ROBUST AND NONPARAMETRIC METHODS

Jana Jurečková

Contents

Chapter 1

Rank tests in linear regression model

1.1 Properties of ranks and order statistics

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector of observations; denote $X_{n:1} \leq X_{n:2} \ldots \leq X_{n:n}$ the components of $\mathbf{X}$ ordered according to increasing magnitude. The vector
$\mathbf{X}_{(.)} = (X_{n:1}, \ldots, X_{n:n})$ is called the *vector of order statistics* and $X_{n:i}$ is called the ith *order statistic*.
Assume that the components of $\mathbf{X}$ are different and define the *rank* of X_i as $R_i = \sum_{j=1}^{n} I[X_j \leq X_i]$. Then the vector $\mathbf{R}$ of ranks of $\mathbf{X}$ takes on the values in the set $\mathcal{R}$ of $n!$ permutations $(r_1, \ldots, r_n)$ of (1,. . . , n).

1.1.1 The distribution of $\mathbf{X}_{(.)}$ and of $\mathbf{R}$:

Lemma 1.1.1 *If* $\mathbf{X}$ *has density* $p_n(x_1, \ldots, x_n)$, *then the vector* $\mathbf{X}_{(.)}$ *of order statistics has the distribution with the density*

$$\bar{p}(x_{n:1}, \ldots, x_{n:n}) = \begin{cases} \sum_{r \in \mathcal{R}} p(x_{n:r_1}, \ldots, x_{n:r_n}) & \ldots x_{n:1} \leq \ldots \leq x_{n:n} \\ 0 & \text{otherwise.} \end{cases}$$

(ii) The conditional distribution of R *given* $\mathbf{X}_{(.)} = \mathbf{x}_{(.)}$ *has the form*

$$I\!P(R = r | \mathbf{X}_{(.)} = \mathbf{x}_{(.)}) = \frac{p(x_{n:r_1}, \ldots x_{n:r_n})}{\bar{p}(x_{n:1}, \ldots, x_{n:n})}$$

for any $r \in \mathcal{R}$ *and any* $x_{n:1} \leq \ldots \leq x_{n:n}$.

Proof. For any Borel set $B \in \mathcal{X}_{(.)}$ should hold

$$I\!P(\mathbf{X}_{(.)} \in B) = \sum_{r \in \mathcal{R}} I\!P(\mathbf{X}_{(.)} \in B, R = r) = \sum_{r \in \mathcal{R}} \int_{\mathbf{x}_{(.)} \in B, R = r} \ldots \int p(x_1, \ldots, x_n) dx_1, \ldots, dx_n$$
$$= \sum_{r \in \mathcal{R}} \int_B \ldots \int p(x_{n:r_1}, \ldots, x_{n:r_n}) dx_{n:1}, \ldots, x_{n:n} = \int_B \ldots \int \bar{p}(x_{n:1}, \ldots, x_{n:n}) dx_{n:1}, \ldots, x_{n:n},$$

what proves (i). Similarly,

$$\begin{aligned}
I\!P(\mathbf{X}_{(.)} \in B, R = r) &= \int_B \cdots \int p(x_{n:r_1}, \ldots, x_{n:r_n}) dx_{n:1}, \ldots, dx_{n:n} \\
&= \int_B \cdots \int \frac{p(x_{n:r_1}, \ldots, x_{n:r_n})}{\bar{p}(x_{n:1}, \ldots, x_{n:n})} \bar{p}(x_{n:1}, \ldots, x_{n:n}) dx_{n:1}, \ldots, dx_{n:n} \\
&= \int_B \cdots \int I\!P(R = r \big| \mathbf{X}_{(.)} = \mathbf{x}_{(.)}) \bar{p}(x_{n:1}, \ldots, x_{n:n}) dx_{n:1}, \ldots, dx_{n:n},
\end{aligned}$$

what proves (ii). □

We say that the random vector $\mathbf{X}$ satisfies the **h**ypothesis of randomness $\mathbf{H}_0$, if it has a probability distribution with density of the form

$$p(\mathbf{x}) = \prod_{i=1}^{n} f(x_i), \ \mathbf{x} \in I\!R^n$$

where f is an arbitrary one-dimensional density. Otherwise speaking, $\mathbf{X}$ satisfies the hypothesis of randomness provided its components are a random sample from an absolutely continuous distribution. We say that the random vector $\mathbf{X}$ satisfies the **h**ypothesis of exchangeability $\mathbf{H}_*$, if

$$p(x_1, \ldots, x_n) = p\left(x_{r_1}, \ldots, x_{r_n}\right)$$

for every permutation $(r_1, \ldots, r_n)$ of $1, \ldots, n$. If $\mathbf{X}$ satisfies $\mathbf{H}_0$, then it obviously satisfies $\mathbf{H}_*$. The following Lemma follows from Lemma 1.1.1.

Lemma 1.1.2 *If $\mathbf{X}$ satisfies $\mathbf{H}_0$ or $\mathbf{H}_*$, then $\mathbf{X}_{(.)}$ and $\mathbf{R}$ are independent, the vector of ranks $\mathbf{R}$ has the uniform discrete distribution*

$$I\!P(\mathbf{R} = r) = \frac{1}{n!}, \ r \in \mathcal{R}$$

and the distribution of $\mathbf{X}_{(.)}$ has the density

$$\bar{p}(x_{n:1}, \ldots, x_{n:n}) = \begin{cases} n! p(x_{n:1}, \ldots, x_{n:n}) & \ldots x_{n:1} \leq \ldots \leq x_{n:n} \\ 0 & \ldots \ \textit{otherwise.} \end{cases}$$

1.1.2 Marginal distributions of the random vectors R and $\mathbf{X}_{(.)}$ under $\mathbf{H}_0$:

Lemma 1.1.3 *Let $\mathbf{X}$ satisfy the hypothesis $\mathbf{H}_0$. Then*

(i) $\Pr(R_i = j) = \frac{1}{n} \ \forall i, j = 1, \ldots, n.$

(ii) $\Pr(R_i = k, R_j = m) = \frac{1}{n(n-1)}$
for $1 \leq i, j, k, m \leq n, i \neq j, \ k \neq m.$

(iii) $I\!ER_i = \frac{n+1}{2}, \ i = 1, \ldots, n.$

(iv) var $R_i = \frac{n^2-1}{12},\ i = 1, \ldots, n.$

(v) cov$(R_i, R_j) = -\frac{n+1}{12},\ 1 \le i, j \le n,\ i \ne j.$

(vi) If $\mathbf{X}$ *has density* $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$, *then* $X_{n:k}$ *has the distribution with density*

$$f_{(n)}(x) = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} (F(x))^{k-1}(1 - F(x))^{n-k} f(x), \quad x \in I\!R^1$$

where $F(x)$ *is the distribution function of* $X_1, \ldots, X_n$.

(vii) If $\mathbf{X}$ *has uniform* $R[0,1]$ *distribution, then* $X_{n:i}$ *has beta* $B(i, n-i+1)$ *distribution with the expectation and variance*

$$I\!EX_{n:i} = \frac{i}{n+1}, \quad \text{Var } X_{n:i} = \frac{i(n-i+1)}{(n+1)^2(n+2)}.$$

Proof. Lemma follows immediately from Lemma 1.1.2. □

1.2 Locally most powerful rank tests

We want to test a hypothesis of randomness $\mathbf{H}_0$ on the distribution of $\mathbf{X}$. The rank test is characterized by test function $\Phi(\mathbf{R})$. The most powerful rank $\alpha-$test of $\mathbf{H}_0$ against a simple alternative $\mathbf{K} : \{Q\}$ [that $\mathbf{X}$ has the fixed distribution Q] follows directly from the Neyman-Pearson Lemma:

$$\Phi(r) = \begin{cases} 1 & \ldots n!\ Q(R = r) > k_\alpha \\ 0 & \ldots n!\ Q(R = r) < k_\alpha \\ \gamma & \ldots n!\ Q(R = r) = k_\alpha,\ r \in \mathcal{R} \end{cases}$$

where k_α and γ are determined so that

$$\#\{r : n!\ Q(R = r) > k_\alpha)\} + \gamma\#\{r : n!\ Q(R = r) = k_\alpha\} = n!\alpha,\ 0 < \alpha < 1.$$

If we want to test against a composite alternative and the uniformly most powerful rank tests do not exist, then we look for a rank test, *most powerful locally* in a neighborhood of the hypothesis.

