Appendix A
Proofs

A.1 Proof of Theorem 1.1

We first provide a lemma (see, e.g., Lemma 17.1 in van der Vaart (1998)).

Lemma A.1. *Assume* $\boldsymbol{Z} \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma})$*, where the* $p \times p$ *matrix* $\boldsymbol{\Sigma}$ *has eigenvalues* $\lambda_1, \ldots, \lambda_p$*. Let* $X_1, \ldots, X_p$ *denote i.i.d. standard normal variates. The quadratic form* $\boldsymbol{Z}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}$ *is then equivalent in distribution with the random variable*

$$\sum_{i=1}^{p} \lambda_i X_i^2.$$

Throughout the proof, we assume that H_0 holds true.

First, write

$$X_n^2 = n \sum_{j=1}^{k} \frac{(\hat{p}_j - \pi_{0j})^2}{\pi_{0j}},$$

where $\hat{p}_j = N_j/n$ is an unbiased and consistent estimator of π_{0j}, and let $\hat{\boldsymbol{p}}_n^t = (\hat{p}_1, \ldots, \hat{p}_k)$. Let $\boldsymbol{D}_{\pi_0} = \text{diag}(\boldsymbol{\pi}_0)$. With this new notation, we may write $X_n^2 = n(\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0)^t D_{\pi_0}^{-1} (\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0)$, which is a quadratic form in $\boldsymbol{Z}_n = \sqrt{n}(\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0)$. By the multivariate central limit theorem (see, e.g., Theorem 5.4.4 in Lehmann (1999)), as $n \to \infty$,

$$\boldsymbol{Z}_n \xrightarrow{d} \text{MVN}(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0 \boldsymbol{\pi}_0^t$. Because X_n^2 is a quadratic form in $\boldsymbol{Z}_n$, Lemma A.1 gives, as $n \to \infty$,

$$X_n^2 \xrightarrow{d} \sum_{j=1}^{k} \lambda_j Z_j^2,$$

where $Z_1, \ldots, Z_k$ are i.i.d. $N(0,1)$, and $\lambda_1 \leq \cdots \leq \lambda_k$ are the eigenvalues of

$$\boldsymbol{L} = \boldsymbol{D}_{\pi_0}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{D}_{\pi_0}^{-1/2} = \boldsymbol{I}_k - \sqrt{\boldsymbol{\pi}_0} \sqrt{\boldsymbol{\pi}_0^t}$$

O. Thas, *Comparing Distributions*,
DOI 10.1007/978-0-387-92710-7, © Springer Science+Business Media, LLC 2010

with $\sqrt{\boldsymbol{\pi}_0^t} = (\sqrt{\pi_{01}}, \ldots, \sqrt{\pi_{0k}})$. It can be shown that $\lambda_1 = 0$ and $\lambda_2 = \cdots = \lambda_k = 1$. This completes the proof. □

A.2 Proof of Theorem 1.2

Because $\hat{\boldsymbol{\beta}}$ is BAN, we obtain

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0(\boldsymbol{\beta}))\boldsymbol{D}_{\pi_0}^{-1/2}\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1} + o_p(n^{-1/2}),$$

where the matrix $\boldsymbol{A}$ has (i,j)th element $(i = 1, \ldots, k; j = 1, \ldots, p)$,

$$\frac{1}{\sqrt{\pi_{0i}}}\frac{\partial \pi_{0i}}{\partial \beta_j}(\boldsymbol{\beta}).$$

By Birch's regularity conditions, we find

$$\boldsymbol{\pi}_0(\hat{\boldsymbol{\beta}}) - \boldsymbol{\pi}_0(\boldsymbol{\beta}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\frac{\partial \boldsymbol{\pi}_0}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}) + o_p(n^{-1/2}).$$

Hence,

$$\boldsymbol{\pi}_0(\hat{\boldsymbol{\beta}}) - \boldsymbol{\pi}_0(\boldsymbol{\beta}) = (\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0(\boldsymbol{\beta}))\boldsymbol{L} + o_p(n^{-1/2}),$$

where $\boldsymbol{L} = \boldsymbol{D}_{\pi_0}^{-1/2}\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{D}_{\pi_0}^{1/2}$. Write

$$\boldsymbol{M}_n = \begin{bmatrix} \hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0 \\ \boldsymbol{\pi}_0(\hat{\boldsymbol{\beta}}) - \boldsymbol{\pi}_0 \end{bmatrix} = (\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0(\boldsymbol{\beta}))\begin{bmatrix} \boldsymbol{I}_k \\ \boldsymbol{L} \end{bmatrix} + o_p(n^{-1/2}).$$

Because $\sqrt{n}(\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0)$ is asymptotically multivariate normal, we may conclude that $\boldsymbol{M}_n$ is also asymptotically multivariate normal with zero mean and variance–covariance matrix equal to

$$\begin{bmatrix} \boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^t & (\boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^t)\boldsymbol{L} \\ \boldsymbol{L}^t(\boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^t) & \boldsymbol{L}^t(\boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^t)\boldsymbol{L} \end{bmatrix}.$$

Hence, $\sqrt{n}(\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0(\hat{\boldsymbol{\beta}})) = \sqrt{n}(\hat{\boldsymbol{p}}_n - \boldsymbol{\pi}_0) - \sqrt{n}(\boldsymbol{\pi}_0(\hat{\boldsymbol{\beta}}) - \boldsymbol{\pi}_0)$ is also asymptotically zero mean multivariate normal with variance–covariance matrix (after some simple algebra)

$$\boldsymbol{\Sigma} = \boldsymbol{D}_{\pi_0} - \boldsymbol{\pi}_0\boldsymbol{\pi}_0^t - \boldsymbol{D}_{\pi_0}^{1/2}\boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t\boldsymbol{D}_{\pi_0}^{1/2}. \quad \text{(A.1)}$$

The asymptotic null distribution of $\hat{X}_n^2$ is now again obtained by applying Lemma A.1. This time we need the eigenvalues of

$$\boldsymbol{D}_{\pi_0}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}_{\pi_0}^{-1/2} = \boldsymbol{I} - \sqrt{\boldsymbol{\pi}_0}\sqrt{\boldsymbol{\pi}_0^t} - \boldsymbol{A}(\boldsymbol{A}^t\boldsymbol{A})^{-1}\boldsymbol{A}^t.$$

It can be shown that this matrix has $k - p - 1$ eigenvalues equal to 1, and the remaining $p + 1$ eigenvalues equal to 0. □

