W

Weak Convergence of Probability Measures

Milan Merkle
Professor, Faculty of Electrical Engineering
University of Belgrade, Belgrade, Serbia

Weak Convergence of Probability Measures on $\mathbb{R}^d$

Among several concepts of convergence that are being used in Probability theory, the weak convergence has a special role, as it is related not to values of random variables, but to their probability distributions. In a simplest case of a sequence $\{X_n\}$ of real valued random variables (or vectors with values in $\mathbb{R}^d$, $d \geq 1$) defined on probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$, we say that a sequence $\{X_n\}$ converges *weakly* (or *in law*) to a random variable X if

$$\lim_{n\to+\infty} F_n(x) = F(x) \tag{1}$$

for each $x \in \mathbb{R}$ where the function F is continuous. Here F_n and F are distribution functions of X_n and X, respectively. The notation for this kind of convergence is $X_n \Rightarrow X$, or $X_n \stackrel{\mathcal{L}}{\to} X$. The convergence defined by (1) can be as well thought of as being a convergence of corresponding distributions, i.e., probability measures defined on $(\mathbb{R}^d, \mathcal{B}^d)$ by $\mu_n(B) = P_n(\{\omega \in \Omega_n \mid X_n(\omega) \in B\})$, where $B \in \mathcal{B}$ and $\mathcal{B}$ is a Borel sigma-field on $\mathbb{R}$. Hence we say also that a sequence of probability measures μ_n on $(\mathbb{R}^d, \mathcal{B}^d)$ converges weakly to μ, in notation $\mu_n \Rightarrow \mu$.

The following result is known as Lévy-Cramér Continuity Theorem.

Theorem 1 *Let μ_n be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}^d)$. Then $\mu_n \Rightarrow \mu$ if and only if the corresponding ►characteristic functions converge pointwise:*

$$\lim_{n\to+\infty} \mathrm{E}\, e^{i\langle t, X_n\rangle} = \mathrm{E}\, e^{i\langle t, X\rangle} \quad \textit{for every } t \in \mathbb{R}^d,$$

where X_n, X are random vectors with distributions μ_n and μ, respectively.

In $d = 1$, the Lévy's metric d_L (see Lévy 1937) is defined as a distance between two univariate distribution functions

$$\begin{aligned} d_L(F,G) = \inf\{\varepsilon > 0 \mid F(x-\varepsilon) - \varepsilon \leq G(x) \\ \leq F(x+\varepsilon) + \varepsilon \quad \text{for all } x \in \mathbb{R}\}. \end{aligned}$$

If φ_F and φ_G are characteristic functions that correspond to F, G, respectively, then

$$d_L(F,G) \leq \frac{1}{\pi}\int_0^T |\varphi_F(t) - \varphi_G(t)|\frac{\mathrm{d}t}{t} + 2e\frac{\log T}{T}, \quad T > e.$$

The weak convergence $X_n \Rightarrow X$ is implied by convergence in probability, and consequently with all stronger notions of convergence (with probability one and in the pth mean, $p \geq 1$). To see that the weak convergence does not imply nearness of values of corresponding random variables, we may recall that for any symmetric random variable ($\mathcal{N}(0,1)$, say), X and $-X$ have the same distribution. However, for any given sequence μ_n of d-dimensional distributions such that $\mu_n \Rightarrow \mu$, there exists a probability space $(\Omega, \mathcal{F}, P)$ and random mappings X_n and X from $(\Omega, \mathcal{F})$ to $(\mathbb{R}^d, \mathcal{B}^d)$ such that μ_n and μ are distributions of X_n and X and also $\lim_{n\to+\infty} X_n = X$ almost surely. This result on separable metric spaces is obtained by Skorohod and later generalized to nets by Wichura (1970).

Some Typical Roles of Weak Convergence in Probability

The weak convergence appears in Probability chiefly in the following classes of problems.

- Knowing that $\mu_n \Rightarrow \mu$, we may replace μ_n by μ for n large enough. A typical example is the Central Limit Theorem (any of its versions), which enables us to conclude that the properly normalized sum of random variables has approximately a unit Gaussian law.
- Conversely, if $\mu_n \Rightarrow \mu$ then we may approximate μ with μ_n, for n large enough. A typical example of this sort is the approximation of Dirac's delta function (understood as a density of a point mass at zero) by, say triangle-shaped functions.
- It is not always easy to construct a measure with specified properties. If we need to show just its existence, sometimes we are able to construct a sequence

Miodrag Lovric (ed.), *International Encyclopedia of Statistical Science*, DOI 10.1007/978-3-642-04898-2,

(or a net) of measures which can be proved to be weakly convergent and that its limit satisfies the desired properties. For example, this procedure is usually applied to show the existence of the Wiener measure (the measure induced by Brownian Motion process (see ▸Brownian Motion and Diffusions) on the space of continuous functions).

The last mentioned example is related to measures in infinitely dimensional spaces, the case which usually arises in the context of assigning measures to the set of trajectories of a stochastic process. In fact, what is called weak convergence in Probability theory, is inherited from so called weak-star convergence in Topology, where it can be defined in duals of arbitrary topological spaces. In Probability theory we do not need such a generality, as we are interested only in spaces of measures. Since the spaces of measures always appear as dual spaces of continuous functions, the most general definition of weak convergence of probability measures is the following.

Definition 1 Let $\mathcal{X}$ be a topological space. Let $\mu_n(n = 1, 2, \ldots)$ and μ be probability measures defined on the Borel sigma field generated by open subsets of $\mathcal{X}$. We say that the sequence $\{\mu_n\}$ converges *weakly* to μ, in notation $\mu_n \Rightarrow \mu$ if

$$\lim_{n\to+\infty} \int_{\mathcal{X}} f(x)\, \mathrm{d}\mu_n(x) = \int_{\mathcal{X}} f(x)\, \mathrm{d}\mu(x)\,,$$

for every continuous and bounded real valued function $f : \mathcal{X} \mapsto \mathbb{R}$. The set of these functions is denoted by $\mathbf{C}(\mathcal{X})$. □

In terms of random variables, let X_n $(n = 1, 2, \ldots)$ and X be $\mathcal{X}$-valued random variables and let μ_n and μ be corresponding distributions. Then we say that the sequence X_n converges weakly to X and write $X_n \Rightarrow X$ if and only if $\mu_n \Rightarrow \mu$. A setup that yields infinitely dimensional spaces $\mathcal{X}$ is when X_n is a sequence of random processes and $\mathcal{X}$ is a space of functions where paths of X_n belong. Finally, in a general situation, we may think of nets $\{X_d\}$ and $\{\mu_d\}$ instead of sequences.

