Unbiasedness and efficiency of non-parametric and UMVUE estimators of the probabilistic index and related statistics

Unbiasedness and efficiency of non-parametric and UMVUE estimators of the probabilistic index and related statistics

Journal Title
XX(X):3–42

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Johan Verbeeck[1], Vaiva Deltuvaite-Thomas[2], Ben Berckmoes[3], Tomasz Burzykowski[1,2], Marc Aerts[1] , Olivier Thas[1,4,5], Marc Buyse [2,6] and Geert Molenberghs[1,7]

Prepared using sagej.cls

Abstract

In reliability theory, diagnostic accuracy, and clinical trials, the quantity $P(X > Y) + 1/2P(X = Y)$, also known as the Probabilistic Index (PI), is a common treatment effect measure when comparing two groups of observations. The quantity $P(X > Y) - P(Y > X)$, a linear transformation of PI known as the net benefit, has also been advocated as an intuitively appealing treatment effect measure. Parametric estimation of PI has received a lot of attention in the past 40 years, with the formulation of the Uniformly Minimum-Variance Unbiased Estimator (UMVUE) for many distributions. However, the non-parametric Mann-Whitney estimator of the PI, is also known to be UMVUE in some situations. To understand this seeming contradiction, in this paper a systematic comparison is performed between the non-parametric estimator for the PI and parametric UMVUE estimators in various settings. We show that the Mann–Whitney estimator is always an unbiased estimator of the PI with univariate, completely observed data, while the parametric UMVUE is not when the distribution is misspecified. Additionally, the Mann-Whitney estimator is the UMVUE when observations belong to an unrestricted family. When observations come from a more restrictive family of distributions, the loss in efficiency for the non-parametric estimator is limited in realistic clinical scenarios. In conclusion, the Mann-Whitney estimator is simple to use and a reliable estimator for the PI and net benefit in realistic clinical scenarios.

Keywords

Completeness, Relative efficiency, Net benefit, Probabilistic Index, UMVUE, Unbiased, Wilcoxon–Mann–Whitney.

1. Introduction

In randomized clinical trials, much attention has been paid to measures of treatment effect that use all pairwise comparisons between the two groups of observations. These measures derive from the Mann-Whitney formulation of the Wilcoxon test, in which the observations of a continuous variable in a new treatment group (X) are paired with the observations of the same variable in a reference treatment group (Y)[1,2]. Using such pairwise comparisons, a natural measure of treatment effect is the probability that the response of a random subject given the new treatment is better than the response of a random subject given the reference treatment, $P(X > Y) + 1/2P(X = Y)$[3,4]. This measure, called probabilistic index (PI) by Acion et al.[5], has received various names in the literature, depending on the field of application, including the probability of a superior outcome[6], concordance index c[7], proportion of similar responses[6], among others[8–11]. The PI is extensively studied in stress-strength models in reliability theory[8,12,13] and Receiver Operating Characteristics (ROC) curve analysis in diagnostic accuracy[7,14]. Recently, the PI has gained renewed interest in medical applications[15], and models have been proposed for this measure[16]. Another closely related measure, $P(X > Y) - P(Y > X)$, called the net benefit, was proposed by Buyse[17]

[1]Data Science Institute (DSI), Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Hasselt, Belgium.
[2]International Drug Development Institute (IDDI), 1341 Louvain-la-Neuve, Belgium.
[3]Department of Mathematics, University of Antwerp, Antwerp, Belgium.
[4]National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong, New South Wales, Australia.
[5]Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium.
[6]International Drug Development Institute (IDDI), San Francisco, CA, USA.
[7]Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), KU Leuven, Leuven, Belgium.

Corresponding author:
Johan Verbeeck, Data Science Institute (DSI), Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Agoralaan building D, 3590 Hasselt, Belgium.
Email: johan.verbeeck@uhasselt.be

Prepared using sagej.cls

and advocated as a clinically meaningful and versatile treatment effect measure[18–20]. This measure is the probability that the response of a random subject given the new treatment is better than the response of a random subject given the reference treatment, minus the probability of the opposite situation. The advantage of this measure is that it does not contain $1/2P(X = Y)$. In diagnostic accuracy, this measure is known as Somers' D[21]. In the remainder of our paper, we will focus on the PI, given the extensive literature on this measure, but all the developed results apply to any linear transformation of the PI, including the net benefit.

In order to estimate the PI, both parametric and non-parametric statistics have been proposed[12]. Since it was observed that a parametric maximum likelihood estimator can be biased, several authors derived the optimal uniformly minimum-variance unbiased estimator (UMVUE) for many types of distributions of the observations[12,22], including the normal distribution[23], the exponential distribution[24,25], and the Poisson distribution[26].

On the other hand, a non-parametric estimator for the PI was proposed by Mann and Whitney[1]. Consider two samples of observations $X_i, \ldots, X_n$ and $Y_j, \ldots, Y_m$ ($i = 1, \ldots, n$ and $j = 1, \ldots, m$), with outcome vectors X and Y. Additionally, consider the function $h(X_i, Y_j)$, which maps a pair of observations, one observation of each group, to a score U_{ij}, according to the scoring algorithm:

$$h(X_i, Y_j) = U_{ij} = \begin{cases} 1 & X_i > Y_j, \\ 0 & X_i < Y_j, \\ 1/2 & \text{otherwise.} \end{cases}$$

The Mann–Whitney statistic is the sum of the U_{ij} scores:

$$\widehat{U}_{XY} = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij},$$

and the PI is then estimated by:

$$\widehat{PI}_e = \frac{\widehat{U}_{XY}}{nm} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} U_{ij}, \tag{1}$$

which is, additionally, a generalized U-statistic[27].

While this non-parametric Mann–Whitney estimator of the PI does not make any assumption about the distribution of the observations and is thus generally applicable, it may be less efficient than a parametric UMVUE. Despite the rich literature on the PI, however, this potential loss in efficiency has not been formally evaluated.

Interestingly, several sources prove the unbiasedness, sufficiency, and completeness for the Mann–Whitney statistic[28,29]. This suggests, upon applying the Lehmann–Scheffé[30,31] and Rao–Blackwell theorem[32,33], that the Mann–Whitney estimator is also the UMVUE of the PI. Moreover, generalized U-statistics theory[27] states that a U-statistic is the UMVUE in a sufficiently large family of distributions. In other words, it must be true that the Mann–Whitney estimator equals at least some of the parametric UMVUE for the PI. It is unknown, however, for which distributions of the observations this is true.

The goal of this paper is to systematically compare the basic statistical properties and the efficiency of the parametric UMVUE and the non-parametric Mann–Whitney estimator for the PI in a univariate outcome data setting, without missing or censored observations. In Section 2, unbiasedness, sufficiency, and completeness of the parametric UMVUE and the Mann–Whitney estimator are compared for several distributions. Next, in Section 3, the efficiency of the Mann–Whitney estimator is compared to the UMVUE, both in theory, where feasible, and by means of simulation otherwise. In Section 4, the consequences of our results

for statistics related to the PI are reviewed. Finally, both the parametric UMVUE and the Mann-Whitney estimator are applied to a real data set in Section 5.

2. Unbiasedness, sufficiency and completeness of the Mann-Whitney estimator

A statistic is the UMVUE of a parameter when the statistic is an unbiased estimator of the parameter and when the statistic is sufficient and complete[30–33]. First we will evaluate the unbiasedness of the non-parametric Mann–Whitney estimator (1) and compare it to the parametric UMVUE of the PI. Next, we evaluate the sufficiency and completeness for various distributions of the observations. If the UMVUE coincides with the Mann–Whitney estimator, then the estimator in (1) is also complete and has uniformly minimum variance. If the two estimators are different, then the Mann–Whitney estimator is less efficient than the UMVUE.

To compute both estimators, assume X_i $(i = 1, \ldots, n)$ are independent observations with common distribution $P \in \mathcal{P}$ and Y_j $(j = 1, \ldots, m)$ are independent observations with common distribution $Q \in \mathcal{Q}$. Additionally, assume no dependence between X_i and Y_j. The Mann–Whitney estimator expressed in parameters of P and Q is then obtained by computing the $P(X > Y) + 1/2P(X = Y)$ for P and Q. The UMVUE of the probabilistic index is obtained by taking an estimator (K_n, L_m) that is sufficient and complete for $\mathcal{P} \times \mathcal{Q}$ and by computing the conditional expectation of the PI given the estimator:

$$\widehat{UMVUE} = \mathbb{E}[1_{\{X_1 > Y_1\}} + \frac{1}{2}1_{\{Y_1 > X_1\}} \mid K_n, L_m].$$

Note that the Mann–Whitney estimator does not need any assumption for the distribution of the observations, as $\widehat{PI}_e$ is distribution independent. On the other hand, the $\widehat{UMVUE}$, always requires a distributional assumption.

It is easy to see that the $\widehat{PI}_e$ in (1) is always an unbiased estimator of the PI, in the sense that for all $(P, Q) \in \mathcal{P} \times \mathcal{Q}$

$$\mathbb{E}_{P,Q}\left[\widehat{PI}_e\right] = PI.$$

Additionally, it should be clear that a parametric UMVUE estimator for the PI is only unbiased, and thus the UMVUE, when the observations follow indeed the assumed parametric distribution.

To evaluate the sufficiency and completeness of the Mann–Whitney estimator, we make use of the results on the UMVUE for the PI in the literature, which we have unified for various choices of $\mathcal{P}$ and $\mathcal{Q}$, with $\mathcal{P} = \mathcal{Q}$, in Appendix I. If the UMVUE $= \widehat{PI}_e$, the estimator in (1) is complete and has uniformly minimum variance over the family $\mathcal{P} \times \mathcal{Q}$ in the sense that, for any statistic $T(X_i, Y_j)$ that is unbiased for PI, it holds for all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ that

$$\text{Var}_{P,Q}[\widehat{PI}_e] \leq \text{Var}_{P,Q}[T].$$

Calculations (Appendix I) show that if the family $\mathcal{P} = \mathcal{Q}$ is unrestricted in the sense that it consists of all distributions if the support is finite, or that it consists of all absolutely continuous distributions if the support is the entire real line, then the UMVUE $= \widehat{PI}_e$ and the Mann–Whitney estimator is the UMVUE for the PI. The Bernoulli distribution is an example of such an unrestricted family. Parametric families such as those consisting of normal distributions or exponential distributions typically fail to be unrestricted.