A.3 Proof of Theorem 4.1

(1) To obtain the score statistic, we first need to specify the log-likelihood function. From Equation (4.1), we find

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log\left(\prod_{i=1}^{n} g_k(X_i;\boldsymbol{\theta})\right) \\ &= n\log C(\boldsymbol{\theta}) + \sum_{i=1}^{n}\log g(X_i) + \sum_{j=1}^{k}\theta_j\sum_{i=1}^{n}h_j(X_i). \end{aligned}$$

The score function for parameter θ_j is given by

$$\begin{aligned} u_j(\boldsymbol{\theta}) &= \frac{\partial l(\boldsymbol{\theta})}{\partial\theta_j} \\ &= n\frac{\partial\log C(\boldsymbol{\theta})}{\partial\theta_j} + \sum_{i=1}^{n}h_j(X_i). \end{aligned}$$

For the construction of the score test statistics, we need to evaluate the score function under the null hypothesis. This gives $u_j(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\mathbf{0}} = \sum_{i=1}^{n} h_j(X_i)$, where we have used

$$\frac{\partial\log C(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\mathbf{0}} = 0.$$

As for all score functions, $\mathrm{E}_0\{u_j\} = 0$. This could also be directly seen from the orthogonality property of the h_j; i.e., $\mathrm{E}_0\{u_j\} = n\int_{\mathcal{S}} h_j(x)g(x)\, dx = n < h_1, 1 >_g = 0$.

The variance–covariance matrix of the vector $\boldsymbol{U}^t = (1/\sqrt{n})\,(u_1,\ldots,u_k)_{\boldsymbol{\theta}=\mathbf{0}}$ involves the covariances

$$\begin{aligned} \mathrm{Cov}_0\{h_i(X),h_j(X)\} &= \int_{\mathcal{S}} h_i(x)h_j(x)g(x)dx \\ &= < h_i, h_j >_g \\ &= \delta_{ij}, \end{aligned}$$

where δ_{ij} is Kronecker delta. Hence, $\mathrm{Var}_0\{\boldsymbol{U}\} = \boldsymbol{I}$, the $k \times k$ identity matrix. The multivariate central limit theorem gives that

$$\boldsymbol{U} \xrightarrow{d} MVN(\mathbf{0},\boldsymbol{I}),$$

and we therefore have also the convergence of the quadratic form (Lemma 17.1 in van der Vaart (1998))

$$T_k = \boldsymbol{U}^t \boldsymbol{U} \xrightarrow{d} \chi^2_k.$$

(2) To prove the second part of the theorem, we only have to show that the score function based on the Barton model is the same as u_j when restricted under the null hypothesis $\boldsymbol{\theta} = \mathbf{0}$.

The log-likelihood function becomes

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log g_k(X_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log g(X_i) + \sum_{i=1}^{n} \log \left(1 + \sum_{j=1}^{k} \theta_j h_j(X_i) \right),$$

and the score function for θ_j

$$u_j(\boldsymbol{\theta}) = \frac{\partial l}{\partial \theta_j} = \sum_{i=1}^{n} \frac{h_j(X_i)}{1 + \sum_{j=1}^{k} \theta_j h_j(X_i)}.$$

Hence,

$$u_j|_{\boldsymbol{\theta}=\mathbf{0}} = \sum_{i=1}^{n} h_j(X_i),$$

which is exactly the same as what we found in part (1) of the proof. □

A.4 Proof of Lemma 4.1

Straightforward calculations give

$$\begin{aligned}
\mathrm{E}_k \left\{ (X-\mu)^m \right\} &= \int_{\mathcal{S}} (x-\mu)^m \left(1 + \sum_{j=1}^{k} \theta_j h_j(x) \right) g(x) dx \\
&= \mu_m + \sum_{j=1}^{k} \theta_j < (x-\mu)^m, h_j(x) >_g \\
&= \mu_m + \sum_{j=1}^{k} \theta_j < (x-\mu)^m - \mu_m, h_j(x) >_g, \qquad \text{(A.2)}
\end{aligned}$$

where the last step makes use of $< \mu_m 1, h_j(x) >_g = \mu_m < 1, h_j(x) >_g = 0$. Because $\mathrm{E}_0 \left\{ (X-\mu)^m - \mu_m \right\} = < (x-\mu)^m - \mu_m, 1 >_g = 0$, we may write the degree m polynomial $(x-\mu)^m - \mu_m$ in terms of the m base functions $h_1, \ldots, h_m$,

$$(x-\mu)^m - \mu_m = \sum_{j=1}^{m} c_j h_j(x), \qquad \text{(A.3)}$$

where $c_1, \ldots, c_m$ are constants. After substituting (A.3) into (A.2), we get

$$\begin{aligned} \mathrm{E}_k\left\{(X-\mu)^m\right\} &= \mu_m + \sum_{j=1}^{k} \theta_j < \sum_{i=1}^{m} c_i h_i, h_j >_g \\ &= \mu_m + \theta_j c_j . \end{aligned}$$

Hence, if $\theta_j = 0$ then $\mathrm{E}_k\left\{(X-\mu)^m\right\} = \mu_m$. The $\Leftarrow$ part of the proof is also true because $(x-\mu)^m - \mu_m$ is a polynomial of exactly degree m, and thus $c_m \neq 0$, and, therefore, $\mathrm{E}_k\left\{(X-\mu)^m\right\} = \mu_m$ if and only if $\theta_j = 0$. □

A.5 Proof of Lemma 4.2

Because $h_0(x) = 1$, we get $\mathrm{E}_0\left\{h_j(X)\right\} =< h_j, 1 >_g = 0$ for all j. To stress that the lemma imposes a restriction on the polynomials, we use the notation $h_j^\star$ whenever they are of the form of Equation (4.13). Under the null hypothesis, $\mu_j = \mathrm{E}_0\left\{(X-\mu)^j\right\}$. Hence, also $\mathrm{E}_0\left\{h_j^\star(X)\right\} =< h_j^\star, 1 >_g = 0$.

It is always possible to write

$$h_j(x) = h_j^\star(x) + z(x),$$

where z is a polynomial of degree $\leq j$. The lemma is proven if we can show that $z(x) \equiv 0$.