Definition 1.2.1 *Let* $d(Q)$ *be a measure of distance of alternative* $Q \in K$ *from the hypothesis* $\mathbf{H}$. *The* $\alpha-$*test* Φ_0 *is called the locally most powerful in the class* $\mathcal{M}$ *of* $\alpha-$*tests of* $\mathbf{H}$ *against* $\mathbf{K}$ *if, given any other test* $\Phi \in \mathcal{M}$, *there exists* $\varepsilon > 0$ *such that the power-functions of* Φ_0 *and* Φ *satisfy the inequality*

$$\beta_{\Phi_0}(Q) \ge \beta_\Phi(Q) \quad \forall Q \quad \textit{satisfying} \quad 0 < d(Q) < \varepsilon.$$

1.3 Structure of the locally most powerful rank tests of $\mathbf{H}_0$:

Theorem 1.3.1 *Let A be a class of densities, $A = \{g(x,\theta) : \theta \in \mathcal{J}\}$ such that*

$$\mathcal{J} \subset \mathbb{R}^1 \text{ is an open interval, } \mathcal{J} \ni 0.$$
$$g(x,\theta) \text{ is absolutely continuous in } \theta \text{ for almost all } x.$$

Moreover, let for almost all x there exist the limit

$$\dot{g}(x,0) = \lim_{\theta\to 0} \frac{1}{\theta}[g(x,\theta) - g(x,0)]$$
$$\text{and } \lim_{\theta\to 0} \int_{-\infty}^{\infty} |\dot{g}(x,\theta)|dx = \int_{-\infty}^{\infty} |\dot{g}(x,0)|dx.$$

Consider the alternative $\mathbf{K} = \{q_\Delta : \Delta > 0\}$, where

$$q_\Delta(x_1,\ldots,x_n) = \prod_{i=1}^{n} g(x_i, \Delta c_i),$$

$c_1,\ldots,c_n$ given numbers. *Then the test with the critical region*

$$\sum_{i=1}^{n} c_i a_n(R_i, g) \geq k$$

is the locally most powerful rank test of $\mathbf{H}_0$ against $\mathbf{K}$ on the significance level $\alpha = P(\sum_{i=1}^n c_i a_n(R_i,g) \geq k)$, where P is any distribution satisfying $\mathbf{H}_0$,

$$a_n(i,g) = \mathbb{E}\left[\frac{\dot{g}(X_{n:i},0)}{g(X_{n:i},0)}\right], \; i = 1,\ldots,n \quad \text{are the scores}$$

where $X_{n:1},\ldots,X_{n:n}$ are the order statistics corresponding to the random sample of size n from the population with the density $g(x,0)$.

Proof. Of Q_Δ is the probability distribution with the density q_Δ, then, for any permutation $\mathbf{r} \in \mathcal{R}$,

$$\lim_{\Delta\to 0} \frac{1}{\Delta}[n!\; Q_\Delta(\mathbf{R} = \mathbf{r}) - 1] = \sum_{i=1}^{n} c_i \; a_n(r_i, g). \qquad (1.3.1)$$

If (1.3.1) is true, then there exists an $\varepsilon > 0$ such that

$$\sum_{i=1}^{n} c_i \; a_n(r_i,g) > \sum_{i=1}^{n} c_i \; a_n(r_i',g) \implies Q_\Delta(\mathbf{R} = \mathbf{r}) > Q_\Delta(\mathbf{R} = \mathbf{r}')$$

for all $\Delta \in (0,\varepsilon)$ and for different $\mathbf{r},\mathbf{r}' \in \mathcal{R}$; then we reject Q_Δ for $\mathbf{r} \in \mathcal{R}$ such that $\sum_{i=1}^n c_i \; a_n(r_i,g) > k$ for a suitable k. So we must prove (1.3.1), what we shall do as follows: We can write

$$\frac{1}{\Delta}\left[Q_\Delta(\mathbf{R}=\mathbf{r})-Q_0(\mathbf{R}=\mathbf{r}\right]=\int_{\mathbf{R}=\mathbf{r}}\cdots\int\frac{1}{\Delta}\left[\prod_{i=1}^n g(x_i,\Delta c_i)-\prod_{i=1}^n g(x_i,0)\right]dx_1,\dots,dx_n$$
$$=\sum_{i=1}^n\int_{\mathbf{R}=\mathbf{r}}\cdots\int\frac{1}{\Delta}\left(g(x_i,\Delta c_i)-g(x_i,0)\right)\prod_{j=1}^{i-1}g(x_j,\Delta c_j)\prod_{k=i+1}^n g(x_k,0)dx_1,\dots,dx_n$$

where we used the identity

$$\prod_{i=1}^n A_i-\prod_{j=1}^n B_j=\sum_{i=1}^n(A_i-B_i)\prod_{j=1}^{i-1}A_j\prod_{k=i+1}^n B_k.$$

If $c_i>0$, then

$$\limsup_{\Delta\to 0}\int_{\mathbf{R}=\mathbf{r}}\cdots\int\frac{1}{\Delta}\left(g(x_i,\Delta c_i)-g(x_i,0)\right)\prod_{j=1}^{i-1}g(x_j,\Delta c_j)\prod_{k=i+1}^n g(x_k,0)dx_1,\dots,dx_n$$
$$\le c_i\int_{\mathbf{R}=\mathbf{r}}\cdots\int|\dot g(x_i,0)|\prod_{j\neq i}g(x_j,0)dx_1,\dots,dx_n,$$

analogously for $c_i<0$. This, combining with the Fatou lemma, leads to

$$\lim_{\Delta\to 0}\sum_{i=1}^n\int_{\mathbf{R}=\mathbf{r}}\cdots\int\frac{1}{\Delta}\left(g(x_i,\Delta c_i)-g(x_i,0)\right)\prod_{j=1}^{i-1}g(x_j,\Delta c_j)\prod_{k=i+1}^n g(x_k,0)dx_1,\dots,dx_n$$
$$=\sum_{i=1}^n\int_{\mathbf{R}=\mathbf{r}}\cdots\int c_i\dot g(x_i,0)\prod_{j\neq i}g(x_j,0)dx_1,\dots,dx_n$$
$$\sum_{i=1}^n c_i\int_{\mathbf{R}=\mathbf{r}}\cdots\int\frac{\dot g(x_i,0)}{g(x_i,0)}\prod_{j=1}^n g(x_j,0)dx_1,\dots,dx_n=\frac{1}{n!}\sum_{i=1}^n c_i\mathbb{E}\left[\frac{\dot g(X_i,0)}{g(X_i,0)}\Big|\mathbf{R}=\mathbf{r}\right]$$
$$=\frac{1}{n!}\sum_{i=1}^n c_i\ a_n(r_i,g).$$

regarding that $g(x,0)=0$ and $\dot g(x,0)\neq 0$ can happen simultaneously only on the set of measure 0. This implies (1.3.1). □

1.3.1 Special cases

I. *Two-sample alternative of the shift in location*: $\mathbf{K}_1:\{q_\Delta:\Delta>0\}$ where

$$q_\Delta(x_1,\dots,x_N)=\prod_{i=1}^m f(x_i)\prod_{i=m+1}^N f(x_i-\Delta)$$

with f being a fixed absolutely continuous density such that $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$. Then the locally most powerful rank $\alpha-$test of $\mathbf{H}_0$ against $\mathbf{K}$ has the critical region

$$\sum_{i=m+1}^{N} a_N(R_i, f) \geq k$$

where k satisfies the condition $P(\sum_{i=m+1}^{N} a_N(R_i, f) \geq k) = \alpha,\ P \in \mathbf{H}_0$ and

$$a_N(i,f) = I\!\!E\left[-\frac{f'(X_{N:i})}{f(X_{N:i})}\right],\ i = 1, \ldots, N$$

where $X_{N:1} < \ldots < X_{N:N}$ are the order statistics corresponding to the sample of size N from the distribution with the density f. The *scores* may be also written as

$$a_N(i,f) = I\!\!E\varphi(U_{N:i}, f),\ i = 1, \ldots, N$$

where $\varphi(u,f) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))},\ 0 < u < 1$ and $U_{N:1}, \ldots, U_{N:N}$ are the order statistics corresponding to the sample of size N from the uniform $R(0,1)$ distribution. Another form of the scores is

$$a_N(i,f) = N \binom{N-1}{i-1} \int_{-\infty}^{\infty} f'(x) F^{i-1}(x)(1-F(x))^{N-i} dx.$$

Remark 1.3.1 *The computation of the scores is difficult for some densities; if there are no tables of the scores at disposal, they are often replaced by the* approximate scores

$$a_N(i,f) = \varphi\left(\frac{i}{N+1}\right) = \varphi(I\!\!EU_{N:i}, f), i = 1, \ldots, N, \quad i = 1, \ldots, N.$$

The asymptotic critical values coincide for both types of scores.