Weak Convergence of Measures on Metric Spaces

Let now $\mathcal{X}$ be a metric space and let $\mathcal{B}$ be the sigma field of Borel subsets of $\mathcal{X}$. Let $\mathcal{M}_1(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$.

Theorem 2 *Let μ_d be a net of probability measures on $\mathcal{X}$ and let μ_0 be a probability measure on $\mathcal{X}$. The following statements are equivalent* (Billingsley 1999; Stroock 1993):

(i) $\mu_d \Rightarrow \mu_0$, i.e., $\lim_d \int f\, \mathrm{d}\mu_d = \int f\, \mathrm{d}\mu_0$, *for each* $f \in \mathbf{C}(\mathcal{X})$.

(ii) $\lim_d \int f\, \mathrm{d}\mu_d = \int f\, \mathrm{d}\mu_0$ *for each* $f \in \mathbf{C}_u(\mathcal{X})$ *(uniformly continuous and bounded functions)*.

(iii) $\overline{\lim}\, \mu_d(F) \leq \mu_0(F)$ *for any closed set* $F \subset \mathcal{X}$.

(iv) $\underline{\lim}\, \mu_d(G) \geq \mu_0(G)$ *for each open set* $G \subset \mathcal{X}$.

(v) $\lim \mu_d(A) = \mu_0(A)$ *for each continuity set A for* μ_0 (that is, $\mu_0(\partial A) = 0$, *where ∂A is the boundary* of A).

(vi) $\overline{\lim} \int f\, \mathrm{d}\mu_d \leq \int f\, \mathrm{d}\mu_0$ *for each upper semicontinuous and bounded from above function* $f : \mathcal{X} \mapsto \mathbb{R}$.

(vii) $\underline{\lim} \int f\, \mathrm{d}\mu_d \geq \int f\, \mathrm{d}\mu_0$ *for each lower semicontinuous and bounded from below function* $f : \mathcal{X} \mapsto \mathbb{R}$.

(viii) $\lim \int f\, \mathrm{d}\mu_d = \int f\, \mathrm{d}\mu_0$ *for each μ_0–a.e. continuous function* $f : \mathcal{X} \mapsto \mathbb{R}$.

In concrete metric spaces, the conditions can be checked to hold only for some special families of sets, so called *convergence determining families*. For example, a convergence determining family in $\mathbb{R}$ is a family of sets of the form $(-\infty, b]$, $b \in \mathbb{R}$, and using this family in the condition (v), we get the standard definition from the beginning of section "▸Weak Convergence of Probability Measures on $\mathbb{R}^d$". Similarly, it can suffice to check condition (i) only for special families of functions – Theorem 1 gives an example of such a family.

If $\mathcal{X}$ is a separable metric space, the topology of weak convergence of probability measures is metrizable by the metric

$$d(P, Q) = \inf\{\varepsilon > 0 \mid Q(B) \leq P(B^\varepsilon) + \varepsilon,\ P(B) \leq Q(B^\varepsilon) + \varepsilon, \quad B \in \mathcal{B}\},$$

where $B^\varepsilon = \{x \in \mathcal{X} \mid d(x, B) < \varepsilon\}$, and $\mathcal{B}$ is the Borel sigma algebra on $\mathcal{X}$. This metric is called Prohorov's metric, and it is a generalization of Lévy's metric from section "▸Weak Convergence of Probability Measures on $\mathbb{R}^d$". There are metrics which are known to be equivalent to Prohorov's metrics (see, for example, [Stroock 1993, p.117]).

Relative Compactness, Tightness and Prohorov's Theorem

Let $\mathcal{X}$ be a metric space, $\mathcal{B}$ a Borel sigma-algebra generated by open subsets of $\mathcal{X}$. In infinitely dimensional metric spaces, the weak convergence of finite dimensional distributions alone is not sufficient condition for weak convergence of measures. The additional condition is relative compactness.

Definition 2 We say that a set $\mathcal{P}$ of probability measures on $(\mathcal{X}, \mathcal{B})$ is relatively compact if any sequence of probability measures $P_n \in \mathcal{P}$ contains a subsequence P_{n_k} which converges weakly to a probability measure in $\mathcal{M}_1(X)$.

Theorem 3 *Let $\{\mu_n\}$ be a relatively compact sequence of probability measures on $\mathcal{X}$. If all finite-dimensional distributions converge weakly to corresponding finite-dimensional distributions of a measure μ, then $\mu_n \Rightarrow \mu$.*

Hence, an usual procedure to show weak convergence on a metric space is to first show convergence of finite dimensional distributions (via ►characteristic functions), and then to prove relative compactness. If $\mathcal{X}$ is compact, then any set $\mathcal{P}$ of probability measures is relatively compact. Otherwise, we need some conditions which are easier to check, a convenient tool is the notion of *tightness*.

Definition 3 Let $\mathcal{P}$ be a set of probability measures on $(\mathcal{X}, \mathcal{B})$. We say that $\mathcal{P}$ is *tight* if for any $\varepsilon > 0$ there is a compact set $K \subset \mathcal{X}$ such that $\mu(K') \leq \varepsilon$ for any $\mu \in \mathcal{P}$. □

Next theorem links tightness with relative compactness.

Theorem 4 (Prohorov 1956) *(a) Any tight set of measures in arbitrary metric space is relatively compact. (b) If X is a complete separable metric space, then any relatively compact set of probability measures is tight.*

In particular metric spaces, it is useful to have simpler equivalent conditions for tightness. For example, observe the metric space $C[0,1]$ of continuous functions defined on $[0,1]$, with the metric of uniform convergence, $d(x,y) = \sup_{t \in [0,1]} |x(t) - y(t)|$. Then a sequence $\{\mu_n\}$ of probability measures (defined on Borel sets of this metric space) is tight if an only if

$$\lim_{\substack{K \to +\infty \\ n \to +\infty}} \mu_n\{x \in C[0,1] \mid |x(0)| \geq K\} = 0 \qquad \text{and}$$

$$\lim_{\delta \to 0} \limsup_{n \to +\infty} \mu_n\{x \in C[0,1] \mid w_x(\delta) \geq \varepsilon\} = 0,$$

$$\text{for each } \varepsilon > 0,$$

where w_x is defined as

$$w_x(\delta) = \sup_{|s-t| \leq \delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1$$

(modulus of continuity of x).

Similar conditions exist in the space $D[0,1]$ of all right - continuous functions with left limits (càdlàg functions), equipped with Skorohod's metric (Billingsley 1999).