We know that $< h_i, h_j >_g = 0$ for all $i \neq j$, thus we get

$$\begin{aligned} 0 &= < h_i, h_j >_g \\ &= < h_i, h_j^\star + z >_g \\ &= < h_i, h_j^\star >_g + < h_i, z >_g . \end{aligned}$$

Hence, $< h_i, z >_g = - < h_i, h_j^\star >_g$. This holds for all $i \neq j$, and since the h_i form a base in a Hibert space, therefore we may conclude that $z = -h_j^\star$ or $z = 0$. However, the former implies $h_j(x) = 0$ which is a contradiction. Therefore, $z = 0$. □

A.6 Proof of Lemma 4.3

(1) Because all first j moments agree with g, Lemma 4.2 implies that $\mathrm{E}\left\{h_j(X)\right\} = 0$. Hence,

$$\begin{aligned} \mathrm{Var}\left\{U_j\right\} &= \mathrm{Var}\left\{h_j(X)\right\} \\ &= \mathrm{E}\left\{\left(h_j(X) - \mathrm{E}\left\{h_j(X)\right\}\right)^2\right\} \qquad (A.4) \\ &= \mathrm{E}\left\{h_j^2(X)\right\}, \end{aligned}$$

where h_j^2 is a polynomial of degree $2j$ which may be written as $h_j^2(x) = \sum_{l=0}^{2j} c_l h_l(x)$. Note that this is a sum of polynomials of degrees corresponding to moments which all agree with g, and, again according to Lemma 4.2, these polynomials have expectation equal to zero. Hence, $\mathrm{E}\left\{h_j^2(X)\right\} = c_0$. Because the same result would have been found under the null hypothesis, we find $c_0 = \mathrm{Var}_0\left\{U_j\right\} = 1$.

(2) We start from (A.4), in which now $\mathrm{E}\{h_j(X)\}$ is not necessarily zero. Write

$$\begin{aligned}
\mathrm{Var}\{U_j\} &= \mathrm{E}\left\{h_j^2(X)\right\} - \left(\mathrm{E}\{h_j(X)\}\right)^2 \\
&= \mathrm{E}\left\{1 + \sum_{l=1}^{2j} c_l h_l(X)\right\} - \left(\mathrm{E}\{h_j(X)\}\right)^2 \\
&= 1 + \sum_{l=m}^{2j} c_l\, \mathrm{E}\{h_l(X)\} - \left(\mathrm{E}\{h_j(X)\}\right)^2 . \qquad \text{(A.5)}
\end{aligned}$$

Lemma 4.2 tells again when the last or the two last terms in (A.5) are zero. This gives the statement in (4.14). □

A.7 Proof of Theorem 4.10

First we introduce some matrix notation. Let $\boldsymbol{H}^t$ the $m \times k$ matrix with the (i,j)th element equal to h_{ij}; i.e., the jth column corresponds to the jth orthonormal vector. We may now write $\boldsymbol{U} = (1/\sqrt{n})\boldsymbol{H}\boldsymbol{N}$, and the orthonormality condition becomes $\boldsymbol{H}\boldsymbol{D}_{\pi_0}\boldsymbol{H}^t = \boldsymbol{I}$, where $\boldsymbol{D}_{\pi_0} = \mathrm{diag}(\boldsymbol{\pi}_0)$. The restriction $\sum_{i=1}^m h_{ij} N_i = 0$ for all $j = 1, \ldots, k$ now becomes $\boldsymbol{H}\boldsymbol{\pi}_0 = \boldsymbol{0}$. This latter restriction allows us to write the equality

$$\boldsymbol{U} = \frac{1}{\sqrt{n}}\boldsymbol{H}\boldsymbol{N} = \frac{1}{\sqrt{n}}\boldsymbol{H}\left(\boldsymbol{N} - n\boldsymbol{\pi}_0\right) = \sqrt{n}\boldsymbol{H}\left(\hat{p} - \boldsymbol{\pi}_0\right). \qquad \text{(A.6)}$$

With this notation the order k smooth test statistic becomes

$$T_k = \boldsymbol{U}^t\boldsymbol{U} = n\left(\hat{p} - \boldsymbol{\pi}_0\right)^t \boldsymbol{H}^t\boldsymbol{H}\left(\hat{p} - \boldsymbol{\pi}_0\right).$$

Because $k = m - 1$ and because $\boldsymbol{H}\boldsymbol{D}_{\pi_0}\boldsymbol{H}^t = \boldsymbol{I}$, we find $\boldsymbol{H}^t\boldsymbol{H} = \boldsymbol{D}_{\pi_0}^{-1}$. Substituting this equality in Equation (A.6) completes the proof. □

A.8 Proof of Theorem 4.2

In this section we acually give the proof of a more general theorem which states the asymptotic distribution of $\hat{\boldsymbol{V}}$ under a sequence of local alternatives. First some notation is introduced.

As a sequence of local alternatives to g, we consider model (4.1) with

$$\boldsymbol{\theta} = \boldsymbol{\theta}_n = n^{-1/2}\boldsymbol{\delta}, \tag{A.7}$$

where $\boldsymbol{\delta}$ is a vector of k positive nonzero constants and $\delta^2 = \boldsymbol{\delta}^t\boldsymbol{\delta} < \infty$. The null hypothesis corresponds to $\boldsymbol{\delta} = \mathbf{0}$. The density or model of the local alternatives is now denoted as

$$g_{nk}(x) = g_{nk}(x; \boldsymbol{\theta}_n, \boldsymbol{\beta}) = C(\boldsymbol{\theta}_n, \boldsymbol{\beta}) \exp\left(\sum_{j=1}^{k} \theta_{nj} h_j(x; \boldsymbol{\beta})\right) g(x; \boldsymbol{\beta}). \tag{A.8}$$

The next two lemmas are needed.

