theta), U_B(\theta))$ is said to be a $(\beta, 1-\alpha)$ Bayesian TI if

$$P_{\theta|X}[[F(U_B(\theta)) - F(L_B(\theta))] \geq \beta] = 1-\alpha$$

(Krishnamoorthy & Mathew, 2009; Young et al., 2016).

Moreover, an equal-tailed $(\beta, 1-\alpha)$ TI $(L'(X), U'(X))$ is widely used in applications that controls the percentages in both tails ensuring that no more than a proportion $(1-\beta)/2$ of the population is below the lower tolerance limit and no more than a proportion $(1-\beta)/2$ of the population is above the upper tolerance limit with a confidence level $1-\alpha$. In this interval, $L'(X)$ and $U'(X)$ can be set to be a lower $((1-\beta)/2, 1-\alpha)$ tolerance limit and an upper $((1+\beta)/2, 1-\alpha)$ tolerance limit, respectively. The equal-tailed $(\beta, 1-\alpha)$, β-expectation, and Bayesian TIs have been discussed for various distributions (Guo et al., 2021; Hoang-Nguyen-Thuy & Krishnamoorthy, 2021; Liu et al., 2021). There may be a need to weigh the two-tail probabilities of a TI when the cost of failing to detect larger (smaller) values is higher than the cost of failing to detect smaller (larger) values. Two-sided TIs with equal or unequal tail probabilities, also known as balanced or unbalanced TIs, could be chosen depending on the objective at hand, and both might have their place in practice (Alqurashi et al., 2023).

In addition to the parametric continuous and discrete univariate distributions, TIs for censored data have been studied. Exact methods for constructing TIs using pivotal quantities under complete or censored data for symmetric distributions in the location-scale family were proposed (Krishnamoorthy & Xie, 2011). The problem of computing two-sided TIs and equal-tailed TIs for a location-scale family of distributions was investigated (Hoang-Nguyen-Thuy & Krishnamoorthy, 2021). Yuan et al. developed a general procedure to compute the exact factors of TIs for both symmetric and nonsymmetric distributions in the log-location-scale family, based on complete or censored data (Yuan et al., 2018). Approximate tolerance limits under the log-location-scale regression models in the presence of censoring were proposed, and a bias-correction technique via the jackknife method was applied to improve small sample accuracy (Emura & Wang, 2010).

TIs for more applications have been established including combing different types of data, record value data, and a mixture of different populations. TIs based on the combination of high-resolution and low-resolution data were proposed (Wang & Tsung, 2017). In many cases, only record values can be observed, and an exact two-sided TI based on record values for the exponential distribution was constructed (Guo et al., 2020). The equal-tailed and shortest Bayesian TIs that could control percentages in both tails of the exponential distribution based on k-record values were built (Kiapour & Qomi, 2017). Simultaneous tolerance limits and TIs for several normal populations with a common unknown variance were proposed (Krishnamoorthy & Chakraberty, 2022). Zoom-in and quasi-independent procedures for estimating TIs of discrete-time and covariance-stationary stochastic processes having a continuous marginal distribution with some autocorrelation property were proposed (Chen & Kelton, 2006).

The normal mixture distribution is more suitable than the normal distribution to fit the data in many applications including audio signal recognition and sports data clustering (Andriyanov, 2020; Wang, 2021). Tolerance limits for the normal mixture distribution have been constructed (Zimmer et al., 2016). Further, the one-sided and two-sided TIs based on the bootstrap and sample quantile methods for the mixture of normal distributions were proposed (Chen & Wang, 2020a). This paper focuses on the review of the β-content TI for the normal mixture distribution.

3 | MIXTURE DISTRIBUTION

Let $X_1, \ldots, X_n$ be a random sample following a k-component mixture distribution F with the probability density function [Equation (1)]

$$f_\theta(x) = \sum_{j=1}^{k} p_j g_j\left(x, \underline{\eta}_j\right), \quad -\infty < x < \infty \tag{1}$$

where $g_j\left(x, \underline{\eta}_j\right)$ denotes a specific distribution with parameter $\underline{\eta}_j$, p_j denotes the weight of the jth component, and $\sum_{j=1}^{k} p_j = 1$. When $g_j\left(x, \underline{\eta}_j\right)$ is the normal distribution, F is the normal mixture distribution, also known as the Gaussian mixture distribution.

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The normal mixture distribution has been successfully used to model various real data. Wireless channels were analyzed and modeled by the mixture of normal distributions (Selim et al., 2016). Brain tumor features were extracted from magnetic resonance imaging using the normal mixture models (Chaddad, 2015). An automatic railway visual detection system for railway surface defects based on a combined filter and improved normal mixture model method could not only improve the detection ability for faint defects, but also reduce the processing time cost (Zhang et al., 2018). In addition to the normal mixture distribution, other mixture distributions are also useful tools for analyzing various data problems. A mixture Gamma distribution was used to model the signal-to-noise ratio of wireless channels and analyze physical layer security problems (Atapattu et al., 2011; Lei et al., 2016). The insurance claims data were modeled by a mixture of exponential distributions (Lee et al., 2012). A Poisson mixture model was proposed to cluster count-based digital gene expression profiles (Rau et al., 2015). A binomial mixture distribution was adopted to model the number of credits gained by university freshmen during the first year (Grilli et al., 2015).

The maximum likelihood estimator (MLE) is usually used to estimate the parameters of the normal mixture distribution, and the expectation–maximization (EM) algorithm is a widely-used method to derive the MLE of the normal mixture distribution (McLachlan & Peel, 2000). However, it suffers from the local maxima problem and the initialization dependence problems. Methods to improve the EM algorithm for mixture distributions have been widely discussed (Scrucca, 2021; Xiang et al., 2020). In many applications, the components of a mixture distribution are skewed or heavy-tailed. The parameter estimation of a mixture of skewed distributions via the EM algorithm was proposed (Castillo-Barnes et al., 2020).