Finally, let us mention that in a Hilbert space H with an inner product $\langle . , . \rangle$, we may define characteristic function of a random variable X with a probability distribution μ, in the same way as in the finite dimensional spaces:

$$\varphi(x) = \int_H e^{i\langle x,y \rangle} \, d\mu(y), \qquad x \in H.$$

Theorem 5 *Let $\{P_n\}$ be a sequence of probability measures on H and let φ_n be the corresponding characteristic functions. Let P and φ be a probability measure and its characteristic function. If $P_n \Rightarrow P$ then $\lim_n \varphi_n(x) = \varphi(x)$ for all $x \in H$.*

Conversely, if a sequence P_n of probability measures on H is relatively compact and $\lim_n \varphi_n(x) = \varphi(x)$ for all $x \in H$, then there exists a probability measure P such that φ is its characteristic function and $P_n \Rightarrow P$.

Cross References

►Almost Sure Convergence of Random Variables
►Bootstrap Asymptotics
►Central Limit Theorems
►Convergence of Random Variables
►Measure Theory in Probability
►Nonparametric Estimation Based on Incomplete Observations
►Role of Statistics in Advancing Quantitative Education

References and Further Reading

Billingsley P (1999) Convergence of probability measures. Wiley, New York

Lévy P (1937) Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris

Parthasarathy KR (1967) Probability measures on metric spaces. Academic Press, New York

Prohorov YuV (1956) Shodimost slučainih processov i predelynie teoremi teorii veroyatnostei. Teor ver primenen 1:177–238

Stroock D (1993) Probability theory, an analytic view. Cambridge University Press, Cambridge

Wichura MJ (1970) On the construction of almost uniformly convergent random variables with given weakly convergent image laws. Ann Stat 41:284–291

Weibull Distribution

William J. Padgett
Distinguished Professor Emeritus of Statistics
University of South Carolina, Columbia, SC, USA

W

The Weibull family of probability distributions (*see* also ►Generalized Weibull Distributions) is one the most widely used parametric families of distributions for modeling failure times or lifetimes. This is especially true in engineering and science applications (as suggested originally by Weibull 1951) and is mainly due to the variety of shapes of its density function and the behaviors of its failure rate function. Literally thousands of references to the Weibull distribution can be found in the scientific literature. See Johnson et al. (1994) or a more recent treatment by Rinne (2008) for a detailed comprehensive overview of this family of distributions.

Let T denote a random variable (rv) representing the failure time or lifetime of an item under study. This rv has a Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$ if its probability density function (pdf) is $f(t) = \alpha t^{\alpha-1}\beta^{-\alpha}\exp[-(t/\beta)^{\alpha}]$ for $t \geq 0$. The cumulative distribution function (cdf) is then $F(t) = 1 - \exp[-(t/\beta)^{\alpha}]$, $t \geq 0$, and the survival or reliability function is $R(t) = 1 - F(t)$. Then the failure rate (or hazard rate) function is $h(t) = f(t)/R(t) = \alpha\beta^{-\alpha}t^{\alpha-1}$. For shape parameter $\alpha = 1$, the Weibull reduces to the *exponential* distribution with scale β, and when $\alpha = 2$ the resulting Weibull distribution is referred to as the *Rayleigh* distribution (*see* "Generalized Rayleigh distribution").

A major reason that the Weibull distributions are so useful is that the failure rate function can be increasing if the shape $\alpha > 1$, decreasing if $\alpha < 1$, or constant for $\alpha = 1$. An increasing failure rate function corresponds to the common assumption that the item whose lifetime is under study fails due to wearout over time, that is, an "ageing process" occurs where failure becomes more likely as time increases. The case of decreasing failure rate is less common but sometimes holds for types of items that tend to fail early due to defects or low quality and that tend to last longer if no defects are present, perhaps with very slow ageing. The constant failure rate corresponds to random failures occurring over time, which is the "memoryless" property of the exponential distribution. That is, there is no ageing process so that an item is always as good as new over time. Although the ageless property might seem to be unrealistic, some high-quality electronic items often approximately satisfy such an assumption for a period of time. So, Weibull distributions provide good models over a wide variety of "ageing" scenarios.

For integer $r > 0$ the r^{th} moment of a Weibull rv T is $E(T^r) = \beta^r\Gamma(1 + r/\alpha)$, where $\Gamma(c) = \int_0^{\infty} x^{c-1}e^{-x}dx$ is the gamma function. Therefore, the mean of the Weibull distribution is $\mu = E(T) = \beta\Gamma(1+1/\alpha)$ and the variance is $\sigma^2 = E(T^2) - \mu^2 = \beta^2\Gamma(1+2/\alpha) - \beta^2\Gamma^2(1+1/\alpha)$. These expressions are generally not very easy to use, but they can be obtained by computing approximate values of the gamma function. Since calculation of the mean lifetime is not very user-friendly, the value of the scale parameter itself is often used as a measure of "typical" lifetime, referred to as the *characteristic life* of the item. Since $R(\beta) = \exp(-1) \approx 0.37$, the characteristic life β is approximately the 63rd percentile of the distribution. Also, the variance is proportional to the square of the characteristic life.

The Weibull distribution arises also as the limiting distribution of the first order statistic from some probability distribution, so the Weibull is a limit of *extreme-value* distributions in this sense. That is, let $X_{(1)}$ denote the first order statistic from n independent identically distributed (iid) random variables, $X_1, \ldots, X_n$, from a specified cdf. Then as $n \to \infty$, the distribution of $X_{(1)}$ approaches a Weibull distribution (see Mann et al. 1974, or Rinne 2008). In fact, the Weibull distribution satisfies the important "weakest-link" property which is another reason for its applicability. Suppose that $X_1, \ldots, X_n$ are n iid random variables each with the Weibull cdf $F(t)$ and reliability function $R(t) = \exp[-(t/\beta)^{\alpha}]$, $t \geq 0$. Then the reliability function of the first order statistic, i.e. the "weakest" or smallest observation, $X_{(1)} = \min\{X_1, \ldots, X_n\}$, is $R_1(t) = \Pr[X_i > t \text{ for all } i = 1, \ldots, n] = [R(t)]^n = \{\exp[-(t/\beta)^{\alpha}]\}^n = \exp[-(t/\beta n^{-1/\alpha})^{\alpha}]$. Thus, $X_{(1)}$ also has the Weibull distribution with the same shape parameter α and new scale parameter $\beta n^{-1/\alpha}$. This implies that in the increasing failure rate case, $\alpha > 1$, a long chain of "links" has a higher probability of failure than a shorter chain. This idea is important in modeling the failure of materials (see, for example, Smith 1991, and Wolstenholme 1995), as well as the failure of a series system of n iid components.