Lemma A.2 (Local Asymptotic Normality (LAN)). *Consider the sequence of alternatives given in (A.7) and model (A.8). Then, the log-likelihood ratio admits the following asymptotic expansion*

$$\log\left(\frac{g_{nk}(x; \boldsymbol{\theta}_n; \boldsymbol{\beta})}{g(x; \boldsymbol{\beta})}\right) = \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t\boldsymbol{h}(x; \boldsymbol{\beta}) - \frac{1}{2}\frac{1}{n}\boldsymbol{\delta}^t\boldsymbol{\delta} + o(\delta^2/n), \tag{A.9}$$

and, as $n \to \infty$,

$$\log \prod_{i=1}^{n} \left(\frac{g_{nk}(X; \boldsymbol{\theta}_n; \boldsymbol{\beta})}{g(X; \boldsymbol{\beta})}\right) \xrightarrow{d} N\left(-\frac{1}{2}\boldsymbol{\delta}^t\boldsymbol{\delta}, \boldsymbol{\delta}^t\boldsymbol{\delta}\right) \tag{A.10}$$

Proof. To prove Equation (A.9) we start with substituting g_{nk} and g into the log-likelihood ratio

$$\begin{aligned}\log \frac{g_{nk}(x; \boldsymbol{\theta}_n; \boldsymbol{\beta})}{g(x; \boldsymbol{\beta})} &= \log C(\boldsymbol{\theta}_n) - \log(C(\mathbf{0})) + \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t\boldsymbol{h}(x) \\ &= \log C(\boldsymbol{\theta}_n) + \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t\boldsymbol{h}(x).\end{aligned}$$

This can be further simplified by applying a Taylor series expansion on $\log C(\boldsymbol{\theta}_n)$,

$$\begin{aligned}\log C(\boldsymbol{\theta}_n) &= \log C(\mathbf{0}) + \frac{1}{\sqrt{n}}\boldsymbol{\delta}^t \left.\frac{\partial \log C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\mathbf{0}} + \frac{1}{2}\frac{1}{n}\boldsymbol{\delta}^t \left.\frac{\partial^2 \log C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\right|_{\boldsymbol{\theta}=\mathbf{0}} \boldsymbol{\delta} + o(\delta^2/n) \\ &= \frac{1}{2}\frac{1}{n}\boldsymbol{\delta}^t \,\mathrm{E}_0\left\{-\boldsymbol{h}(X)\boldsymbol{h}^t(X)\right\}\boldsymbol{\delta} + o(\delta^2/n) \\ &= -\frac{1}{2}\frac{1}{n}\boldsymbol{\delta}^t\boldsymbol{I}\boldsymbol{\delta} + o(\delta^2/n) \\ &= -\frac{1}{2}\frac{1}{n}\boldsymbol{\delta}^t\boldsymbol{\delta} + o(\delta^2/n).\end{aligned}$$

The convergence in Equation (A.10) follows from

$$\log \prod_{i=1}^{n} \frac{g_{nk}(X;\boldsymbol{\theta}_n;\boldsymbol{\beta})}{g(X;\boldsymbol{\beta})} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\delta}^t \boldsymbol{h}(X_i;\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\delta}^t\boldsymbol{\delta} + o_P(1), \tag{A.11}$$

where $(1/\sqrt{n})\sum_{i=1}^{n} \boldsymbol{h}(X_i;\boldsymbol{\beta})$ converges according to the multivariate central limit theorem to a multivariate normal distribution with mean

$$\mathrm{E}_0\left\{\boldsymbol{h}(X)\right\} = \boldsymbol{0},$$

and variance–covariance matrix

$$\mathrm{Var}_0\left\{\boldsymbol{h}(X)\right\} = \mathrm{E}_0\left\{\boldsymbol{h}(X)\boldsymbol{h}^t(X)\right\} = \boldsymbol{I}.$$

Using this result and applying Slutsky's lemma completes the proof.

Lemma A.3. *Let* $\boldsymbol{w}(x;\boldsymbol{\beta})$ *be a vector-valued function that satisfies the regularity conditions, and for which* $\mathrm{E}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta})\right\} = \boldsymbol{0}$. *Then*

$$\mathrm{E}_0\left\{\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} = -\,\mathrm{Cov}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta}), \boldsymbol{u}_\beta(X;\boldsymbol{\beta})\right\} = <\boldsymbol{w}, \boldsymbol{u}_\beta> .$$

Proof. It is assumed that

$$\mathrm{E}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta})\right\} = \boldsymbol{0} = \int_{-\infty}^{+\infty} \boldsymbol{w}(x;\boldsymbol{\beta}) g(x;\boldsymbol{\beta}) dx.$$

Differentiating both sides of this equation yields

$$\begin{aligned}
\int_{-\infty}^{+\infty} \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\beta}}(x;\boldsymbol{\beta}) f(x;\boldsymbol{\beta}) dx + \int_{-\infty}^{+\infty} \boldsymbol{w}(x;\boldsymbol{\beta}) \frac{\partial g}{\partial \boldsymbol{\beta}}(x;\boldsymbol{\beta}) dx &= 0 \\
\mathrm{E}_0\left\{\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} + \int_{-\infty}^{+\infty} \boldsymbol{w}(x;\boldsymbol{\beta}) \frac{\partial \log g}{\partial \boldsymbol{\beta}}(x;\boldsymbol{\beta}) g(x;\boldsymbol{\beta}) dx &= 0 \\
\mathrm{E}_0\left\{\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} + \mathrm{E}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta}) \frac{\partial \log g}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} &= 0.
\end{aligned}$$

Because

$$\mathrm{E}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta})\right\} = \mathrm{E}_0\left\{\frac{\partial \log g}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} = \boldsymbol{0},$$

we obtain

$$\mathrm{E}_0\left\{\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\} = -\,\mathrm{Cov}_0\left\{\boldsymbol{w}(X;\boldsymbol{\beta}), \frac{\partial \log g}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\},$$

which completes the proof. □

Theorem A.1. *Under the sequence of local alternatives given in (A.7), the vector $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ converges, as $n \to \infty$, in distribution to a multivariate normal distribution with variance–covariance matrix*

$$\boldsymbol{\Sigma}_{\hat{v}} = \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_{v\beta}\boldsymbol{\Sigma}_{b\beta}^{-1}\boldsymbol{\Sigma}_{bb}\boldsymbol{\Sigma}_{\beta b}^{-1}\boldsymbol{\Sigma}_{\beta v} - \boldsymbol{\Sigma}_{vb}\boldsymbol{\Sigma}_{\beta b}^{-1}\boldsymbol{\Sigma}_{\beta v} - \boldsymbol{\Sigma}_{v\beta}\boldsymbol{\Sigma}_{b\beta}^{-1}\boldsymbol{\Sigma}_{bv}, \quad \text{(A.12)}$$

and mean

$$\boldsymbol{\mu}_{\hat{v}} = \left(\boldsymbol{\Sigma}_{vh} - \boldsymbol{\Sigma}_{v\beta}\boldsymbol{\Sigma}_{b\beta}^{-1}\boldsymbol{\Sigma}_{bh}\right)\boldsymbol{\delta}.$$

Proof. The proof consists of two parts. First the asymptotic null distribution of $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ is found. Then the joint null distribution of $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ and the log-likelihood ratio statistic is proven, from which by means of Le Cam's third lemma the theorem immediately follows.

