Lecture Notes on Nonparametric Methods

January 19, 2010

Contents

Chapter 1

Introduction and basic definitions

Definition 1.1 Let $X_i : (\Omega_i, \mathcal{A}_i) \to \mathbb{R}$, $i = 1, \dots, n$, denote real-valued random variables. Further let $f : \mathbb{R}^n \to \mathbb{R}^m$, $m \in \mathbb{N}$, denote a measurable function. Then the random variable $T = f(X_1, \dots, X_n)$ is called a **statistic** based on $X_1, \dots, X_n$.

Remark 1.1 As random variables, the $X_i : (\Omega_i, \mathcal{A}_i) \to (\mathbb{R}, \mathcal{B})$ are measurable functions, with $\mathcal{B}$ denoting the Borel σ-algebra. Then the composition

$$f \circ (X_1, \dots, X_n) : \Omega_1 \times \dots \times \Omega_n \to \mathbb{R}^m$$

is measurable ($\mathcal{A}_1 \times \dots \times \mathcal{A}_n \to \mathcal{B}^m$), where $\mathcal{B}^m$ denotes the Borel σ-algebra on $\mathbb{R}^m$. Thus $T = f(X_1, \dots, X_n)$ is indeed a random variable.

Example 1.1 Let $m = 1$ and $f(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i$. Then

$$\bar{X} = f(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

is the **sample mean** of $X_1, \dots, X_n$.

Example 1.2 Let $m = 1$ and define $f(x_1, \dots, x_n) := \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$, where $\mu = \mu(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n x_i$. Then

$$S = f(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

is the **sample variance** of $X_1, \dots, X_n$.

Example 1.3 Define the function $\Psi : \mathbb{R} \to [0, 1]$ by

$$\Psi(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Let $m = 1$ and $f(x_1, \ldots, x_n) := \sum_{i=1}^{n} \Psi(x_i)$. Then

$$B = f(X_1, \ldots, X_n) = \sum_{i=1}^{n} \Psi(X_i)$$

is called the **sign test statistic**.

Example 1.4 Let $m = n$ and define $f(x_1, \ldots, x_n) := (x_{(1)}, \ldots, x_{(n)})$, where $x_{(1)} \leq \ldots \leq x_{(n)}$. Then

$$T = f(X_1, \ldots, X_n) = (X_{(1)}, \ldots, X_{(n)})$$

is called the **order statistic** of $X_1, \ldots, X_n$.

Definition 1.2 Let $\mathcal{C}$ denote a class of distributions on $\mathbb{R}^n$. A statistic T is called **distribution-free** over $\mathcal{C}$ if $T = f(X_1, \ldots, X_n)$ has the same distribution for all distributions $F \in \mathcal{C}$ of $X_1, \ldots, X_n$.

Example 1.5 Let $\mathcal{C} := \{N(\mu_0, \sigma^2)^n : \sigma^2 > 0\}$ denote the class of joint distributions of n independent and identically distributed (or shortly **iid**) random variables $X_1, \ldots, X_n$ that have a normal distribution with known mean $\mu_0 \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. Let $\bar{X}$ be the sample mean and S be the sample variance. Then the statistic $T = \sqrt{n}(\bar{X} - \mu_0)/\sqrt{S}$ has a t-distribution with $n - 1$ degrees of freedom for all $\sigma^2 > 0$. Thus T is distribution-free over $\mathcal{C}$. It is called the **t-statistic**. T is used to test the null hypothesis $H_0 : \mu = \mu_0$. Note that H_0 coincides with $\mathcal{C}$.

Definition 1.3 Let T denote a statistic that is distribution-free over a class $\mathcal{C}$ of distributions. If $\mathcal{C}$ is finite-dimensional, then T is called **parametric**. If $\mathcal{C}$ has infinite dimension, then T is called **nonparametric**.