Several generalizations of the Weibull distribution have been proposed, three of which will be mentioned here. A frequently used version is obtained by adding a "shift parameter," γ, also referred to as a "guarantee time." That is, the Weibull pdf and cdf are shifted from zero to γ, so $f(t) = \alpha(t-\gamma)^{\alpha-1}\beta^{-\alpha}\exp\left[-\left(\frac{t-\gamma}{\beta}\right)^{\alpha}\right]$ and $F(t) = 1 - \exp[-((t-\gamma)/\beta)^{\alpha}]$, $t \geq \gamma$, to obtain the *three-parameter Weibull* distribution. Another generalization introduced by Mudholkar and Srivastava (1993) is known as the *exponentiated Weibull* distribution which has cdf $F_{EW}(t) = \{1 - \exp[-(t/\beta)^{\alpha}]\}^{\theta}$, $t \geq 0$, with another parameter $\theta > 0$. For $\alpha = 2$ and $\beta = 1$, this exponentiated Weibull distribution reduces to the *Burr type X* distribution (see Burr 1942). The third generalization mentioned here is called the *brittle fracture* distribution (see Black et al. 1990), whose reliability function is of the form $R_{BF}(t) = \exp[-\delta t^{2\rho}\exp(-\theta/t^2)]$, $t > 0$. This distribution was found to provide good model fits specifically to observed breaking stress data for boron fibers and carbon fibers. Taking $\theta = 0$ and $2\rho = a$ yields the usual two-parameter Weibull distribution with shape parameter α.

Estimation of the parameters for fitting the Weibull distribution to observed failure data is typically accomplished by the method of maximum likelihood. This involves numerical solution since the likelihood equations yield nonlinear functions of the shape parameter. For example, see Mann et al. (1974) and Rinne (2008). *Weibull plotting* is a graphical technique that is often used for quick (not maximum likelihood) estimation of the parameters (for examples, refer to Rinne (2008) and Wolstenholme (1999)). Tests of hypotheses for the parameters, interval estimation, and other inferences for the Weibull model are discussed by Mann et al. (1974) and Rinne (2008) as well as by many other authors. There are several available statistical or engineering software packages that include Weibull modeling procedures. Among others, two dedicated software packages for Weibull analysis of lifetime data may be found at http://Weibull.ReliaSoft.com and http://www.relex.com/products/weibull.asp.

About the Author

Professor Padgett was Department Chairman (1985–1993) and (1996–2001). He was awarded the Donald S. Russell Award for Creative Research in Science and Engineering, University of South Carolina (1975) and Paul Minton Service Award, Southern Regional Council on Statistics (2003), among others. He is a Fellow of the American Statistical Association and of the Institute of Mathematical Statistics, and an Elected member of the International Statistical Institute. Professor Padgett was an Associate editor for many international journals including, *Technometrics* (1987–1992), *Journal of Nonparametric Statistics* (1989–2004), *Journal of Statistical Theory and Applications* (2001–2007), *Lifetime Data Analysis* (1994–2003), *Journal of Applied Statistical Science* (1992–2001) and *Journal of Statistical Computation and Simulation* (1980–1986). He was also Coordinating Editor of *Journal of Statistical Planning and Inference* (1995–1997).

Cross References

►Extreme Value Distributions
►Generalized Extreme Value Family of Probability Distributions
►Generalized Rayleigh Distribution
►Generalized Weibull Distributions
►Modeling Survival Data
►Multivariate Statistical Distributions
►Statistical Distributions: An Overview
►Statistics of Extremes
►Survival Data

References and Further Reading

Black CM, Durham SD, Padgett WJ (1990) Parameter estimation for a new distribution for the strength of brittle fibers: a simulation. Comm Stat Simulat Comput 19:809–825

Burr IW (1942) Cumulative frequency functions. Ann Math Stat 13:215–222

Johnson NL, Kotz S, Balakrishnan N (1994) Continuous univariate distributions, vol 1, 2nd edn. Wiley Series in Probability and Mathematical Statistics, New York

Mann NR, Shafer RE, Singpurwalla ND (1974) Methods for statistical analysis of reliability and lifetime data. Wiley, New York

Mudholkar GS, Srivastava DK (1993) Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Trans Reliab 42:299–302

Rinne H (2008) The Weibull distribution: a handbook. CRC Press, Boca Raton

Smith RJ (1991) Weibull regression models for reliability data. Reliab Eng Syst Saf 34:35–57

Weibull W (1951) A statistical distribution function of wide applicability. J Appl Mech 18:293–297

Wolstenholme LC (1995) A non-parametric test of the weakest-link property. Technometrics 37:169–175

Wolstenholme LC (1999) Reliability modelling: a statistical approach. Chapman & Hall/CRC, Boca Raton

Weighted Correlation

Joaquim F. Pinto da Costa
Professor
University of Porto, Porto, Portugal

Introduction

Weighted correlation is concerned with the use of weights assigned to the subjects in the calculation of a correlation coefficient (see ►Correlation Coefficient) between two variables X and Y. The weights can either be naturally available beforehand or chosen by the user to serve a specific purpose. For instance, if there is a different number of measurements on each subject, it is natural to use these numbers as weights and calculate the correlation between the subject means. On the other hand, if the variables X and Y represent, for instance, the ranks of preferences of two human beings over a set of n items, one might want to give larger weights to the first preferences, as these are more accurate. In another situation, if we want to calculate the correlation between two stocks in a stock exchange market during last year, we might want to favor (larger weight) the more recent observations, as these are more important for the present situation. Suppose that X_i and Y_i are the pair of values corresponding to observation i in each sample and w_i the weight attributed to this observation, such that $\sum_{i=1}^{n} w_i = 1$. Then, the sample weighted correlation coefficient is given by the formula

$$r_w = \frac{\sum w_i (X_i - \overline{X}_w)(Y_i - \overline{Y}_w)}{\sqrt{\sum w_i (X_i - \overline{X}_w)^2}\sqrt{\sum w_i (Y_i - \overline{Y}_w)^2}} = \frac{\sum w_i X_i Y_i - \sum w_i X_i \sum w_i Y_i}{\sqrt{\sum w_i X_i^2 - (\sum w_i X_i)^2}\sqrt{\sum w_i Y_i^2 - (\sum w_i Y_i)^2}}, \quad (1)$$

where the sums are from $i = 1$ to n and $\overline{X}_w = \sum w_i X_i$ and $\overline{Y}_w = \sum w_i Y_i$ are the weighted means. When all the w_i are equal they cancel out, giving the usual formula for the Pearson product–moment correlation coefficient.

