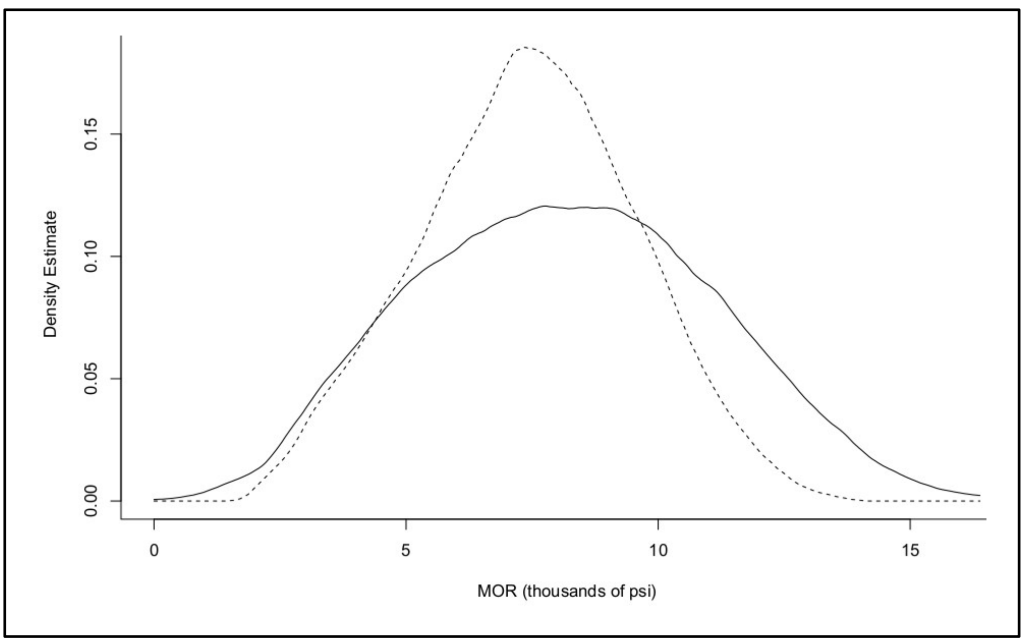
United States Department of Agriculture

Small-Sample Solution to the Two-Sample Problem for Quantiles Using Melded Random Confidence Intervals

Matthew A. Arvanitis



Forest Service | Forest Products Laboratory | Research Paper FPL–RP–713 | March 2022

Abstract

A new nonparametric solution to the two-sample problem for quantiles is proposed. This solution is applicable to small samples and/or extreme quantiles, and it prioritizes limiting Type I error to the indicated level of significance over optimizing power or confidence interval width. Aside from continuity, no assumptions about the distributions are made. In this approach, nonparametric random confidence intervals obtained from the samples are "melded," resulting in a nonparametric confidence interval for the difference between the two population quantiles with a specified confidence level. Simulations and an application to lumber strength characteristics are exhibited.

Keywords: nonparametric, quantile, two-sample problem, small sample, melded confidence interval, exact random nonparametric confidence interval, lumber quantile comparison

Contents

March 2022 [Errata: Equation (4.1) revised February 2025.]

Arvanitis, Matthew A. 2022. Small-sample solution to the two-sample problem for quantiles using melded random confidence intervals. Research Paper FPL-RP-713. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 15 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53726-2398. This publication is also available online at www.fpl.fs.fed.us. Laboratory publications are sent to hundreds of libraries in the United States and elsewhere.

The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin.

Small-Sample Solution to the Two-Sample Problem for Quantiles Using Melded Random Confidence Intervals

Matthew A. Arvanitis, Research Mathematical Statistician
USDA Forest Products Laboratory, Madison, Wisconsin, USA

1 Introduction

The two-sample problem for an arbitrary quantile (or quantiles) has received moderate attention, and several nonparametric solutions have been successful. Many of these solutions focus their attention on the median, though some are easily extended to arbitrary quantiles. Su and Wei (1993) apply a simple procedure with convenient asymptotic properties that permits the presence of censored observations. Wang and Hettmansperger (1990) applied a method to the Kaplan-Meier estimator for median survival times. Kosorok (1999) developed a method for approaching this problem for multiple quantiles simultaneously. Goldman and Kaplan (2018) use fractional order statistics and corrected confidence levels for the endpoints of individual confidence bounds to avoid conservatism in the interval for the quantile difference. Others apply numerical methods to form confidence intervals for quantiles and quantile differences (see Egloff et al. (2010) and Yang and Zhao (2017)). The current method applies a process of "melding" (Fay et al. 2015) to random nonparametric confidence intervals for quantiles (Zieliński and Zieliński 2005). It requires minimal assumptions and no random or bootstrap processes, nor does it require defining smoothing or other parameters that typically depend on the specific underlying distributions.

Consider random samples $X_1, X_2, ..., X_m \overset{iid}{\sim} F_X$, and $Y_1, Y_2, ..., Y_n \overset{iid}{\sim} F_Y$, where F_X and F_Y are continuous. Set $\boldsymbol{X}_m = (X_1, X_2, ..., X_m)$, and $\boldsymbol{Y}_n = (Y_1, Y_2, ..., Y_n)$, and assume $\boldsymbol{x}_m = (x_1, x_2, ..., x_m)$, and $\boldsymbol{y}_n = (y_1, y_2, ..., y_n)$ to be realizations of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, respectively. Set $\{X_{(i)}, i \in \{1, 2, ..., m\}\}$ and $\{Y_{(i)}, i \in \{1, 2, ..., n\}\}$ to be the sets of order statistics of $\boldsymbol{X}_m$ and $\boldsymbol{Y}_n$, respectively, while $\{x_{(i)}, i \in \{1, 2, ..., m\}\}$ and $\{y_{(i)}, i \in \{1, 2, ..., n\}\}$ are the realizations corresponding to $\{X_{(i)}, i \in \{1, 2, ..., m\}\}$ and $\{Y_{(i)}, i \in \{1, 2, ..., n\}\}$, respectively. This notation will be used throughout the paper.