II. *Alternative of simple linear regression:* $\mathbf{K}_2 = \{q_\Delta : \Delta > 0\}$ where $q_\Delta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i - \Delta c_i)$ with a fixed absolutely continuous density f and with given constants $c_1, \ldots, c_n,\ \sum_{i=1}^{n} c_i^2 > 0$. Then the locally most powerful rank α-test has the critical region

$$\sum_{i=1}^{n} c_i\ a_n(R_i, f) \geq k \tag{1.3.2}$$

with the the same scores as in case I, and with k determined by the condition

$$I\!\!P\left(\sum_{i=1}^{n} c_i\ a_n(R_i, f) > k\right) + \gamma I\!\!P\left(\sum_{i=1}^{n} c_i\ a_n(R_i, f) > k\right) = \alpha.$$

In the practice we most often use the test with the Wilcoxon scores: Put $\varphi(u) = u - \frac{1}{2}$ and reject $\mathbf{H}_0$ provided

$$W_n = \sum_{i=1}^{n} c_i\ R_i > k, \quad \text{where } k \text{ is such that}$$

$$P\left(\sum_{i=1}^{n} c_i\ R_i > k \middle| \mathbf{H}_0\right) + \gamma P\left(\sum_{i=1}^{n} c_i\ R_i = k \middle| \mathbf{H}_0\right) = \alpha,\ 0 \le \gamma < 1.$$

This test is the locally most powerful against $\mathbf{K}_2$ with F logistic with the density

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2},\ x \in I\!R$$

but is rather efficient also for other alternatives. For large n we use the *normal approximation* of W_n : If $n \to \infty$, then W_n has asymptotically normal distribution under $\mathbf{H}_0$ in the following sense:

$$\lim_{n\to\infty} P_{H_0}\left\{\frac{W_n - I\!EW_n}{\sqrt{\text{var } W_n}} < x\right\} = \Phi(x),\ x \in I\!R^1,$$

where Φ is the standard normal distribution function.

To be able to use the normal approximation, we must know the expectation and variance of W_n under $\mathbf{H}_0$. The following Lemma gives the expectation and the variance of a more general linear rank statistic, covering the Wilcoxon as well other rank tests.

Lemma 1.3.1 *Let the random vector* $(R_1, \ldots, R_n)$ *have the discrete uniform distribution on the set* $\mathcal{R}$ *of all permutations of numbers* $1, \ldots, n$, *i.e.* $I\!P(\mathbf{R} = \mathbf{r}) = \frac{1}{n!}$, $\mathbf{r} \in \mathcal{R}$; *let* $c_1, \ldots, c_N$ *and* $a_1 = a(1), \ldots, a_n = a(n)$ *are arbitrary constants. Then the expectation and variance of the linear statistic* $S_n = \sum_{i=1}^n c_i\ a(R_i)$ *are*

$$I\!ES_N = \frac{1}{n}\sum_{i=1}^{n} c_i \sum_{j=1}^{n} a_j$$

$$\text{var } S_n = \frac{1}{n-1}\sum_{i=1}^{n}(c_i - \bar{c})^2 \sum_{j=1}^{n}(a_j - \bar{a})^2,$$

where $\bar{c} = \frac{1}{n}\sum_{i=1}^n c_i, \quad \bar{a} = \frac{1}{n}\sum_{i=1}^n a_i.$

Proof. The proposition follows from the distribution of $\mathbf{R}$ under $\mathbf{H}_0$.

1.4 Rank tests for simple regression model with nonrandom regressors

Let $X_1, \ldots, X_N$ be independent random variables with continuous distribution funtions $F_1, \ldots, F_N$, where

$$F_i(x) = F(x - \beta_0 - \beta c_i), \quad i = 1, \ldots, N,\ x \in \mathbb{R},$$

F is continuous, $\mathbf{c}_N = (c_1, \ldots, c_n)'$ is a vector of (known) regression constants (not all equal), and (β_0, β) are unknown parameters; we call β_0 an *intercept* of the regression line and β is called the *regression coefficient*. Our first hypothesis is that there is no regression,

$$\mathbf{H}_0^{(1)} :\ \beta = 0 \ \text{ against } \ \mathbf{K}^{(1)} :\ \beta \neq 0 \ \text{ or } \ \mathbf{K}_+^{(1)} :\ \beta > 0, \qquad (1.4.1)$$

where β_0 is considered as a nuisance parameter. We may be also interested in the joint hypothesis

$$\mathbf{H}_0^{(2)}: \ (\beta_0, \beta) = \mathbf{0} \ \text{ against } \ \mathbf{K}^{(2)}: \ (\beta_0, \beta) \neq \mathbf{0}. \tag{1.4.2}$$

The third hypothesis is

$$\mathbf{H}_0^{(3)}: \ \beta_0 = 0 \ \text{ against } \ \mathbf{K}^{(3)}: \ \beta_0 \neq 0 \ \text{ or } \ \mathbf{K}_+^{(3)}: \ \beta_0 > 0, \tag{1.4.3}$$

where β is treated as a nuisance parameter.

In either case there exists a *distribution-free* rank test, whose critical values do not depend on F. We can also consider $\beta = \beta^*$ or $(\beta_0, \beta) = (\beta_0^*, \beta^*)$; then we work with $X_i^* = X_i - \beta_0^* - \beta^* c_i, \ i = 1, \ldots, N$.

1.4.1 Rank tests for $\mathbf{H}_0^{(1)}$

Let $\mathbf{R}_N = (R_{N1}, \ldots, R_{NN})$ be the ranks of $X_1, \ldots, X_N$. Choose some nondecreasing *score function* $\varphi : (0,1) \mapsto \mathbb{R}$ and put

$$S_N = \sum_{i=1}^N (c_i - \bar{c}_N) a_N(R_{Ni}), \quad \bar{c}_N = \frac{1}{N} \sum_{i=1}^N c_i \tag{1.4.4}$$

where the scores have the form

$$a_N(i) = \mathbb{E}\varphi(U_{N:i}) \quad \text{ or } \quad \varphi\left(\frac{i}{N+1}\right), \quad 1 \leq i \leq N, \tag{1.4.5}$$

where $U_{N:1} \leq \ldots U_{N:N}$ are the order statistics corresponding to the sample $U_1, \ldots, U_N$ from the uniform $R(0,1)$ distribution. Under $\mathbf{H}_0^{(1)}$, it holds $F_1(x) = \ldots = F_N(x) = F(x - \beta_0) = F_0(x)$ (say), where F_0 is continuous. Because the ties between $X_1, \ldots, X_N$ can happen with probability 0, we have

$$\mathbb{P}\left\{\mathbf{R}_N = \mathbf{r}_N \middle| \mathbf{H}_0^{(1)}\right\} = \frac{1}{N!} \quad \forall \mathbf{r}_N \in \mathcal{R}_N \quad \text{(permutations)},$$

hence

$$\mathbb{P}\{R_{Ni} = k | \ \mathbf{H}_0^{(1)}\} = \tfrac{1}{N} \quad \forall i, k, \ 1 \leq i, k \leq N$$

$$\mathbb{P}\{R_{Ni} = k, R_{Nj} = \ell | \ \mathbf{H}_0^{(1)}\} = \tfrac{1}{N(N-1)} \quad \forall i, j, k, \ell, \ 1 \leq i \neq j, k \neq \ell \leq N.$$

Hence,

$$\mathbb{E}\{S_N | \ \mathbf{H}_0^{(1)}\} = \sum_{i=1}^N (c_i - \bar{c}_N) \mathbb{E}\{a_N(R_{Ni}) | \ \mathbf{H}_0^{(1)}\} = \frac{1}{N} \sum_{i=1}^N (c_i - \bar{c}_N) \sum_{j=1}^N a_N(i) = 0,$$

$$\text{Var}\ \{S_N | \ \mathbf{H}_0^{(1)}\} = \frac{1}{N-1} \sum_{i=1}^N (c_i - \bar{c}_N)^2 \sum_{j=1}^N (a_N(i) - \bar{a}_N)^2$$