Chapter 2

Counting Statistics

Example 2.1 Let $X_1, \ldots, X_n$ be independent random variables with a common p-quantile θ, i.e. $\mathbb{P}(X_i \leq \theta) = p$ for all $i = 1, \ldots, n$. Define the function

$$\Psi_\theta(x) := \Psi(x - \theta) = \begin{cases} 1, & x > \theta \\ 0, & x \leq \theta \end{cases}$$

Then the statistic $B := \sum_{i=1}^n \Psi_\theta(X_i)$ has a binomial distribution with parameter $1 - p$ and n degrees of freedom. Thus B is nonparametric distribution-free over the class $\mathcal{C} = \{\prod_{i=1}^n F_i : F_i \text{ is a distribution function}, F_i(\theta) = p\}$. B is called the **sign test statistic**. This generalises example 1.3, where $\theta = 0$.

This observation translates immediately to the following

Theorem 2.1 *Let $\mathcal{C}$ denote a class of distributions such that the events $A_1, \ldots, A_n$ are independent and $P(A_i) = p$ for all $i = 1, \ldots, n$ and $P \in \mathcal{C}$. Then the statistic $B := \sum_{i=1}^n 1_{A_i}$ has a binomial distribution with parameter p and n degrees of freedom for all $P \in \mathcal{C}$. B is called* **counting statistic**.

Exercise 2.1 Let $X_1, \ldots, X_n$ be iid continuous random variables which are symmetric around 0. Let F denote the distribution function of X_1. Describe a test based on a counting statistic for the null hypothesis $H_0 : F(1) - F(-1) = 1/2$.

Exercise 2.2 Consider the regression model $Y_i = \beta c_i + E_i$ for $i = 1, \ldots, n$, where the Y_i are the observations, $\beta \geq 0$ is an unknown parameter, the $c_i > 0$ are known constants, and the E_i are iid continuous random variables with median 0. Describe a test based on a counting statistic for the null hypothesis $H_0 : \beta = 0$.

Chapter 3

Ranking Statistics

Let $Z_1, \ldots, Z_n$ denote iid real-valued random variables with a continuous distribution function F. Let $Z_{(1)} \leq \ldots \leq Z_{(n)}$ denote the order statistic of $(Z_1, \ldots, Z_n)$, see example 1.4.

Remark 3.1 The event $Z_i = Z_j$ for $j \neq i$ is called a **tied rank**. For independent random variables with continuous distributions, tied ranks are null events, i.e. they have probablity 0. We can thus assume without loss of generality that there are no ties and $Z_{(1)} < \ldots < Z_{(n)}$.

Definition 3.1 The number $R_i^* \in \{1, \ldots, n\}$ such that $Z_{(R_i^*)} = Z_i$ is called the **rank** of Z_i. Let $R^* = (R_1^*, \ldots, R_n^*)$ denote the vector of ranks.

Theorem 3.1 *Let $Z_1, \ldots, Z_n$ denote iid real-valued random variables with a continuous distribution function and $R^* = (R_1^*, \ldots, R_n^*)$ be the corresponding vector of ranks. Then R^* is uniformly distributed over the set $\mathcal{R}$ of all permutations of $\{1, \ldots, n\}$.*

Proof: Let $\pi = (\pi_1, \ldots, \pi_n) \in \mathcal{R}$. Since $Z_1, \ldots, Z_n$ are iid, we obtain

$$(Z_{\pi_1}, \ldots, Z_{\pi_n}) \stackrel{d}{=} (Z_1, \ldots, Z_n)$$

where $\stackrel{d}{=}$ means equality in distribution. Hence

$$\mathbb{P}(R^* = \pi) = \mathbb{P}(Z_{\pi_1} < \ldots < Z_{\pi_n}) = \mathbb{P}(Z_1 < \ldots < Z_n) = \mathbb{P}(R^* = (1, \ldots, n))$$

Since π can be chosen arbitrarily, the statement follows.
□

Exercise 3.1 Show that $\mathbb{P}(R_i^* = k) = 1/n$ and

$$\mathbb{P}(R_i^* = k, R_j^* = l) = \frac{1}{n(n-1)}$$

for all $k \neq l \in \{1, \dots, n\}$ and $i \neq j \in \{1, \dots, n\}$.

Corollary 3.1 *Let $Z_1, \dots, Z_n$ denote iid real-valued random variables. Further let $R^* = (R_1^*, \dots, R_n^*)$ be the corresponding vector of ranks. Then any statistic $V(R^*)$ based on $R_1^*, \dots, R_n^*$ is distribution-free over the class of joint distributions for $Z_1, \dots, Z_n$.*

Definition 3.2 A statistic $V(R^*)$ based on the ranks $R_1^*, \dots, R_n^*$ of iid real-valued random variables is called a **rank statistic**.