However, if $\mathcal{P} = \mathcal{Q}$ is a restricted parametric family, like, for example, Poisson, normal, or exponential distributions, then the UMVUE in general does not coincide with $\widehat{PI}_e$, from which it follows that the Mann–Whitney estimator cannot be complete nor sufficient.

3. Efficiency of the Mann-Whitney estimator

As shown in the previous section, whether the Mann–Whitney estimator is the UMVUE or not depends on the assumed distribution of the observations. The practical implication of the Mann–Whitney estimator not being complete when observations come from certain specified distributions, is investigated theoretically in this section and by comparing the variances of the UMVUE and of the non-parametric estimator by means of the relative efficiency (RE). The relative efficiency is defined with the UMVUE in the numerator.

Theorem 1, presented in Appendix II, which holds for any assumed distribution of the observations, shows that, when m and n are moderately large, the extra variance of $\widehat{PI}_e$ on top of the variance of the UMVUE, is negligible and vanishes with rate $1/n + 1/m$. Theorem 1 holds even if P and Q belong to different families of distributions. In particular, for any $m, n > 1$,

$$\mathrm{Var}_{P,Q}[\widehat{PI}_e] \leq \frac{1}{4}\left(\frac{1}{mn} + \frac{1}{n}\frac{m-1}{m} + \frac{1}{m}\frac{n-1}{n}\right) = \frac{m+n-1}{4mn}.$$

Given the invariance of Theorem 1 to the assumed distribution, it follows that although the always unbiased Mann-Whitney estimator fails to be the UMVUE in the restricted families of distributions, the loss in efficiency as compared to the UMVUE, is bounded. On the other hand, the parametric UMVUE is unbiased and the most precise estimator, only when the observations come indeed from the specified distribution. Simulations

of 10,000 samples with sample size 200 and PI values ranging from 0.5 to 0.95 from exponential and Weibull distributions with the same scale parameters (Figure 1) and simulations from a normal and lognormal distribution (Figure 2) show that, in contrast to the Mann-Whitney estimator, the parametric UMVUE is sensitive to misspecification of the distribution.

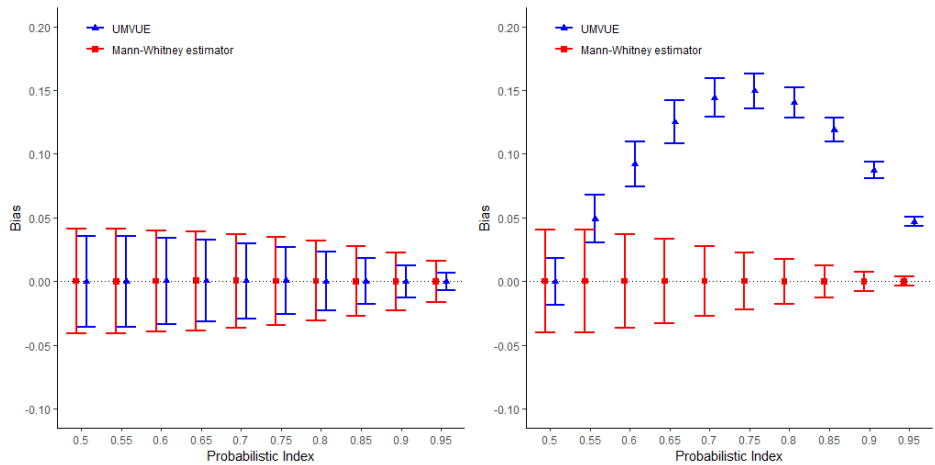


Figure 1. Average of 10,000 estimators (dot) $\pm$ standard deviation (error bars). (a) Bias and standard deviation of the UMVUE based on an exponential distribution (Appendix I) and the Mann–Whitney estimator for exponentially distributed simulated data ($N = 200$). (b) Bias and standard deviation of the UMVUE based on an exponential distribution (Appendix I) and the Mann-Whitney estimator for Weibull distributed simulated data with shape parameter equal to 2 ($N = 200$).

To quantify the loss in efficiency of the Mann–Whitney estimator when the observations come from a particular restricted family distribution, the relative efficiency is evaluated in this particular restricted family.

We will start with the simple setting of a univariate normal case with known variance and treat subsequently the normal case with unknown variance. In the univariate normal case with unit variance, the explicit formulas for Var$[UMVUE]$ and Var$[\widehat{PI}_e]$ can be derived (Appendix

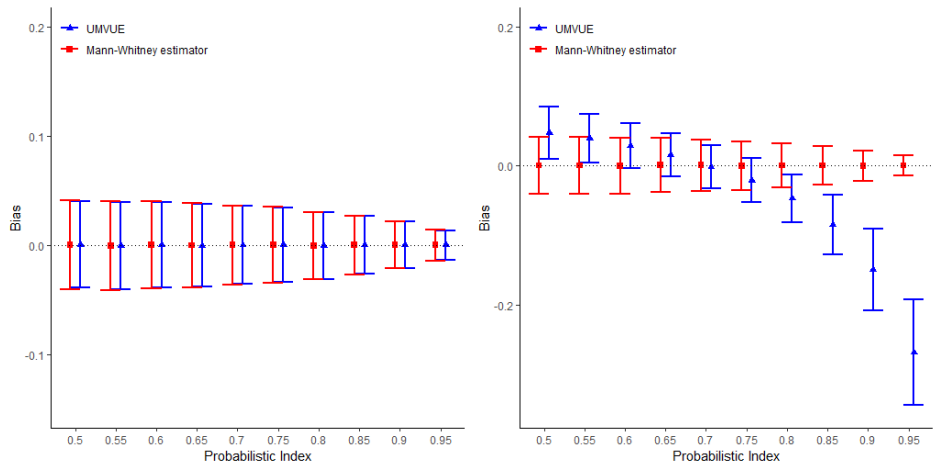


Figure 2. Average of 10,000 estimators (dot) ± standard deviation (error bars). (a) Bias and standard deviation of the UMVUE based on a normal distribution with unknown variance and the Mann–Whitney estimator for normal distributed simulated data ($N = 200$). (b) Bias and standard deviation of the UMVUE based on a normal distribution with unknown variance and the Mann-Whitney estimator for lognormal distributed simulated data ($N = 200$).

III). This allows us to formulate the RE and study the properties of the asymptotic relative efficiency (ARE). In the univariate normal case with unit variance, the ARE equals

$$\lim_{m,n\to\infty} \frac{\text{Var}[UMVUE]}{\text{Var}[\widehat{PI}_e]} = \frac{\exp\left[-\frac{1}{2}(\mu_1-\mu_2)^2\right]}{8\pi\left[T\left(\frac{\mu_1-\mu_2}{\sqrt{2}},1\right)-T\left(\frac{\mu_1-\mu_2}{\sqrt{2}},\frac{1}{\sqrt{3}}\right)\right]}, \quad (2)$$

with T denoting Owen's T-function[34], (see Appendix IV).

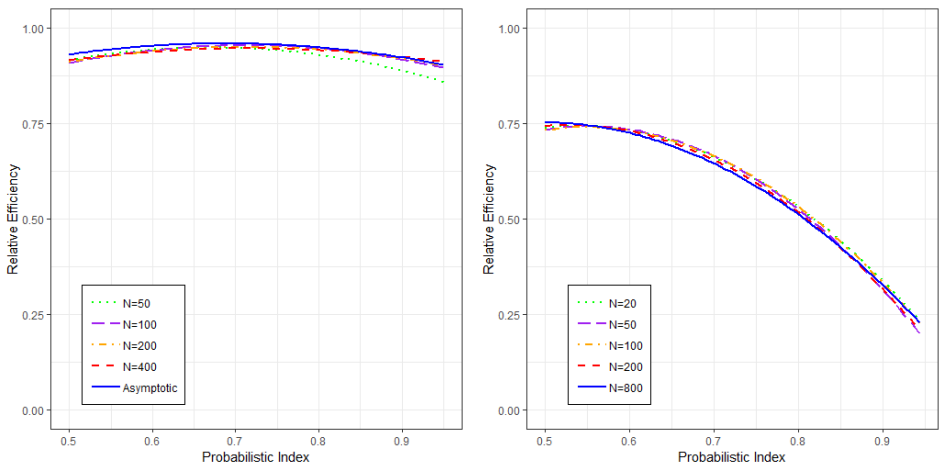
From (2) it follows that if there is no treatment effect, that is if $\mu_1 = \mu_2$, the ARE for the univariate normal case is maximal and equal to

$$ARE = \frac{1}{4\left[\arctan(1)-\arctan\left(1/\sqrt{3}\right)\right]} = \frac{3}{\pi} \approx 0.955.$$

Note that the ARE of the UMVUE relative to the Mann–Whitney estimator is exactly equal to the efficiency loss of the Mann–Whitney

test relative to the t-test for normally distributed data[35]. With increasing treatment effect, that is if $|\mu| = |\mu_1 - \mu_2| > 0$, the ARE will decrease. Thus, ARE $\leq 3/\pi$.

Simulations of 10,000 samples (Figures 3a, 3b, and Appendix V) show that the impact of the sample size on the RE is minimal compared to the impact of the treatment effect. When observations come from a normal distribution with unit variance, the RE remains, depending on the sample size, between 83.5% and 86.5%, with a PI of 0.76 or an effect size equal to 1[17]. In clinical trials, effect sizes are often not larger than the unit standard deviation. In reliability studies and ROC curve analyses, the PI is however much closer to 1, where the efficiency of the Mann–Whitney estimator rapidly becomes very poor.