The distribution of S_N under $\mathbf{H}_0^{(1)}$ does not depend on F and on β_0, hence we reject $\mathbf{H}_0^{(1)}$ in favor of $\{\mathbf{K}_+^{(1)} : \ \beta > 0\}$ when $S_N > k_\alpha^+$ and reject with probability γ when $S_N = k_\alpha^+$, where k_α^+ is determined so that

$$I\!P\{S_N > k_\alpha^+ | \ \mathbf{H}_0^{(1)}\} + \gamma I\!P\{S_N = k_\alpha^+ | \ \mathbf{H}_0^{(1)}\} = \alpha$$

and $\alpha = 0.05$ or 0.01, for instance. Similarly, we reject $\mathbf{H}_0^{(1)}$ in favor of $\{\mathbf{K}^{(1)} : \ \beta \neq 0\}$ when $|S_N| > k_\alpha$ and reject with probability $\gamma \in [0,1)$ when $|S_N| = k_\alpha$, where k_α is determined so that

$$I\!P\{|S_N| > k_\alpha | \ \mathbf{H}_0^{(1)}\} + \gamma I\!P\{|S_N| = k_\alpha | \ \mathbf{H}_0^{(1)}\} = \alpha.$$

For small N we can calculate the critical values k_α^+ and k_α; but for large N we must use an asymptotic approximation. The asymptotic distribution of S_N under $\mathbf{H}_0^{(1)}$ is based on the following theorems, proved by Hájek (1961):

Theorem 1.4.1 *Let* $\mathbf{R}_N = (R_{N1}, \ldots, R_{NN})$ *be a random vector such that*

$$I\!P\{\mathbf{R} = \mathbf{r}\} = \frac{1}{N!} \quad \forall \mathbf{r} \in \mathcal{R}$$

and let $\{a_N(i), \ 1 \le i \le N\}$ *and* $\{c_N(i), \ 1 \le i \le N\}$ *be two sequences of real numbers such that, as* $N \to \infty$,

$$\max_{1\le i\le N} \frac{(a_N(i) - \bar{a}_N)^2}{\sum_{j=1}^N (a_N(j) - \bar{a}_N)^2} \to 0, \quad \max_{1\le i\le N} \frac{(c_N(i) - \bar{c}_N)^2}{\sum_{j=1}^N (c_N(j) - \bar{c}_N)^2} \to 0 \quad \textit{(Noether condition).} \tag{1.4.6}$$

Then

$$I\!P\left\{\frac{S_N - I\!E S_N}{\sqrt{\operatorname{Var} S_N}} \le x\right\} \to \Phi(x) \quad \text{as} \ \ N \to \infty \quad \forall x \in \mathbb{R}$$

where Φ *is the standard normal distribution function, if and only if, for every* $\varepsilon > 0$,

$$\lim_{N\to\infty} \left\{\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \kappa_{N,\ ij}^2 \ I[|\kappa_{N,\ ij}| > \varepsilon]\right\} = 0 \quad \textit{(Lindeberg condition)} \tag{1.4.7}$$

and

$$\kappa_{N,\ ij} = \frac{(a_N(i) - \bar{a}_N)(c_N(j) - \bar{c}_N)}{\left\{N^{-1} \sum_{k=1}^N (a_N(k) - \bar{a}_N)^2 \sum_{\ell=1}^N (c_N(\ell) - \bar{c}_N)^2\right\}^{1/2}}, \quad i, j = 1, \ldots, N.$$

Theorem 1.4.2 (Projection theorem). *If* $a_N(1) \le \ldots \le a_N(N)$ *and*

$$\max_{1\le i\le N} \frac{(a_N(i) - \bar{a}_N)^2}{\sum_{j=1}^N (a_N(j) - \bar{a}_N)^2} \to 0 \quad \text{as} \ \ N \to \infty,$$

then S_N *is asymptotically equivalent in the quadratic mean to the statistic*

$$T_N = \sum_{i=1}^N (c_N(i) - \bar{c}_N) a_N^0(U_i) + N \bar{c}_N \bar{a}_N$$

in the sense that

$$\lim_{N\to\infty} I\!E \left[\frac{(S_N - T_N)^2}{\text{Var } S_N}\right] = 0.$$

Here

$$a_N^0(i) = a_N(i) \quad \textit{for} \quad \frac{i-1}{N} < u \le \frac{i}{N}, \quad i = 1, \ldots, N$$

and $U_1, \ldots, U_N$ is a random sample from the uniform $R(0,1)$ distribution.

Corollary 1.4.1 *Let*

$$\kappa_{N,\ ij} = \frac{(a_N(i) - \bar{a}_N)(c_i - \bar{c}_N)}{A_N C_N}, \quad i, j = 1, \ldots, N,$$

$$A_N^2 = (N-1)^{-1} \sum_{k=1}^{N} (a_k - \bar{a}_N)^2, \quad C_N^2 = \sum_{\ell=1}^{N} (c_\ell - \bar{c}_N)^2,$$

and let the sequences $\{a_N(1) \ldots, a_N(N)\}$ and $\{c_1, \ldots, c_N\}$ satisfy the Noether condition (1.4.6). Then

$$\lim_{N\to\infty} I\!P \left\{ \frac{S_N}{A_N C_N} \le x \middle| \mathbf{H}_0^{(1)} \right\} = \Phi(x) \quad \forall x \in \mathbb{R}.$$

The asymptotic rank test rejects $\mathbf{H}_0^{(1)}$ in favor of $\mathbf{K}_+^{(1)}$ on the significance level α if

$$\frac{S_N}{A_N C_N} \ge \Phi^{-1}(1-\alpha)$$

and in favor of $\mathbf{K}^{(1)}$ if

$$\frac{|S_N|}{A_N C_N} \ge \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big),$$

respectively.

1.4.2 Rank tests for $\mathbf{H}_0^{(2)}$

The hypothesis

$$\mathbf{H}_0^{(2)} : \ (\beta_0, \beta) = \mathbf{0}$$

we shall test under the condition of symmetry on F, i.e.

$$F(x) + F(-x) = 1 \quad \text{for} \ \ x \in \mathbb{R}.$$

Because the ranks are invariant to the shift in location, the test should also involve the signs of observations. Let R_{Ni}^+ be the rank of $|X|_{Ni}$ among $|X|_{N1}, \ldots, |X|_{NN},\ i = 1, \ldots, N$. Choose a score-generating function $\varphi^* : (0,1) \mapsto [0, \infty)$ and the scores $a_N^*(1), \ldots, a_N^*(N)$ generated by $\varphi*$ in the same manner as in (1.4.5). Under the hypothesis $\mathbf{H}_0^{(2)}$, the observations are independent and identically distributed with a continuous distribution function F, symmetric about 0. Consider two statistics

$$S_{N,1}^+ = \sum_{i=1}^{N} a_N^*(R_{Ni}^+) \text{sign } X_i, \quad S_{N,2}^+ = \sum_{i=1}^{N} c_i a_N^*(R_{Ni}^+) \text{sign } X_i, \quad \mathbf{S}_N = (S_{N,1}^+, S_{N,2}^+)'$$

and denote

$$\lambda_{11}^{(N)} = N, \quad \lambda_{12}^{(N)} = \sum_{i=1}^{N} c_i, \quad \lambda_{22}^{(N)} = \sum_{i=1}^{N} c_i^2, \quad \mathbf{\Lambda}^{(N)} = \left\| \lambda_{ij}^{(N)} \right\|_{i,j=1,2}.$$

Under $\mathbf{H}_0^{(2)}$ and under symmetry of F, the vector (sign $X_1 \cdot R_{N1}^+, \ldots,$ sign $X_N \cdot R_{NN}^+$) can take on $N!2^N$ values, each with probability $1/(N!2^N)$, and sign X_i is independent of $R_{Ni}^+,\ i = 1, \ldots, N$. Hence,

$$\begin{aligned}
&I\!E(\mathbf{S}_N^+ | \mathbf{H}_0^{(2)}) = \mathbf{0}, \\
&I\!E(\mathbf{S}_N^+ \mathbf{S}_N^{+\prime} | \mathbf{H}_0^{(2)}) = A_N^{*2} \mathbf{\Lambda}^{(N)}, \\
&A_N^{*2} = \frac{1}{N} \sum_{i=1}^{N} \left(a_N^*(i)\right)^2.
\end{aligned}$$

Consider the following test criterion

$$W_N^+ = \mathbf{S}_N^{+\prime} \left(I\!E_{\mathbf{H}_0^{(2)}} \mathbf{S}_N^+ \mathbf{S}_N^{+\prime} \right)^{-1} \mathbf{S}_N^+ = \left(\mathbf{S}_N^{+\prime} \mathbf{\Lambda}_N^{-1} \mathbf{S}_N \right) / A_N^{*2}. \qquad (1.4.8)$$