Consider the two-sample problem for an arbitrary quantile, $q \in (0, 1)$. We wish to perform the hypothesis test

$$H_0 : F_X^{-1}(q) = F_Y^{-1}(q)$$
$$\text{versus}$$
$$H_1 : F_X^{-1}(q) \neq F_Y^{-1}(q)$$

at the $\alpha \in (0, 1)$ level of significance. The present work details an approach by which a nonparametric confidence interval for $F_Y^{-1}(q) - F_X^{-1}(q)$ is constructed, where assuring coverage probability is at least $1 - \alpha$ is given priority over minimizing interval width. This means that priority is placed on limiting the probability of Type I error–rejecting H_0 when it holds–over minimizing the probability of Type II error–failing to reject H_0 when it does not hold.

This paper is organized as follows. Section 2 reviews the methods that are applied in the current procedure. Section 3 details the current procedure. In Section 4, simulations reflecting the performance of this method for some specific distributions are presented. In Section 5, the utility of the method is shown through an engineering example comparing lower tail properties of hypothetical and actual lumber test data. Concluding remarks are presented in Section 6.

2 Background

To approach this problem thoroughly, relevant previous work is placed into the context of the current problem. First, a brief review of nonparametric confidence bounds is presented, followed by a discussion of exact random confidence bounds. Finally, melded confidence intervals will be explained in a general context before proceeding to Section 3 for their specific application to the current problem.

2.1 Univariate Non-Parametric Confidence Bounds

A well-known result first established by Thompson (1936) and expounded upon by Wilks (1941) provides that

$$P(X_{(r)} < F_X^{-1}(q)) = \sum_{j=r}^{m} \binom{m}{j} q^j (1-q)^{m-j} \tag{2.1}$$

This result is a commonly applied tool for obtaining nonparametric tolerance bounds and confidence intervals for quantiles. To obtain a lower tolerance bound from (2.1), recognize that $P(X_{(r)} < F_X^{-1}(q))$ is the probability that the interval $\left(X_{(r)}, \infty\right)$ contains $F_X^{-1}(q)$; that is, $\left(X_{(r)}, \infty\right)$ is a $100P(X_{(r)} < F_X^{-1}(q))\%$ confidence interval for $F_X^{-1}(q)$, and thus, $X_{(r)}$ is a lower tolerance bound for $100(1-q)\%$ of the population, with probability $P(X_{(r)} < F_X^{-1}(q))$. The probability (2.1) is exact and does not depend on F_X.

2.2 Exact Random Confidence Bounds

Although confidence levels associated with the intervals described in Section 2.1 are known exactly, these confidence intervals may not, in general, be constructed with arbitrarily specified confidence levels because the probability in (2.1) may be computed for only a finite number of values that depend entirely on the sample size. To deal with this problem, Zieliński and Zieliński (2005) proposed random confidence intervals. In the context of Section 2.1, if $p \in (0,1)$ is the desired confidence level for a lower quantile confidence bound, and $p_1 = P\left(X_{(r)} < F_X^{-1}(q)\right) > p > P\left(X_{(r+1)} < F_X^{-1}(q)\right) = p_2$ for some integer r satisfying $1 \le r < m$, then defining the random variable,

$$K \sim \text{Bernoulli}\left(\frac{p_1 - p}{p_1 - p_2}\right) \tag{2.2}$$

we have $P\left(X_{(r+K)} < F_X^{-1}(q)\right) = E_K\left[P\left(X_{(r+K)} < F_X^{-1}(q)\right)\right] = p$, so that $\left(X_{(r+K)}, \infty\right)$ is a random $100p\%$ confidence interval for $F_X^{-1}(q)$. Therefore, this method admits a means to construct $100p\%$ confidence bounds for $F_X^{-1}(q)$, for all values of p as high as $P\left(X_{(1)} < F_X^{-1}(q)\right)$ for lower bounds and $P\left(X_{(m)} > F_X^{-1}(q)\right)$ for upper bounds.

2.3 Melded Confidence Interval

For some parameter, θ_x, of F_X and any $a \in (0,1)$, define $L_{\theta_x}(\boldsymbol{x}_m, a)$ to be a lower $100a\%$ confidence bound for θ_x, based on $\boldsymbol{x}_m$; that is, $P\left(L_{\theta_x}(\boldsymbol{X}_m, a) < \theta_x\right) = a$. Similarly, define $U_{\theta_x}(\boldsymbol{x}_m, a)$ to be an upper $100a\%$ confidence bound for θ_x, based on $\boldsymbol{x}_m$; that is, $P\left(U_{\theta_x}(\boldsymbol{X}_m, a) > \theta_x\right) = a$. Define in a similar way $L_{\theta_y}(\boldsymbol{y}_n, b)$ and $U_{\theta_y}(\boldsymbol{y}_n, b)$, $b \in (0,1)$, for some parameter, θ_y, of F_Y. Under some mild regularity conditions, Fay et al. (2015)[1] constructed a method to 'meld' these intervals into a single interval for the difference between the two population parameters, $\theta_y - \theta_x$. In particular, the interval, (L, U), where L is the $\frac{\alpha}{2}$th quantile of

$$L_{\theta_y}(\boldsymbol{y}_n, B) - U_{\theta_x}(\boldsymbol{x}_m, A)$$

and U is the $\left(1 - \frac{\alpha}{2}\right)$ th quantile of

$$U_{\theta_y}(\boldsymbol{y}_n, B) - L_{\theta_x}(\boldsymbol{x}_m, A)$$

such that A and B are independent Uniform$(0,1)$ random variables, is a $100(1-\alpha)\%$ confidence interval for $\theta_y - \theta_x$.

3 Confidence Intervals for Quantile Differences

An immediate observation from Section 2.3 is that for the melded procedure to work, the lower and upper confidence bounds of a parameter must be known (or at least able to be censored left and right of certain values) for all confidence levels in the unit interval. For quantiles, Section 2.1 provides these for only a finite number of confidence levels, but Section 2.2 provides them for all confidence levels except those on short intervals at the lower and upper extremes of the unit interval. So, rather than constructing conservative nonparametric bounds for all confidence levels, we now have a means with which we must do so only for confidence levels on these short extreme subintervals of the unit interval, but these unknown values can usually be effectively censored without consequence by using the melded quantile confidence interval procedure, as follows.