1. A first-order Taylor expansion of $\boldsymbol{v}(\hat{\boldsymbol{\beta}})$ gives

$$\boldsymbol{v}(x;\hat{\boldsymbol{\beta}}) = \boldsymbol{v}(x;\boldsymbol{\beta}) + \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(x;\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_P(n^{-1/2}).$$

Substituting this into $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ and recognising that $\boldsymbol{\beta}$ is an asymptotic linear estimator it becomes

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{v}(X_i;\boldsymbol{\beta}) + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X_i;\boldsymbol{\beta})\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Psi}(X_i;\boldsymbol{\beta})\right) + o_P(1).$$

This is further simplified by applying the law of large numbers on

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X_i;\boldsymbol{\beta}),$$

resulting in

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{v}(X_i;\boldsymbol{\beta}) + \mathrm{E}_0\left\{\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta})\right\}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Psi}(X_i;\boldsymbol{\beta})\right) + o_P(1). \quad \text{(A.13)}$$

Under the null hypothesis, the multivariate central limit theorem gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{v}(X_i;\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_v),$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Psi}(X_i;\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\Psi}),$$

where

$$\boldsymbol{\Sigma}_{\Psi} = \boldsymbol{\Sigma}_{b\beta}^{-1} \boldsymbol{\Sigma}_b \boldsymbol{\Sigma}_{\beta b}^{-1}.$$

The joint distribution of these two random vectors is obtained by applying the Cramér–Wald device. In particular it is a multivariate normal distribution with mean $\mathbf{0}$ and variance–covariance matrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_v & \boldsymbol{\Sigma}_{v\Psi} \\ \boldsymbol{\Sigma}_{\Psi v} & \boldsymbol{\Sigma}_{\Psi} \end{bmatrix},$$

where $\boldsymbol{\Sigma}_{v\Psi} = \mathrm{Cov}_0 \{\boldsymbol{v}(X;\boldsymbol{\beta}), \boldsymbol{\Psi}(X;\boldsymbol{\beta})\}$. Using Lemma A.3 we find

$$\mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} = -\,\mathrm{Cov}_0 \{\boldsymbol{v}(X), \boldsymbol{u}_{\beta}\} = -\boldsymbol{\Sigma}_{v\beta},$$

and using Slutsky's lemma, we find that the limiting null disitribution of $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ is a multivariate normal distribution with mean $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}_{\hat{v}}$ as stated in Equation (A.12).

2. The proof of the joint null distribution of $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ and the log-likelihood ratio statistic is along the same lines as van der Vaart (1998), p. 219. We only need to calculate the covariance between the two random vectors,

$$\mathrm{Cov}_0 \left\{ \boldsymbol{V}(\hat{\boldsymbol{\beta}}), \log \prod_{i=1}^{n} \frac{g_{nk}(X_i;\boldsymbol{\theta}_n;\boldsymbol{\beta})}{g(X_i;\boldsymbol{\beta})} \right\}.$$

The solution is obtained by substituting $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ and the log-likelihood ratio statistic by their respective asymptotic expansions (Equations (A.13) and (A.11)):

$$\begin{aligned}
&\mathrm{Cov}_0 \left\{ \boldsymbol{V}(\hat{\boldsymbol{\beta}}), \log \prod_{i=1}^{n} \frac{g_{nk}(X_i;\boldsymbol{\theta}_n;\boldsymbol{\beta})}{g(X_i;\boldsymbol{\beta})} \right\} \\
&= \mathrm{Cov}_0 \Bigg\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{v}(X_i;\boldsymbol{\beta}) + \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\Psi}(X_i;\boldsymbol{\beta}) \right) \\
&\quad + o_P(1), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\delta}^t \boldsymbol{h}(X_i;\boldsymbol{\beta}) - \frac{1}{2}\delta^2 + o_P(1) \Bigg\} \\
&= \mathrm{Cov}_0 \left\{ \boldsymbol{v}(X;\boldsymbol{\beta}) + \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} \boldsymbol{\Psi}(X;\boldsymbol{\beta}), \boldsymbol{\delta}^t \boldsymbol{h}(X;\boldsymbol{\beta}) \right\} + o(1) \\
&= \mathrm{Cov}_0 \{\boldsymbol{v}(X;\boldsymbol{\beta}), \boldsymbol{h}(X;\boldsymbol{\beta})\}\, \boldsymbol{\delta} \\
&\quad + \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} \mathrm{Cov}_0 \{\boldsymbol{\Psi}(X;\boldsymbol{\beta}), \boldsymbol{h}(X;\boldsymbol{\beta})\}\, \boldsymbol{\delta} + o(1) \\
&= \boldsymbol{\Sigma}_{vh} \boldsymbol{\delta} + \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} \boldsymbol{\Sigma}_{\Psi h} \boldsymbol{\delta} + o(1) \\
&= \boldsymbol{\Sigma}_{vh} \boldsymbol{\delta} + \mathrm{E}_0 \left\{ \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{\beta}}(X;\boldsymbol{\beta}) \right\} \mathrm{E}_0 \left\{ -\dot{\boldsymbol{b}}(X) \right\}^{-1} \boldsymbol{\Sigma}_{bh} \boldsymbol{\delta} + o(1).
\end{aligned}$$

Applying Lemma A.3 to the last equation gives

$$\mathrm{Cov}_0\left\{\boldsymbol{V}(\hat{\boldsymbol{\beta}}), \log\prod_{i=1}^{n}\frac{g_{nk}(X_i;\boldsymbol{\theta}_n;\boldsymbol{\beta})}{g(X_i;\boldsymbol{\beta})}\right\} = \boldsymbol{\Sigma}_{vh}\boldsymbol{\delta} - \boldsymbol{\Sigma}_{v\beta}\boldsymbol{\Sigma}_{b\beta}^{-1}\boldsymbol{\Sigma}_{bh}\boldsymbol{\delta} + o(1).$$

Now that the joint distribution of $\boldsymbol{V}(\hat{\boldsymbol{\beta}})$ and the log-likelihood ratio statistic are known, we can directly apply Le Cam's third lemma which immediately completes the proof.