Under $\mathbf{H}_0^{(2)}$ and under symmetry of F, the distribution of W_N^+ does not depend on the unknown F. However, the exact distribution of W_N^+ is very laborious to calculate, hence we should again use the asymptotic approximation. The asymptotic behavior is described in the following theorem:

Theorem 1.4.3 *Assume that the sequences* $\{a_N(i),\ 1 \le i \le N\}$ *and* $\{c_{Ni},\ 1 \le i \le N\}$ *satisfy, as* $N \to \infty$,

$$\frac{\max_{1 \le i \le N} a_N^2(i)}{\sum_{j=1}^{N} a_N^2(j)} \to 0, \qquad \frac{\max_{1 \le i \le N} c_{Ni}^2}{\sum_{j=1}^{N} c_{Nj}^2} \to 0.$$

Denote

$$\kappa_{N,ij} = \frac{a_N(i) c_{Nj}}{\left[N^{-1} \sum_{k=1}^{N} a_N^2(k) \sum_{\ell=1}^{N} c_{N\ell}^2 \right]^{1/2}}, \quad i, j = 1, \ldots, N.$$

Then, under $\mathbf{H}_0^{(2)}$ *and under symmetry of* F*, the sequence* $(S_{N2}^+ - I\!E S_{N2}^+)/\sqrt{\mathrm{Var} S_{N2}^+}$ *is asymptotically normally distributed* $N(0,1)$ *if and only if, for every* $\varepsilon > 0$,

$$\lim_{N \to \infty} \left\{ \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \kappa_{N,ij}^2 I[|\kappa_{N,ij}| > \varepsilon] \right\} = 0 \qquad \textit{(Lindeberg condition).}$$

If we further apply Theorem 1.4.3 to $c_{ni} = 1,\ i = 1, \ldots, N$, we conclude that the random vector $\mathbf{S}_N^+$ has asymptotically a bivariate normal distribution $\mathcal{N}_2\left(\mathbf{0},\ A_N^* \mathbf{\Lambda}^{(N)}\right)$. This implies that under $\mathbf{H}_0^{(2)}$ and under symmetry of F, W_N^+ has asymptotically χ^2 distribution with 2 degrees of freedom. Hence, the asymptotic test rejects $\mathbf{H}_0^{(2)}$ in favor $\mathbf{K}^{(2)}$ if $W_N^+ \ge \chi_{2,\alpha}^2$.

1.4.3 Example

A group of students, boys and girls, graduated in a summer language course. They passed two tests, before and after the course. The responses in the table are differences in the tests scores for each individual; $c_i = 1$ for a boy and $c_i = -1$ for a girl.

#	response	c_i	R_{Ni}	R_{Ni}^+	$c_i R_{Ni}$	sign $X_i R_{Ni}^+$
1	5.2	1	19	19	19	19
2	-0.7	1	6	63	6	-6
3	-2.3	1	2	13	2	-13
4	3.2	1	16	15	16	15
5	-1.5	1	4	9	4	-9
6	4.7	1	18	18	18	18
7	1.8	1	14	12	14	12
8	-0.4	1	8	3	8	-3
9	0.6	1	11	5	11	5
10	6.6	1	20	20	20	20
11	-0.9	-1	5	8	-5	-8
12	1.7	-1	13	11	-13	11
13	-0.3	-1	9	2	-9	-2
14	2.4	-1	15	14	-15	146
15	4.2	-1	17	16	-17	16
16	-1.6	-1	3	10	-3	-10
17	-4.3	-1	1	17	-1	-17
18	0.8	-1	12	7	-12	7
19	-0.5	-1	7	4	-7	-4
20	-0.2	-1	10	1	-10	-1

We want to test whether the course had an effect and whether there is a difference between the performance of boys and girls. We take the Wilcoxon scores, $a_N(i) = a_N^*(i) = \frac{i}{21},\ i = 1, \ldots, 20$ and get

$$\frac{S_N}{A_N C_N} = 0.9826 < 1.96 = \Phi^{-1}(0.95),$$

$$W_N^+ = 2.368 < 5.99 = \chi_2^2(0.95).$$

Hence, we cannot reject either of the hypotheses.

1.5 Rank tests for some multiple linear regression models

Consider the linear regression model

$$Y_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + e_i, \quad i = 1, \dots, N \tag{1.5.1}$$

where $\beta_0 \in \mathbb{R}_1$, $\boldsymbol{\beta} \in \mathbb{R}_p$ are unknown parameters and $e_i, \dots, e_N$ are independent errors, identically distributed according to a continuous d.f. F and $\mathbf{x}_i \in \mathbb{R}_p$ are given regressors, $i = 1, \dots, N$. Denote

$$\mathbf{X}_N = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_N' \end{bmatrix}$$

the regression matrix. We shall first consider the hypotheses

$$\mathbf{H}_0^{(1)} : \ \boldsymbol{\beta} = \mathbf{0} \ \text{ versus } \ \mathbf{K}^{(1)} : \ \boldsymbol{\beta} \neq \mathbf{0}$$

and

$$\mathbf{H}_0^{(2)} : \boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}')' = \mathbf{0} \ \text{ versus } \ \mathbf{K}^{(2)} : \ \boldsymbol{\beta}^* \neq \mathbf{0}.$$

The hypotheses and tests are extensions of those for the regression line.

1.5.1 Rank tests for $\mathbf{H}_0^{(1)}$

Let $R_{N1}, \dots, R_{NN}$ be the ranks of $Y_1, \dots, Y_N$ and let $a_N(1), \dots, a_N(N)$ be the scores generated by a nondecreasing, square-integrable score function $\varphi : \ (0,1) \mapsto \mathbb{R}_1$ so that $a_N(i) = \varphi\left(\frac{i}{N+1}\right), \ i = 1, \dots, N$.

Consider the linear rank statistics

$$S_{Nj} = \sum_{i=1}^{N} (x_{ij} - \bar{x}_{Nj}) a_N(R_{Ni}), \quad \bar{x}_{Nj} = \frac{1}{N} \sum_{i=1}^{N} x_{ij}, \quad j = 1, \dots, N$$

and the vector

$$\mathbf{S}_N = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}}_N) a_N(R_{Ni}) = (S_{N1}, \dots, S_{Np})'.$$

The distribution function of observation Y_i under $\mathbf{H}_0^{(1)}$ is $F(y - \beta_0), \ i = 1, \dots, N$. Hence, $(R_{N1}, \dots, R_{NN})$ assumes all possible permutations of $(1, 2, \dots, N)$ with equal probability $\frac{1}{N!}$. Hence, the expectation and covariance matrix of $\mathbf{S}_N$ under $\mathbf{H}_0^{(1)}$ are

$$\mathbb{E}(\mathbf{S}_N | \mathbf{H}_0^{(1)}) = \mathbf{0} \quad \text{and} \quad \mathbb{E}(\mathbf{S}_N \mathbf{S}_N' | \mathbf{H}_0^{(1)}) = A_N^2 \mathbf{Q}_N,$$

where

$$A_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} (a_N(i) - \bar{a}_N)^2, \quad \mathbf{Q}_N = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}}_N)(\mathbf{x}_i - \bar{\mathbf{x}}_N)'.$$

Our test for $\mathbf{H}_0^{(1)}$ is based on the quadratic form

$$\mathcal{S}_N = A_N^{-2}\left(\mathbf{S}_N'\mathbf{Q}_N^{-1}\mathbf{S}_N\right), \tag{1.5.2}$$

where $\mathbf{Q}_N^{-1}$ is replaced by the generalized inverse $\mathbf{Q}_N^{-}$ if $\mathbf{Q}_N$ is singular. We reject $\mathbf{H}_0^{(1)}$ if $\mathcal{S}_N > k_\alpha$ where k_α is a suitable critical value.