For $r \in \{1, 2, ..., m\}$ and $s \in \{1, 2, ..., n\}$, set

$$u_r = P\left(X_{(r)} < F_X^{-1}(q)\right) \text{ and } v_s = P\left(Y_{(s)} < F_Y^{-1}(q)\right) \tag{3.1}$$

according to (2.1), with the added conventions that $u_0 = v_0 = 1$ and $u_{m+1} = v_{n+1} = 0$. We begin with the lower bounds, $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$, $a \in (0,1)$. We are interested in guaranteeing coverage probability of the resultant interval for all values of a. To do this, we set these bounds to (possibly random) values so that $P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) \geq a$ while $P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) - a$ is minimized over all a. First, for $a \leq u_m$, the highest value to set $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$ to and still satisfy the first constraint with certainty is $x_{(m)}$, so it is set as such. For $a > u_m$, we use the exact random confidence interval procedure detailed in Section 2.2. Set $x_{(0)} = \sup\{x : F_X(x) = 0\}$, if it exists (and is known), or $x_{(0)} = -\infty$

[1]Important supporting information accompanying this paper can be found at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4852749/

otherwise. Then clearly $x_{(0)} < F_X^{-1}(q)$ with probability 1, so that even though the interval $(x_{(0)}, x_{(1)})$ is outside the range of the data, the random confidence interval procedure applied to $u_0 > a > u_1$ still yields $P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) = a$. Set the random variable

$$L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a) = x_{(K_a)} \tag{3.2}$$

such that

$$K_a = J_a + r \tag{3.3}$$

where $r = \sum_{j=1}^{m} I\{a \leq u_r\}$, and $J_a \sim \text{Bernoulli}\left(\frac{u_r - a}{u_r - u_{r+1}}\right)$. Here, in contrast to the case where $a \leq u_m$, $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$ is a random variable with exactly two possible values, both from $\left\{x_{(i)}, i \in \{0, 1, 2, ..., m\}\right\}$, and associated probabilities completely determined by a and m. Therefore, under this regime, $L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a)$ is well-defined for all $a \in (0, 1)$, and for any $a \geq u_m$,

$$P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) = a \tag{3.4}$$

and for $a < u_m$,

$$P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) - a > 0 \tag{3.5}$$

but minimized, as desired.

As a consequence of this definition of $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$, if $A \sim \text{Uniform(0,1)}$, then $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, A)$ has a discrete distribution:

$$\begin{aligned}
&P(L_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) = x_{(0)}) \\
&= \int_0^1 P(L_{F_X^{-1}(q)}(\boldsymbol{x}_m, t) = x_{(0)}) dF_A(t) \\
&= \int_{u_1}^{u_0} \left(\frac{t - u_1}{u_0 - u_1}\right) dt \\
&= \frac{u_0 - u_1}{2} = \frac{1 - u_1}{2}
\end{aligned}$$

for $r \in \{1, ..., m-1\}$,

$$\begin{aligned}
&P(L_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) = x_{(r)}) \\
&= \int_0^1 P(L_{F_X^{-1}(q)}(\boldsymbol{x}_m, t) = x_{(r)}) dF_A(t) \\
&= \int_{u_{r+1}}^{u_r} \left(\frac{t - u_{r+1}}{u_r - u_{r+1}}\right) dt + \int_{u_r}^{u_{r-1}} \left(\frac{u_{r-1} - t}{u_{r-1} - u_r}\right) dt \\
&= \frac{u_{r-1} - u_{r+1}}{2}
\end{aligned}$$

and, as required above, the remaining probability is attributed to $x_{(m)}$. The resulting distribution is thus given by

$$P(L_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) = x) = \begin{cases} \frac{1-u_1}{2} & x = x_{(0)} \\ \frac{u_{r-1} - u_{r+1}}{2} & x = x_{(r)}, r \in \{1, 2, ..., m-1\} \\ \frac{u_{m-1} + u_m}{2} & x = x_{(m)} \\ 0 & \text{Otherwise} \end{cases} \tag{3.6}$$

Similarly, setting $y_{(0)} = \sup\{y : F_Y(y) = 0\}$, if it exists, or $y_{(0)} = -\infty$ otherwise, the same regime is applied to $L_{F_Y^{-1}(q)}(\boldsymbol{y}_n, B)$, $B \sim \text{Uniform}(0,1)$, B independent of A.

$$P(L_{F_Y^{-1}(q)}(\boldsymbol{y}_n, B) = y) = \begin{cases} \frac{1-v_1}{2} & y = y_{(0)} \\ \frac{v_{r-1}-v_{r+1}}{2} & y = y_{(r)}, r \in \{1,2,...,n-1\} \\ \frac{v_{n-1}+v_n}{2} & y = y_{(n)} \\ 0 & \text{Otherwise} \end{cases} \tag{3.7}$$

For the upper bounds, set $x_{(m+1)} = \inf\{x : F_X(x) = 1\}$, if it exists, or $x_{(m+1)} = \infty$ otherwise, and set $y_{(n+1)} = \inf\{y : F_Y(y) = 1\}$, if it exists, or $y_{(n+1)} = \infty$ otherwise. Here, the conservative bounds are applied in cases where $a \geq u_1$ or $b \geq v_1$, and the random confidence interval is applied for all other values. As a result, a similar strategy yields the upper bound distributions that also guarantee coverage probability for all values of A and B.

$$P(U_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) = x) = \begin{cases} 1 - \frac{u_2+u_1}{2} & x = x_{(1)} \\ \frac{u_{r-1}-u_{r+1}}{2} & x = x_{(r)}, r \in \{2,3,...,m\} \\ \frac{u_m}{2} & x = x_{(m+1)} \\ 0 & \text{Otherwise} \end{cases} \tag{3.8}$$

and

$$P(U_{F_Y^{-1}(q)}(\boldsymbol{y}_n, B) = y) = \begin{cases} 1 - \frac{v_2+v_1}{2} & y = y_{(1)} \\ \frac{v_{r-1}-v_{r+1}}{2} & y = y_{(r)}, r \in \{2,3,...,n\} \\ \frac{v_n}{2} & y = y_{(n+1)} \\ 0 & \text{Otherwise} \end{cases} \tag{3.9}$$