□

A.9 Heuristic Proof of Theorem 5.2

(1) Because both $\{h_j \circ G\}$ and $\{k_j^a\}$ are systems of orthonormal functions in $L_2(\mathcal{S}, G)$, there exists a set of constants $\{a_{ij}\}$ so that for all $x \in \mathcal{S}$,

$$\sum_i a_{ij} v_i(x) = k_j^a(x). \tag{A.14}$$

Let $\boldsymbol{A}$ denote the matrix with (i,j)th element equal to a_{ij}, and assume that $\boldsymbol{A}$ has an inverse $\boldsymbol{A}^{-1}$. Equation (A.14) may now be written as $\boldsymbol{A}^t\boldsymbol{v}(x) = \boldsymbol{k}_a(x)$. We now project both sides of the equation onto $\boldsymbol{v}$, resulting in $\boldsymbol{A}^t\boldsymbol{\Sigma}_{\hat{v}} =< \boldsymbol{k}_a, \boldsymbol{v} >_g$, from which we find

$$\boldsymbol{A} = \boldsymbol{\Sigma}_{\hat{v}}^{-1} < \boldsymbol{v}, \boldsymbol{k}_a >_g. \tag{A.15}$$

We now simplify this expression for $\boldsymbol{A}$ by looking for an alternative representation of $\boldsymbol{\Sigma}_{\hat{v}}$.

Denote the (i,j)th element of $\boldsymbol{A}^{-1}$ as a^{ij}. From $\boldsymbol{A}^t\boldsymbol{v}(x) = \boldsymbol{k}_a(x)$ we find $\boldsymbol{v}(x) = \boldsymbol{A}^{-t}\boldsymbol{k}_a(x)$, or $v_i(x) = \sum_j a^{ji}k_j^a(x)$.

The (i,j)th element of $\boldsymbol{\Sigma}_{\hat{v}} =< \boldsymbol{v}, \boldsymbol{v} >_g$ is given by

$$< v_i, v_j >_g = \sum_m\sum_n a^{mi}a^{nj}\int_{\mathcal{S}} k_m^a(x)k_n^a(x)dG(x) = \sum_m a^{mi}a^{mj},$$

which is the (i,j)th element of $\boldsymbol{A}^{-t}\boldsymbol{A}^{-1}$. Hence,

$$\begin{aligned}\boldsymbol{\Sigma}_{\hat{v}} &= \boldsymbol{A}^{-t}\boldsymbol{A}^{-1} \\ &= \left(< \boldsymbol{v}, \boldsymbol{l} >_g^{-1} \boldsymbol{\Sigma}_{\hat{v}}\right)^t \left(< \boldsymbol{v}, \boldsymbol{k}_a >_g^{-1} \boldsymbol{\Sigma}_{\hat{v}}\right) \\ &= \boldsymbol{\Sigma}_{\hat{v}} < \boldsymbol{v}, \boldsymbol{k}_a >_g^{-t} < \boldsymbol{v}, \boldsymbol{k}_a >_g \boldsymbol{\Sigma}_{\hat{v}}.\end{aligned}$$

Solving this equation for $\boldsymbol{\Sigma}_{\hat{v}}$ gives $\boldsymbol{\Sigma}_{\hat{v}} =< \boldsymbol{v}, \boldsymbol{k}_a >_g< \boldsymbol{k}_a, \boldsymbol{v} >_g$. We now substitute this expression into (A.15),

$$\boldsymbol{A} = \boldsymbol{\Sigma}_{\hat{v}}^{-1} < \boldsymbol{v}, \boldsymbol{k}_a >_g =< \boldsymbol{v}, \boldsymbol{k}_a >_g^{-1} = \boldsymbol{\Sigma}_{\hat{v}}^{-1/2}.$$

(2) By the definition of the $\{l_j\}$ and the $\{\gamma_j\}$, we have

$$\int_{\mathcal{S}} c(x,y) l_j(x) dG(x) = \gamma_j l_j(y).$$

We now project both sides of the equation onto l_j,

$$\int_{\mathcal{S}} l_j(y) \int_{\mathcal{S}} c(x,y) l_j(x) dG(x) dG(y) = \gamma_j.$$

Equation (5.11) is found by substituting $l_j(x) = \boldsymbol{a}_j^t \boldsymbol{v}(x)$.

A.10 Proof of Theorem 9.1

We provide only a sketch of the proof.

Write

$$\boldsymbol{R}^t = (R_{11}, R_{12}, \ldots, R_{1n_2}, R_{21}, \ldots, R_{Kn_K}),$$

which is the vector of ranks, ordered according the usual convention. From Lemma 7.3 we know that

$$\mathrm{Var}\{\boldsymbol{R}\} = \boldsymbol{\Sigma}_R = \frac{n+1}{12}\left(n\boldsymbol{I} - \boldsymbol{J}\boldsymbol{J}^t\right)$$

with $\boldsymbol{I}$ and $\boldsymbol{J}$ the $n \times n$ identity matrix and the n-unit vector, respectively. Define the n-dimensional vectors $\boldsymbol{c}_s$ as vectors with all entries equal to zero, except the entries at the positions corresponding to the elements of the sth sample in $\boldsymbol{R}$; these entries are equal to

$$\frac{\sqrt{12}}{\sqrt{n_s n(n+1)}}$$

($s = 1, \ldots, K$). Let $\boldsymbol{C}$ denote an $n \times K$ matrix with sth column equal to $\boldsymbol{c}_s^t$. Note that the columns of this matrix are orthogonal. In a similar fashion we also construct a matrix $\boldsymbol{D}$ which only differs from $\boldsymbol{C}$ by the absence of the factor $\sqrt{12}/\sqrt{n(n+1)}$. The columns of this matrix are orthonormal.