4 | TOLERANCE INTERVAL ON MIXTURE DISTRIBUTIONS

4.1 | Distribution-free tolerance interval

For mixture distributions, it is difficult to derive simple formulas for approximate or exact TIs. The distribution-free TI based on a nonparametric method can be applied in the mixture distributions. Let $X_{(1)},\cdots,X_{(n)}$ denote the order statistics of $X_i, i=1,\cdots,n$. An interval $\left[X_{(r)},X_{(s)}\right]$ is said to be a $(\beta, 1-\alpha)$ distribution-free TI if r and s satisfy $P(Y \leq s-r-1) \geq 1-\alpha$, where Y follows a binomial distribution $\text{Bin}(n,\beta)$ (Krishnamoorthy & Mathew, 2009).

The distribution-free TI does not have a good performance for the normal mixture distribution when the sample size is small. The coverage probability (CP) of the distribution-free TI has been studied (Chen & Wang, 2020a). The CPs of the (0.99, 0.95) one-sided and two-sided distribution-free TIs for several normal mixture models when the sample size is 20, 100, and 200 were tabulated in Table 1. These CPs are much lower than the nominal level of 0.95.

4.2 | One-sided tolerance limit for the mixture normal distribution

As shown in Table 1, the performance of the distribution-free TI is not satisfactory when the sample size is small. Since the distribution-free TI is based on the nonparametric method, TIs constructed by a parametric method can be expected

TABLE 1 Coverage probabilities of the (0.99, 0.95) distribution-free upper tolerance limits and TIs for the normal mixture distribution (Chen & Wang, 2020a).

Models	Sample size	CP of the upper tolerance limit	CP of the two-sided TI
$\frac{1}{3}N(0,1)+\frac{2}{3}N(0.5,1)$	20	0.173	0.021
	100	0.634	0.263
	200	0.864	0.596
$\frac{1}{2}N(0,1.2)+\frac{1}{2}N(4,1.5)$	20	0.183	0.017
	100	0.637	0.267
	200	0.869	0.601
$\frac{1}{4}N(0,1)+\frac{1}{2}N(1,1)+\frac{1}{4}N(2,1)$	20	0.182	0.022
	100	0.633	0.268
	200	0.868	0.597

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to have better performance than the distribution-free TI. A bootstrap-based method was proposed to construct the TI for the normal mixture distribution (Chen & Wang, 2020a). In general, the bootstrap method is preferred over the distribution-free method unless the components of the mixture distribution have a large overlap because the EM algorithm may not precisely converge to the true parameter value. However, the bootstrap-based method relies on heavy calculation. Another method based on the asymptotic distribution of the sample quantile was proposed to construct TIs for the normal mixture distribution (Chen & Wang, 2020a). Let q_r and $\widehat{q}_r$ denote the rth quantile and its MLE of the normal mixture distribution, respectively. The empirical distribution based on a sample $X_1,\cdots,X_n$ is defined as [Equation (2)]

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x) \tag{2}$$

The rth sample quantile of $F_n(x)$ is defined as [Equation (3)]

$$\widetilde{q}_r \equiv F_n^{-1}(r) = \inf\{\mathrm{t}|F_n(t) \ge r\}. \tag{3}$$

An asymptotic distribution of $\widetilde{q}_r$ has the form:

$\sqrt{n}(\widetilde{q}_r - q_r) \to N\left(0, \frac{r(1-r)}{f_\theta(q_r)^2}\right)$

(Serfling, 2009)

The upper and lower $(\beta, 1-\alpha)$ tolerance limits derived by the sample quantile method are [Equations (4) and (5)]

$$U = \widetilde{q}_\beta + z_{1-\alpha}\left(\frac{\beta(1-\beta)}{nf_{\widehat{\theta}}\left(\widehat{q}_\beta\right)^2}\right)^{\frac{1}{2}} \tag{4}$$

and

$$L = \widetilde{q}_{1-\beta} - z_{1-\alpha}\left(\frac{\beta(1-\beta)}{nf_{\widehat{\theta}}\left(\widehat{q}_{1-\beta}\right)^2}\right)^{\frac{1}{2}}, \tag{5}$$

respectively, where $\widehat{\theta}$ is MLE of (1). These tolerance limits based on the sample quantile method are easier to be calculated and have better performance than the bootstrap-based method (Table 2) (Chen & Wang, 2020a).

Tolerance limits for the mixture normal distribution can also be constructed from the perspective of confidence intervals for percentiles (Zimmer et al., 2016). Percentiles are descriptions of quantiles relative to 100. An upper $(\beta, 1-\alpha)$ tolerance limit is a $100(1-\alpha)\%$ upper confidence limit for the $100 \times \beta$th percentile of the distribution. Let $\widehat{\xi}_\beta$ denote the MLE of the $100 \times \beta$th percentile ξ_β of the normal mixture distribution. A $100(1-\alpha)\%$ upper confidence limit for ξ_β was obtained using the asymptotic normality of $\widehat{\xi}_\beta$, where the asymptotic standard error of $\widehat{\xi}_\beta$ was obtained by the delta method (Zimmer et al., 2016). An algorithm for obtaining an upper tolerance limit was provided by Zimmer et al. (2016) (Zimmer et al., 2016). A simulation study showed that the CP of this upper confidence limit for ξ_β is close to the nominal level for some models (Table 3) (Zimmer et al., 2016).

4.3 | Two-sided tolerance interval for the normal mixture distribution

To construct two-sided TIs, it is common that the two-sided interval is built from both the upper and the lower tolerance limits. The upper or lower tolerance limits constructed by the quantile or sample quantile methods in both papers have good performance in terms of the CP (Chen & Wang, 2020a; Zimmer et al., 2016). But the two-sided TI directly constructed by the upper and lower tolerance limits did not have a satisfactory result (Chen & Wang, 2020a). Zimmer et al. (2016) only presented the upper tolerance limit case and did not mention the two-sided TI case. It can be expected

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Coverage probability of the (0.99, 0.95) upper tolerance limits for the mixture normal distribution (Chen & Wang, 2020a).