Example 3.1 Two-sample location problem
Let $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ denote iid real-valued random variables with continuous distribution functions $F(x)$ and $G(x) = F(x - \Delta)$, respectively, where $\Delta \in \mathbb{R}$ is an unknown shift parameter. Let $R^* = (Q_1, \dots, Q_m, R_1, \dots, R_n)$ denote the vector of ranks for $(X_1, \dots, X_m, Y_1, \dots, Y_n)$. Then

$$W = \sum_{i=1}^{n} R_i \qquad\qquad U = \sum_{i=1}^{m} \sum_{j=1}^{n} \Psi(Y_j - X_i)$$

are called the **Wilcoxon** and the **Mann-Whitney** rank sum statistics, respectively. Under the null hypothesis $H_0 : \Delta = 0$, W is a rank statistic.

Exercise 3.2 Show that $W = U + \frac{n(n+1)}{2}$. This implies that under $H_0 : \Delta = 0$, U is a rank statistic, too.

Theorem 3.2 *Under $H_0 : \Delta = 0$, the distribution of W is given by*

$$P_0(W = k) = \frac{t_{m,n}(k)}{\binom{n+m}{n}}$$

where $t_{m,n}(k) = |\{A = \{k_1, \dots, k_n\} \subset \{1, \dots, n+m\} : \sum_{i=1}^{n} k_i = k\}|$.

Remark 3.2 $t_{m,n}(k)$ is the number of subsets of $\{1, \dots, n+m\}$ with n elements for which the sum of all elements equals k.

Proof: Under $H_0 : \Delta = 0$, the vector $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ consists of iid random variables with a continuous distribution function. By theorem 3.1 the vector of ranks $R^* = (Q_1, \ldots, Q_m, R_1, \ldots, R_n)$ is then uniformly distributed over all permutations of $\{1, \ldots, n+m\}$. Since $\binom{n+m}{n}$ is the number of all subsets of $\{1, \ldots, n+m\}$ with n elements, we obtain the statement.
□

Exercise 3.3 Determine the distribution of W for $m = 3$ and $n = 2$.

Theorem 3.3 *Under $H_0 : \Delta = 0$, the distribution of W is symmetric around $\mu = n \cdot (m + n + 1)/2$.*

Proof: Under H_0, the random variables $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are iid. Thus, the random variables $-X_1, \ldots, -X_m, -Y_1, \ldots, -Y_n$ are iid, too. This implies that

$$(Q_1, \ldots, Q_m, R_1, \ldots, R_n) \stackrel{d}{=} (N+1-Q_1, \ldots, N+1-Q_m, N+1-R_1, \ldots, N+1-R_n)$$

where $N = m + n$. From this we obtain

$$\sum_{i=1}^{n} R_i \stackrel{d}{=} \sum_{i=1}^{n} (N + 1 - R_i) = n \cdot (N + 1) - \sum_{i=1}^{n} R_i$$

which translates to

$$W - \frac{n}{2}(N + 1) \stackrel{d}{=} \frac{n}{2}(N + 1) - W$$

□

Chapter 4

Counting and ranking combined

Recall the function Ψ defined in example 1.3.

Lemma 4.1 *Let Z be a continuous random variable that is symmetric around 0. Then the random variables $|Z|$ and $\Psi(Z)$ are independent.*

Proof: For all $x > 0$ we obtain

$$\begin{aligned}\mathbb{P}(\Psi(Z) = 1, |Z| \leq x) = \mathbb{P}(Z > 0, |Z| \leq x) = \mathbb{P}(0 < Z \leq x) \\ = \frac{1}{2} \cdot \mathbb{P}(-x \leq Z \leq x)\end{aligned}$$

since Z is continuous and symmetric around 0. Since Z has median 0, the last expression equals

$$\frac{1}{2} \cdot \mathbb{P}(|Z| \leq x) = \mathbb{P}(\Psi(Z) = 1) \cdot \mathbb{P}(|Z| \leq x)$$

The proof for $\mathbb{P}(\Psi(Z) = 0, |Z| \leq x) = \mathbb{P}(\Psi(Z) = 0) \cdot \mathbb{P}(|Z| \leq x)$ is analogous and left as an exercise.
□

Definition 4.1 Let $Z_1, \ldots, Z_n$ be real-valued random variables. The **absolute rank** of Z_i is the rank of $|Z_i|$ among $|Z_1|, \ldots, |Z_n|$. It shall be denoted by R_i^+. The **signed rank** of Z_i is $\Psi(Z_i)R_i^+$. A statistic based on $\Psi(Z_1)R_1^+, \ldots, \Psi(Z_n)R_n^+$ is called a **signed rank statistic**.