Notice that $\mathbf{S}_N$ depends only on $\mathbf{x}_1, \ldots, \mathbf{x}_N$, on the scores $a_N(1), \ldots, a_N(N)$ and on the ranks $R_{N1}, \ldots, R_{NN}$. Hence, the distribution of $\mathbf{S}_N$ and thus also that of $\mathcal{S}_N$ under the hypothesis $\mathbf{H}_0^{(1)}$ does not depend on the distribution function F of the errors. For small N, the critical value can be calculated numerically, but it would become laborious with increasing N. Hence, again, we should use the large-sample approximation. This can be derived under some conditions on the matrix $\mathbf{X}_N$, and on the scores:

Theorem 1.5.1 *Assume that*

(i) *the matrix $\mathbf{Q}_N$ is regular for $N > N_0$ and*

$$\max_{1\le i\le N}(\mathbf{x}_i - \bar{\mathbf{x}}_N)'\mathbf{Q}_N^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_N) \to 0 \quad \text{as} \quad N \to \infty,$$

(ii) *the scores satisfy the Noether condition, i.e.*

$$\max_{1\le i\le N} \frac{(a_N(i) - \bar{a}_N)^2}{\sum_{j=1}^N (a_N(j) - \bar{a}_N)^2} \to 0 \quad \text{as} \quad N \to \infty,$$

(iii)

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \delta_{N,ijk}^2 I[|\delta_{N,ijk}| > \varepsilon]\right] = 0 \quad \text{for every} \ \ \varepsilon > 0, \ \forall k = 1, \ldots, p,$$

where

$$\delta_{N,ijk} = \frac{(a_N(i) - \bar{a}_N)(x_{jk} - \bar{x}_k)}{\left[N^{-1}\sum_{i=1}^N (a_N(i) - \bar{a}_N)^2 \sum_{j=1}^N (x_{jk} - \bar{x}_k)^2\right]^{1/2}}, \quad k = 1, \ldots, p, \ i, j = 1, \ldots, N.$$

Then, under $\mathbf{H}_0^{(1)}$, the criterion $\mathcal{S}_N$ in (1.5.2) has asymptotically χ^2 distribution with p degrees of freedom.

Remark 1.5.1 *We reject hypothesis $\mathbf{H}_0^{(1)}$ on the significance level α if*

$$\mathcal{S}_N > \chi_p^2(1-\alpha),$$

where $\chi_p^2(1-\alpha)$ is the $(1-\alpha)$ quantile of the χ^2 distribution with p degrees of freedom.

Sketch of the proof. It suffices to show that under $\mathbf{H}_0^{(1)}$ the asymptotic distribution of $\mathbf{S}_N$ is p-dimensional normal with expectation equal to $\mathbf{0}$ and dispersion matrix $A_N^2\mathbf{Q}_N$. Then the quadratic form $\mathcal{S}_N$ will have asymptotically the $\chi^2(p)$. To prove the asymptotic normality of $\mathbf{S}_N$, we must prove that, for any vector $\boldsymbol{\lambda} \in \mathbb{R}_p,\ \boldsymbol{\lambda} \neq \mathbf{0}$, the scalar product $\boldsymbol{\lambda}'\mathbf{S}_N$ has asymptotically normal distribution $\mathcal{N}(0, \boldsymbol{\lambda}' A_N^2 \mathbf{Q}_N \boldsymbol{\lambda})$. But

$$\boldsymbol{\lambda}'\mathbf{S}_N = \sum_{i=1}^{N} [\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}}_N)] a_N(R_{Ni})$$

and its coefficients $\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}}_N)$ satisfy the Noether condition (1.4.6), because

$$\max_{1\le i\le N} \frac{[\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}}_N)]^2}{\sum_{j=1}^N [\boldsymbol{\lambda}'(\mathbf{x}_j - \bar{\mathbf{x}}_N)]^2} = \max_{1\le i\le N} \frac{\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}}_N)(\mathbf{x}_i - \bar{\mathbf{x}}_N)'\boldsymbol{\lambda}}{\boldsymbol{\lambda}'\mathbf{Q}_N\boldsymbol{\lambda}}$$

$$\le \max_{1\le i\le N} \|\mathbf{x}_i - \bar{\mathbf{x}}\|^2 \cdot \kappa_{max}(\mathbf{Q}_N^{-1}) = \max_{1\le i\le N} \kappa_{max}\{(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{Q}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\} \to 0.$$

Moreover, we can show by some arithmetics that the entities

$$\delta_{N,ij}(\boldsymbol{\lambda}) = \frac{\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}})(a_N(j) - \bar{a}_N)}{N^{-1}\sum_{i=1}^N [\boldsymbol{\lambda}'(\mathbf{x}_i - \bar{\mathbf{x}})]^2 \sum_{j=1}^N (a_N(j) - \bar{a}_N)^2}$$

satisfy the Lindeberg condition (1.4.7). Then the asymptotic normality of the scalar product will follow from Theorem 1.4.3 for every $\boldsymbol{\lambda} \neq \mathbf{0}$. □

1.5.2 Rank tests for $\mathbf{H}_0^{(2)}$

Consider again the model $Y_i = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + e_i, \quad i = 1, \dots, N$, and assume that the errors e_i have a symmetric distribution function, $F(x) + F(-x) = 1\ \forall x$. Let $R_{N1}^+, \dots, R_{NN}^+$ be the ranks of $|Y_1|, \dots, |Y_N|$. Choose a score-generating function $\varphi^* : (0,1) \mapsto [0, \infty)$ and the scores $a_N^*(1), \dots, a_N^*(N)$ generated by φ^*. Put $x_{i0} = 1,\ i = 1, \dots, N$, and for $j = 0, 1, \dots, p$ consider the signed-rank statistics

$$S_{N,j}^+ = \sum_{i=1}^{N} x_{ij} \text{ sign } Y_i\ a_N^*(R_{Ni}^+)$$

and the vector

$$\mathbf{S}_N^+ = (S_{N,0}^+, S_{N,1}^+, \dots, S_{N,p}^+)'.$$

Then, under $\mathbf{H}_0^{(2)}$,

$$I\!E\left(\mathbf{S}_N^+ | \mathbf{H}_0^{(2)}\right) = \mathbf{0} \quad \text{and} \quad I\!E\left(\mathbf{S}_N^+ \mathbf{S}_N^{+\prime} | \mathbf{H}_0^{(2)}\right) = A_N^{*2}\mathbf{Q}_N^*,$$

where $A_N^{*2} = \frac{1}{N}\sum_{i=1}^N [a_N^*(i)]^2$ and

$$\mathbf{Q}_N^* = \sum_{i=1}^{N} \mathbf{x}_i^* \mathbf{x}_i^{*\prime} = \left[\sum_{i=1}^{N} x_{ij} x_{ij'}\right]_{j,j'=0,1,\dots,p}$$

and $\mathbf{x}_i^* = (x_{i0,}, x_{i1}, \ldots, x_{ip})'$.

The test criterion will be the quadratic form

$$\mathcal{S}_N^+ = A_N^{*-2}\left(\mathbf{S}_N^{+\prime}(\mathbf{Q}_N^*)^{-1}\mathbf{S}_N^+\right).$$

The distribution of $\mathbf{S}_N^+$ (and hence of $\mathcal{S}_N^+$) is generated by $N!2^N$ equally probable realizations of (sign $Y_1, \ldots,$ sign Y_N) and $(R_{N1}^+, \ldots, R_{NN}^+)$.

The asymptotic distribution of $\mathcal{S}_N^+$ under $\mathbf{H}_0^{(2)}$ will be $\chi^2(p+1)$, provided

$$\max_{1\leq i\leq N} \mathbf{x}_i^{*\prime}(\mathbf{Q}_N^*)^{-1}\mathbf{x}_i^* \to 0 \quad \text{as} \quad N \to \infty,$$

$(a_N^*(1), \ldots, a_N^*(N))$ satisfy the Noether condition (1.4.6), and under the Lindeberg condition (1.4.7) on some mixed terms corresponding to $\mathbf{x}_i^*$ and $a_N^*(i)$, analogously as under the regression line.

1.6 Rank estimation in simple linear regression models

1.6.1 Estimation of the slope β of the regression line

Let $Y_1, \ldots, Y_N$ be independent random variables, Y_i have a distribution function

$$F_i(y) = F(y - \beta_0 - \beta(x_i - \bar{x}_N)), \quad i = 1, \ldots, N$$

where F is continuous. We want to estimate the parameter β with the aid of ranks.

Denote

$$Y_i(b) = Y_i - (x_i - \bar{x}_N)b, \quad 1 \leq i \leq N, \quad b \in \mathbb{R}_1.$$

Let $T_N(Y_1, \ldots, Y_N)$ be a test statistics for testing $\mathbf{H}_0 : \ \beta = 0$ and assume that under $\mathbf{H}_0$ the distribution of T_N is symmetric about μ_N or that $\mathbb{E}_{\mathbf{H}_0} T_N = \mu_N$.