Model	Sample size	Bootstrap method CP for the upper limit	Sample quantile method CP for the upper limit
$\frac{1}{3}N(0,1)+\frac{2}{3}N(0.5,1)$	20	N	0.701
	100	0.807	0.939
	200	0.854	0.948
	400	0.925	0.951
	1000	0.929	0.956
$\frac{1}{2}N(0,1.2)+\frac{1}{2}N(4,1.5)$	20	0.801	0.759
	100	0.904	0.958
	200	0.913	0.955
	400	0.945	0.959
	1000	0.943	0.957
$\frac{1}{4}N(0,1)+\frac{1}{2}N(1,1)+\frac{1}{4}N(2,1)$	20	N	N
	100	0.766	0.901
	200	0.815	0.918
	400	0.897	0.933
	1000	0.912	0.939

Note: "N" denotes that the EM algorithm does not converge.

TABLE 3 Coverage probabilities of the (0.99, 0.99) upper tolerance limit for the mixture normal distribution (Zimmer et al., 2016).

Model	Sample size	CP of the upper tolerance limit
$\frac{1}{5}N(5,0.5)+\frac{4}{5}N(9,1)$	40	0.9840
	100	0.9880
$\frac{3}{10}N(1100,130)+\frac{7}{10}N(1650,155)$	40	0.972
	100	0.987
$\frac{7}{10}N(1100,130)+\frac{3}{10}N(1650,155)$	40	0.938
	100	0.979

that the two-sided TI case directly implemented using the tolerance limits cannot achieve a good result. A modified method was proposed to construct two-sided TIs and this method could have a significant improvement (Chen & Wang, 2020a).

This modified method uses one of the two one-sided tolerance limits as a tolerance limit first, and then adjusts the other one. For example, the conventional tolerance limits of a $(\beta, 1-\alpha)$ two-sided TI built from the formulas (4) and (5) of one-sided tolerance limits are [Equations (6) and (7)]

$$U^* = \widetilde{q}_{\beta_U} + z_{1-\alpha/2}\left(\frac{\beta_U(1-\beta_U)}{nf_{\widehat{\theta}}\left(\widehat{q}_{\beta_U}\right)^2}\right)^{\frac{1}{2}} \tag{6}$$

and

$$L^* = \widetilde{q}_{\beta_L} - z_{1-\alpha/2}\left(\frac{\beta_L(1-\beta_L)}{nf_{\widehat{\theta}}\left(\widehat{q}_{\beta_L}\right)^2}\right)^{\frac{1}{2}}, \tag{7}$$

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\beta_U = (1+\beta)/2$ and $\beta_L = (1-\beta)/2$. Then we can use L^* as a lower limit of the modified two-sided TI, and modify the upper tolerance limit to be

$$\widetilde{q}_{\beta_U^*} + z_{1-\alpha/2}\left(\frac{\beta_U^*\left(1-\beta_U^*\right)}{nf_{\widehat{\theta}}\left(\widehat{q}_{\beta_U^*}\right)^2}\right)$$

which replaces $\beta_U = (1+\beta)/2$ with $\beta_U^* = F_{\widehat{\theta}}(L^*)+\beta$ in (6). This modification method adjusts the upper limit or lower limit such that the two-sided interval can cover β proportion of the population. By a similar argument, this method can use (6) as an upper tolerance limit and replace $\beta_L = (1-\beta)/2$ by $\beta_L^* = F_{\widehat{\theta}}(U^*)-\beta$ in (7). Although this modified method has only been studied for the normal mixture distribution in the literature, it may be applied to general mixture distributions. Procedures 1 and 2 summarize this modified method. The feasibility of applying this method to general mixture distributions needs to be verified by future studies.

Procedure 1. Construct a level $(\beta, 1-\alpha)$ two-sided TI for the mixture distribution F by adjusting the upper tolerance limit.

Step 1: Find the MLE $\widehat{\theta}$ of the parameter θ.
Step 2: Construct a level $((1-\beta)/2, 1-\alpha)$ lower tolerance limit L_1^*.
Step 3: Let $\beta_1 = F_{\widehat{\theta}}\left(L_1^*\right)+\beta$. If β_1 is greater than 1, then set β_1 to 1.
Step 4: Construct a level $(\beta_1, 1-\alpha/2)$ upper tolerance limit U_1^*.
Step 5: The interval $\left(L_1^*, U_1^*\right)$ is the modified two-sided TI.

Procedure 1 is the method to construct a two-sided TI by first using a level $((1-\beta)/2, 1-\alpha/2)$ lower tolerance limit, and then adjusting the upper tolerance limit. Similarly, this method can first fix a level $((1+\beta)/2, 1-\alpha/2)$ upper tolerance limit, and then adjust the lower tolerance limit as stated in Procedure 2.

Procedure 2. Construct a level $(\beta, 1-\alpha)$ two-sided TI for mixture distribution F by adjusting the lower tolerance limit.

Step 1: Find the MLE $\widehat{\theta}$ of the parameter θ.
Step 2: Construct a level $((1+\beta)/2, 1-\alpha/2)$ upper tolerance limit U_2^*.
Step 3: Let $\beta_2 = F_{\widehat{\theta}}\left(U_2^*\right)-\beta$. If β_2 is less than 0, then set β_2 to 0.
Step 4: Construct a level $(\beta_2, 1-\alpha/2)$ lower tolerance limit L_2^*.
Step 5: The interval $\left(L_2^*, U_2^*\right)$ is the modified two-sided TI.

A more detailed algorithm by applying Procedure 1 with Equations (6) and (7) for the normal mixture distribution is provided in Algorithm 1.

Algorithm 1. Use Equations (6) and (7) to construct a level $(\beta, 1-\alpha)$ two-sided TI for the normal mixture distribution by adjusting the upper tolerance limit.