Remark 4.1

$$\Psi(Z_i)R_i^+ = \begin{cases} R_i^+, & Z_i > 0 \\ 0, & Z_i \leq 0 \end{cases}$$

Theorem 4.1 *Let $Z_1, \dots, Z_n$ be iid continuous real-valued random variables that are symmetric around 0. Let $R^+ = (R_1^+, \dots, R_n^+)$ denote the vector of absolute ranks of $Z_1, \dots, Z_n$. Then the random variables $\Psi(Z_1), \dots, \Psi(Z_n), R^+$ are independent. Each $\Psi(Z_i)$ has a Bernoulli distribution with parameter $p = 1/2$ and R^+ is uniformly distributed over the set $\mathcal{R}$ of permutations on $\{1, \dots, n\}$.*

Proof: By lemma 4.1, independence of $Z_1, \dots, Z_n$ implies independence of the random variables $\Psi(Z_1), |Z_1|, \dots, \Psi(Z_n), |Z_n|$. Since R^+ depends on $|Z_1|, \dots, |Z_n|$ only, $\Psi(Z_1), \dots, \Psi(Z_n), R^+$ are independent, too. Since each Z_i is continuous and symmetric around 0, it is clear that $\Psi(Z_i) \sim Be(1/2)$ for all $i = 1, \dots, n$. Finally, theorem 3.1 yields $R^+ \sim U(\mathcal{R})$ as the $|Z_1|, \dots, |Z_n|$ are independent.
□

Corollary 4.1 *Assume that $Z_1, \dots, Z_n$ are iid random variables with $Z_1 \sim F$. Let S denote a statistic based on $\Psi(Z_1), \dots, \Psi(Z_n), R^+$. Then S is distribution-free over the class of joint distributions $\prod_{i=1}^n F$, where F is any continuous distribution function with $F(-x) = 1 - F(x)$ for all $x > 0$.*

Example 4.1 Let $X_1, \dots, X_n$ be iid continuous real-valued random variables that are symmetric around $\theta \in \mathbb{R}$, where θ is unknown. For some $\theta_0 \in \mathbb{R}$, define $Z_i := X_i - \theta_0$ for all $i = 1, \dots, n$. Let $R^+ = (R_1^+, \dots, R_n^+)$ denote the vector of absolute ranks for $Z_1, \dots, Z_n$. Then

$$W^+ := \sum_{i=1}^n \Psi(Z_i) R_i^+$$

is called the **Wilcoxon signed rank statistic**.

Theorem 4.2 *Under the null hypothesis $H_0 : \theta = \theta_0$, the distribution of W^+ is*

$$P_0(W^+ = k) = \frac{c_n(k)}{2^n}$$

for all $k = 0, 1, \dots, n(n+1)/2$, where $c_n(k) = |\{A \subset \{1, \dots, n\} : \sum_{i \in A} i = k\}|$.

Remark 4.2 $c_n(k)$ is the number of subsets of $\{1, \dots, n\}$ for which the sum of all elements equals k.

Proof: For $\psi \in \{0, 1\}^n$ and $r \in \mathcal{R}$ a permutation of $\{1, \dots, n\}$, the vector (ψ, r) is called a constellation. Theorem 4.1 yields

$$\mathbb{P}((\Psi(Z_1), \dots, \Psi(Z_n), R^+) = (\psi, r)) = \frac{1}{2^n} \cdot \frac{1}{n!}$$

for all constellations (ψ, r). Let $A \subset \{1, \ldots, n\}$ with $\sum_{i \in A} i = k$, and write $q := |A|$. Further write $\psi = (\psi_1, \ldots, \psi_n)$ and $r = (r_1, \ldots, r_n)$. There are $\binom{n}{q} q!(n-q)! = n!$ constellations such that $A = \{r_i : \psi_i = 1\}$. The definition of $c_n(k)$ yields

$$\mathbb{P}(W^+ = k) = c_n(k) \cdot n! \cdot \frac{1}{2^n} \cdot \frac{1}{n!} = \frac{c_n(k)}{2^n}$$

□

Exercise 4.1 Compute the distribution of W^+ for $n = 3$.