(a) **(b)**

Figure 3. (a) Relative efficiency of the empirical variance of the UMVUE for normally distributed data with unknown variance versus the theoretical variance of the Mann–Whitney estimator, with varying PI and sample size. (b) Relative efficiency of the empirical variance of the UMVUE for exponentially distributed data versus the theoretical variance of the Mann–Whitney estimator, with varying PI and sample size. Note that a smoothed line was drawn through the empirical results for visual clarity.

It is very tedious to derive the theoretical UMVUE variance for other distributions. Therefore, we rely on simulations to evaluate the variance and the RE in these settings. Since we are interested in a clinical application, the distributions we focus on are distributions frequently present in clinical data, such as a normal distribution with unknown variance, exponential distribution, Weibull distribution, and Poisson distribution. Varying the sample size and the PI, the variance is estimated by the empirical variance of 10,000 samples of each combination.

In clinical trials often the variance is not known, so the evaluation of the RE for the normal distribution is repeated, but now assuming an unknown variance. The UMVUE for the normal case with unknown variance is known[22]. With an unknown variance the RE is very stable and does not decrease below 92%, except for values of the PI larger than 0.9 (Figure 3a, Appendix V). Neither the sample size nor the treatment effect impact the RE much (Figure 3a, Appendix V).

The RE for exponential data is in general much lower than with normal data and it does not exceed 78% (Figure 3b). Similar to the normal case, the sample size does not seem to influence the RE for exponential data, while increasing treatment effect decreases the RE (Figure 3b and Appendix V). However, for values of the PI between 0.5 and 0.6, which correspond to a reduction in hazard from 0 to 0.33[17], the RE remains stable between 70 and 78% (Figure 3b and Appendix V).

For the Weibull distribution the simulations could not be performed, since calculation of the UMVUE of the PI supersedes the numerical limitation of frequently used statistical software. The UMVUE of the Weibull[22] contains factorials, that take arguments that depend on the sample size, but statistical software put limits on the argument that a factorial can take. For total sample sizes of 100 observations, which

is unrealistically low for clinical trials, the numerical limit is already surpassed. So, for observations belonging to the Weibull distribution, only the non-parametric estimator can be used in practice for clinical trials.

When observations come from a Poisson distribution, the simulations suffer as well from the numerical limitation for the UMVUE from total sample sizes above 200 observations. Up until a total sample size of 200 observations, the relative efficiency follows a similar pattern as for normal data (Appendix V).

The simulations show that the loss in efficiency for the always unbiased Mann–Whitney estimator is limited to 30% in realistic clinical scenarios.

4. Related measures of treatment effect

Recently, measures of treatment effect related to the PI have been suggested in the literature[17,36,37]. As mentioned in the introduction, the net benefit (Δ) was defined by Buyse[17] as $P(X > Y) - P(X < Y)$. It is a linear transformation of the probabilistic index:

$$\Delta = 2P(X > Y) - 1.$$

Pocock et al.[38] suggested the win ratio (Ψ), defined as $P(X > Y)/P(X < Y)$. Very recently, Brunner[39] proposed the success-odds, defined as $[P(X > Y) + 1/2P(X = Y)]/[P(X < Y) + 1/2P(X = Y)]$.

The proposed estimators for the net benefit and win ratio are, just as the Mann–Whitney estimator, non-parametric statistics based on pairwise comparisons. These estimators, however, use different functions $h(X_i, Y_j)$, to map a pair of observations, again one observation of each

group, to a score U_{ij} $(i = 1, .., n$ and $j = 1, .., m)$. The net benefit estimator uses the function:

$$h(X_i, Y_j) = U_{ij} = \begin{cases} 1 & X_i > Y_j, \\ -1 & X_i < Y_j, \\ 0 & \text{otherwise.} \end{cases}$$

The net benefit is then estimated by the following U-statistic:

$$\widehat{\Delta}_e = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}.$$

It is lineary related to the Mann–Whitney statistic ($\widehat{U}_{XY}$), and thus also to the Wilcoxon statistic[5], if X and Y are univariate and without missing or censored data:

$$\widehat{\Delta}_e = 2 \left(\frac{\widehat{U}_{XY}}{nm} \right) - 1. \tag{3}$$

The PI $\in [0, 1]$ and is centered around 1/2, while the $\Delta \in [-1, 1]$ and is centered around 0.

Alternatively, the net benefit estimator can be expressed as the difference between two U-statistics $\widehat{U}^T$ and $\widehat{U}^C$, where $\widehat{U}^T = 1/(nm) \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}^T$ and $\widehat{U}^C = 1/(nm) \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}^C$, with U_{ij}^T equal to 1 if U_{ij} is 1 and 0 otherwise, and U_{ij}^C equal to 1 if U_{ij} is -1 and 0 otherwise. Note that the non-parametric estimator of the PI for continuous distributions equals $\widehat{U}^T$.

An estimator of the win ratio is given by the ratio of two U-statistics $\widehat{U}^T$ and $\widehat{U}^C$:

$$\widehat{\Psi} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}^T}{\sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}^C}. \tag{4}$$

Due to its linearity, the results on the basic properties and efficiency of the Mann-Whitney estimator extend to the net benefit estimator. However, they do not extend to a non-linear function such as the win ratio estimator. Specifically, the win ratio statistic defined in (4) is not unbiased, because the expectation of the win ratio is not equal to the ratio of the expectations. The delta method approximation of the expectation of the win ratio equals

$$\mathbb{E}\left[\widehat{\Psi}\right] = \mathbb{E}\left[\frac{\widehat{U}^T}{\widehat{U}^C}\right] \approx \frac{\mathbb{E}[\widehat{U}^T]}{\mathbb{E}[\widehat{U}^C]} - \frac{\text{Cov}[\widehat{U}^T, \widehat{U}^C]}{\left(\mathbb{E}[\widehat{U}^C]\right)^2} + \frac{\text{Var}[\widehat{U}^C]\mathbb{E}[\widehat{U}^T]}{\left(\mathbb{E}[\widehat{U}^C]\right)^3}.$$

Hence, the win ratio is nearly unbiased only for large values of $\mathbb{E}[\widehat{U}^C]$. This will only be the case when the sample size is sufficiently large and the difference between the two sets of observations is not too large.

Inference for the net benefit can be based on its permutation distribution[17], bootstrap distribution, or on asymptotic U-statistic theory[37]. Verbeeck et al.[40] compared different inferential methods for the net benefit and concluded that the exact permutation variance is a good estimator for the variance under the null hypothesis $\Delta = 0$, even in very small sample sizes. Under the alternative hypothesis, the exact bootstrap variance is preferred. The notation 'exact' refers to the fact that there is a relatively simple closed form formula for the variance of the distribution of the Mann-Whitney or the net benefit estimators that one would obtain if every possible permutation or bootstrap sample were considered once[40]. Since the net benefit and PI are linearly related, the exact permutation and bootstrap tests are also valid inferential methods for the PI. See Appendix V for some simulated examples.

5. Example

The Age Related Macular Degeneration (ARMD) trial was a randomized, multicentre, doubled-blind, placebo-controlled study evaluating interferon-α versus placebo in age-related macular degeneration subjects[41]. The disease causes progressively loss of vision in the subjects. The progression of the disease is monitored by evaluating the visual acuity in the subjects, which is measured in the trial by the ability of the subjects to read lines of letters on standardized vision charts. We will evaluate the difference between the visual acuity at 52 weeks and the baseline visual acuity. The outcome can be expressed either by a continuous or a binary variable. The difference in number of letters read at 52 weeks versus baseline is the continuous outcome.

The binary outcome is obtained by assigning a negative outcome when less letters are read at 52 weeks then at baseline and a positive outcome when equal or more letters are read. Equal letters read is considered positive, since there is no further degeneration of the vision.

The ARMD dataset contains 481 patients, but here we consider only 240 patients randomized equally to interferon-α administered subcutaneously at 6 million international units (MIU) three times a week, or corresponding placebo. Only 195 patients are used in our analyses (90 on interferon-α, 105 on placebo) since 45 subjects did not have a visual acuity measurement at 52 weeks. The hypothesis test is based on the exact permutation method and the confidence interval (CI) is obtained from the exact bootstrap method. We use the two-sided significance level of 5%.

The Mann–Whitney estimator of the PI for the continuous visual acuity outcome is equal to 0.4135 (95%CI [0.3321,0.4950], p = 0.038), while the parametric UMVUE based on a normal distribution of the observations is equal to 0.4239 (95%CI [0.3449,0.5029], p = 0.059). Thus, the non-parametric estimator suggests some harm of interferon-α

in terms of visual acuity, while the parametric UMVUE just fails to show a difference between interferon-α and placebo. The variance of the parametric UMVUE is indeed smaller than the variance of the non-parametric estimator (0.00162 respectively 0.00173), although the difference is minimal. The net benefit estimator, obtained from the Mann–Whitney statistic by (3), is equal to -0.1729 (95%CI [-0.0100,-0.3358], p = 0.038). It is worth noting that the estimated value of the net benefit, unlike for the PI, is negative, which immediately signals the negative effect of interferon-α.

To understand the difference between the non-parametric estimator and the parametric UMVUE, it is important to remember, that the parametric UMVUE is based on the assumption that the data are normally distributed. Formal tests (the p-values of Shapiro-Wilk tests for the interferon-α and placebo arm are respectively p = 0.02 and p = 0.18) and QQ-plots (Figure 5) and boxplot of the data suggest modest deviations from normality. Moreover, the Mann-Whitney estimator for the PI is always unbiased with complete data. This means that the non-parametric test and estimator may be preferred to the parametric test.