Step 1: Fit data with a normal mixture distribution. Use the EM algorithm to find the MLE $\widehat{\theta}$ of the parameter θ.
Step 2: Calculate the empirical distribution from the data using Equation (2). Find the $(1-\beta)/2$th sample quantile $\widetilde{q}_{\frac{1-\beta}{2}}$ from this empirical distribution.
Step 3: Find the $(1-\beta)/2$th quantile $\widehat{q}_{\frac{1-\beta}{2}}$ of the fitted normal mixture distribution, and then calculate the value of the density function $f_{\widehat{\theta}}\left(\widehat{q}_{\frac{1-\beta}{2}}\right)$ of this fitted normal mixture distribution at this quantile value $\widehat{q}_{\frac{1-\beta}{2}}$.
Step 4. Using the values calculated in Steps 2 and 3 and Equation (7), a level $((1-\beta)/2, 1-\alpha/2)$ lower tolerance limit L_1^* can be obtained.
Step 5. Let $\beta_1 = F_{\widehat{\theta}}\left(L_1^*\right)+\beta$. If β_1 is greater than 1, then set β_1 to 1.

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Step 6. Use the empirical distribution derived in Step 2 to find the $\beta_1 th$ sample quantile $\tilde{q}_{\beta_1}$. Also, find the β_1th quantile $\hat{q}_{\beta_1}$ of the fitted normal mixture distribution, and then calculate the value of the density function $f_{\hat{\theta}}\left(\hat{q}_{\beta_1}\right)$ of this fitted normal mixture distribution at this quantile value $\hat{q}_{\beta_1}$.
Step 7. Using the values calculated in Step 6 and Equation (6), a level $(\beta_1, 1-\alpha/2)$ upper tolerance limit U_1^* can be obtained.
Step 8. The interval $\left(L_1^*, U_1^*\right)$ is the modified two-sided TI.

A small example is provided to illustrate Algorithm 1.

Example 1. The 12 data 0.7708, 12.9807, 1.3233, 2.9906, 1.7710, 0.0802, 8.1795, 0.8446, 0.6032, −1.0528, 0.2842, and −0.9290 are used to construct a level $(\beta, 1-\alpha) = (0.99, 0.95)$ TI using Algorithm 1. The TI can be obtained by adhering to the procedures outlined in Algorithm 1, as follows.

Step 1. These data are fitted to a normal mixture distribution. The fitted normal mixture distribution is $0.8328 \times N(0.6672, 1.3066^2) + 0.1672 \times N(10.5553, 5.9350^2)$.
Step 2. Calculate $(1-0.99)/2$th sample quantile from the empirical distribution based on the 12 data, which is −1.0528.
Step 3. Calculate the $(1-0.99)/2$th quantile of the fitted normal mixture distribution, which is −2.2020, and calculate the value of the density function of this fitted normal mixture distribution at this quantile value, which is 0.0125.
Step 4. Using Equation (7), a level $((1-\beta)/2, 1-\alpha/2) = (0.005, 0.975)$ lower tolerance limit $L_1^* = -1.4090$ is obtained.
Step 5. Calculate $F_{\hat{\theta}}\left(L_1^*\right)$, which is 0.0289. Then $\beta_1 = 0.0289 + 0.99 = 1.0189$, which is greater than 1. Set β_1 to 1.
Step 6. Find $\beta_1 = 1$st sample quantile (100th percentile) from the empirical distribution, which is 12.9807. Find the $\beta_1 = 1$st quantile of the fitted normal mixture distribution, which is 21.6066, and find the value of the density function of this fitted normal mixture distribution at this quantile value, which is 9.3042e-07.
Step 7. Based on Equation (6), a level $(\beta_1, 1-\alpha/2) = (0.995, 0.975)$ upper tolerance limit $U_1^* = 12.9807$ is obtained. Then (−1.4090, 12.9807) is a modified two-sided level $(0.99, 0.95)$ TI derived by the 12 data.

In addition to equal-tailed TIs, not equal-tailed TIs can be constructed based on these procedures. In Step 2 of Procedure 1, a level $((1-\beta)/2, 1-\alpha/2)$ lower tolerance limit is considered. In some situations, if the TIs are not limited to equal-tailed TIs, then the value $(1-\beta)/2$ can be replaced by another specified value. This not equal-tailed TI method can be applied to Procedure 2. A more general procedure that is an extension of Procedure 1 is summarized in a flowchart in Figure 1.

Other bootstrap methods can be applied to improve the CPs of TIs. Two content-adjusted procedures for TIs based on bootstrap were proposed to improve the CPs of TIs for some non-normal distributions or the normal mixture distribution (Jiao et al., 2022). Furthermore, a researcher might be also interested in exploring TIs for specific subpopulations as well as the entire population when assuming that the population has a normal mixture distribution. Accordingly, individual TIs for subpopulations were investigated based on generalized fiducial inferences (Tsai, 2020).

5 | CONTROL CHART APPLICATION

One of the useful applications of TI is to construct control charts. A simple control chart for monitoring the mean of samples is the $\overline{X}$ chart, and the widely-used control charts for monitoring the standard deviation are R and S charts (Montgomery, 2020). TI control limits were developed for $\overline{X}$, R, and S charts for the normal distribution (Hamada, 2003). Control charts that are based on the assumption of a particular form of parametric distribution such as the normal distribution are called parametric control charts. The Bayesian TI has also been used to construct control charts for parametric distributions such as the exponential distribution (Ali et al., 2016; Demirhan & Hamurkaroglu, 2014). Bayesian TI control limits which have the advantage of controlling the probability content at a specified level with given confidence were proposed (Hamada, 2002). In addition to parametric distributions, in many industrial applications, the