Theorem 4.3 *Under $H_0 : \theta = \theta_0$, the distribution of W^+ is symmetric around $\mu = n \cdot (n+1)/4$.*

Proof: By theorem 4.1, the random variables $\Psi(Z_1), \ldots, \Psi(Z_n), R^+$ are independent and $\Psi(Z_i) \sim Be(1/2)$ for all $i = 1, \ldots, n$. This implies that the random variables $1 - \Psi(Z_1), \ldots, 1 - \Psi(Z_n), R^+$ are independent, too, and

$$(\Psi(Z_1), \ldots, \Psi(Z_n), R^+) \stackrel{d}{=} (1 - \Psi(Z_1), \ldots, 1 - \Psi(Z_n), R^+)$$

Hence we obtain

$$\sum_{i=1}^{n} \Psi(Z_i) R_i^+ \stackrel{d}{=} \sum_{i=1}^{n} (1 - \Psi(Z_i)) R_i^+ = \frac{n(n+1)}{2} - \sum_{i=1}^{n} \Psi(Z_i) R_i^+$$

which means that

$$W^+ - \frac{n(n+1)}{4} \stackrel{d}{=} \frac{n(n+1)}{4} - W^+$$

□

Exercise 4.2 For random variables $X_1, \ldots, X_n$ and any $i \leq j \in \{1, \ldots, n\}$, the average $(X_i + X_j)/2$ of X_i and X_j is called a **Walsh average**. Show that W^+ equals the number of Walsh averages that are greater than θ_0.

Exercise 4.3 Let X and Y be two real-valued random variables and denote the difference by $Z := Y - X$. Assume that X and Y are symmetric around a and b, respectively. Show that then Z is symmetric around $b - a$.

Example 4.2 treatment effect (paired replicates)
Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid continuous bivariate random variables. Define $Z_i := Y_i - X_i$ for all $i = 1, \ldots, n$ and let $\theta \in \mathbb{R}$ denote the unknown median of Z_1. Then the Wilcoxon signed rank statistic W^+ may be used to test the null hypothesis $H_0 : \theta = \theta_0$. Often the X_i are called the **pre-treatment variables** and the Y_i are called the **post-treatment variables**. Then θ is called the **treatment effect**.

Chapter 5

Distribution of the order statistic

Let $X_1, \ldots, X_n$ denote iid real-valued random variables with an absolutely continuous distribution function F, i.e. there is a density function $f(x) = F'(x)$ for all $x \in \mathbb{R}$. Let $X_{(1)} < \ldots < X_{(n)}$ be the order statistic of $X_1, \ldots, X_n$ (see example 1.4) and recall remark 3.1.

Theorem 5.1 *The joint density function of the order statistic is given by*

$$\mathbb{P}(X_{(1)} \in dx_1, \ldots, X_{(n)} \in dx_n) = n! f(x_1) \ldots f(x_n)\, dx_1 \ldots dx_n$$

for all $x_1 < \ldots < x_n \in \mathbb{R}$.

Proof: The random variables $X_1, \ldots, X_n$ are independent, and there are $n!$ possible permutations of $\{1, \ldots, n\}$.
□

Theorem 5.2 *For all* $i \in \{1, \ldots, n\}$ *and* $x \in \mathbb{R}$,

$$\mathbb{P}(X_{(i)} \in dx) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}(1 - F(x))^{n-i} f(x) dx$$

Proof: The event $X_{(i)} \in dx$ implies $X_{(k)} < x$ for all $k = 1, \ldots, i-1$ and $X_{(k)} > x + dx$ for all $k = i+1, \ldots n$. This is the result of a trinomial experiment with probability

$$\frac{n!}{(i-1)! \cdot 1! \cdot (n-i)!}(F(x))^{i-1}(F(x+dx) - F(x))(1 - F(x))^{n-i}$$

The statement now follows from $F(x + dx) - F(x) = f(x)dx + o(dx)$.
□

Definition 5.1 Let F denote the distribution function of a real-valued random variable X. Define its **generalised inverse** by

$$F^{-1}(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}$$

for $0 < y < 1$. The random variable $F(X)$ is called the **probability integral transform** of X.