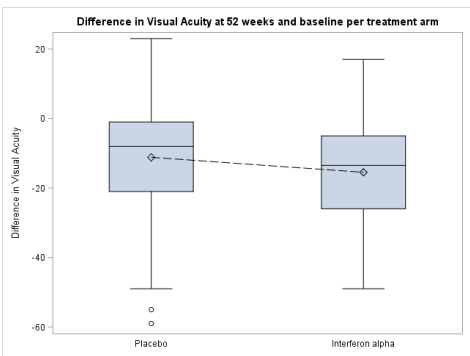
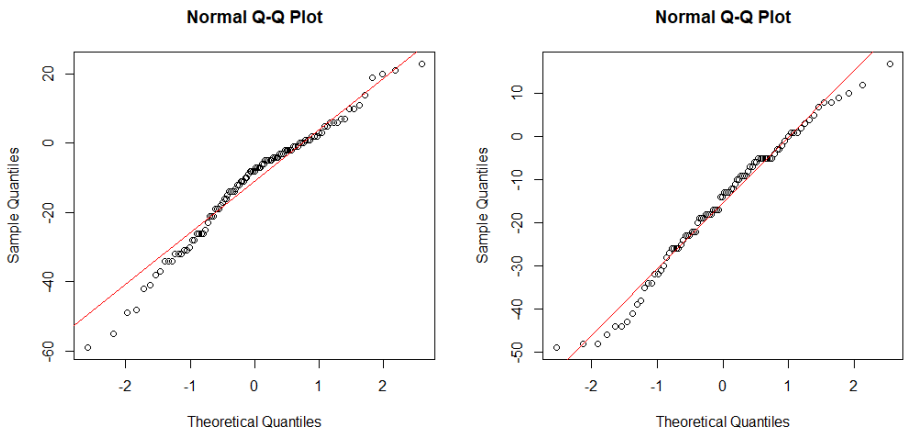


Figure 4. Box-plot for the placebo and interferon alpha arm of the contiunuous difference in visual acuity at 52 weeks and baseline of the ARMD trial.

Using the relationship between the PI and the difference in means (Table 1, Appendix II), a treatment difference of -4.92 letters (95%CI [-9.8, -0.28], p = 0.038) is obtained from the Mann-Whitney estimator. The non-parametric analysis thus shows that interferon-α significantly worsens the mean visual acuity at 52 weeks. For comparison, a t-test, which also assumes a normal distribution of the data, fails to detect a treatment effect, with a difference in visual acuity of -4.30 letters (95%CI [-8.79, 0.20], p= 0.061).



(a) **(b)**

Figure 5. (a) Normal QQ-plot for the interferon alpha arm of the ARMD trial. (b) Normal QQ-plot for the placebo arm of the ARMD trial.

The PI for the binary visual acuity outcome does not show any statistically-signifcant effect of interferon-α on the visual acuity after 52 weeks (PI = 0.4683 [0.4129, 0.5236], p = 0.26). This PI is much closer to the null value of 0.5, which reflects the loss of information when dichotomizing the continuous outcome. This loss is clearly noted by evaluating the number of pairs that contribute to the resulting PI. In the continuous outcome analysis, 3819 pairs indicate a benefit for interferon-α, 5453 pairs a better outcome for placebo, while only 178

pairs are tied. For the binary outcome a lot more pairs are evaluated as a tie (6554 pairs), while 1148 pairs favor interferon-α and 1748 pairs favor placebo. The net benefit for the binary visual acuity outcome is equal to -0.0635 [-0.1730, 0.046], p = 0.26. Interestingly, the net benefit for a binary outcome is exactly equal to the difference in proportions between the treatment groups[17]: 15.56% of patients lost at least 1 letter of visual acuity in the interferon-α group versus 21.90% in the placebo group. The Chi-squared test for the difference in proportions (-0.0635 [-0.1724, 0.045], p = 0.26) coincides well with the test based on the net benefit.

6. Discussion

We have studied the efficiency, unbiasedness, sufficiency, and completeness of the non-parametric and parametric estimators for the PI and net benefit, in the case of complete and uncensored observations. It has been shown that when observations come from an unrestricted family of distributions, such as the Bernoulli distribution, the non-parametric Mann–Whitney estimator for the PI is the UMVUE. In contrast, when observations come from a restricted parametric family of distributions, such as a normal, Poisson, or exponential distribution, the Mann–Whitney estimator is unbiased, but neither sufficient nor complete.

The additional variance due to the non-parametric estimator not being complete, converges to zero with rate $1/n + 1/m$. In contrast, the UMVUE depends on the assumed distribution and is only unbiased and has minimal variance when the observations come from the specified distribution. The Mann–Whitney estimator is thus robust against misspecification of the distribution and its loss in efficiency compared to the UMVUE is limited.

When observations do come from the assumed distribution, the relative efficiency of the UMVUE and the non-parametric estimator is shown to be acceptable in common clinical scenarios, where the PI is often close to 0.5. For reliability studies and ROC curve analyses, the efficiency of the Mann–Whitney estimator is worse, since the PI is often much closer to 1. When observations come from a normal or Poisson distribution and the PI is not larger than 0.76, which corresponds to the unit effect size (difference in means equals the standard deviation), the RE is between 85 and 95.5%. In clinical trials, an effect size is usually not larger than the standard deviation, and thus the Mann–Whitney estimator will generally have acceptable precision. Moreover, when the variance is considered unknown for the normal distribution, the loss in efficiency is not larger than 8%. When observations come from an exponential distribution, which is linked to survival data, the reported PI is not larger than 0.6[20,42–44], corresponding to a hazard reduction of 0.33. In clinical practice, the expected hazard reduction often does not exceed 30%, for which simulations show that the expected maximum loss in efficiency, will be between 22-30%. Note however that the work presented here ignores the impact of censoring.

The decision to use the UMVUE or the non-parametric estimator thus depends on the underlying distribution, the expected treatment effect, and the complexity of the UMVUE. The variance for the non-parametric statistic has always the same structure, while for the UMVUE it differs per the underlying distribution and it can be quite cumbersome to calculate. For some distributions, such as the exponential distribution, no closed form expression for the variance is available. In this case, the variance can be obtained by re-sampling techniques or numerical integration.

In practice, the variance of the Mann–Whitney estimator is estimated and not based on a distributional assumption. Consequently, the efficiency of the Mann–Whitney estimator will deviate minimally from the theoretical values. In reliability and diagnostic accuracy studies, many inferential methods have been proposed[45,46] to estimate the variance of the Mann–Whitney estimator. Due to the linearity between the net benefit and the Mann-Whitney estimator, the exact permutation and exact bootstrap tests are additional very precise inferential methods for the PI.

The general results on the properties of the non-parametric estimator for the PI presented in this paper coincide with the evidence found in the literature. The proofs that the Mann–Whitney estimator is the UMVUE for the PI are either based on ranks[28] or restricted to the Bernoulli setting[29], while in the U-statistics theory the unrestricted family setting is exploited to prove that U-statistics are UMVUE. As such, all evidence in the literature falls in the unrestricted family setting, where it is shown in this paper that the non-parametric estimator is the UMVUE.

In practice, the PI or net benefit can be used in much more complex settings than those presented here. Data may be incomplete or censored, several outcomes may be of interest, and thresholds of clinical relevance can be used in generalized pairwise comparisons[17]. The same theoretical properties should be investigated in these more realistic scenarios to get a full picture of the theoretical properties of the Mann-Whitney estimator for the PI or net benefit.

Acknowledgements

We thank the Pharmacological Therapy for Macular Degeneration Study Group and the company F. Hoffmann LaRoche for granting permission to use the ARMD dataset, a copy of which can be found online (https://ibiostat.be/online-resources) in the datasets2005.zip file.

Declaration of conflicting interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work is partly funded by an unrestricted grant provided by the European Cardiovascular Research Institute (ECRI). BB is post doctoral fellow at the Fund for Scientific Research of Flanders (FWO).

References

1. Mann H and Whitney D. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947; 18(1): 50–60.
2. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin* 1945; 1: 80–83.
3. Fay M, Brittain E, Shih J et al. Causal estimands and confidence intervals associated with Wilcoxon–Mann–Whitney tests in randomized experiments. *Statistics in Medicine* 2018; 37(20): 2923–2937.
4. Greenland S, Fay M, Brittain E et al. On causal inferences for personalized medicine: How hidden causal assumptions led to erroneous causal claims about the D–value. *The American Statistician* 2019; DOI:10.1080/00031305.2019.15757712.
5. Acion A, Peterson J, Temple S et al. Probabilistic index: an intuitive non-parametric approach to measuring the size of the treatment effects. *Statistics in Medicine* 2006; 25: 591–602.
6. Grissom R. Probability of the superior outcome of one treatment over another. *Journal of Applied Psychology* 1994; 79: 314–316.
7. Harrell F, Califf R, Pryor D et al. Evaluating the yield of medical tests. *Journal of the American Medical Association* 1982; 247: 2543–2546.
8. Church JD and Harris B. The estimation of reliability from stress-strength relationships. *Technometrics* 1970; 12: 49–54.
9. McGraw K and Wong S. A common language effect size statistic. *Psychological Bulletin* 1992; 111: 361–365.
10. Senn S. Testing for individual and population equivalence based on the proportion of similar responses. *Statistics in Medicine* 1997; 16: 1303–1306.
11. Lehmann E. *Nonparametrics. Statistical methods based on ranks*. Prentice Hall, Upper Saddle River, New Jersey, USA, 1998.