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

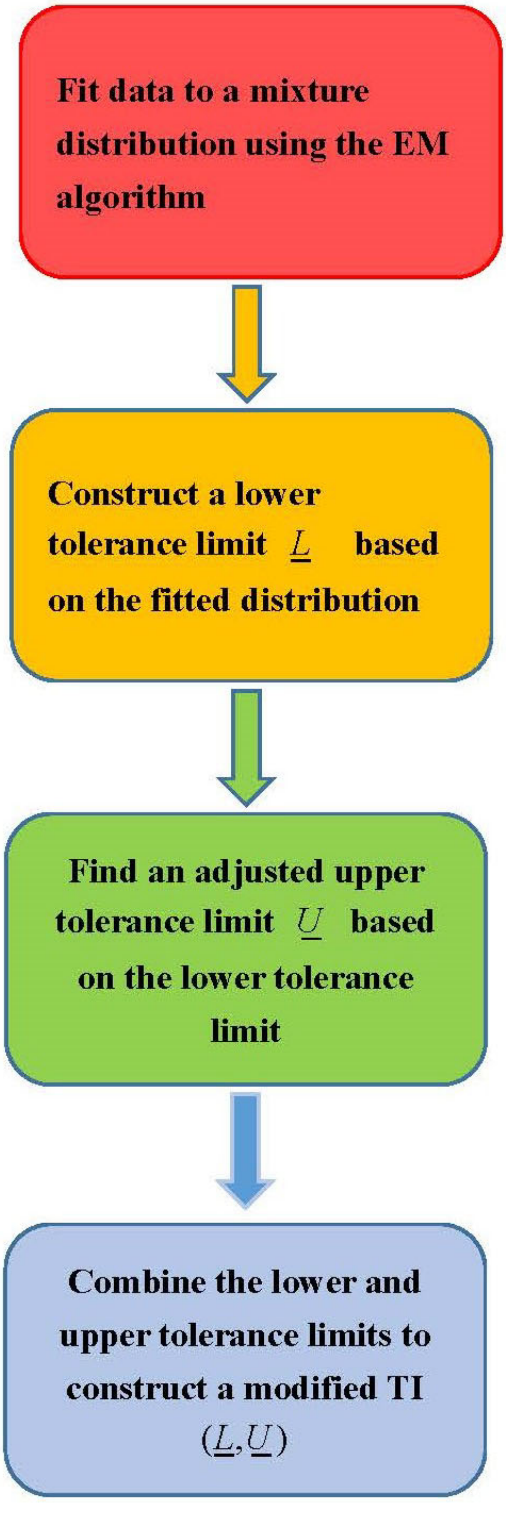

FIGURE 1 Flowchart of the construction of a modified two-sided TI.

distribution assumption is hard to be justified. In this case, nonparametric or distribution-free control charts can be used for a wider variety of purposes (Chakraborti, 2004; Chakraborti & Graham, 2019a). Both univariate and multivariate nonparametric control charts have been reviewed and discussed (Chakraborti & Graham, 2019b).

In manufacturing processes, several factors may cause defects in the final output. That is, the engineering processes may consist of several phenomena leading to defects with multiple causes. In this situation, the mixture models are more suitable to model such types of multiple-cause defect systems instead of the simple probability models. Control

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

charts constructed for the mixture distribution have been proposed in the literature. A control chart for the two-component mixture of inverse Rayleigh distribution was proposed (Ali & Riaz, 2014). Suppose that a population of defective items is divided into two subpopulations. Then the total number of items inspected follows a two-component mixture of geometric distributions. A control chart named the mixture cumulative count control chart was constructed for the distribution function of a two-component mixture of geometric distributions using the number of items inspected until a defective item was observed (Majeed et al., 2013). An industrial multi-stream process was a process in which items of the same type were manufactured in parallel in multiple output streams (Epprecht, 2015). Construction of control limits with a mixture of probability distribution applications in the multiple-stream processes was proposed, and it presented an alternative solution to Shewhart charts in cases where there were different machines and product production flows (Vicentin et al., 2018). Nonparametric control charts including multivariate signed-rank and multivariate sign control charts were applied on a mixture of multivariate normal and t-distributions (Udom et al., 2021). In sensor networks and manufacturing technology, multivariate processes face a new challenge with high-dimensional data. Control charts based on the variable-selection (VS) algorithms have been developed. Since in manufacturing processes, data can have multimodal properties. A VS-based control chart with a Gaussian mixture distribution was proposed to handle the problem of high-dimensionality and multimodality (Yan et al., 2019). The VS-based control chart framework was also suggested to be extended to the process with other multimodal distributions (Yan et al., 2019).

The control charts with the mixture distribution have various applications. The TIs for the mixture distributions in Section 4 can be used to construct limits of control charts. In some situations, the characteristic of interest in the manufactured process might be an observation rather than the mean of several observations. In this case, the tolerance limits can be directly used as control limits. Sometimes only the upper or lower control limits of the characteristic of interest may be required, not both. For example, only an upper control limit is necessary when monitoring the number of defective products. In situations where only one control limit is needed, the upper or lower tolerance limit can be used as the control limit. For the case of requiring both upper and lower control limits, the modified two-sided TIs in Procedures 1 and 2 can be considered as control limits. Since both Procedures 1 and 2 can be considered, we may be interested in which one is better. It is suggested that when the lower control limit is more important, then we can adopt Procedure 1 by fixing the lower tolerance limit and then adjusting the upper tolerance limit, and vice versa.

6 | REAL DATA EXAMPLE

A coronavirus disease 2019 (COVID-19) dataset is used to illustrate the TI method for the mixture normal distribution and the connection to the control chart. The COVID-19 pandemic is a global outbreak of coronavirus since the end of 2019 (Chen & Wang, 2020b; Wang, 2022). To date, more than 600 million confirmed cases of COVID-19 have been reported to the World Health Organization. Monitoring the number of daily new confirmed cases can help predict the peak of the epidemic. The data of daily new COVID-19 confirmed cases were downloaded from the website of Our World in Data https://github.com/owid/covid-19-data/tree/master/public/data on November 29, 2022. The Taiwan data in 2020 and 2021 are used in this study. There are 351 and 365 data recorded for the 2 years of 2020 and 2021, respectively. The date range for 2020 data are from January 16 to December 31, 2020, and the date range for 2021 data are from January 1 to December 31, 2021. In 2020, there is no COVID-19 outbreak in Taiwan. The major outbreak was around April and May 2021. Thus, the data of the 2020 year are used as historical data to construct a control chart for monitoring the 2021 data. To predict the epidemic peak, we only need to set up a control chart with an upper control limit but no lower control limit, because a small number of daily new COVID-19 confirmed cases is irrelevant to the prediction of the epidemic peak.