Remark 5.1 If F is strictly increasing, then F^{-1} coincides with the usual inverse function of F. If F is continuous, then $F(F^{-1}(y)) = y$ for all $0 < y < 1$. In general, $F(F^{-1}(y)) \geq y$ as the following example shows.

Example 5.1 Let F be the distribution function of the Dirac measure δ_0 on 0, i.e.

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Then $F^{-1}(1/2) = 0$ and $F(F^{-1}(1/2)) = F(0) = 1 > 1/2$.

Theorem 5.3 *Let X be a continuous random variable with distribution function F. Then the probability integral transform $F(X)$ is uniformly distributed over $[0, 1]$.*

Proof: Choose any $0 < y < 1$. Since F is continuous and increasing, the implication

$$X \leq F^{-1}(y) \quad \Rightarrow \quad F(X) \leq F(F^{-1}(y)) = y$$

holds certainly, i.e. $\{X \leq F^{-1}(y)\} \subset \{F(X) \leq y\}$. We further observe that $\mathbb{P}(F(X) = y) = 0$, since X is continuous. By definition of F^{-1}, the implication

$$X > F^{-1}(y) \quad \Rightarrow \quad F(X) \geq y$$

holds certainly, i.e. $\{F(X) < y\} \subset \{F^{-1}(y) \geq X\}$. Altogether we obtain

$$\mathbb{P}(F(X) \leq y) = \mathbb{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$$

Since $0 < y < 1$ can be chosen arbitrarily, this yields the statement.
□

Theorem 5.4 *Let $X_1, \ldots, X_n$ be iid absolutely continuous random variables with distribution function F and $X_{(1)} < \ldots < X_{(n)}$ their order statistic. Define $Z_{(i)} := F(X_{(i)})$ for all $i = 1, \ldots, n$. Then the density function of $Z_{(i)}$ is given by*

$$\mathbb{P}(Z_{(i)} \in dz) = \frac{n!}{(i-1)!(n-i)!} z^{i-1}(1-z)^{n-i} dz$$

for $0 < z < 1$.

Proof: Theorem 5.3 states that $F(X_1), \ldots, F(X_n)$ are iid with $Z_i := F(X_i) \sim U(0,1)$ for all $i = 1, \ldots, n$. Since F is increasing, $Z_{(1)} < \ldots < Z_{(n)}$ is the order statistic of $Z_1, \ldots, Z_n$. Now the statement follows from theorem 5.2.
□

Remark 5.2 The distribution in the above theorem is a beta distribution with positive integer parameters.

Chapter 6

The empirical distribution

Let $X_1, \ldots, X_n$ be iid real-valued random variables with an unknown distribution function G. Their **empirical distribution function** is defined as

$$\hat{F}_n(x) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}$$

for all $x \in \mathbb{R}$, where

$$1_A := \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases}$$

is the **indicator function** of an event A.

Remark 6.1 The X_i can be seen as independent observations from a population with distribution function G. Then $\hat{F}_n(x)$ is the number of observations of size at most x divided by the sample size n.

Remark 6.2 Under a null hypothesis $H_0 : G = F$, where F is a known distribution function, we obtain $P_0(X_i \leq x) = F(x)$ for all $x \in \mathbb{R}$ and $i = 1, \ldots, n$. Thus for a fixed $x \in \mathbb{R}$, the random variable $n \cdot \hat{F}_n(x)$ has a binomial distribution with parameter $F(x)$ and n degrees of freedom. This implies $\mathbb{E}(\hat{F}_n(x)) = F(x)$ and $\mathbb{V}ar(\hat{F}_n(x)) = F(x) \cdot (1 - F(x))/n$.