Define the random functions $G_{XY}(z)$ and $H_{XY}(z)$ to be the distribution functions of $L_q(A, B|\boldsymbol{X}_m, \boldsymbol{Y}_n)$ and $U_q(A, B|\boldsymbol{X}_m, \boldsymbol{Y}_n)$, respectively, where, for specific $(\boldsymbol{x}_m, \boldsymbol{y}_n)$,

$$L_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n) = L_{F_Y^{-1}(q)}(\boldsymbol{y}_n, B) - U_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) \tag{3.10}$$

and

$$U_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n) = U_{F_Y^{-1}(q)}(\boldsymbol{y}_n, B) - L_{F_X^{-1}(q)}(\boldsymbol{x}_m, A) \tag{3.11}$$

Then, $L_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n)$ and $U_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n)$ have a total number of possible values less than or equal to $(m+1)(n+1)$, with equality (with probability 1) whenever $x_{(0)}, y_{(0)}, x_{(m+1)}$, and $y_{(n+1)}$ are all finite. The confidence bounds for the quantile difference, $L = G_{XY}^{-1}\left(\frac{\alpha}{2}|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n\right)$ and $U = H_{XY}^{-1}\left(1 - \frac{\alpha}{2}|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n\right)$, may be obtained by taking the greatest value, l, such that $P(L_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n) \leq l) \leq \frac{\alpha}{2}$, and the smallest value, u, such that $P(U_q(A, B|\boldsymbol{X}_m = \boldsymbol{x}_m, \boldsymbol{Y}_n = \boldsymbol{y}_n) \geq u) \leq \frac{\alpha}{2}$, respectively, as prescribed in Section 2.3. It should be noted that one or both of the bounds may be infinite using this procedure. If this is the case, insufficient data exist to nonparametrically obtain a finite-length confidence interval at the specified level, just as it is occasionally the case in the one-sample problem for quantiles. However, even in this case, the hypothesis test in Section 1 may still result in a rejection of H_0, particularly if the interval does not contain 0.

One counterintuitive attribute of this method is that even though the confidence bounds for $F_X^{-1}(q)$ and $F_Y^{-1}(q)$ are (mostly) random, the resultant bounds for $F_Y^{-1}(q) - F_X^{-1}(q)$ are fixed, nonrandom. For intuition, the randomness does not disappear; rather, it remains and is simply consolidated with that of A and B into G_{XY} or H_{XY}. The randomness is removed only when specific quantiles of these distributions are computed in accordance with the melded procedure.

Remark Revisiting the setting of $x_{(0)}$, it should now be clear that as long as $x_{(0)}$ satisfies $F_X\left(x_{(0)}\right) = 0$, $P\left(L_{F_X^{-1}(q)}(\boldsymbol{X}_m, a) < F_X^{-1}(q)\right) = a$ for all $a > u_1$, even though these values of a correspond to confidence bounds that are necessarily outside the range of the observed data. We choose $\sup\{x : F_X(x) = 0\}$ because it maximally exploits the available information to obtain the lower bound. However, the reason the same procedure is not used for $a < u_m$ is that doing so would potentially result in ambiguous values (e.g., $\infty - \infty$) for the calculation in (3.11), effectively precluding the possibility of censoring them appropriately. To avoid this, we choose $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$ conservatively whenever $a < u_m$, and, likewise, we choose $U_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$ conservatively whenever $a > u_1$. The reader might realize a handy deduction from this: if one or more of $x_{(0)}, y_{(0)}, x_{(m+1)}$, or $y_{(n+1)}$ are (justifiably) assigned finite values, then it is possible to use the random confidence interval method over the entire unit interval for a and/or b for some subset of the four bound distributions to potentially reduce conservatism of the resultant confidence interval without giving up the guaranteed coverage probabilities.

Remark This method is, of course, easily adapted to form a one-tailed test against H_0 by omitting either L_q or U_q and adjusting the tail size from $\frac{\alpha}{2}$ to α. It can also be approached from the opposite perspective: specifying a value for a bound and determining its associated tolerance level, which is exhibited in Section 5.

Remark Rather than using exact random confidence bounds, an alternative method of computing $L_{F_X^{-1}(q)}(\boldsymbol{x}_m, a)$ is interpolated nonparametric confidence intervals (Hettmansperger and Sheather 1986, Beran and Hall 1993). For some distributions, particularly well-behaved distributions, the use of this method performs slightly better than the exact random method. However, by its design, this method cannot guarantee a minimum coverage probability for the individual bounds (Frey and Zhang 2017) in the way the exact random method can.

Remark Note that q need not be the same for both X and Y. The method may be easily adapted to compare one quantile of X to a different quantile of Y.

Remark For large sample sizes (that is, when both m and n are large), this method can become computationally expensive. However, it also exhibits convenient asymptotic properties that may be exploited. These properties, including asymptotic distributions and rates of convergence, will be investigated in a subsequent report.

4 Simulations

To test this method, two other methods are presented for comparison: normal approximation and re-sampling. These have been chosen because of their popularity and ease of implementation.

First, the normal approximation method simply assumes the distribution of the difference

Case	X	Y	Challenges
1	$\exp(3)$	$1.514964 - \exp(2)$	Different shapes Asymmetric
2	$\text{Logistic}(1, 0)$	$\text{Logistic}(1.944439, 2)$	Heavy tails Different Variances
3	$\Gamma(7, 3)$	$\Gamma(9, 4)$	Asymmetric
4	$\Gamma(3, 3)$	$\text{Cauchy}(4, 1)$	Lack of finite moments for Cauchy

Table 1: Distributions for simulations.

of the sample quantiles is given by

[Equation (4.1) revised February 2025.]

$$\begin{aligned}&\hat{F}_Y^{-1}(q) - \hat{F}_X^{-1}(q)\\ &\qquad \dot{\sim}\ \mathrm{N}\left(F_Y^{-1}(q) - F_X^{-1}(q), \frac{q(1-q)}{m f_X\left(F_X^{-1}(q)\right)^2} + \frac{q(1-q)}{n f_Y\left(F_Y^{-1}(q)\right)^2}\right)\end{aligned} \tag{4.1}$$

density estimation is applied to obtain estimates of f_X and f_Y (the "density" function in R using the Epanechnikov kernel), and the linearly interpolated empirical distribution function (type 4 of the "quantile" function in R) given by

$$\hat{F}_X^{-1}(q) = (1 - mq + \lfloor mq \rfloor)X_{(\lfloor mq \rfloor)} + (mq - \lfloor mq \rfloor)X_{(\lfloor mq \rfloor + 1)} \tag{4.2}$$

is applied to estimate each quantile. The re-sampling method forms samples, with replacement, of the same sizes from the two original samples. It then obtains estimates of the two quantiles using (4.2) and takes the difference. Repeating this process 100 times results in a sample of quantile differences, which is then assumed to be normally distributed. The sample mean and variance are used to form a confidence interval for the difference, based on that normality assumption.