In this scenario, the characteristic of interest is the number of the daily new confirmed case, which is an observation, but not a parameter of a distribution. Hence, an upper tolerance limit for the distribution of the number of daily new confirmed cases can be directly used as an upper control limit for monitoring the number of daily new confirmed cases. First, the 351 data from 2020 are used to construct an upper tolerance limit. A normal mixture distribution was used to fit the data by the Matlab software. The result showed that a normal mixture distribution with two components could fit the data well. Thus, the fitted normal mixture distribution $0.7570 \times N(0.6767, 0.8002^2) + 0.2430 \times N(7.5284, 42.0552^2)$ is used in this study.

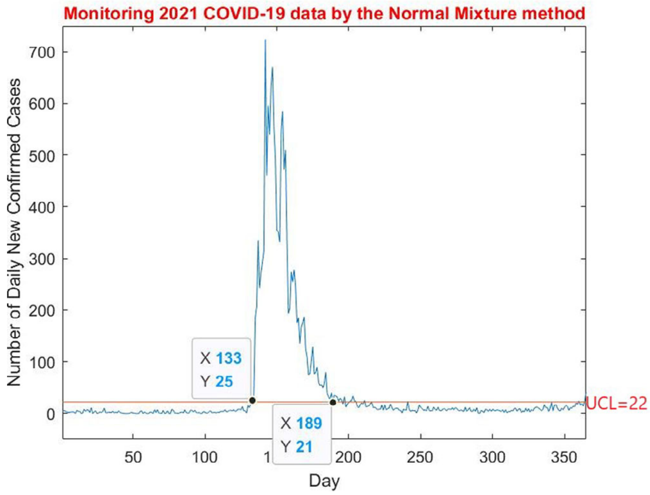

FIGURE 2 2021 data monitoring using an upper control limit constructed by the normal mixture distribution.

To construct a level $1-\alpha=0.95$ and $\beta=0.99$ upper tolerance limit based on Equation (4), it needs to calculate $\tilde{q}_\beta$ and $f_{\hat{\theta}}\left(\hat{q}_\beta\right)$. By calculation $\tilde{q}_\beta=22$, $f_{\hat{\theta}}\left(\hat{q}_{\beta_U}\right)=0.0032$ and $z_{1-\alpha}=1.6449$. As a result, the upper tolerance limit is [Equation (8)]

$$U=22+1.6449\times\left(\frac{0.99(1-0.99)}{351\times0.0032}\right)^{1/2}=22.1466 \tag{8}$$

This upper control limit is used as an upper control limit to monitor the 2021 data. Figure 2 displays the 365 daily new confirmed cases of 2021 data. The red line in Figure 2 is the upper control limit of 22.15. The first data exceeding the control limit is the 133rd day of 2021 (May 13, 2021). On May 15, 2021, Taipei City was announced to enter the third level of alert. On May 19, 2021, Taiwan's epidemic prevention alert level was raised to the third level. According to this control chart monitor, the first out-of-control signal appeared 2 days earlier than the announcement of the third level. After the peak period, the first data to reach the upper control limit is the 189th day (July 8, 2021). Since then, the data gradually stabilized. It shows that this control limit also can predict the end of the peak accurately. On June 8, 2021, the Moderna vaccine started and some people were given priority to receive the vaccine in Taiwan. This result shows that this control chart can accurately predict the start and end of the peak period. It is also consistent with the policy of epidemic prevention measures. Additionally, it is noted that the daily new confirmed cases may depend on past data, but in this case, the control char method leads to a good result, so the dependence problem can be neglected. The data and Matlab code for this real example and Example 1 are available at http://hwang.stat.nctu.edu.tw/TI.htm or https://drive.google.com/drive/folders/1khbZ68_Eb6AaWO7Dn8cD1T1P8_to3ZHF?usp=sharing

The upper control limit using the tolerance limit constructed by the sample quantile method performs well in this case. We can compare this upper control limit with the one constructed by the distribution-free tolerance limit. Using the *R* package "tolerance," the level $1-\alpha=0.95$ and $\beta=0.99$ distribution-free upper tolerance limit is 27. From Figure 2, if an upper control limit of 27 is used, it leads to a conservative result, predicting the start of the peak period later and the end of the peak period earlier than the method using the normal mixture distribution.

7 | CONCLUSION

TIs are a useful tool to estimate the interval that can cover a fixed proportion of a population with a specified confidence level for the common continuous and discrete distributions. TIs have various applications in quality control. In multiple-stream manufacturing processes, a mixture distribution of some suitable probabilistic model might be a better choice to be used to fit the data than a simple model. In this paper, TIs of normal mixture distributions and their applications to control charts are reviewed. A rule of constructing a modified two-sided TI for general mixture distributions

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is suggested based on the rule for the normal mixture distribution. The potential use of TI to construct control charts with mixture distributions is also discussed.

AUTHOR CONTRIBUTIONS

Hsiuying Wang: Conceptualization (equal); investigation (equal); methodology (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

The author thanks the Editors for the invitation and two reviewers for their valuable comments.

FUNDING INFORMATION

This research was funded by the Ministry of Science and Technology 109-2118-M-009-005-MY2, Taiwan.

CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

The data and Matlab code for this example are available on http://hwang.stat.nctu.edu.tw/TI.htm.

ORCID

Hsiuying Wang https://orcid.org/0000-0002-5863-0756

RELATED WIREs ARTICLE

On the number of components in a Gaussian mixture model

REFERENCES

Ali, S., Pievatolo, A., & Göb, R. (2016). An overview of control charts for high-quality processes. *Quality Reliability Engineering International*, *32*(7), 2171–2189.

Ali, S., & Riaz, M. (2014). Cumulative quantity control chart for the mixture of inverse Rayleigh process. *Computers & Industrial Engineering*, *73*, 11–20.

Alqurashi, M., Chakraborti, S., & Holcombe, C. (2023). Exact tolerance interval with specified tail probabilities and a control chart for the sample variance. *Quality and Reliability Engineering International*, *39*(1), 283–301.

Andriyanov, N. (2020). Comparative analysis of football statistics data clustering algorithms based on deep learning and Gaussian mixture model. *CEUR Workshop Proceedings*, *2667*, 71–74.

Atapattu, S., Tellambura, C., & Jiang, H. (2011). A mixture gamma distribution to model the SNR of wireless channels. *IEEE Transactions on Wireless Communications*, *10*(12), 4193–4203.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, *16*(2), 101–133.