Weighted Rank Correlation

Rank correlation coefficients are nonparametric statistics that are less restrictive than others (e.g., Pearson's correlation coefficient), because they do not try to fit one particular kind of relationship, linear or other, to the data. Their objective is to assess the degree of monotonicity between two series of paired data. Common rank correlation coefficients are Spearman's and Kendall's (Neave and Worthington 1992). One interesting fact about rank correlation is that, contrary to other correlation methods, it can be used not only on numerical data but on any data that can be ranked.

Blest (2000) proposed an alternative weighted measure of rank correlation that gives more importance to the first ranks but has some drawbacks because it is not a symmetric function of the two vectors of ranks. Later, Pinto da Costa and Soares (Pinto da Costa and Soares 2005; Soares et al. 2001) presented a new weighted rank correlation coefficient that gives larger weight to the first ranks and does not have the problems of Blest's coefficient. This coefficient is

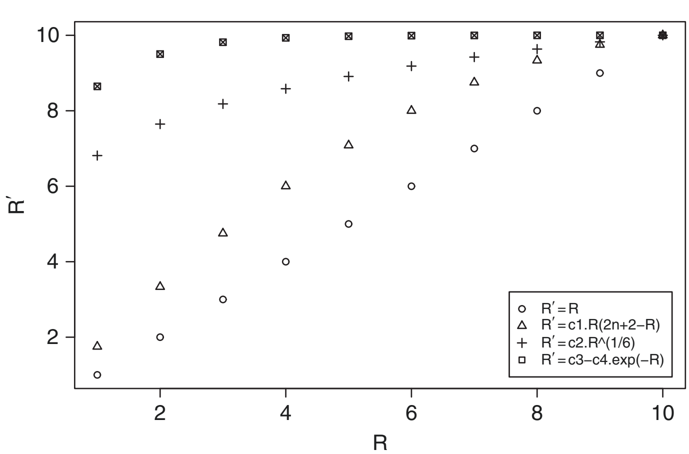
$$r_W = 1 - \frac{6\sum_{i=1}^{n}(R_i - Q_i)^2(2n + 2 - R_i - Q_i)}{n^4 + n^3 - n^2 - n}, \quad (2)$$

where R_i is the rank corresponding to the ith observation of the first variable, X, and Q_i is the rank corresponding to the ith observation of the second variable, Y. r_W, which yields values between −1 and +1, uses a linear weight function: $2n + 2 - R_i - Q_i$. Some properties of the distribution of the statistic r_W, including its sample distribution, are analyzed in Pinto da Costa and Soares (2005) and Pinto da Costa and Roque (2006); in particular, the expected value of this statistic is zero when the two variables are independent, and its sampling distribution converges to the Gaussian when the sample size increases. Later, Pinto da Costa and Soares (2007) introduced a new weighted rank correlation coefficient that uses a quadratic weight function:

$$r_{W2} = 1 - \frac{90\sum_{i=1}^{n}(R_i - Q_i)^2(2n + 2 - R_i - Q_i)^2}{n(n-1)(n+1)(2n+1)(8n+11)}. \quad (3)$$

A New Way of Developing Weighted Correlation Coefficients

It can be proved that the coefficient r_{W2} is equal to the Pearson's correlation coefficient of the transformed ranks $R'_i = R_i(2n + 2 - R_i)$ and $Q'_i = Q_i(2n + 2 - Q_i)$ and this suggests a new and easy way of developing weighted correlation coefficients. In fact, by applying a transformation to the ranks so that the first ones are favored and then computing the Pearson's correlation coefficient of the transformed ranks, we can define many new measures of weighted correlation (Pinto da Costa and Soares 2007). In Fig. 1 we can see four different cases. The first, when $R' = R$, corresponds to Spearman's coefficient and so it



Weighted Correlation. Fig. 1 Scatterplot for four different rank transformations does not correspond to a weighted measure; when $R' = R(2n+2-R)$ we have r_{W2}. We can now use other functions such as $R' = R^{1/6}$ or $R' = -e^{-R}$. In order to be able to represent the four cases in the same diagram some of the transformations had to be multiplied by a constant and in the last case another constant was also added, but these operations do not change the value of Pearson's correlation. Thus, the importance given to the first ranks is larger when $R' = -e^{-R}$ and smaller when $R' = R$. This means that the ranks that are in a flatter region are given smaller weight.

From this perspective, and in case we want to give larger weight to the first ranks, all that is needed is that the transformation is monotone and the last ranks are more flattened by the transformation compared with the first ranks. However, if we want to give larger weight to other ranks, not the first, we just have to find an appropriate transformation to do that; one that is less flat where the weights are to be larger. This in turn has two additional advantages. First, we can use different transformations to each variable and so we are not obliged to give the same set of weights to the two variables. Secondly, this strategy can be used with the original data, not only ranks, and so many new measures of weighted correlation can be developed.

About the Author

Joaquim Costa, University of Porto, is an elected member of the International Statistical Institute. He is also a member of the Portuguese Statistical Society and of the Portuguese Classification Society. He completed his PhD in 1996 under the supervision of Israel Lerman from University of Rennes 1. He has published in statistical and machine learning journals and his main contributions are in weighted measures of correlation, weighted principal component analyses, and also new methods of supervised classification for ordered classes.

Cross References

►Correlation Coefficient
►Kendall's Tau
►Measures of Dependence
►Rank Transformations

References and Further Reading

Blest D (2000) Rank correlation – an alternative measure. Aust N Z J Stat 42(1):101–111

Neave H, Worthington P (1992) Distribution-free tests. Routledge, London

Pinto da Costa J, Roque L (2006) Limit distribution for the weighted rank correlation coefficient, rW. REVSTAT – Stat J 4(3):189–200

Pinto da Costa J, Soares C (2005) A weighted rank measure of correlation. Aust N Z J Stat 47(4):515–529

Pinto da Costa J, Soares C (2007) Rejoinder to letter to the editor from C. Genest and J-F. Plante concerning Pinto da Costa J & Soares C (2005) A weighted rank measure of correlation. Aust N Z J Stat 49(2):205–207

Soares C, Pinto da Costa J, Brazdil P (2001) Improved statistical support for matchmaking: rank correlation taking rank importance into account. In: JOCLAD 2001: VII Jornadas de Classificação e Análise de Dados, Porto, Portugal, pp 72–75

Weighted *U*-Statistics

Aluísio de Souza Pinheiro
Professor, Head of Department of Statistics
University of Campinas, Campinas, SP, Brazil

Following the seminal papers of Hoeffding (1948, 1961), let T_n be a linear combination defined by

$$T_n = \sum_{i_1,\ldots,i_m}^{1,n} \eta_{n,i_1\cdots i_m}\phi(\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_m}), \qquad (1)$$

such that: (a) $\eta_{n,i_1\cdots i_m}$ are weight functions, (b) $\sum_{i_1,\ldots,i_m}^{1,n}$ is taken on all strictly ordered permutations of $1,\ldots,n$, (c) $\phi(\cdot,\ldots,\cdot)$ is a kernel of degree m, stationary of order r $(1 \le r \le m)$, for which we let $\theta = \mathrm{E}\phi(X_1,\ldots,X_m)$, and (d) $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are i.i.d. random vectors of dimension K, not necessarily quantitative in nature (Pinheiro et al. 2009).