Let $W_s = \boldsymbol{c}_s^t(\boldsymbol{R} - ((n+1)/2)\boldsymbol{J})$. With this notation the KW statistic becomes

$$KW = \sum_{s=1}^{K} W_s^2 = \left(\boldsymbol{R} - \frac{n+1}{2}\boldsymbol{J}\right)^t \boldsymbol{C}\boldsymbol{C}^t \left(\boldsymbol{R} - \frac{n+1}{2}\boldsymbol{J}\right).$$

Asymptotic multivariate normality of $\boldsymbol{W} = \boldsymbol{C}^t \left(\boldsymbol{R} - \frac{n+1}{2}\boldsymbol{J}\right)$ can be shown easily. It has mean zero and its covariance matrix equals

$$\begin{aligned}
\text{Var}\{\boldsymbol{W}\} &= \boldsymbol{C}^t \boldsymbol{\Sigma}_R \boldsymbol{C} \\
&= \boldsymbol{D}^t \left(\boldsymbol{I} - (\boldsymbol{J}/\sqrt{n})(\boldsymbol{J}/\sqrt{n})^t\right) \boldsymbol{D} \\
&= \boldsymbol{D}^t \boldsymbol{E}\boldsymbol{\Gamma}\boldsymbol{E}^t \boldsymbol{D},
\end{aligned}$$

where $\boldsymbol{E}$ has rows equal to the eigenvectors of $\boldsymbol{I} - (\boldsymbol{J}/\sqrt{n})(\boldsymbol{J}/\sqrt{n})^t$, and $\boldsymbol{\Gamma}$ is the diagonal matrix with the eigenvalues. Note that this particular matrix has exactly one zero eigenvalue and $n-1$ eigenvalues equal to one (see also Appendix A.1). For convenience we set this zero at the first diagonal position of $\boldsymbol{\Gamma}$. We now write

$$\text{Var}\{\boldsymbol{W}\} = \boldsymbol{D}^t(\boldsymbol{E}\boldsymbol{\Gamma}^{1/2})(\boldsymbol{E}\boldsymbol{\Gamma}^{1/2})^t\boldsymbol{D}.$$

The zero eigenvalue implies that all entries in the first column of $\boldsymbol{E}\boldsymbol{\Gamma}^{1/2}$ are zero. Moreover, by the orthonormality of $\boldsymbol{D}$ and the $n-1$ eigenvalues 1 in $\boldsymbol{\Gamma}$, we may conclude that Var $\{\boldsymbol{W}\}$ has also one eigenvalue equal to zero, and $K-1$ eigenvalues equal to one. On using Lemma A.1 we may conclude that KW has asymptotically a χ^2_{K-1} distribution under the general K-sample null hypothesis. □

Appendix B
The Bootstrap and Other Simulation Techniques

B.1 Simulation of EDF Statistics Under the Simple Null Hypothesis

In traditional univariate statistics, many test statistics have a limiting standard normal null distribution. For instance, let T_n denote such a test statistic; then the asymptotic results may be denoted by $T_n \xrightarrow{d} N(0,1)$, or $T_n \xrightarrow{d} Z$, where $Z \sim N(0,1)$. A one-sided α-level test may be performed by comparing the observed test statistic with the $1-\alpha$ quantile of the standard normal distribution, which can be found in tables in many textbooks. When working with empirical processes, however, we will often encounter test statistics which have a limiting distribution that has no explicit distribution function. The limiting distribution is often expressed as a function of a Gaussion process. In this case, the criticial values will often have to be esimated by means of simulations of the empirical process. The next R-code generates a realization of a Brownian bridge at frequency=1000 equally spaced points between 0 and 1. The larger the frequency, the better the realization approximates a true continuous process.

```
> B<-rbridge(frequency=1000)
```

The asymptotic null distribution of the KS test can now be simulated by the following lines.

```
> ks<-rep(NA,10000)
> for(i in 1:10000) {
+   ks[i]<-max(abs(rbridge(frequency=1000)))
+ }
```

We can use ks for instance to compute the p-value of the PRG example.

```
> length(ks[ks>sqrt(100000)*0.0029])/10000
[1] 0.3394
```


O. Thas, *Comparing Distributions*,
DOI 10.1007/978-0-387-92710-7,

A better approximation can be obtained by increasing the frequency and the number of Monte Carlo simulation runs.

B.2 The Parametric Bootstrap for Composite Null Hypotheses

The parametric bootstrap may be used for testing a full parametric null hypothesis, whether simple or composite. Here we describe the method for testing for a composite null hypothesis, but it can be applied to simple null hypotheses too by simply fixing the $\boldsymbol{\beta}$ nuisance parameter throughout the algorithm.

Consider the null hypothesis

$$H_0 : F \in \{G(.;\boldsymbol{\beta}) : \boldsymbol{\beta} \in B\}$$

(see Section 4.2.2 for more details on this type of composite null hypothesis). Let $\boldsymbol{X}^t = (X_1, \ldots, X_n)$ denote the sample of n i.i.d. observations, and $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$ under H_0. Suppose the test statistic is denoted by $T = T(\boldsymbol{X}, \hat{\boldsymbol{\beta}})$.

The parametric bootstrap procedure consists in sampling B times n i.i.d. observations from the distribution $G(.;\hat{\boldsymbol{\beta}})$. The jth sample is denoted by $\boldsymbol{X}_j^*$, and the estimator of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_j^*$. For each bootstrap sample the test statistic is recalculated, which is denoted by $T_j^* = T(\boldsymbol{X}_j^*, \hat{\boldsymbol{\beta}}_j^*)$. The empirical distribution of the B bootstrapped test statistics, $T_1^*, \ldots, T_B^*$, serves as an approximation of the asymptotic null distribution of T.

B.3 A Modified Nonparametric Bootstrap for Testing Semiparametric Null Hypotheses

The method described here was proposed by Bickel and Ren (2001). See also Bickel et al. (2006).

Let $\mathcal{F}$ denote a class of density functions for which the distribution of the test statistic behaves well, and let $\boldsymbol{X}^t = (X_1, \ldots, X_n)$ denote the vector of the n i.i.d. sample observations. Let $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{X})$ denote a k-dimensional statistic. Consider test statistics of the form

$$T = \boldsymbol{U}^t(\boldsymbol{X})\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X})\boldsymbol{U}(\boldsymbol{X}),$$

where $\hat{\boldsymbol{\Sigma}}(\boldsymbol{X})$ is an estimator of $\mathrm{Var}\{\boldsymbol{U}\}$ that is $\sqrt{n}$-consistent for all $f \in \mathcal{F}$. Consider a semiparametric null hypothesis formulated as

$$H_0 : f \in \mathcal{F}_0,$$

where

$$\mathcal{F}_0 = \{f \in \mathcal{F} : \mathrm{E}_f \{\boldsymbol{U}\} = 0\}.$$

Consider now a nonparametric bootstrap procedure in which $\boldsymbol{X}_j^*$ denotes the jth bootstrap sample. For each bootstrap sample the test statistic is calculated as

$$T_j^* = \left(\boldsymbol{U}(\boldsymbol{X}_j^*) - \boldsymbol{U}(\boldsymbol{X})\right)^t \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_j^*) \left(\boldsymbol{U}(\boldsymbol{X}_j^*) - \boldsymbol{U}(\boldsymbol{X})\right).$$

When B bootstrap simulations are performed, the empirical distribution of $T_1^*, T_2^*, \ldots, T_B^*$ is used as an approximation of the null distribution of T.