Cai, T. T., & Wang, H. (2009). Tolerance intervals for discrete distributions in exponential families. *Statistica Sinica*, *19*(3), 905–923.

Castillo-Barnes, D., Martínez-Murcia, F. J., Ramírez, J., Górriz, J., & Salas-Gonzalez, D. J. N. (2020). Expectation–Maximization algorithm for finite mixture of α-stable distributions. *Neurocomputing*, *413*, 210–216.

Chaddad, A. (2015). Automated feature extraction in brain tumor by magnetic resonance imaging using Gaussian mixture models. *International Journal of Biomedical Imaging*, 8–8.

Chakraborti, S. (2004). Nonparametric (distribution-free) quality control charts. *Encyclopedia of Statistical Sciences*, 1–27.

Chakraborti, S., & Graham, M. A. (2019a). Nonparametric (distribution-free) control charts: An updated overview and some results. *Quality Engineering*, *31*(4), 523–544.

Chakraborti, S., & Graham, M. A. (2019b). *Nonparametric statistical process control*. John Wiley & Sons.

Chen, C., & Wang, H. (2020a). Tolerance interval for the mixture normal distribution. *Journal of Quality Technology*, *52*(2), 145–154.

Chen, E. J., & Kelton, W. D. (2006). Quantile and tolerance-interval estimation in simulation. *European Journal of Operational Research*, *168*(2), 520–540.

Chen, Y. H., & Wang, H. Y. (2020b). Exploring diversity of COVID-19 based on substitution distance. *Infection and Drug Resistance*, *13*, 3887–3894.

Chou, Y. M., & Mee, R. W. (1984). Determination of sample sizes for setting Beta-content tolerance limits controlling both tails of the normal-distribution. *Statistics & Probability Letters*, *2*(5), 311–314.

Demirhan, H., & Hamurkaroglu, C. (2014). Bayesian X control limits for exponentially distributed measurements. *Journal of Statistical Computation Simulation*, *84*(3), 628–643.

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive , Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Dong, X. Y., Tsong, Y., Shen, M. Y., & Zhong, J. L. (2015). Using tolerance intervals for assessment of pharmaceutical quality. *Journal of Biopharmaceutical Statistics*, *25*(2), 317–327.

Emura, T., & Wang, H. (2010). Approximate tolerance limits under log-location-scale regression models in the presence of censoring. *Technometrics*, *52*(3), 313–323.

Epprecht, E. K. (2015). Statistical control of multiple-stream processes: A literature review. *Frontiers in Statistical Quality Control*, *11*, 49–64.

Grilli, L., Rampichini, C., & Varriale, R. (2015). Binomial mixture modeling of university credits. *Communications in Statistics-Theory and Methods*, *44*(22), 4866–4879.

Guo, B. C., Wang, W., Wu, Z. C., He, X. X., & Zhu, Y. M. (2021). Exact equal-tailed beta-expectation tolerance intervals for sample variances. *Communications in Statistics-Simulation and Computation*, 1–18.

Guo, B. C., Zhu, N. F., Wang, W., & Wang, H. (2020). Constructing exact tolerance intervals for the exponential distribution based on record values. *Quality and Reliability Engineering International*, *36*(7), 2398–2410.

Hamada, M. (2002). Bayesian tolerance interval control limits for attributes. *Quality Reliability Engineering International*, *18*(1), 45–52.

Hamada, M. (2003). Tolerance interval control limits for x, R, and S charts. *Quality Engineering*, *15*(3), 471–487.

Hoang-Nguyen-Thuy, N., & Krishnamoorthy, K. (2021). A method for computing tolerance intervals for a location-scale family of distributions. *Computational Statistics*, *36*(2), 1065–1092.

Jiao, J. J., Zhao, X., & Cheng, W. H. (2022). Content-adjusted tolerance intervals via bootstrap calibration. *Stat*, *11*(1), e449.

Kiapour, A., & Qomi, M. N. (2017). Equal-tailed and shortest Bayesian tolerance intervals based on exponential k-records. *Communications in Statistics-Theory and Methods*, *46*(8), 3949–3956.

Krishnamoorthy, K., & Chakraberty, S. (2022). Construction of simultaneous tolerance intervals for several normal distributions. *Journal of Statistical Computation Simulation*, *92*(1), 101–114.

Krishnamoorthy, K., & Mathew, T. (2009). *Statistical tolerance regions: Theory, applications, and computation*. John Wiley & Sons.

Krishnamoorthy, K., & Xie, F. (2011). Tolerance intervals for symmetric location-scale families based on uncensored or censored samples. *Journal of Statistical Planning and Inference*, *141*(3), 1170–1182.

Lee, D., Li, W. K., & Wong, T. S. T. (2012). Modeling insurance claims via a mixture exponential model combined with peaks-over-threshold approach. *Insurance Mathematics & Economics*, *51*(3), 538–550.

Lei, H. J., Zhang, H., Ansari, I. S., Gao, C., Guo, Y. C., Pan, G. F., & Qaraqe, K. A. (2016). Performance analysis of physical layer security over generalized-K fading channels using a mixture gamma distribution. *IEEE Communications Letters*, *20*(2), 408–411.

Little, T. A. (2016). Essentials in tolerance design and setting specification limits. *BioPharm International*, *29*(6), 41–45.

Liu, W., Bretz, F., & Cortina-Borja, M. (2021). Reference range: Which statistical intervals to use? *Statistical Methods in Medical Research*, *30*(2), 523–534.

Majeed, M. Y., Aslam, M., & Riaz, M. (2013). Mixture cumulative count control chart for mixture geometric process characteristics. *Quality Quantity*, *47*(4), 2289–2307.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons.

Mee, R. W. (1984). β-Expectation and β-content tolerance limits for balanced one-way ANOVA random model. *Technometrics*, *26*(3), 251–254.

Meeker, W. Q., Hahn, G. J., & Escobar, L. A. (2017). *Statistical intervals: A guide for practitioners and researchers* (Vol. 541). John Wiley & Sons.