Some configurations of $\eta_{n,i_1\cdots i_m}$ lead to special classes of (generalized) ►*U*-statistics, as follows: If $\eta_{n,i_1\cdots i_m} \equiv \binom{n}{m}^{-1}$ and $r \ge 1$, T_n is a degenerate U-statistics of degree m whose projection variances are such that $0 = \sigma_1^2 = \cdots = \sigma_r^2 < \sigma_{r+1}^2$; then T_n has a degeneracy of order r and $n^{(r+1)/2}(T_n - \theta)$ converges to a (possibly) infinite linear combination of independent random variables, each distributed accordingly to a $(r+1)$-dimensional Wiener integral (Dynkin and Mandelbaum 1983).

If $K = 1$ and the $\eta_{n,i_1\cdots i_m}$ assume 0 or 1 values only, T_n is said to be an *incomplete* U-statistic (Janson 1984). Asymptotic distribution of T_n will be either a linear combination of independent Wiener integrals or a mixture of such a distribution with an independent normal r.v., under suitable sampling conditions (Janson 1984). For a class of *conditional* U-statistics, where the weights can be decomposed as $\eta_{n,i_1\cdots i_m} = e(i_1)\cdots e(i_m)$, $e(\cdot)$ being the marginal weight function, ►asymptotic normality follows from Stute (1991). Moreover, the conditional nature of the class derives from the fact that weights are defined as random functions of another set of r.v.'s.

For $K = 1$, O'Neil and Redner (1993) and Major (1994) present asymptotic results in a more general setup for the class of *weighted* U-statistics, defined by (1). The case $m = 2$ using moment matching techniques to determine the asymptotic distribution of T_n is discussed in O'Neil and Redner (1993). Under some regularity conditions on $\eta_{n,i_1\cdots i_m}$, a non-normal limit is proven for either $r = 1$ or $r = 0$. For $r = 0$, a class of weighted U-statistics is proved to be asymptotically normal under a second set of conditions on weights. ►Asymptotic normality is also established for $r = 1$ and *incomplete designs*. The common idea behind all weight-designs is the orthogonality on the set of (possibly random) weights. Major (1994) points out that the aforementioned approach cannot be adapted for $m \geq 3$; Poisson approximation is then used to pursue asymptotic behavior of T_n.

A class of *quasi U-statistics* having the novelty that it can be applied to any i.i.d. random vectors of arbitrary (and even increasing) dimension K, is introduced in Pinheiro et al. (2009). The proposed class is constructed in such a way that, although ϕ can be degenerate, the chosen weights lead to a contrast, i.e., such that $\sum_{i_1,\ldots,i_m}^{1,n} \eta_{n,i_1\cdots i_m} = 0$, providing asymptotically normal distributions. For the quasi U-statistics, the aforementioned contrast condition is an essential requirement. Otherwise, for degenerate U-statistics the asymptotic distribution is non-normal.

About the Author

Dr. Aluisio Pinheiro has been Chair of the Department of Statistics at the University of Campinas since 2007. He has coauthored several papers on U statistics (with P. K. Sen and H. P. Pinheiro), and a book (with H. P. Pinheiro) on the use of Quasi-statistics in Genetic Data published by the Brazilian Society of Mathematics in 2007.

Cross References

►Asymptotic Normality
►U-Statistics

References and Further Reading

Dynkin EB, Mandelbaum A (1983) Symmetric statistics, Poisson point processes and multiple Wiener integrals. Ann Stat 11(3):739–745

Hoeffding W (1948) A class of statistics with asymptotically normal distribution. Ann Math Stat 19(3):293–325

Hoeffding W (1961) The strong law of large numbers for U-statistics. Institute of Statistics Mimeo Series No. 302. University of North Carolina, Chapel Hill

Janson S (1984) The asymptotic distribution of incomplete U-statistics. Z Wahrsch Verw Geb 66(4):495–505

Major P (1994) Asymptotic distributions for weighted U-statistics. Ann Prob 22(3):1514–1535

O'Neil KA, Redner RA (1993) Asymptotic distributions of weighted U-statistics of degree 2. Ann Probab 21(2):1159–1169

Pinheiro A, Sen PK, Pinheiro HP (2009) Decomposability of high-dimensional diversity measures: quasi U-statistics, martingales and nonstandard asymptotics. J Multivar Anal 100:1645–1656

Stute W (1991) Conditional U-statistics. Ann Probab 19(2):812–823

Wilcoxon–Mann–Whitney Test

Markus Neuhäuser
Professor
Koblenz University of Applied Sciences, Remagen, Germany

The Wilcoxon–Mann–Whitney (WMW) test was proposed by Frank Wilcoxon in 1945 ("Wilcoxon rank sum test") and by Henry Mann and Donald Whitney in 1947 ("Mann–Whitney U test"). However, the test is older: Gustav Deuchler introduced it in 1914 (see Kruskal 1957). Nowadays, this test is a commonly used nonparametric test for the two-sample location problem. As with many other nonparametric tests, this is based on ranks rather than on the original observations.

The sample sizes of the two groups or random samples are denoted by n and m. The observations within each sample are independent and identically distributed, and we assume independence between the two samples. The null hypothesis, H_0, is one of no difference between the two groups.

Let F and G be the distribution functions corresponding to the two samples. Then we have the null hypothesis $H_0 : F(t) = G(t)$ for every t. Under the two-sided alternative there is a difference between F and G. Often, it is assumed that F and G are identical except a possible shift in location (location-shift model), i.e., $F(t) = G(t - \theta)$ for every t. Then, the null hypothesis states $\theta = 0$, and the two-sided alternative is $H_1 : \theta \neq 0$. Of course, one-sided alternatives are possible, too.

Let $V_i = 1$ when the ith smallest of the $N = n + m$ observations is from the first sample and $V_i = 0$ otherwise. The Wilcoxon rank sum is a linear rank statistic defined by $W = \sum_{i=1}^{N} i \cdot V_i$. Hence, W is the sum of the n ranks of group 1; the ranks are determined based on the pooled sample of all N values.