4.1 Test Distributions

The distributions for X and Y, values for q, and the values of m and n (equal in all cases tested) are selected to pose an assortment of challenges, including small samples, distributions with different shapes near the quantiles, and extreme quantiles. Table 1 displays the four pairs of distributions on which each of the three methods were tested. The combinations of (m, n, q) that were tested are the same for all four pairs of distributions: $(250, 250, 0.05)$, $(100, 100, 0.25)$, and $(50, 50, 0.5)$. Note that as the test sample size decreases, the test quantile becomes more central; this is necessary because it becomes increasingly impractical to estimate extreme quantiles nonparametrically as sample sizes decrease. Therefore, small but still practical sample sizes are used for each selected quantile. Though in some cases $F_X^{-1}(q)$ and $F_Y^{-1}(q)$ are not equal, the confidence intervals were still tested against the actual quantile differences. All confidence intervals are at the 95% level ($\alpha = 0.05$), and all simulations were repeated 10,000 times.

4.2 Results

For all 36 sampling conditions, Table 2 details the success rate (the fraction of cases in which the interval contained the actual quantile difference) and the mean 95% confidence interval width. These results suggest that the melded method is capable of achieving its prescribed confidence level as well as the other methods, and it appears to be the only one of the three that can consistently do so for extreme quantiles, particularly in Case 1, where the shapes of F_X and F_Y, at their respective quantiles of interest, are arguably the most distinct. With only a minor loss in efficiency (in terms of confidence interval width), the melded method also works well with small sample sizes and more central quantiles. The higher confidence levels and wider intervals are to be expected because, unlike the other two methods, it is designed to be a procedure that will sacrifice interval width in favor of limiting Type I error to α. As a side note, the normal approximation method result for Case 4, $q = 0.05$, is, in fact, accurate. It is due in large part to the substantial variability of $\hat{f}_Y$ at the already highly variable extreme quantile estimate of the Cauchy distribution.

5 Application: Dimension Lumber

Dimension lumber used for construction purposes has measurable characteristics that must be monitored over time to ensure that parent populations have maintained their engineering properties (Kretschmann et al. 2014). Two such characteristics are the modulus of elasticity (MOE), which quantifies its resistance to flexure, and modulus of rupture (MOR), which quantifies its resistance to breakage, both under bending force. For illustration, a ceramic floor tile will have a high MOE but a relatively low MOR; whereas a commercial aircraft wing will have the opposite characteristics (relatively speaking). Benchmark thresholds, called design values, for MOE and MOR are used for comparing temporally separated populations of specific species or species groups. The focus in this section is the design value of MOR for dimension lumber, which is based on the 5th percentile. Applied to dimension lumber, MOE and MOR exhibit a weak positive correlation, and because dimension lumber is often graded based on characteristics such as knot size or MOE,[2] that is, placed into certain categories defined by firm limits, the within-grade distribution of MOR is "softly" truncated. Verrill et al. (2015) introduced a model that takes into account this truncation, called pseudo-truncation, and which ultimately led to the introduction of a new univariate distribution: the pseudo-truncated Weibull distribution.

This section focuses on achieving two objectives. First, as a result of its fitness to actual lumber data, a set of specific cases of the pseudo-truncated Weibull distribution will be used to generate data to which the melded random quantile difference method will be applied in order to demonstrate its applicability and effectiveness in terms of both Type I and Type II errors. Second, an example using actual data from the dimension lumber "Original In-Grade" and subsequent "In-Grade resource monitoring" programs will be presented.

[2] This is done through a process called machine stress rating.

Case	Method	$m = n$	q	Success Rate	Mean Width
1	Melded	250	0.05	0.9468	0.5612
1	Normal	250	0.05	0.9155	0.5336
1	Re-Sampling	250	0.05	0.9299	0.5447
1	Melded	100	0.25	0.9596	0.3681
1	Normal	100	0.25	0.9368	0.3402
1	Re-Sampling	100	0.25	0.9445	0.3543
1	Melded	50	0.5	0.9715	0.3624
1	Normal	50	0.5	0.9417	0.3275
1	Re-Sampling	50	0.5	0.9533	0.3469
2	Melded	250	0.05	0.9662	2.866
2	Normal	250	0.05	0.9275	2.428
2	Re-Sampling	250	0.05	0.9455	2.579
2	Melded	100	0.25	0.9687	2.158
2	Normal	100	0.25	0.9558	2.082
2	Re-Sampling	100	0.25	0.9551	2.081
2	Melded	50	0.5	0.9688	2.664
2	Normal	50	0.5	0.9647	2.687
2	Re-Sampling	50	0.5	0.9508	2.553
3	Melded	250	0.05	0.9749	0.4088
3	Normal	250	0.05	0.9486	0.3673
3	Re-Sampling	250	0.05	0.9591	0.3798
3	Melded	100	0.25	0.9697	0.5368
3	Normal	100	0.25	0.9648	0.5436
3	Re-Sampling	100	0.25	0.9545	0.5216
3	Melded	50	0.5	0.9731	0.8432
3	Normal	50	0.5	0.9592	0.8230
3	Re-Sampling	50	0.5	0.9568	0.8038
4	Melded	250	0.05	0.9454	9.144
4	Normal	250	0.05	0.8333	92.38
4	Re-Sampling	250	0.05	0.9278	8.251
4	Melded	100	0.25	0.9596	1.260
4	Normal	100	0.25	0.9405	1.057
4	Re-Sampling	100	0.25	0.9486	1.176
4	Melded	50	0.5	0.9643	1.075
4	Normal	50	0.5	0.9804	1.175
4	Re-Sampling	50	0.5	0.9575	1.026

Table 2: Simulation Results.

Year	Variable	Distribution	5th Percentile
0	X	PTW$(0.6454, 4, 0.3, 0.25, 0.75)$	1.866
5	$Y^{(1)}$	PTW$(0.5437, 3.3, 0.25, 0.25, 0.75)$	1.866
10	$Y^{(2)}$	PTW$(0.5437, 3.3, 0.25, 0.15, 0.65)$	1.700
15	$Y^{(3)}$	PTW$(0.5437, 2.9, 0.25, 0.05, 0.75)$	1.430

Table 3: Simulated MOR Distributions.