12. Kotz S, Lumelskii Y and Pensky M. *The Stress-Strength Model and its Generalizations*. Singapore: World Scientific Press, 2003.
13. Birnbaum Z. On a use of Mann-Whitney statistics. *Proc Third Berkeley Symp in Math Statist Probab* 1956; 1: 13–17.
14. Pepe S. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, 2010.
15. Brown P and Ezekowitz J. Composite end points in clinical trials of heart failure therapy. how do we measure the effect size? *Circulation Heart Failure* 2017; 10(1). DOI:10.1161/CIRCHEARTFAILURE.116.003222.
16. Thas O, De Neve J, Clement L et al. Probabilistic index models. *Journal of the Royal Statistical Society - Series B* 2012; 74: 623–671.
17. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine* 2010; 29: 3245–3257.
18. Moser B and MacCann M. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clinical Trials* 2008; 5: 248–252.
19. Buyse M. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clinical Trials* 2008; 5: 641–642.
20. Péron J, Roy P, Ozenne B et al. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. *JAMA Oncology* 2016; 2(7): 901–905.
21. Somers R. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 1962; 27(6): 799–811.
22. Obradovic M, Jovanovic M and Milosevic B. Optimal unbiased estimates of $P(X < Y)$ for some families if distributions. *Metodoloski zveski* 2014; 11(1): 21–29.
23. Downtown F. On the estimation of $Pr(Y < X)$ in the normal case. *Technometrics* 1973; 15: 551–558.
24. Tong H. A note on the estimation of $Pr(Y < X)$ in the exponential case. *Technometrics* 1974; 16: 625.
25. Tong H. On the estimation of $Pr(Y < X)$ for exponential families. *IEEE Transactions on Reliability* 1977; 26: 54–56.
26. Belyaev Y and Lumelskii Y. Multidimensional poisson walks. *Journal of Mathematical Sciences* 1988; 40: 162–165.
27. Lee A. *U-Statistics: Theory and Practice*. Chapman & Hall/CRC, 1990.
28. Lehmann E and Romano J. *Testing statistical hypotheses*. Springer, 2005.

29. Klotz J. The Wilcoxon, ties, and the computer. *Journal of the American Statistical Association* 1966; 61(315): 772–787.
30. Lehmann E and Scheffé H. Completeness, similar regions, and unbiased estimation. i. *Sankhya* 1950; 10(4): 305–340.
31. Lehmann E and Scheffé H. Completeness, similar regions, and unbiased estimation. ii. *Sankhya* 1955; 15(3): 219–236.
32. Blackwell D. Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statistics* 1947; 18(1): 105–110.
33. Rao C. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society* 1945; 37(3): 81–91.
34. Owen D. Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics* 1956; 27: 1075–1090.
35. Lehmann E. *Elements of Large-Sample Theory*. Springer, 1999.
36. Finkelstein D and Schoenfeld D. Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine* 1999; 18: 1341–1354.
37. Ramchandani R, Schoenfeld D and Finkelstein D. Global rank tests for multiple, possibly censored, outcomes. *Biometrics* 2016; 72: 926–935.
38. Pocock S, Ariti C, Collier T et al. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal* 2012; 33: 176–182.
39. Brunner E. Success-odds: An improved win-ratio. *arXiv:200209273 [statME]* 2020; .
40. Verbeeck J, Ozenne B and Anderson W. Evaluation of inferential methods for the net benefit and win ratio statistics. *Journal of Biopharmaceutical Statistics* 2020; DOI:10.1080/10543406.2020.1730873.
41. Pharmacological therapy for macular degeneration study group. Interferon α-iia is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. results of a prospective randomized placebo-controlled clinical trial. *Archives of Ophthalmology* 1997; 115: 865–872.
42. Péron J, Roy P, Ding K et al. Assessing the benefit-risk of new treatments using generalised pairwise comparisons: the case of erlotinib in pancreatic cancer. *Britisch Journal of Cancer* 2015; 112(6): 971–976.
43. Péron J, Roy P, Conroy T et al. An assessment of the benefit-risk balance of folfirinox in metastatic pancreatic adenocarcinoma. *Oncotarget* 2016; 7(50): 82953–82960.
44. Péron J, Lambert A, Munier S et al. Assessing long-term survival benefits of immune checkpoint inhibitors using the net survival benefit. *Journal of National Cancer Institute*

2019; 111(11): 1186–1191.

45. Zhou W. Statistical inference for $P(X < Y)$. *Statistics in Medicine* 2008; 27: 257–279.
46. Newcombe R. Confidence intervals for an effect size measure based on the Mann– Whitney statistic. part 2: asymptotic methods and evaluation. *Statistics in Medicine* 2006; 25: 559–573.
47. Morris C and Lock K. Unifying the named natural exponential families and their relatives. *The American Statistician* 2009; 63(3): 247–53.

Appendix I. Explicit formulas for the PI and the UMVUE in various univariate cases

Let $A \subset \mathbb{R}$ be a Borel set and $\mathcal{P}$ a family of probability measures on A. Furthermore, $X_1, \ldots, X_n$ are independent observations with law $P \in \mathcal{P}$ and $Y_1, \ldots, Y_m$ independent observations with law $Q \in \mathcal{P}$, with no dependence between X_i and Y_j. Here we will derive explicit formulas for the Probabilistic Index

$$PI(P,Q) = \mathbb{P}_{P,Q}[X_1 > Y_1] + \frac{1}{2}\mathbb{P}_{P,Q}[Y_1 = X_1]$$

and its UMVUE

$$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$$

for various choices of A and $\mathcal{P}$. Table 1 summarizes all the formulas. The calculations follow below.

The cases where $\mathcal{P}$ is unrestricted

We will call a family $\mathcal{P}$ of probability measures on A *unrestricted* if

a) either $A = \mathbb{R}$ and $\mathcal{P}$ consists of all absolutely continuous distributions on A, that is the distributions with a continuous density with respect to the Lebesgue measure on A,

b) or A is finite and $\mathcal{P}$ contains all probability distributions on A.

Notice that parametric families such as those consisting of normal distributions or exponential distributions typically fail to be unrestricted. Throughout, we fix a Borel set $A \subset \mathbb{R}$ and an unrestricted family $\mathcal{P}$ of probability measures on A.

Consider the crude statistic

$$T(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = U_{11},$$

with

$$U_{11} = 1_{\{X_1 > Y_1\}} + \frac{1}{2} 1_{\{Y_1 = X_1\}}.$$

It is easily seen that T is unbiased for $PI(P, Q)$.

Now consider the order statistics $\left(X_{(1)}, \ldots, X_{(n)}\right)$ and $\left(Y_{(1)}, \ldots, Y_{(m)}\right)$. Then, because the family $\mathcal{P}$ is unrestricted, $\left(X_{(1)}, \ldots, X_{(n)}; Y_{(1)}, \ldots, Y_{(m)}\right)$ is complete and sufficient for the family $\mathcal{P} \times \mathcal{P}$ p. 3[27,28] p. 37 and p. 118.

Then, the Lehmann-Scheffé theorem shows that

$$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \mathbb{E}_{P,Q}\left[T \mid \left(X_{(1)}, \ldots, X_{(n)}\right), \left(Y_{(1)}, \ldots, Y_{(m)}\right)\right]$$

is the unique UMVUE of $PI(P, Q)$.

Recall that the PI can be estimated by

$$\widehat{PI}_e(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij},$$

with

$$U_{ij} = 1_{\{X_i > Y_j\}} + \frac{1}{2} 1_{\{Y_j = X_i\}}.$$

It is a function of $\left(X_{(1)}, \ldots, X_{(n)}\right)$ and $\left(Y_{(1)}, \ldots, Y_{(m)}\right)$ because it is symmetric in the X_i and symmetric in the Y_j. More formally, it is a generalized U-statistic[27] p. 37. Finally, for an event $E \subset$

$\sigma\left(\left(X_{(1)},\ldots,X_{(n)}\right),\left(Y_{(1)},\ldots,Y_{(m)}\right)\right)$,

$$\mathbb{E}[\widehat{PI}_e 1_E] = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}\left[U_{ij}1_E\right] = \mathbb{E}[U_{11}1_E] = \mathbb{E}[T1_E],$$

from which it follows that $UMVUE = \widehat{PI}_e$.

The case where $\mathcal{P}$ is the Poisson family

Now we consider the Poisson family by taking $A = \mathbb{N}$ and $\mathcal{P} = \left\{\text{Pois}(\lambda_1) \mid \lambda_1 \in \mathbb{R}_0^+\right\}$. So the X_i have distribution $\text{Pois}(\lambda_1)$ and the Y_j distribution $\text{Pois}(\lambda_2)$.

So the PI can be estimated by

$$\begin{aligned} & PI(\lambda_1,\lambda_2) \\ &= \mathbb{P}_{\lambda_1,\lambda_2}[X_1 > Y_1] + \frac{1}{2}\mathbb{P}_{\lambda_1,\lambda_2}[Y_1 = X_1] \\ &= \left(\sum_{i\in\mathbb{N},0\le j<i}\frac{\lambda_1^i\lambda_2^j}{i!j!} + \frac{1}{2}\sum_{(i=j)\in\mathbb{N}}\frac{\lambda_1^i\lambda_2^j}{i!j!}\right)\exp(-(\lambda_1+\lambda_2)). \end{aligned}$$

Put

$$K_n = \sum_{i=1}^{n} X_i$$

and

$$L_m = \sum_{j=1}^{m} Y_j.$$

Then it follows from classical theory that the estimator (K_n, L_m) is complete and sufficient for (λ_1,λ_2).

As before it now follows from the Lehmann-Scheffé theorem that the UMVUE for $PI(\lambda_1,\lambda_2)$ is given by

$$\widehat{UMVU}E(X_1,\ldots,X_n,Y_1,\ldots,Y_m) = \mathbb{E}_{\lambda_1,\lambda_2}[1_{\{X_1>Y_1\}} + \frac{1}{2}1_{\{Y_1>X_1\}} \mid K_n, L_m],$$

which, by sufficiency, does not depend on λ_1 and λ_2.