The Mann–Whitney statistic U is defined as $U = \sum_{i=1}^{n}\sum_{j=1}^{m} \psi(X_i, Y_j)$ where X_i (Y_j) is an observation from group 1 (group 2), and

$$\psi(X_i, Y_j) = \begin{cases} 1 & \text{if} \quad X_i > Y_j \\ 0.5 & \text{if} \quad X_i = Y_j \\ 0 & \text{if} \quad X_i < Y_j. \end{cases}$$

Because of $W = U + \frac{n}{2}(n+1)$, the tests based on W and U are equivalent.

The standardized statistic Z_W can be computed as $Z_W = \frac{W - E_0(W)}{\sqrt{Var_0(W)}}$ with $E_0(W) = \frac{n(N+1)}{2}$ and $Var_0(W) = \frac{nm(N+1)}{12}$. In the presence of ties mean ranks can be recommended for tied observations. Then, the variance changes, in this case we have

$$Var_0(W) = \frac{nm}{12}\left(N + 1 - \frac{\sum_{i=1}^{g}(t_i - 1)t_i(t_i + 1)}{N(N-1)}\right),$$

where g is the number of tied groups and t_i the number of observations within the ith tied group. An untied value is regarded as a tied group with $t_i = 1$ (Hollander and Wolfe 1999, p. 109).

Under H_0, the standardized Wilcoxon statistic asymptotically follows a standard normal distribution. This result can be used to carry out the test and to calculate an asymptotic p-value. According to Brunner and Munzel (2002, p. 63) the normal approximation is acceptable in case of $\min(n, m) \geq 7$, if there were no ties. The two-sided asymptotic WMW test can reject H_0 if $|Z_W| \geq z_{1-\alpha/2}$, the corresponding p-value can be computed as $2(1 - \Phi(|Z_W|))$, where $z_{1-\alpha/2}$ and Φ denote the $(1 - \alpha/2)$-quantile and the distribution function, respectively, of the standard normal distribution.

Alternatively, the exact permutation null distribution of W can be determined and used for inference (see the chapter about ►permutation tests). Some monographs include tables of critical values for the permutation test, but these tables can only be used if there were no ties. A permutation test, however, is also possible in the presence of ties, because the exact conditional distribution of W can be obtained.

As a rank test the WMW test does not use all the available information; despite this, it is quite powerful. If the normal distribution is a reasonable assumption, little is lost by using the Wilcoxon test instead of the parametric t test. On the other hand, when the assumption of normality is not satisfied, the nonparametric Wilcoxon test may have considerable advantages in terms of efficiency. To be precise, the asymptotic relative efficiency (ARE) of the WMW test in comparison to Student's t test (see ►Student's t-Tests) cannot be smaller than 0.864. However, there is no upper limit. If the data follow a normal distribution the ARE is $3/\pi = 0.955$ (Hodges and Lehmann 1956).

The two-sided WMW test is consistent against all alternatives with $\Pr(X_i < Y_j) \neq 0.5$. However, the WMW test can give a significant result for a test at the 5% level with much more than 5% probability when the population medians are identical, but the population variances differ. A generalization exists that can be applied for testing a difference in location irrespective of a possible difference in variability (Brunner and Munzel 2000).

About the Author

Dr Markus Neuhäuser is a Professor, Department of Mathematics and Technique, Koblenz University of Applied Sciences, Remagen, Germany. He was Senior Lecturer at the Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand (2002–2004). He has authored and co-authored more than 100 papers and 2 books, including *Computer-intensive und nichtparametrische statistische Tests* (Oldenbourg, 2010). Currently, he is an Associate Editor for the *Journal of Statistical Computation and Simulation*, *Communications in Statistics - Theory and Methods*, and *Communications in Statistics – Simulation and Computation*.

Cross References

►Asymptotic Relative Efficiency in Testing
►Continuity Correction
►Explaining Paradoxes in Nonparametric Statistics
►Nonparametric Rank Tests
►Nonparametric Statistical Inference
►Presentation of Statistical Testimony
►Rank Transformations
►Ranks
►Scales of Measurement and Choice of Statistical Methods
►Sequential Ranks
►Statistical Fallacies: Misconceptions, and Myths
►Student's t-Tests
►Wilcoxon-Signed-Rank Test

References and Further Reading

Brunner E, Munzel U (2000) The nonparametric Behrens-Fisher problem: asymptotic theory and a small sample approximation. Biom J 42:17–25

Brunner E, Munzel U (2002) Nichtparametrische Datenanalyse. Springer, Berlin

Hodges JL, Lehmann EL (1956). The efficiency of some nonparametric competitors of the t-test. Ann Math Stat 27:324–335

Hollander M, Wolfe DA (1999) Nonparametric statistical methods, 2nd edn. Wiley, New York
Kruskal WH (1957) Historical notes on the Wilcoxon unpaired two-sample test. J Am Stat Assoc 52:356–360

Wilcoxon-Signed-Rank Test

Denise Rey[1], Markus Neuhäuser[2]
[1]Director
Rey Analytical Research, Hürth, Germany
[2]Professor
Koblenz University of Applied Sciences, Remagen, Germany

The Wilcoxon-signed-rank test was proposed together with the Wilcoxon-rank-sum test (see ►Wilcoxon–Mann–Whitney Test) in the same paper by Frank Wilcoxon in 1945 (Wilcoxon 1945) and is a nonparametric test for the one-sample location problem. The test is usually applied to the comparison of locations of two dependent samples. Other applications are also possible, e.g., to test the hypothesis that the median of a symmetrical distribution equals a given constant. As with many nonparametric tests, the distribution-free test is based on ranks.

To introduce the classical Wilcoxon-signed-rank test and also important further developments of it we denote by $D_i = Y_i - X_i$, $i = 1, \ldots, N$ the difference between two paired random variables. The classical Wilcoxon-signed-rank test assumes that the differences D_i are mutually independent and D_i, $i = 1, \ldots, N$ comes from a continuous distribution F that is symmetric about a median θ. The continuity assumption on the distribution of the differences implies that differences which are equal in absolute value may not occur, i.e., the classical Wilcoxon-signed-rank test assumes no ties in the differences $|d_i| \neq |d_j|$ for $i \neq j$ and $1 \leq i, j \leq N$. Moreover, it is assumed that the sample is free of zero differences, i.e., $d_i \neq 0$, $\forall i = 1, \ldots, N$. We further denote by N_0 and M the number of zero and the number of non-zero differences in the sample, respectively. It follows $N = N_0 + M$ with $N_0 = 0$ for the classical Wilcoxon-signed-rank test.