If $T_N(Y_1(b), \ldots, Y_N(b))$ is nonincreasing in $b \in \mathbb{R}_1$, then we can define the estimate of β as

$$\begin{aligned} &\widehat{\beta}_N = \tfrac{1}{2}(\widehat{\beta}_N^- + \widehat{\beta}_N^+), \\ &\widehat{\beta}_N^- = \sup\{b : \ T_N(b) > \mu_N\}, \quad \widehat{\beta}_N^+ = \inf\{b : \ T_N(b) < \mu_N\}. \end{aligned} \tag{1.6.1}$$

If $T_N = \sum_{i=1}^N (x_i - \bar{x}_N)(Y_i - \bar{Y}_N)$, then $\mu_N = 0$ and $T_N(b)$ is linear in b; the estimator is the least-squares estimator 0f β.

Lemma 1.6.1 *Let $T_N = S_N = \sum_{i=1}^N (x_i - \bar{x}_N)a_N(R_{Ni})$ where $a_N(1) \leq \ldots \leq a_N(N)$ (not all equal) and R_{Ni} is the rank of Y_i, $i = 1, \ldots, N$. Then $S_N(b)$ is nonincreasing in b.*

Proof. See Puri and Sen (1985).

The following Lemma shows that S_N is symmetrically distributed under some conditions.

Lemma 1.6.2 *Let either*

$$x_i - \bar{x}_N = \bar{x}_N - x_{N-i+1}, \quad i = 1, \ldots, N \tag{1.6.2}$$

or

$$a_i - \bar{a}_N = \bar{a}_N - a_{N-i+1}, \quad i = 1, \ldots, N. \tag{1.6.3}$$

Then, if $\beta = 0$, the distribution of S_N is symmetric about 0.

Proof. Let (1.6.2) hold. Because $(R_{N1}, \ldots, R_{NN})$ have the same distribution as $(R_{NN}, \ldots, R_{N1})$, then S_N has the same distribution as $\bar{S}_N = \sum_{i=1}^{N}(x_i - \bar{x}_N) a_N(R_{N,N-i+1}) = -S_N$.

Similarly we proceed under (1.6.2). □

Properties of $\widehat{\beta}_N$:

1. $\widehat{\beta}_N(Y_1 + x_1 b, \ldots, Y_N + x_N b) = \widehat{\beta}_N(Y_1, \ldots, Y_N) + b \quad \forall b \in \mathbb{R}_1$.

2. $\widehat{\beta}_N(cY_1, \ldots, cY_N) = c\widehat{\beta}_N(Y_1, \ldots, Y_N) \quad \forall c > 0$.

3. $I\!P(\widehat{\beta}_N < a) \leq I\!P(S_N(a) < \mu_n) \leq I\!P(S_N(a) \leq \mu_N) \leq I\!P(\widehat{\beta}_N \leq a)$

Asymptotic normality of $\widehat{\beta}_N$:

Theorem 1.6.1 *Assume that $\{x_{N1}, \ldots, x_{NN}\}$ satisfy the conditions*

$$0 < \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_{Ni} - \bar{x}_N)^2 = C_0^2 < \infty, \tag{1.6.4}$$

$$\max_{1 \leq i \leq N} \frac{1}{N}(x_{Ni} - \bar{x}_N)^2 \to 0 \quad \text{as} \quad N \to \infty.$$

Let $a_N(i) = I\!E\varphi(U_{N:i})$ or $= \varphi\left(\frac{i}{N+1}\right)$, $\quad i = 1, \ldots, N$, where φ is nondecreasing on $(0,1)$ and

$$A_\varphi^2 = \int_0^1 \varphi^2(u)du < \infty, \ \int_0^1 \varphi(u)du = 0.$$

Let F have finite Fisher's information, i.e.

$$A_\psi^2 = \int_0^1 \psi^2(u)du, \quad \text{where} \quad \psi(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}, \ 0 < u < 1.$$

Then $\left\{N^{1/2}(\widehat{\beta}_N - \beta)\right\}_{N=1}^{\infty}$ is asymptotically normally distributed

$$\mathcal{N}\left(0, \ \frac{A_\varphi^2}{C_0^2 \gamma^2(\varphi, F)}\right), \quad \gamma(\varphi, F) = \int_0^1 \varphi(u)\psi(u)du.$$

1.6.2 Estimation in multiple regression model

Let $Y_1, \dots, Y_N$ be independent observations, Y_i have distribution function

$$F_i(y) = F(y - \beta_0 - (\mathbf{x}_i - \bar{\mathbf{x}}_N)'\boldsymbol{\beta}), \quad \mathbf{x}_i \in \mathbb{R}_p, \quad 1 \le i \le N.$$

Consider the (vector) linear rank statistic

$$\mathbf{S}_N(\mathbf{b}) = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}}_N) a_N(R_{Ni}(\mathbf{b})) = (S_{N1}(\mathbf{b}), \dots, S_{NN}(\mathbf{b}))',$$

where $R_{Ni}(\mathbf{b})$ is the rank of $Y_i - \mathbf{x}'\mathbf{b},\ i = 1, \dots, N$, and the scores are nondecreasing. Obviously $\mathbb{E}\mathbf{S}_N(\mathbf{0}) = \mathbf{0}$. Define

$$\mathcal{D}_N = \left\{ \mathbf{b} : \ \|\mathbf{S}_N(\mathbf{b})\| = \min,\ \mathbf{b} \in \mathbb{R}_p \right\}$$

where $\|\cdot\|$ is either L_1 or the L_2-norm. If $\mathcal{D}_N$ is a convex set, then we can define the center of gravity of $\mathcal{D}_N$ as an estimator $\widehat{\boldsymbol{\beta}}_N$ of $\boldsymbol{\beta}$.

Assume that $\mathbf{x}_{Ni}$ satisfy the (Noether) condition

$$\max_{1 \le i \le N} (\mathbf{x}_{Ni} - \bar{x}_N)'\mathbf{Q}_N^{-1}(\mathbf{x}_{Ni} - \bar{x}_N) \to 0 \quad \text{as} \quad N \to \infty,$$

where $\mathbf{Q}_N = \sum_{i=1}^{N} (\mathbf{x}_{Ni} - \bar{x}_N)(\mathbf{x}_{Ni} - \bar{x}_N)'$. If F has the finite Fisher's information, then $\left\{ N^{1/2}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \right\}$ is asymptotically normally distributed

$$\mathcal{N}_p\left(\mathbf{0},\ \frac{A_\varphi^2}{\gamma^2(\varphi, F)} \left(\tfrac{1}{N}\mathbf{Q}_N \right)^{-1} \right).$$

1.7 Aligned rank tests about the intercept

1.7.1 Regression line

Let $Y_1, \dots, Y_N$ are independent, Y_i has distribution function

$$F_i(y) = \mathbb{P}(Y_i \le y) = F(y - \beta_0 - (x_i - \bar{x}_N)\beta),\ 1 \le i \le N,\ y \in \mathbb{R}.$$

Consider the hypothesis

$$\mathbf{H}_0 : \ \beta_0 = 0 \quad \text{versus} \quad \mathbf{K}^+ : \ \beta_0 > 0 \quad \textit{or} \quad \mathbf{K} : \ \beta_0 \neq 0$$

where β is treated as a nuisance parameter. If $\beta \neq 0$, then $Y_1, \dots, Y_N$ are not identically distributed, and we cannot use their ranks. If we have an estimate $\widehat{\beta}_N$ of β, we can consider the ranks of the residuals $|Y_i - (x_i - \bar{x}_N)\widehat{\beta}_N|,\ i = 1, \dots, N$ *(aligned ranks)* and an (aligned) signed rank statistics based on them. Under some conditions, such statistic is asymptotically *distribution-free*, i.e. under the hypothesis $\mathbf{H}_0 : \ \beta_0 = 0$, its asymptotic distribution does not depend on F.

Let $\widehat{\beta}_N$ be the rank estimate (1.6.1) based on the linear rank statistic

$$\sum_{i=1}^{N}(x_i - \bar{x}_N)a_N(R_{Ni}(b)),\ b \in \mathbb{R}_1.$$

$\widehat{Y}_i = Y_i - (x_i - \bar{x}_N)\widehat{\beta}_N,\ i = 1, \ldots, N$ and the aligned signed rank statistic

$$\widehat{S}_N = \sum_{i=1}^{N} \text{sign}\ \widehat{Y}_i\ a_N^*(R_{Ni}^+),$$

where R_{Ni}^+ is the rank of $|Y_i - (x_i - \bar{x}_N)\widehat{\beta}_N|,\ i = 1, \ldots, N$. The test criterion for $\mathbf{H}_0$ will be

$$T_N = \frac{N^{-1/2}\widehat{S}_N}{A_N^*}, \quad (A_N^*)^2 = \frac{1}{N}\sum_{i=1}^{N}(a_N^*(i))^2.$$

We reject $\mathbf{H}_0$ in favor of $\mathbf{K}^+$ if $T_N > k_\alpha^+$, and reject $\mathbf{H}_0$ in favor of $\mathbf{K}$ if $|T_N| > k_\alpha$. The critical values k_α^+ and k_α are determined from the asymptotic normal distribution of T_N.