It is well known[47] that, for $0 \leq x \leq k$ and $0 \leq y \leq l$, the following explicit expression holds for the conditional density:

$$\mathbb{P}[X_i = x, Y_j = y \mid K_n = k, L_m = l] = \binom{k}{x}\left(\frac{1}{n}\right)^x\left(1-\frac{1}{n}\right)^{k-x}\binom{l}{y}\left(\frac{1}{m}\right)^y\left(1-\frac{1}{m}\right)^{l-y}.$$

We conclude that

$$\begin{aligned} &UMVUE(X_1,\ldots,X_n,Y_1,\ldots,Y_m) \\ &= \sum_{0\leq x\leq K_n, 0\leq y<\min\{x-1,L_m\}} \binom{K_n}{x}\left(\frac{1}{n}\right)^x\left(1-\frac{1}{n}\right)^{K_n-x}\binom{L_m}{y}\left(\frac{1}{m}\right)^y \\ &\quad \left(1-\frac{1}{m}\right)^{L_m-y} + \frac{1}{2}\sum_{0\leq y=x\leq\min\{K_n,L_m\}} \binom{K_n}{x}\left(\frac{1}{n}\right)^x\left(1-\frac{1}{n}\right)^{K_n-x} \\ &\quad \binom{L_m}{y}\left(\frac{1}{m}\right)^y\left(1-\frac{1}{m}\right)^{L_m-y} \end{aligned}$$

is the UMVUE of $PI(\lambda_1, \lambda_2)$.

The case where $\mathcal{P}$ is a normal family

Now consider the normal setting by taking $A = \mathbb{R}$ and $\mathcal{P} = \{N(\mu, 1) \mid \mu \in \mathbb{R}\}$. Thus the X_i have distribution $N(\mu_1, 1)$ and the Y_j distribution $N(\mu_2, 1)$.

Straightforward calculation shows that the PI is

$$\begin{aligned} PI(\mu_1,\mu_2) &= \mathbb{P}_{\mu_1,\mu_2}[X_1 > Y_1] + \frac{1}{2}\mathbb{P}_{\mu_1,\mu_2}[Y_1 = X_1] = \mathbb{P}_{\mu_1,\mu_2}[X_1 > Y_1] \\ &= \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right), \end{aligned}$$

with Φ the standard normal CDF.

Put

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\overline{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j.$$

Then it follows from classical theory that the estimator $(\overline{X}_n, \overline{Y}_m)$ is complete and sufficient for (μ_1, μ_2).

As before it now follows from the Lehmann-Scheffé theorem that the UMVUE for $PI(\mu_1, \mu_2)$ is given by

$$\widehat{UMVUE}(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \mathbb{E}_{\mu_1,\mu_2}[1_{\{X_1 > Y_1\}} \mid \overline{X}_n, \overline{Y}_m],$$

which, by sufficiency, does not depend on μ_1 and μ_2.

It is well known[47] that, for $n, m \geq 2$, the following explicit expression holds for the conditional density:

$$f_{X_i,Y_j|\overline{X}_n,\overline{Y}_m}(x, y \mid k, l) = \sqrt{\frac{n}{n-1}}\phi\left(\frac{x-k}{\sqrt{\frac{n-1}{n}}}\right)\sqrt{\frac{m}{m-1}}\phi\left(\frac{y-l}{\sqrt{\frac{m-1}{m}}}\right),$$

with ϕ the standard normal density. It follows that, for $n, m \geq 2$,

$$\begin{aligned}
&\mathbb{E}_{\mu_1,\mu_2}[1_{\{X_1>Y_1\}} \mid \overline{X}_n, \overline{Y}_m] \\
&\quad = \mathbb{P}_{\mu_1,\mu_2}\left[X_1 > Y_1 \mid \overline{X}_n, \overline{Y}_m\right] \\
&\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{s}\sqrt{\frac{n}{n-1}}\phi\left(\frac{s-k}{\sqrt{\frac{n-1}{n}}}\right)\sqrt{\frac{m}{m-1}}\phi\left(\frac{t-l}{\sqrt{\frac{m-1}{m}}}\right)dtds,
\end{aligned}$$

which is calculated to equal

$$\Phi\left(\frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{n-1}{n} + \frac{m-1}{m}}}\right).$$

We conclude that, for $n, m \geq 2$,

$$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \Phi\left(\frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{n-1}{n} + \frac{m-1}{m}}}\right).$$

The case where $\mathcal{P}$ is the exponential family

Now consider the exponential family setting by taking $A = \mathbb{R}^+$ and $\mathcal{P} = \{\text{Exp}(\lambda) \mid \lambda \in \mathbb{R}_0^+\}$. Thus the X_i have distribution $\text{Exp}(\lambda_1)$ and the Y_j distribution $\text{Exp}(\lambda_2)$

A straightforward calculation shows that the PI is given by

$$\begin{aligned} PI(\lambda_1, \lambda_2) &= \mathbb{P}_{\lambda_1,\lambda_2}[X_1 > Y_1] + \frac{1}{2}\mathbb{P}_{\lambda_1,\lambda_2}[Y_1 = X_1] = \mathbb{P}_{\lambda_1,\lambda_2}[X_1 > Y_1] \\ &= \frac{\lambda_2}{\lambda_2 + \lambda_1}. \end{aligned}$$

Put

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\overline{Y}_m = \frac{1}{m}\sum_{j=1}^{m} Y_j.$$

Then $(\overline{X}_n, \overline{Y}_m)$ is complete and sufficient for (λ_1, λ_2).

Again the Lehmann-Scheffé theorem entails that the UMVUE for $PI(\lambda_1, \lambda_2)$ is given by

$$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m) = \mathbb{E}_{\lambda_1,\lambda_2}[1_{\{X_1 > Y_1\}} \mid \overline{X}_n, \overline{Y}_m],$$

which, by sufficiency, does not depend on λ_1 and λ_2.

It is well known[47] that, for $n, m \geq 2$, the following explicit expression holds for the conditional density:

$$f_{X_1,Y_1|\overline{X}_n,\overline{Y}_m}(x,y \mid k,l) = \begin{cases} \frac{n-1}{kn}(1-\frac{x}{kn})^{n-2}\frac{m-1}{lm}(1-\frac{y}{lm})^{m-2} & \text{if} \quad 0 \leq x \leq nk, 0 \leq y \leq lm \\ 0 & \text{elsewhere} \end{cases}$$

It follows that, for $n, m \geq 2$,

$$\begin{aligned} &\mathbb{E}_{\mu_1,\mu_2}[1_{\{X_1>Y_1\}} \mid \overline{X}_n, \overline{Y}_m] \\ &= \mathbb{P}_{\mu_1,\mu_2}\left[X_1 > Y_1 \mid \overline{X}_n, \overline{Y}_m\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{s} \frac{n-1}{kn}(1-\frac{s}{kn})^{n-2}\frac{m-1}{lm}(1-\frac{t}{lm})^{m-2}dtds, \end{aligned}$$

which is calculated to equal

$$\begin{cases} 1 - {}_2F_1\left(1-m, 1; n; \frac{n\overline{X}_n}{m\overline{Y}_m}\right) & \text{if} \quad n\overline{X}_n < m\overline{Y}_m \\ 1 - \frac{n-1}{n}\frac{\overline{Y}_m}{\overline{X}_n} \cdot {}_2F_1\left(2-n, 1; m+1; \frac{m\overline{Y}_m}{n\overline{X}_n}\right) & \text{if} \quad n\overline{X}_n > m\overline{Y}_m \end{cases},$$

with ${}_2F_1$ the *hypergeometric function*, i.e.

$${}_2F_1(a,b;c;z) = \frac{1}{B(b,c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a}dt,$$

where B denotes the beta function.

We conclude that, for $n, m \geq 2$,

$$UMVUE(X_1,\ldots,X_n,Y_1,\ldots,Y_m) = \begin{cases} 1 - {}_2F_1\left(1-m, 1; n; \frac{n\overline{X}_n}{m\overline{Y}_m}\right) & \text{if} \quad n\overline{X}_n < m\overline{Y}_m \\ 1 - \frac{n-1}{n}\frac{\overline{Y}_m}{\overline{X}_n} \cdot {}_2F_1\left(2-n, 1; m+1; \frac{m\overline{Y}_m}{n\overline{X}_n}\right) & \text{if} \quad n\overline{X}_n > m\overline{Y}_m \end{cases}.$$

Appendix II. Proof of Theorem 1: The variance of $\widehat{PI}_e$ vanishes with rate $1/n + 1/m$

Theorem 1. *For all $P \in \mathcal{P}$ and $Q \in \mathcal{Q}$ and for all n and m,*

$$Var_{P,Q}\left[\widehat{PI}\right] \leq \frac{1}{4}\left(\frac{1}{mn} + \frac{1}{n}\frac{m-1}{m} + \frac{1}{m}\frac{n-1}{n}\right). \tag{5}$$

Proof. We have

$$\begin{aligned}
&\mathrm{Var}_{P,Q}\left[\widehat{PI}\right] \\
&= \mathrm{Var}_{P,Q}\left[\frac{1}{mn}\sum_{i,j} U_{i,j}\right] \\
&= \frac{1}{m^2n^2}\left(\sum_{i,j}\mathrm{Var}_{P,Q}[U_{i,j}] + 2\sum_{(i,j)\neq(i',j')}\mathrm{Cov}_{P,Q}[U_{i,j}, U_{i'j'}]\right) \\
&= \frac{1}{m^2n^2}\sum_{i,j}\mathrm{Var}_{P,Q}[U_{i,j}] + \frac{2}{m^2n^2}\sum_{i,j\neq j'}\mathrm{Cov}_{P,Q}[U_{i,j}, U_{i,j'}] \\
&\quad + \frac{2}{m^2n^2}\sum_{i\neq i',j}\mathrm{Cov}_{P,Q}[U_{i,j}, U_{i',j}] \\
&= \frac{1}{mn}\mathrm{Var}_{P,Q}[U_{1,1}] + \frac{2n}{m^2n^2}\binom{m}{2}\mathrm{Cov}_{P,Q}[U_{1,1}, U_{1,2}] \\
&\quad + \frac{2m}{m^2n^2}\binom{n}{2}\mathrm{Cov}_{P,Q}[U_{1,1}, U_{2,1}].
\end{aligned} \tag{6}$$