The null hypothesis states that $H_0 : \theta = 0$, i.e., the distibution of the differences is symmetric about zero corresponding to no difference in location between the two samples. The two-sided alternative is $H_1 : \theta \neq 0$. One-sided alternatives are also possible.

The Wilcoxon-signed-rank test statistic is the linear rank statistic $R_+ = \sum_{i=1}^{N} R_i V_i$ where $V_i = 1_{D_i > 0}$, is the indicator for the sign of the difference and R_i is the rank of $|D_i|$, $i = 1, \ldots, N$. Therefore, the test statistic represents the sum of the positive signed ranks. (The test statistic could also be build in terms of the sum of negative signed ranks, R_- or the difference of both $R = R_+ - R_-$. The three statistics are equivalent. For theoretical investigations is R often more suitable. Nevertheless, in literature R_+ and R_- are widespread.) The critical values w_α for the exact distribution of R_+ are tabulated. Reject the null hypothesis at the α level of significance if $R_+ \geq w_{\alpha/2}$ or $R_+ \leq \frac{N(N+1)}{2} - w_{\alpha/2}$.

Nowadays, the exact distribution can be determined by generating all 2^N sign permutations of the ranked differences. For each permutation, the value of the test statistic has to be calculated. The proportion of permutations that give a value as or more extreme than observed, is the p-value of the resulting exact test. Hence, in terms of p-values and due to the symmetry of the distribution, we reject the null hypothesis if the p-value $p = 2P(R_+ \geq r_+) \leq \alpha$ where r_+ is the observed value of the test statistic.

A large-sample approximation uses the asymptotic normal distribution of R_+. Under the null hypothesis we have

$$E_0(R_+) = \frac{N(N+1)}{4}, \quad Var_0(R_+) = \frac{N(N+1)(2N+1)}{4}$$

and the standardized version of R_+ is asymptotically

$$R_+^* = \frac{R_+ - E_0(R_+)}{\sqrt{Var_0(R_+)}} \overset{H_0}{\sim} N(0,1).$$

Reject the null hypothesis if $|R_+^*| \geq z_{1-\alpha/2}$.

In applications, the assumptions of the classical Wilcoxon-signed-rank test of non-zero differences and no ties in the sample are often not fulfilled.

We still assume that zero values are not possible but allow ties among the non-zero differences (the continuity assumption on the distribution of the differences is relaxed). Then one can apply the classical Wilcoxon-signed-rank test on the mean ranks that are associated with the tied group. In the case of ties among the non-zero differences, a conditional test based on the exact conditional distribution of the Wilcoxon signed-rank statistic given the set of tied ranks and by means of mean ranks is possible (Hollander and Wolfe 1999 p. 46).

For the large-sample approximation in the case of non-zero differences but existing tied observations among the non-zero differences, the variance of the test statistic changes to

$$Var_0(R_+) = \frac{1}{24}\left(N(N+1)(2N+1) - \frac{1}{2}\sum_{i=1}^{C} T_i(T_i - 1)(T_i + 1)\right) \tag{1}$$

where we have denoted by C the number of groups with ties and by $T_i \geq 1$, $i = 1, \ldots, C$ the number of observations within tie group i. It holds for this case $N = M = \sum_{i=1}^{C} T_i$. An untied observation is then considered to be a group of size 1 (Hollander and Wolfe 1999 p. 38). We remark that the classical Wilcoxon-signed-rank test assumes $T_i = 1$, $\forall i = 1, \ldots, C$. The test statistic R_+^* adapted by equation (1) is then computed with respect to mean ranks. Under the null hypothesis it is asymptotically normal distributed and corresponding tests can be applied.

In applications zero differences do often exist. Wilcoxon suggested dropping the zeros from the initial data and go on with the test on the reduced data.

Another method for handling zero differences was given by Pratt (Pratt 1959). Pratt suggested to rank all observations, including the zeros, from smallest to largest in absolute magnitude and afterwards drop the ranks of the zeros without changing the ranks of the non-zero values and proceed with the testing. In this case we have $N_0 > 0$ and ranks start by $N_0 + 1, \ldots, N$. Pratt motivated this procedure by showing that contradictory test decisions could occur when zero differences are ignored. More exactly, he showed that dropping the zeros before ranking fails to satisfy a monotonicity requirement: The probability under a test based on the signed rank statistic and randomized to have exact α level of calling a sample significantly positive should be a nondecreasing function of the observations (Pratt 1959, p. 659). Tables of critical values for a conditional exact Pratt test where a certain number of zero differences are allowed and mean ranks for ties are involved are computed by Buck (Buck 1979). Analogously, running through all 2^N sign permutations allows the computation of the exact distribution of the test statistic independent of tabulated values.

Asymptotically, the standardized test statistic where expectation and variance is properly adapted for the modification of Pratt is under the null hypothesis normal distributed and corresponding tests can be applied (Buck 1979).

Conover (Conover 1973, p. 985) showed that there are cases (e.g., the uniform distribution) for which the Wilcoxon test with the Pratt modification for handling zero differences and mean ranks for the non-zero differences has a greater asymptotic efficiency than the classical Wilcoxon test. Moreover he showed that there are also cases (e.g., the ▸binomial distribution) for which the Wilcoxon test with non-zero differences and mean ranks for the non-zero differences gives better asymptotic efficiency than the Pratt method.

Another well-known test for the one-sample location problem is the ▸sign test. Compared to the sign test, the Wilcoxon-signed-rank test has the additional assumption of the symmetry of the distribution but uses the ordering of the differences as additional information. In literature we find that there are advantageous cases with respect to the asymptotic efficiency for both the sign test and the Wilcoxon-signed-rank test (Higgins 2004).

About the Author

For biography *see* the entry ▸Wilcoxon-Mann-Whitney test.

Cross References

▸Asymptotic Normality
▸Asymptotic Relative Efficiency in Testing
▸Nonparametric Rank Tests
▸Nonparametric Statistical Inference
▸Sign Test
▸Student's *t*-Tests
▸Wilcoxon–Mann–Whitney Test

References and Further Reading

Buck W (1979) Signed-rank tests in presence of ties (with extended tables). Biom J 21(6):501–526

Conover WJ (1973) On methods of handling ties in the Wilcoxon signed-rank test. J Am Stat Assoc 68(344):985–988

Higgins JJ (2004) An introduction to modern nonparametric statistics. Brooks/Cole, Pacific Grove

Hollander M, Wolfe DA (1999) Nonparametric statistical methods. 2nd edn. Wiley, New York

Pratt JW (1959) Remarks on zeros and ties in the Wilcoxon signed rank procedures. J Am Stat Assoc 54:655–667

Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1:80–83

This encyclopedia includes no entries for, X, Y & Z.