5.1 Simulated Modulus of Rupture

Consider a random vector, (V, W), with marginal distributions, $V \sim \mathrm{N}(\mu, \sigma^2)$ and $W \sim \mathrm{Weibull}(\gamma, \beta)$, and a Gaussian copula with correlation parameter, ρ. Then X has a pseudo-truncated Weibull distribution with parameters ρ, β, γ, p_1, and p_u, denoted $X \sim \mathrm{PTW}(\rho, \beta, \gamma, p_1, p_u)$, if

$$X = W \left| p_1 < \Phi\left(\frac{V-\mu}{\sigma}\right) < p_u \right. \tag{5.1}$$

where $\Phi(\cdot)$ is the distribution function of a standard normal random variable, with parameter space given by $\rho, p_1, p_u \in (0, 1)$ and $\beta, \gamma \in (0, \infty)$. Suppose, for a specific class of lumber, MOR is given by PTW distributions on years 0 (the benchmark year), 5, 10, and 15. On each of these years, a random sample of size $n = 360$ is collected and tested for MOR.[3] All values are in thousands of pounds per square inch (psi). The (unknown) distributions and 5th percentiles are detailed in Table 3, and the distribution functions associated with them are shown in Figure 1. Based on samples of this type, two questions are of importance:

1. To what degree of confidence, C, can it be claimed that the 5th percentile of MOR has not decreased between year 0 and each of years 5, 10, and 15?

2. Can we construct a $100(1-\alpha)\%$ confidence interval, (L, U), for the magnitude of any increase in MOR from year 0 to each of years 5, 10, and 15?

For convenience, consider C, L, and U to be random variables, based on $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$, where $\boldsymbol{Y}_n$ is a random sample from the appropriate variable listed in Table 3. Then $C = G_{XY}(0|\boldsymbol{X}_n, \boldsymbol{Y}_n)$, $L = G_{XY}^{-1}(\frac{\alpha}{2}|\boldsymbol{X}_n, \boldsymbol{Y}_n)$ and $U = H_{XY}^{-1}(1-\frac{\alpha}{2}|\boldsymbol{X}_n, \boldsymbol{Y}_n)$, as outlined in Section 3. In year 5, when there has been no change in the 5th percentile, we should expect that $C \mathrel{\dot\sim} \mathrm{Uniform}(0, 1)$, so that $P(C < \alpha) \approx \alpha$. For years 10 and 15, we should expect C to be skewed heavily to the right, so that $P(C < \alpha) >> \alpha$. For all three years, we should observe $F_Y^{-1}(q) - F_X^{-1}(q) \in (L, U)$ with probability approximately $1 - \alpha$.

We may generate samples of C, L, and U for each year by simulating $(\boldsymbol{x}_n, \boldsymbol{y}_n)$ from the four PTW distributions and computing them using the definitions of G_{XY} and H_{XY}. The results of these simulations for various values of α are detailed in Table 4. A plot of Type II error, as a function of α, for years 10 and 15 is provided in Figure 2. These results indicate that the melded method performs as expected. Though it has great difficulty detecting a decrease in MOR for year 10 (which is to be expected given the small difference at the

[3]This sample size is chosen since it is typical in the industry.