Exercise 6.1 Let Y be a real-valued random variable and $h(y) \geq 0$ an increasing function. Show that for any $\varepsilon > 0$,

$$\mathbb{P}(Y \geq \varepsilon) \leq \frac{1}{h(\varepsilon)} \mathbb{E}(h(Y))$$

Lemma 6.1 *Let Y and $(Y_n : n \geq 1)$ be real-valued random variables. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) < \infty$$

for all $\varepsilon > 0$, then $Y_n \to Y$ for $n \to \infty$ with probability 1.

Proof: The statement $\mathbb{P}(\lim_{n\to\infty} Y_n = Y) = 1$ holds if for every $\delta, \varepsilon > 0$ there is an $m \in \mathbb{N}$ such that

$$\mathbb{P}(|Y_n - Y| < \varepsilon \ \ \forall n \geq m) > 1 - \delta$$

This is equivalent to

$$\mathbb{P}\left(\bigcup_{n\geq m} \{|Y_n - Y| \geq \varepsilon\}\right) < \delta$$

Choose any $\delta, \varepsilon > 0$. By assumption, $\sum_{n=1}^{\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) < \infty$, which implies that there is an $m \in \mathbb{N}$ such that

$$\sum_{n=m}^{\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon) < \delta$$

Now the statement follows from

$$\mathbb{P}\left(\bigcup_{n\geq m} \{|Y_n - Y| \geq \varepsilon\}\right) \leq \sum_{n=m}^{\infty} \mathbb{P}(|Y_n - Y| \geq \varepsilon)$$

□

Theorem 6.1 *(Borel)*
Under $H_0 : G = F$ and for any $x \in \mathbb{R}$,

$$\hat{F}_n(x) \to F(x) \qquad \textit{for} \quad n \to \infty$$

holds with probability 1.

Proof: Define the random variables $Y_n := |\hat{F}_n(x) - F(x)|$ and the increasing function $h(y) := y^4$ for all $y \geq 0$. Then exercise 6.1 yields

$$\mathbb{P}\left(|\hat{F}_n(x) - F(x)| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^4}\mathbb{E}\left((\hat{F}_n(x) - F(x))^4\right)$$

Write $p := F(x)$ and $q := 1 - F(x)$. Since $Z := n\hat{F}_n(x)$ has a binomial distribution with parameter p and n degrees of freedom, the expectation on the right-hand side is related to the fourth centralised moment of Z as

$$\mathbb{E}\left((\hat{F}_n(x) - F(x))^4\right) = \frac{1}{n^4}\mathbb{E}\left((Z - np)^4\right) = \frac{1}{n^4}\left(3(npq)^2 + npq(1 - 6pq)\right)$$

see e.g. [1], p.110, for the last equality. Thus we obtain a bound

$$\mathbb{E}\left((\hat{F}_n(x) - F(x))^4\right) = \frac{1}{n^2}\left(3(pq)^2 + \frac{pq(1 - 6pq)}{n}\right) \leq \frac{C}{n^2}$$

which holds for all $n \geq 1$ for the constant $C = 3(pq)^2 + pq(1 - 6pq)$. Hence

$$\mathbb{P}\left(|\hat{F}_n(x) - F(x)| \geq \varepsilon\right) \leq \frac{C}{\varepsilon^4 n^2}$$

for all $\varepsilon > 0$ and $n \geq 1$. Since $\sum_{n=1}^{\infty} 1/n^2 < \infty$, the statement follows from lemma 6.1 with $Y_n := \hat{F}_n(x)$ and $Y := F(x)$.
□

Theorem 6.2 *(Glivenko-Cantelli)*
Under $H_0 : G = F$ and if F is continuous, then

$$\lim_{n\to\infty}\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|\right) = 0$$

with probability 1.