Because $U_{i,j}$ takes values between 0 and 1, we have, by Popoviciu's inequality,

$$\mathrm{Var}_{P,Q}[U_{i,j}] \leq \frac{1}{4}[1-0]^2 = 1/4, \tag{7}$$

which, by the Cauchy-Schwarz inequality, leads to

$$|\mathrm{Cov}_{P,Q}[U_{i,j}, U_{i',j'}]| \leq \sqrt{\mathrm{Var}_{P,Q}[U_{i,j}]\mathrm{Var}_{P,Q}[U_{i',j'}]} \leq 1/4. \tag{8}$$

Table 1. Formulas for PI and $UMVUE$ in various univariate cases

$X_1, \ldots, X_n \sim P$ and $Y_1, \ldots, Y_m \sim Q$			
support of P and Q	family $\mathcal{P} = \mathcal{Q}$ of distributions	Probabilistic Index	UMVUE of the Probabilistic Index
$A \subset \mathbb{R}$ finite	$P, Q \in \mathcal{P}(A)$	$PI(P,Q)$ $= \mathbb{P}[X_1 > Y_1] + \frac{1}{2}\mathbb{P}[Y_1 = X_1]$	$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $= \widehat{PI}_e(X_1, \ldots, X_n, Y_1, \ldots, Y_m))$
$\mathbb{N}$	$P = \text{Pois}(\lambda_1)$, $Q = \text{Pois}(\lambda_2)$	$PI(\lambda_1, \lambda_2)$ $= \left(\sum_{i \in \mathbb{N}, 0 \le j < i} \frac{\lambda_1^i \lambda_2^j}{i!j!} + \frac{1}{2} \sum_{i=j \in \mathbb{N}} \frac{\lambda_1^i \lambda_2^j}{i!j!} \right)$ $\exp(-(\lambda_1 + \lambda_2))$	$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $= \sum_{0 \le x \le K_n, 0 \le y < \min\{x-1, L_m\}} \binom{K_n}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{K_n - x} \binom{L_m}{y} \left(\frac{1}{m}\right)^y \left(1 - \frac{1}{m}\right)^{L_m - y}$ $+ \frac{1}{2} \sum_{0 \le x = y \le \min\{K_n, L_m\}} \binom{K_n}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{K_n - x} \binom{L_m}{y} \left(\frac{1}{m}\right)^y \left(1 - \frac{1}{m}\right)^{L_m - y}$
$\mathbb{R}$ absolutely continuous	P, Q $P = N(\mu_1, 1)$, $Q = N(\mu_2, 1)$	$PI(P,Q)$ $= \mathbb{P}[X_1 > Y_1]$ $PI(\mu_1, \mu_2)$ $= \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{2}}\right)$	$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $= \widehat{PI}_e(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $= \Phi\left(\frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{n-1}{n} + \frac{m-1}{m}}}\right)$
$\mathbb{R}^+$	$P = \text{Exp}(\lambda_1)$, $Q = \text{Exp}(\lambda_2)$	$PI(\lambda_1, \lambda_2)$ $= \frac{\lambda_2}{\lambda_2 + \lambda_1}$	$UMVUE(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ $= \begin{cases} 1 - {}_2F_1\left(1 - m, 1; n; \frac{n\overline{X}_n}{m\overline{Y}_m}\right) & \text{if} \quad n\overline{X}_n < m\overline{Y}_m \\ 1 - \frac{n-1}{n} \frac{\overline{Y}_m}{\overline{X}_n} \cdot {}_2F_1\left(2 - n, 1; m+1; \frac{m\overline{Y}_m}{n\overline{X}_n}\right) & \text{if} \quad n\overline{X}_n > m\overline{Y}_m \end{cases}$

Combining (6), (7) and (8), finishes the proof.

Appendix III. Explicit formulas for the variance of UMVUE and $\widehat{PI}_e$

The variance of $\widehat{PI}_e$

Let $X_i \sim N(\mu_1, 1)$, iid, $i = 1, \ldots, n$ and $Y_j \sim N(\mu_2, 1)$, iid, $j = 1, \ldots, m$.

$$\widehat{PI}_e = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$$

It folows that

$$E[\widehat{PI}_e] = E[U_{ij}] = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{2}}\right).$$

Moreover,

$$\mathrm{Var}[\widehat{PI}_e] = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{Var}[U_{ij}] + \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{i' \neq i} \sum_{j' \neq j} \mathrm{Cov}[U_{ij}, U_{i'j'}].$$

It can be shown that:

- $\mathrm{Var}[U_{ij}] = \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right) - \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right)^2$
- $\mathrm{Cov}[U_{ij}, U_{i'j'}] = 0$
- $\mathrm{Cov}[U_{ij}, U_{ij'}] = \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right) - \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right)^2 - 2T\left(\frac{\mu_2-\mu_1}{\sqrt{2}}; \frac{1}{\sqrt{3}}\right)$
- $\mathrm{Cov}[U_{ij}, U_{i'j}] = \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right) - \Phi\left(\frac{\mu_1-\mu_2}{\sqrt{2}}\right)^2 - 2T\left(\frac{\mu_1-\mu_2}{\sqrt{2}}; \frac{1}{\sqrt{3}}\right)$

Then, based on the property of Owen's T-function:

$$T(a, 1) = \frac{\Phi(a) - \Phi^2(a)}{2},$$

we have that:

$$\text{Var}[\widehat{PI}_e] = \frac{2(m+n-1)}{nm} T\left(\frac{\mu_1-\mu_2}{\sqrt{2}}; 1\right) - \frac{2(m-1)}{nm} T\left(\frac{\mu_2-\mu_1}{\sqrt{2}}; \frac{1}{\sqrt{3}}\right) - \frac{2(n-1)}{nm} T\left(\frac{\mu_1-\mu_2}{\sqrt{2}}; \frac{1}{\sqrt{3}}\right).$$

The variance of the UMVUE estimator

Let $X_i \sim N(\mu_1, 1)$, iid, $i = 1, \ldots, n$ and $Y_j \sim N(\mu_2, 1)$, iid, $j = 1, \ldots, m$.

$$\widehat{UMVUE} = \Phi\left(\frac{\bar{X}_n - \bar{Y}_m}{\sqrt{2 - \left(\frac{1}{n} + \frac{1}{m}\right)}}\right)$$

Denote:

$$\bar{X}_n - \bar{Y}_m = z$$
$$\frac{1}{n} + \frac{1}{m} = x$$

Let:

$$\bar{X}_n - \bar{Y}_m = z \sim N(\mu_1 - \mu_2, \frac{1}{n} + \frac{1}{m}) = N(\mu_z, x)$$

$$\text{Var}[\widehat{UMVUE}] = E[\widehat{UMVUE}^2] - E[\widehat{UMVUE}]^2$$

1.
$$E[\widehat{UMVUE}] = \Phi\left(\frac{\mu_z}{\sqrt{2}}\right)$$

2.
$$E[\widehat{UMVUE}^2] = \Phi\left(\frac{\mu_z}{\sqrt{2}}\right) - 2T\left(\frac{\mu_z}{\sqrt{2}}; \frac{1}{\sqrt{\frac{2+x}{2-x}}}\right)$$

The overall variance for $\widehat{UMVUE}$ is then:

$$Var[\widehat{UMVUE}] = 2T\left(\frac{\mu_z}{\sqrt{2}}; 1\right) - 2T\left(\frac{\mu_z}{\sqrt{2}}; \sqrt{\frac{2-x}{2+x}}\right)$$

Appendix IV. Explicit formulas for the relative efficiency and asymptotic relative efficiency of the UMVUE and the non-parametric estimator in the univariate normal case

Relative Efficiency

Assume X_i follows $N(\mu_1, 1)$ and Y_j follows $N(\mu_2, 1)$. Let T be Owen's T-function, $\mu = \mu_1 - \mu_2$, and $x = \frac{1}{n} + \frac{1}{m}$. We denote

$$\tau_\mu(x) = T\left(\frac{\mu}{\sqrt{2}}, \sqrt{\frac{2-x}{2+x}}\right).$$

It was shown in Appendix III that

$$\text{Var}[\widehat{PI}_e] = 2x\left[\tau_\mu(0) - \tau_\mu(1)\right] + \frac{2}{mn}\left[2\tau_\mu(1) - \tau_\mu(0)\right]$$

and

$$\text{Var}[\widehat{UMVUE}] = 2\left[\tau_\mu(0) - \tau_\mu(x)\right].$$

Then, the inverse of the relative efficiency is given by

$$\frac{\text{Var}[\widehat{PI}_e]}{\text{Var}[\widehat{UMVUE}]} = [\tau_\mu(1) - \tau_\mu(0)]\frac{x}{\tau_\mu(x) - \tau_\mu(0)} + R(m,n)$$

with

$$R(m,n) = [\tau_\mu(0) - 2\tau_\mu(1)]\frac{\frac{1}{mn}}{\tau_\mu(x) - \tau_\mu(0)}.$$

Asymptotic relative efficiency

The asymptotic relative efficiency (ARE) studies the asymptotic behavior if $m, n \to \infty$, or equivalently, $x \to 0$. Notice that

$$0 \leq \frac{1}{mn} \leq x^2.$$

Thus,

$$|R(m,n)| \leq |\tau_\mu(0) - 2\tau_\mu(1)| \left| \frac{x}{\tau_\mu(x) - \tau_\mu(0)} \right| x.$$