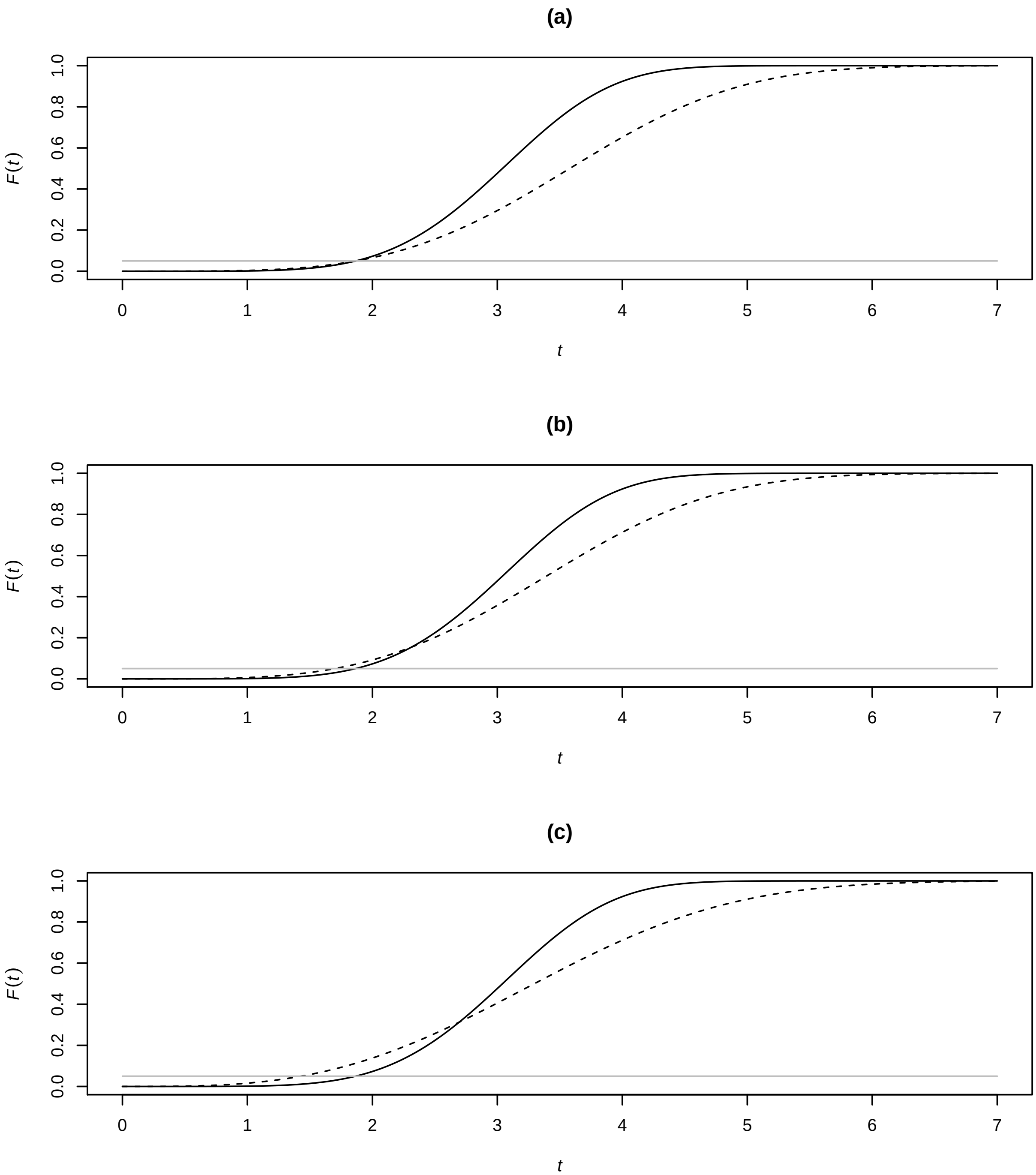


Figure 1: Distribution functions of $X \sim \mathrm{PTW}(0.6454, 4, 0.3, 0.25, 0.75)$ (solid (a), (b), and (c)), $Y^{(1)} \sim \mathrm{PTW}(0.5437, 3.3, 0.25, 0.25, 0.75)$ (dashed (a)), $Y^{(2)} \sim \mathrm{PTW}(0.5437, 3.3, 0.25, 0.15, 0.65)$ (dashed (b)), and $Y^{(3)} \sim \mathrm{PTW}(0.5437, 2.9, 0.25, 0.05, 0.75)$ (dashed (c)). The solid gray line intersects the functions at their respective 5th percentiles.

Year	α	$P(C < \alpha)$	$E(U - L)$	$P\left[L < F_Y^{-1}(q) - F_X^{-1}(q) < U\right]$
5	0.025	0.0135	0.6456	0.9894
5	0.05	0.0338	0.5587	0.9703
5	0.1	0.0797	0.4655	0.9305
5	0.25	0.2230	0.3235	0.7872
10	0.025	0.2027	0.6303	0.9868
10	0.05	0.3173	0.5455	0.9712
10	0.1	0.4718	0.4543	0.9308
10	0.25	0.7252	0.3151	0.7943
15	0.025	0.9638	0.6264	0.9780
15	0.05	0.9847	0.5422	0.9530
15	0.1	0.9955	0.4516	0.8995
15	0.25	0.9998	0.3137	0.7484

Table 4: PTW simulation results.

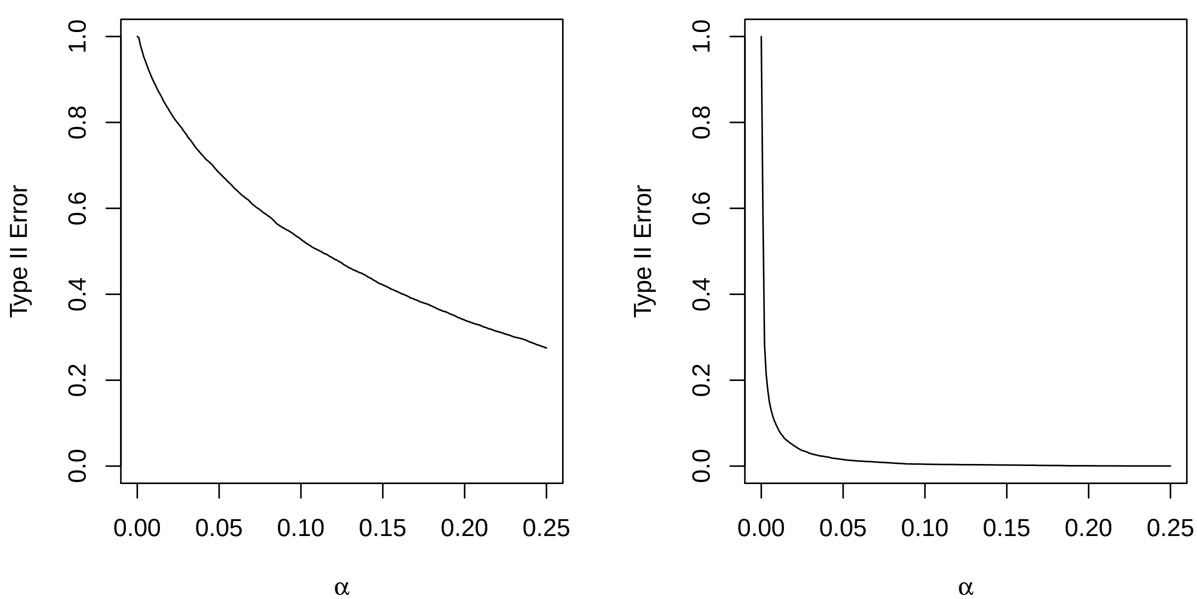


Figure 2: Type II error (failure to detect decrease in property ($C > \alpha$)) as a function of α for year 10 (left), and year 15 (right).

quantile), it appropriately contains Type I error and quickly reduces Type II error as the quantile difference increases.

5.2 Example: Modulus of Rupture

The American Lumber Standard Committee (ALSC) Board of Review is responsible for ensuring that reported MOR design values are representative of those associated with lumber grades according to the National Grading Rule for lumber used as a structural material. Various rules writing grading agencies perform periodic monitoring of species or species groups to test whether or not MOR has decreased over time. For one such ALSC-graded species group, an original sample ($\boldsymbol{x}_{380}$) of specimens was collected and tested for MOR in the late 20th century, and the results (adjusted for moisture, size, and other characteristics) were obtained. Recently, a monitoring sample ($\boldsymbol{y}_{360}$) was collected and tested for MOR and the results adjusted according to the same paradigm as the original sample. Smoothed histograms of the two datasets are presented in Figure 3. Applying definitions similar to