Proof: Let $r \in \mathbb{N}$ and $k \in \{1, \ldots, r-1\}$. Define $x_{r,k} := \inf\{x \in \mathbb{R} : F(x) = k/r\}$ and further $x_{r,r} := \infty$ as well as $x_{r,0} := -\infty$. Then for $k = 1, \ldots, r-2$ and $x \in [x_{r,k}, x_{r,k+1}[$ we obtain the bound

$$\begin{aligned}\hat{F}_n(x) - F(x) &\leq \hat{F}_n(x_{r,k+1}) - F(x_{r,k}) \\ &= \hat{F}_n(x_{r,k+1}) - F(x_{r,k+1}) + F(x_{r,k+1}) - F(x_{r,k}) \\ &\leq \hat{F}_n(x_{r,k+1}) - F(x_{r,k+1}) + \frac{1}{r}\end{aligned}$$

Analogously one arrives at

$$\hat{F}_n(x) - F(x) \geq \hat{F}_n(x_{r,k}) - F(x_{r,k}) - \frac{1}{r}$$

For $x < x_{r,1}$ the same arguments yield

$$-\frac{1}{r} \leq \hat{F}_n(x) - F(x) \leq \hat{F}_n(x_{r,1}) - F(x_{r,1}) + \frac{1}{r}$$

while for $x > x_{r,r-1}$ we obtain

$$\hat{F}_n(x_{r,r-1}) - F(x_{r,r-1}) - \frac{1}{r} \leq \hat{F}_n(x) - F(x) \leq \frac{1}{r}$$

Thus for any $r \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq \max_{1 \leq k \leq r-1} |\hat{F}_n(x_{r,k}) - F(x_{r,k})| + \frac{1}{r} \tag{6.1}$$

For all $1 \leq k \leq r-1$, define the events

$$E_{k,r} := \left\{ \lim_{n \to \infty} |\hat{F}_n(x_{r,k}) - F(x_{r,k})| = 0 \right\}$$

Theorem 6.1 yields $P_0(E_{k,r}) = 1$ for all $1 \leq k \leq r-1$. Defining $E_r := \bigcap_{k=1}^{r-1} E_{k,r}$, this implies $P_0(E_r) = 1$ for all $r \in \mathbb{N}$. For the limit $E := \bigcap_{r=1}^{\infty} E_r$ we then obtain $P_0(E) = 1$. Due to the bound (6.1), this implies the statement.
□

Definition 6.1 The **Kolmogorov-Smirnov** (one-sample) statistics are defined as

$$D_n^+ := \sup_{x \in \mathbb{R}} \left(\hat{F}_n(x) - F(x) \right), \quad D_n^- := \sup_{x \in \mathbb{R}} \left(F(x) - \hat{F}_n(x) \right)$$

and

$$D_n := \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| = \max \left\{ D_n^+, D_n^- \right\}$$

Theorem 6.3 *Assume that F is continuous. Then the statistics D_n^+, D_n^-, D_n are dsitribution-free under $H_0 : G = F$.*

Proof: By definition,

$$\hat{F}_n(x) = \begin{cases} 0, & x < X_{(1)} \\ k/n, & X_{(k)} \leq x \leq X_{(k+1)}, k = 1, \ldots, n-1 \\ 1, & x \geq X_{(n)} \end{cases}$$

where $X_{(1)} < \ldots < X_{(n)}$ is the order statistic of $X_1, \ldots, X_n$. Define $X_{(0)} := -\infty$ and $X_{(n+1)} := \infty$. Then

$$\begin{aligned} D_n^+ &= \sup_{x \in \mathbb{R}} \left(\hat{F}_n(x) - F(x) \right) \\ &= \max_{0 \le k \le n} \sup_{X_{(k)} \le x < X_{(k+1)}} \left(\frac{k}{n} - F(x) \right) \\ &= \max_{0 \le k \le n} \left(\frac{k}{n} - F(X_{(k)}) \right) \end{aligned}$$

since F is increasing. Analogously,

$$D_n^- = \max_{1 \le k \le n+1} \left(F(X_{(k)}) - \frac{k-1}{n} \right)$$

Finally $D_n = \max \{D_n^+, D_n^-\}$. Since F is continuous, theorem 5.4 states that the $F(X_{(k)})$ are distribution-free under H_0, which implies the statement.
□

Remark 6.3 For exact expressions of the null distribution for D_n^+, D_n^-, and D_n, see [3], section 3.5.b.

Bibliography

[1] N. L. Johnson, S. Kotz, and A. W. Kemp. *Univariate discrete distributions, 3rd ed.* Wiley, 2005.

[2] E. Manoukian. *Mathematical nonparametric statistics*. Gordon and Breach, 1986.

[3] R. Randles and D. Wolfe. *Introduction to the theory of nonparametric statistics*. Wiley, 1979.