Theorem 1.7.1 *Assume that*

(i) F *is symmetric about 0 and has an absolutely continuous density* f *and finite and positive Fisher information,* $0 < I(f) = \int \left(\frac{f'(z)}{f(z)}\right)^2 dF(z) < \infty$.

(ii) $\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_N)^2 \to C^2,\ 0 < C < \infty$, *and* $\frac{1}{N}\left[\max_{1\le i\le N}(x_i - \bar{x}_N)^2\right] \to 0$ *as* $N \to \infty$.

(iii) $\varphi(t)$ *is nondecreasing,* $\varphi(1-t) = -\varphi(t), \quad t \in (0,1),$ *and*
$0 < A^2(\varphi) = \int_0^1 \varphi^2(t)dt < \infty$. *Put* $\varphi^*(u) = \varphi\left(\frac{u+1}{2}\right),\ 0 < u < 1$ *and*
$a_N^*(i) = I\!\!E\varphi^*(U_{N:i})$ *or* $a_N^*(i) = \varphi^*\left(\frac{i}{N+1}\right),\ i = 1, \ldots, N$.

Then, under $\mathbf{H}_0 :\ \beta_0 = 0$, *the criterion* T_N *has asymptotically normal distribution with mean 0 and variance 1.*

Sketch of he proof. Because $\lim_{N\to\infty} A_N^* = A^2(\varphi)$ and $N^{1/2}(\widehat{\beta}_N - \beta) = O_p(1)$, it can be proved (not elementary) that under $\mathbf{H}_0$

$$N^{-1/2}[\widehat{S}_N - S_N(\beta)] \xrightarrow{p} 0 \ \text{ as } \ N \to \infty, \tag{1.7.5}$$

where

$$S_N(\beta) = \sum_{i=1}^{N} \text{sign}(Y_i(\beta))\ a_N^*(R_{Ni}^+(\beta)),$$

where $Y_i(\beta) = Y_i - (x_i - \bar{x}_N)\beta$ and $R_{Ni}^+(\beta)$ is the rank of $Y_i(\beta) = Y_i - (x_i - \bar{x}_N)\beta,\ 1 \le i \le N$. Under $\mathbf{H}_0$ are $Y_i(\beta) = Y_i - (x_i - \bar{x}_N)\beta$ independent and identically distributed with d.f. F symmetric about 0. It was shown earlier that

$$N^{-1/2}S_N(\beta) \xrightarrow{d} \mathcal{N}(0, A^2(\varphi)),$$

hence, regarding (1.7.5), also $N^{-1/2}\widehat{S}_N \xrightarrow{d} \mathcal{N}(0, A^2(\varphi))$. □

Remark 1.7.1 *We reject* $\mathbf{H}_0$ *in favor of* $\mathbf{K}^+$ *on the asymptotic significance level* α, *provided* $T_N \ge \Phi^{-1}(1-\alpha)$, *and we reject* $\mathbf{H}_0$ *in favor of* $\mathbf{K}$ *provided* $|T_N| \ge \Phi\left(1 - \frac{\alpha}{2}\right)$.

Powers of the tests against local alternatives:

The tests are consistent in the sense that their powers tend to 1 as $\beta_0 \to \infty$ (or $|\beta_0| \to \infty$). However, important is the power for alternatives close the the hypothesis, namely

$$\mathbf{K}_{1N}: \ \beta_0 = N^{-1/2}\lambda, \quad \lambda \neq 0 \ \text{ fixed }.$$

Such alternative is *contiguous* in the sense of LeCam/Hájek, and it can be shown that the approximation (1.7.5) holds not only under the hypothesis, but also under $\mathbf{K}_{1N}$. Hence, $N^{-1/2}\widehat{S}_N$ has the same asymptotic distribution as $S_N(\beta)$ also under $\mathbf{K}_{1N}$.

Denote $\tau_\alpha = \Phi^{-1}(1-\alpha), \ 0 < \alpha < 1$. The asymptotic power of the aligned rank test is

$$I\!P\{T_N \geq \tau_\alpha | \mathbf{K}_{1N}\} \to 1 - \Phi\left(\tau_\alpha - \frac{\lambda}{A_\varphi}\int_0^1 \varphi(u)\varphi_f(u)du\right) \ \text{one-sided test}$$

Comparison: Classical test of $\mathbf{H}_0$

The least-squares estimator of β_0 is

$$\tilde{\beta}_{0N} = \bar{Y}_N = \frac{1}{N}\sum_{i=1}^N Y_i$$

and the likelihood ratio statistic is

$$\begin{aligned} L_N &= \sqrt{N}\frac{\bar{Y}_N}{s_N}, \quad \text{where} \\ s_N^2 &= \frac{1}{N-2}\sum_{i=1}^N [Y_i - \bar{Y}_N - (x_i - \bar{x}_N)\tilde{\beta}_N]^2, \\ \tilde{\beta}_N &= \frac{\sum_{i=1}^N (x_i - \bar{x}_N)(Y_i - \bar{Y}_N)}{\sum_{i=1}^N (x_i - \bar{x}_N)^2}. \end{aligned}$$

If $\sigma^2 = \int z^2 dF(z) < \infty$, then

$$s_N^2 \xrightarrow{p} \sigma^2, \quad \bar{Y}_N \xrightarrow{p} \beta_0, \quad \tilde{\beta}_N \xrightarrow{p} \beta \ \text{ as } \ N \to \infty.$$

Under $\mathbf{H}_0: \ \beta_0 = 0$, the likelihood ratio is asymptotically $\mathcal{N}(0,1)$. The asymptotic relative efficiency of the aligned signed rank test with respect to the likelihood ratio test is

$$\sigma^2 \frac{\left(\int_0^1 \varphi(u)\varphi_f(u)du\right)^2}{\int_0^1 \varphi^2(u)du} \leq \sigma^2 \mathcal{I}(f).$$

1.7.2 Multiple regression model

Let $Y_1, \ldots, Y_N$ be independent with distribution functions $F_1, \ldots, F_N$ such that

$$F_i(y) = I\!P(Y_i \leq y) = F(y - \beta_0 - (\mathbf{x}_i - \bar{\mathbf{x}}_N)'\boldsymbol{\beta}), \ 1 \leq i \leq N, \ y \in \mathbb{R}_1, \ \boldsymbol{\beta} \in \mathbb{R}_p.$$

We want to test the hypothesis

$$\mathbf{H}_1: \ \beta_0 = 0 \ \text{ versus } \ \mathbf{K}_1^+: \ \beta_0 > 0 \ \text{ or } \ \mathbf{K}_1: \ \beta_0 \neq 0,$$

where $\boldsymbol{\beta}$ is unspecified. We may also partition $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

where $\boldsymbol{\beta}_1 \in \mathbb{R}_{p_1}, \ \boldsymbol{\beta}_2 \in \mathbb{R}_{p_2}, \ p_1 + p_2 = p$. We want to test the hypothesis

$$\mathbf{H}_2 : \boldsymbol{\beta}_2 = \mathbf{0} \ \text{ versus } \ \boldsymbol{\beta}_2 \neq \mathbf{0}$$

where $\beta_0, \ \boldsymbol{\beta}_1$ are unspecified.

Test of H$_1$

Let $\widehat{\boldsymbol{\beta}}_N$ be the estimator of $\boldsymbol{\beta}$. Consider the residuals $\widehat{Y}_i = Y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}, \ i = 1, \dots, N$ and the (aligned) ranks $\widehat{R}_{N1}^+, \dots, \widehat{R}_{NN}^+$ of $|\widehat{Y}_i|, \ i = 1, \dots, N$. Similarly as in the case of the regression line, the test is based on the aligned sign rank statistic

$$\widehat{S}_N = \sum_{i=1}^{N} \operatorname{sign}(\widehat{Y}_i) \ a_N^*(R_{Ni}^+)$$

and the test criterion is

$$T_N^2 = \frac{\widehat{S}_N^2}{N A_N^{*2}}, \quad (A_N^*)^2 = \frac{1}{N} \sum_{i=1}^{N} (a_N^*(i))^2$$

T_N^2 has asymptotically χ^2 distribution with 1 d.f.