Montgomery, D. C. (2020). *Introduction to statistical quality control*. John Wiley & Sons.

Patel, J. K. (1986). Tolerance limits—A review. *Communications in Statistics-Theory and Methods*, *15*(9), 2719–2762.

Pusztai, É., & Kemény, S. (2020). Process capability indices when two sources of variability present, a tolerance interval approach. *Journal of Chemometrics*, *34*(4), e3213.

Rahman, M. S., Yoshida, N., Tsuboi, H., Ishii, Y., Akimoto, Y., & Kimura, K. (2021). Small-scale dissolution test screening tool to select potentially substandard and falsified (SF) medicines requiring full pharmacopoeial analysis. *Scientific Reports*, *11*(1), 12145.

Rau, A., Maugis-Rabusseau, C., Martin-Magniette, M. L., & Celeux, G. (2015). Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. *Bioinformatics*, *31*(9), 1420–1427.

Scrucca, L. (2021). A fast and efficient modal EM algorithm for Gaussian mixtures. *Statistical Analysis and Data Mining*, *14*(4), 305–314.

Selim, B., Alhussein, O., Muhaidat, S., Karagiannidis, G. K., & Liang, J. (2016). Modeling and analysis of wireless channels via the mixture of Gaussian distribution. *IEEE Transactions on Vehicular Technology*, *65*(10), 8309–8321.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*. John Wiley & Sons.

Short, M. (2021). On binomial quantile and proportion bounds: With applications in engineering and informatics. *Communications in Statistics-Theory and Methods*, 1–17.

Tsai, S. F. (2020). Approximate two-sided tolerance intervals for normal mixture distributions. *Australian & New Zealand Journal of Statistics*, *62*(3), 367–382.

Tsong, Y., Dong, X., Shen, M., & Lostritto, R. T. (2015). Quality assurance test of delivered dose uniformity of multiple-dose inhaler and dry powder inhaler drug products. *Journal of Biopharmaceutical Statistics*, *25*(2), 328–338.

Udom, A. U., Ugwu, O. S., Ossai, E. O., & Udoh, N. S. (2021). Multivariate process control chart based on sign and signed-rank tests for mixture distributions. *Journal of Statistics Management Systems*, *24*(8), 1669–1688.

Vicentin, D. S., Silva, B. B., Piccirillo, I., Bueno, F. C., & Oprime, P. C. (2018). Monitoring process control chart with finite mixture probability distribution: An application in manufacture industry. *International Journal of Quality Reliability Management*, 35(2),335-353.

Wald, A. (1943). An extension of Wilks' method for setting tolerance limits. *The Annals of Mathematical Statistics*, *14*(1), 45–55.

Wald, A., & Wolfowitz. (1946). Tolerance limits for a normal distribution. *The Annals of Mathematical Statistics*, *17*, 208–215.
Wang, H. (2007). Estimation of the probability of passing the USP dissolution test. *Journal of Biopharmaceutical Statistics*, *17*(3), 407–413.
Wang, H. (2008). Confidence intervals for the mean of a normal distribution with restricted parameter space. *Journal of Statistical Computation and Simulation*, *78*(9), 829–841.
Wang, H. (2021). The generalized Bayes method for high-dimensional data recognition with applications to audio signal recognition. *Symmetry-Basel*, *13*(1), 19.
Wang, H. (2022). COVID-19, anti-NMDA receptor encephalitis and MicroRNA. *Frontiers in Immunology*, *13*, 825103.
Wang, H., & Tsung, F. (2009). Tolerance intervals with improved coverage probabilities for binomial and Poisson variables. *Technometrics*, *51*(1), 25–33.
Wang, H., & Tsung, F. (2017). Constructing tolerance intervals for the number of defectives using both high-and low-resolution data. *Journal of Quality Technology*, *49*(4), 354–364.
Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *The Annals of Mathematical Statistics*, *12*(1), 91–96.
Wilks, S. S. (1942). Statistical prediction with special reference to the problem of tolerance limits. *The Annals of Mathematical Statistics*, *13*(4), 400–409.
Xiang, W. T., Karfoul, A., Yang, C. F., Shu, H. Z., & Jeannes, R. L. (2020). An exact line search scheme to accelerate the EM algorithm: Application to Gaussian mixture models identification. *Journal of Computational Science*, *41*, 101073.
Yan, D., Zhang, S., & Jung, U. (2019). A variable-selection control chart via penalized likelihood and Gaussian mixture model for multimodal and high-dimensional processes. *Quality and Reliability Engineering International*, *35*(4), 1263–1275.
Young, D. S. (2010). Tolerance: An R package for estimating tolerance intervals. *Journal of Statistical Software*, *36*(5), 1–39.
Young, D. S., Gordon, C. M., Zhu, S. H., & Olin, B. D. (2016). Sample size determination strategies for normal tolerance intervals using historical data. *Quality Engineering*, *28*(3), 337–351.
Yuan, M., Hong, Y. L., Escobar, L. A., & Meeker, W. Q. (2018). Two-sided tolerance intervals for members of the (log)-location-scale family of distributions. *Quality Technology and Quantitative Management*, *15*(3), 374–392.
Zhang, H., Jin, X. T., Wu, Q. M. J., Wang, Y. N., He, Z. D., & Yang, Y. M. (2018). Automatic visual detection system of railway surface defects with curvature filter and improved Gaussian mixture model. *IEEE Transactions on Instrumentation and Measurement*, *67*(7), 1593–1608.
Zimmer, Z., Park, D., & Mathew, T. (2016). Tolerance limits under normal mixtures: Application to the evaluation of nuclear power plant safety and to the assessment of circular error probable. *Computational Statistics Data Analysis*, *103*, 304–315.

How to cite this article: Wang, H. (2023). Tolerance limits for mixture-of-normal distributions with application to COVID-19 data. *WIREs Computational Statistics*, e1611. https://doi.org/10.1002/wics.1611

19390068, 0, Downloaded from https://wires.onlinelibrary.wiley.com/doi/10.1002/wics.1611 by Hsiuying Wang - National Yang Ming Chiao Tung Unive, Wiley Online Library on [04/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License