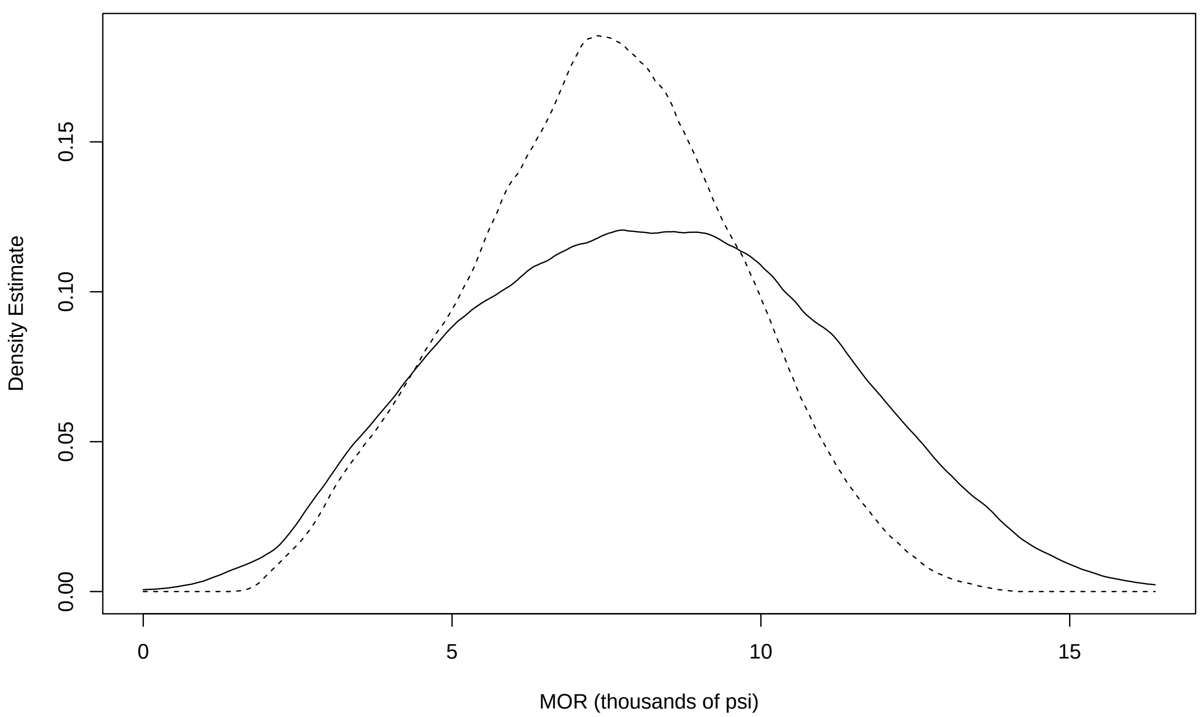


Figure 3: In-Grade MOR: Original (solid) and Recent (dashed).

those in Section 5.1, we obtain $C = 0.6999$, $L = -0.4824$ and $U = 0.7456$. Thus, while the histograms indicate a clear stochastic shift to the left over time, it can still be declared that, with 70% confidence, the 5th percentile has not decreased.

6 Conclusion

In this paper a new method of approaching the two sample problem for quantiles is introduced. It is particularly useful for small datasets or in cases where the quantile of interest is extreme. It is shown that the method provides predictable, yet realistic, results against essentially any underlying (continuous) distributions, and, by design, it prioritizes coverage probability over confidence interval width.[4]

One final remark about the approach, and arguably of the greatest importance, is that this method focuses entirely on the quantile(s) of interest, and, to the extent possible, it (deliberately) ignores all other characteristics of the distributions of X and Y. This results in an obvious benefit: if the researcher's assumptions mirror this, it is essentially *the ideal* approach. If, however, further assumptions can be made about the distributions, especially those which can provide additional inference about the quantile of interest, its statistical power will be handicapped, possibly severely, by that same ignorance. In light of these facts, the melded random method should be considered one of the more conservative tools in the statistician's quantile-difference toolbox to be used when achieving prescribed coverage probabilities is deemed to be significantly more important than confidence interval width.

7 Acknowledgments

The author wishes to thank the reviewers and the FPL editorial staff for their valuable inputs and corrections as well as the ALSC and related rules writing agencies for their cooperation.

References

Beran, R. and Hall, P. (1993). Interpolated nonparametric prediction intervals and confidence intervals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(3):643–652.

Egloff, D., Leippold, M., et al. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, 38(2):1244–1278.

Fay, M. P., Proschan, M. A., and Brittain, E. (2015). Combining one-sample confidence procedures for inference in the two-sample case. *Biometrics*, 71(1):146–156.

Frey, J. and Zhang, Y. (2017). What do interpolated nonparametric confidence intervals for population quantiles guarantee? *The American Statistician*, 71(4):305–309.

Goldman, M. and Kaplan, D. M. (2018). Non-parametric inference on (conditional) quantile differences and interquantile ranges, using l-statistics. *The Econometrics Journal*, 21(2):136–169.

[4]Though Fay et. al.'s melded confidence interval method has not been proven to guarantee coverage if the same for the individual confidence intervals is guaranteed, its performance in simulation to date has strongly suggested it.

Hettmansperger, T. P. and Sheather, S. J. (1986). Confidence intervals based on interpolated order statistics. *Statistics & Probability Letters*, 4(2):75–79.

Kosorok, M. R. (1999). Two-sample quantile tests under general conditions. *Biometrika*, 86(4):909–921.

Kretschmann, D., DeVisser, D., Cheung, K., Browder, B., and Rozek, A. (2014). Maintenance procedures for north american visually-graded dimension lumber design values. In *In: World Conference on Timber Engineering, Quebec City, Canada, August 10-14, 2014; p. 8.*, pages 1–8.

Su, J. Q. and Wei, L. (1993). Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*, pages 603–607.

Thompson, W. R. (1936). On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. *The Annals of Mathematical Statistics*, 7(3):122–128.

Verrill, S. P., Evans, J. W., Kretschmann, D. E., and Hatfield, C. A. (2015). Asymptotically efficient estimation of a bivariate gaussian–weibull distribution and an introduction to the associated pseudo-truncated weibull. *Communications in Statistics-Theory and Methods*, 44(14):2957–2975.

Wang, J.-L. and Hettmansperger, T. P. (1990). Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistical Association*, 85(410):529–536.

Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *The Annals of Mathematical Statistics*, 12(1):91–96.

Yang, H. and Zhao, Y. (2017). Smoothed jackknife empirical likelihood for the difference of two quantiles. *Annals of the Institute of Statistical Mathematics*, 69(5):1059–1073.

Zieliński, R. and Zieliński, W. (2005). Best exact nonparametric confidence intervals for quantiles. *Statistics*, 39(1):67–71.