Now, since

$$\frac{x}{\tau_\mu(x) - \tau_\mu(0)} \to \frac{1}{\tau'_\mu(0)} \text{ when } x \to 0,$$

we infer that

$$R(m,n) \to 0 \text{ when } m, n \to \infty,$$

and

$$\frac{\text{Var}[\widehat{PI}_e]}{\text{Var}[\widehat{UMVUE}]} \to \frac{\tau_\mu(1) - \tau_\mu(0)}{\tau'_\mu(0)} \text{ when } m, n \to \infty.$$

Recalling that

$$T(h,a) = \frac{1}{2\pi} \int_0^a \frac{\exp\left[-\frac{1}{2}h^2(1+x^2)\right]}{1+x^2} dx,$$

it is easily calculated that

$$\tau'_\mu(0) = -\frac{1}{8\pi} \exp\left[-\frac{1}{2}\mu^2\right].$$

The inverse of the asymptotic relative efficiency (ARE) is thus

$$\lim_{m,n\to\infty} \frac{\text{Var}[\widehat{PI}_e]}{\text{Var}[\widehat{UMVUE}]} = 8\pi \frac{T\left(\frac{\mu_1-\mu_2}{\sqrt{2}}, 1\right) - T\left(\frac{\mu_1-\mu_2}{\sqrt{2}}, \frac{1}{\sqrt{3}}\right)}{\exp\left[-\frac{1}{2}(\mu_1-\mu_2)^2\right]}.$$

Appendix V. Simulations of the variances and relative efficiency (RE)

We provide simulations of the variance and the relative efficiency (RE) of UMVUE variance $\widehat{PI}_U$versus the theoretical variance of the non-parametric estimator $\widehat{PI}_e$, versus the exact permutation distribution variance $\widehat{PI}_p$, versus the exact bootstrap distribution variance $\widehat{PI}_b$ or versus the asymptotic U-statistic distribution variance $\widehat{PI}_u$.

Table 2. Variance and relative efficiency in the univariate normal case with unknown means and unit variance, for varying total sample size (N) and Probabilistic Index (PI).

N	**PI**	$Var_{\widehat{PI}_e}$	$Var_{\widehat{PI}_U}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_e}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_p}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_b}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_u}$
50	0.50	0.0068	0.0064	0.9365	0.9365	0.9469	0.9080
	0.60	0.0065	0.0060	0.9254	0.8788	0.9175	0.8806
	0.70	0.0055	0.0049	0.8902	0.7160	0.8795	0.8496
	0.80	0.0039	0.0032	0.8201	0.4677	0.8014	0.7878
	0.90	0.0019	0.0013	0.6832	0.1879	0.6570	0.6866
100	0.50	0.0034	0.0032	0.9455	0.9455	0.9502	0.9309
	0.60	0.0032	0.0030	0.9341	0.8868	0.9293	0.9110
	0.70	0.0027	0.0024	0.8978	0.7210	0.8911	0.8765
	0.80	0.0019	0.0016	0.8256	0.4690	0.8173	0.8112
	0.90	0.0009	0.0006	0.6855	0.1868	0.6738	0.6900
200	0.50	0.0017	0.0016	0.9502	0.9502	0.9523	0.9427
	0.60	0.0016	0.0015	0.9385	0.8909	0.9364	0.9273
	0.70	0.0013	0.0012	0.9016	0.7236	0.8996	0.8925
	0.80	0.0009	0.0008	0.8284	0.4697	0.8269	0.8241
	0.90	0.0005	0.0003	0.6868	0.1862	0.6811	0.6897
400	0.50	0.0008	0.0008	0.9526	0.9526	0.9536	0.9489
	0.60	0.0008	0.0007	0.9408	0.8929	0.9389	0.9344
	0.70	0.0007	0.0006	0.9036	0.7249	0.9021	0.8986
	0.80	0.0005	0.0004	0.8298	0.4701	0.8283	0.8269
	0.90	0.0002	0.0002	0.6874	0.1859	0.6860	0.6904

Table 3. Variance and relative efficiency in the univariate normal case with unknown means and unknown unequal variances for each group, for varying total sample size (N) and probabilistic index (PI). Note, that the variance for the UMVUE statistic in this case cannot be derived and thus the average empirical variance was used.

N	**PI**	$Var_{\widehat{PI}_e}$	$Var_{\widehat{PI}_U}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_e}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_p}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_b}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_u}$
50	0.50	0.0070	0.0064	0.9227	0.9432	0.9386	0.8989
	0.60	0.0065	0.0061	0.9331	0.8907	0.9097	0.8721
	0.70	0.0056	0.0053	0.9337	0.7748	0.9276	0.8944
	0.80	0.0040	0.0037	0.9452	0.5502	0.9118	0.8936
	0.90	0.0020	0.0019	0.9208	0.2722	0.9103	0.9447
100	0.50	0.0034	0.0032	0.9271	0.9492	0.9348	0.9152
	0.60	0.0033	0.0030	0.9276	0.9028	0.9215	0.9027
	0.70	0.0028	0.0026	0.9372	0.7800	0.9369	0.9206
	0.80	0.0020	0.0019	0.9454	0.5636	0.9401	0.9310
	0.90	0.0010	0.0009	0.9304	0.2706	0.9256	0.9444
200	0.50	0.0017	0.0016	0.9303	0.9316	0.9122	0.9027
	0.60	0.0016	0.0015	0.9364	0.9186	0.9388	0.9294
	0.70	0.0014	0.0013	0.9413	0.7879	0.9461	0.9380
	0.80	0.0010	0.0010	0.9585	0.5725	0.9637	0.9594
	0.90	0.0005	0.0005	0.9340	0.2702	0.9252	0.9345
400	0.50	0.0009	0.0008	0.9340	0.9779	0.9541	0.9491
	0.60	0.0008	0.0008	0.9325	0.9221	0.9419	0.9372
	0.70	0.0007	0.0006	0.9203	0.7685	0.9246	0.9207
	0.80	0.0005	0.0005	0.9574	0.5645	0.9495	0.9474
	0.90	0.0002	0.0002	0.9380	0.2711	0.9324	0.9372

Table 4. Variance and relative efficiency in the univariate exponentional case, for varying total sample size (N) and probabilistic index (PI). Note, that the variance for the UMVUE statistic in this case cannot be derived and thus the average empirical variance was used.

N	**PI**	$Var_{\widehat{PI}_e}$	$Var_{\widehat{PI}_U}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_e}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_p}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_b}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_u}$
50	0.50	0.0067	0.0050	0.7478	0.7364	0.7443	0.7137
	0.60	0.0063	0.0047	0.7437	0.6936	0.7259	0.6966
	0.70	0.0055	0.0036	0.6643	0.5332	0.6500	0.6277
	0.80	0.0040	0.0021	0.5256	0.3063	0.5143	0.5045
	0.90	0.0021	0.0007	0.3187	0.1001	0.3200	0.3300
100	0.50	0.0033	0.0025	0.7540	0.7477	0.7513	0.7361
	0.60	0.0032	0.0023	0.7249	0.6831	0.7148	0.7007
	0.70	0.0028	0.0018	0.6480	0.5332	0.6553	0.6444
	0.80	0.0020	0.0010	0.5323	0.3110	0.5251	0.5202
	0.90	0.0010	0.0003	0.3297	0.1003	0.3193	0.3244
200	0.50	0.0017	0.0013	0.7421	0.7541	0.7558	0.7482
	0.60	0.0016	0.0012	0.7298	0.7025	0.7363	0.7291
	0.70	0.0013	0.0009	0.6538	0.5207	0.6404	0.6352
	0.80	0.0010	0.0005	0.5214	0.3096	0.5249	0.5227
	0.90	0.0005	0.0002	0.3241	0.0985	0.3180	0.3205
400	0.50	0.0009	0.0006	0.7538	0.7724	0.7732	0.7693
	0.60	0.0008	0.0006	0.7280	0.6894	0.7241	0.7206
	0.70	0.0007	0.0004	0.6570	0.5262	0.6488	0.6462
	0.80	0.0005	0.0003	0.5296	0.3095	0.5255	0.5244
	0.90	0.0003	0.0001	0.3198	0.0964	0.3136	0.3149

Table 5. Variance and relative efficiency in the univariate Poisson case, for varying total sample size (N) and Probabilistic Index (PI). Note, that the variance for the UMVUE statistic in this case cannot be derived and thus the average empirical variance was used.

N	**PI**	$Var_{\widehat{PI}_e}$	$Var_{\widehat{PI}_U}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_e}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_p}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_b}$	**RE** $\frac{\widehat{PI}_U}{\widehat{PI}_u}$
50	0.50	0.0065	0.0061	0.9408	0.9244	0.9566	1.0025
	0.54	0.0066	0.0062	0.9397	0.9300	0.9711	1.0183
	0.62	0.0069	0.0061	0.9292	0.8542	0.9478	0.9969
	0.71	0.0052	0.0046	0.8965	0.6940	0.9119	0.9680
	0.92	0.0015	0.0010	0.6454	0.1469	0.5973	0.7066
100	0.50	0.0033	0.0031	0.9427	0.9489	0.9652	0.9880
	0.54	0.0032	0.0030	0.9444	0.9177	0.9418	0.9643
	0.62	0.0030	0.0028	0.9344	0.8543	0.9318	0.9555
	0.71	0.0026	0.0023	0.8908	0.6852	0.8912	0.9182
	0.92	0.0008	0.0005	0.6657	0.1500	0.6393	0.6974
200	0.50	0.0017	0.0016	0.9464	0.9691	0.9774	0.9890
	0.54	0.0016	0.0015	0.9480	0.9357	0.9523	0.9636
	0.62	0.0015	0.0014	0.9322	0.8623	0.9346	0.9465
	0.71	0.0013	0.0012	0.8996	0.6992	0.9044	0.9180
	0.91	0.0004	0.0002	0.5191	0.1165	0.5132	0.5363