References

L. Acion, J. Peterson, S. Temple, and S. Arndt. Probabilistic index: An intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in Medicine*, 25:591–602, 2006.

N. Aguirre and M. Nikulin. Goodness-of-fit tests for the family of logistic distributions. *Q uesti'o*, 18:317–335, 1994.

H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. Petrov and F. Csàki, editors, *Second International Symposium on Inference Theory*, pages 267–281, Budapest, 1973. Akadémiai Kiadó.

H. Akaike. A new look at statistical model identification. *I.E.E.E. Transactions on Automatic Control*, 19:716–723, 1974.

M. Akritas and E. Brunner. A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61:249–277, 1997.

W. Alexander. *Boundary Kernel Estimation of the Two-Sample Comparison Density Function*. PhD thesis, Texas A& M University, College Station, Texas, USA, 1989.

D. Allison, G. Page, T. Beasley, and J. E. Edwards. *DNA Microarrays and Related Genomics Techniques : Design, Analysis, and Interpretation of Experiments*. Chapman and Hall, Boca Raton, Florida, USA, 2006.

T. Anderson and D. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23:193–212, 1952.

T. Anderson and D. Darling. A test of goodness-of-fit. *Journal of the American Statistical Association*, 49:765–769, 1954.

A. Ansari and R. Bradley. Rank-sum tests for dispersion. *Annals of Mathematical Statistics*, 31:1174–1189, 1960.

A. Atkinson. Tests of pseudo-random numbers. *Applied Statistics*, 29:164–171, 1980.

G. Babu and A. Padmanabhan. Resampling methods for the non-parametric Behrens-Fisher problem. *Sankhya, Series A*, 64:678–692, 2002.

G. Babu and C. Rao. Goodness-of-fit tests when parameters are estimated. *Sankhya, series A*, 2004.

D. Bamber. The area above the ordinal dominance graph and the area below the receiver operator characteristic graph. *Journal of Mathematical Psychology*, 12:287–415, 1975.

L. Baringhaus and N. Henze. A class of tests for exponentiality based on the empirical Laplace transform. *annals of the institute of statistical mathematics*, 43:551–564, 1991.

L. Baringhaus, N. G Urtler, and N. Henze. Weighted integral test statistics and components of smooth tests of fit. *Australian and New Zealand journal of statistics*, 42:179–192, 2000.

D. Barton. On Neyman's smooth test of goodness of fit and its power with respect to a particular system of alternatives. *Skandinavisk Aktuarietidskrift*, 36:24–63, 1953.

O. Thas, *Comparing Distributions*,
DOI 10.1007/978-0-387-92710-7,

D. Bauer. Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67:687–690, 1972.

T. Bednarski and T. Ledwina. A note on biasedness of tests of fit. *Mathematische Operationsforschung und Statistik, Series Statistics*, 9:191–193, 1978.

K. Behnen and M. Husková. A simple algorithm for the adaptation of scores and power behavior of the corresponding rank tests. *Communications in Statistics - Theory and Methods*, 13:305–325, 1984.

K. Behnen and G. Neuhaus. Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Annals of Statistics*, 11:588–599, 1983.

K. Behnen, G. Neuhaus, and F. Ruymgaart. Two sample rank estimators of optimal nonparametric score-functions and corresponding adaptive rank statistics. *Annals of Statistics*, 11:1175–1189, 1983.

A. Bernard and E. Bos-Levenbach. The plotting of observations on probability paper. *Statistica Neerlandica*, 7:163–173, 1953.

P. Bickel and D. Freedman. Some asymptotic theory for the bootstrap. *annals of statistics*, 9:1196–1217, 1981.

P. Bickel and J. Ren. The *Bootstrap in Hypothesis Testing.* In M. de Gunst, C. Klaassen, and A. van der Vaart, editors, *Festschrift for Willem R. van Zwet*, pages 91–112. IMS, Beachwood, USA, 2001.

P. Bickel, Y. Ritov, and T. Stoker. Tailor-made tests of goodness of fit to semiparametric hypotheses. *Annals of Statistics*, 34:721–741, 2006.

M. Birch. A new proof of the Pearson-Fisher theorem. *AMS*, 35:817–824, 1964.

Z. Birnbaum and O. Klose. Bounds for the variance of the Mann-Whitney statistic. *Annals of Mathematical Statistics*, 23:933–945, 1957.

Y. Bishop, S. Fienberg, and P. Holland. *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA, USA, 1975.

G. Blom. *Statistical Estimates and Transformed Beta Variables.* Wiley, New York, 1958.

M. Bogdan. Data driven version of Pearson's chi-square test for uniformity. *Journal of Statistical Computation and Simulation*, 52:217–237, 1995.

D. Boos. Comparing k populations with linear rank statistics. *Journal of the American Statistical Association*, 81:1018–1025, 1986.

D. Boos. On generalized score tests. *The American Statistician*, 46:327–333, 1992.

J. Box. *R. A. Fisher, the Life of a Scientist.* Wiley, New York, USA, 1978.

S. Buckland. Fitting density functions with polynomials. *Applied Statistics*, 41:63–76, 1992.

C. Carolan and J. Tebbs. Nonparametric tests for and against likelihood ratio ordening in the two-sample problem. *Biometrika*, 92:159–171, 2005.

B. Carvalho, C. Postma, S. Mongera, E. Hopmans, S. Diskin, M. van de Wiel, W. Van Criekinge, O. Thas, A. Matth Ai, M. Cuesta, J. Terhaar, M. Craanen, E. Schr Ock, B. Ylstra, and G. Meijer. Integration of dna and expression microarray data unravels seven putative oncogenes on 20q amplicon involved in colorectal adenoma to carcinoma progression. *Cellular Oncology*, 2008.

N. Cencov. Evaluation of an unknown distribution density from observations. *Soviet. Math.*, 3:1559–1562, 1962.

H. Chernoff and E. Lehmann. The use of maximum-likelihood estimates in χ^2 tests for goodness of fit. *Annals of Mathematical Statistics*, 25:579–586, 1954.

H. Chernoff and I. Savage. Asymptotic normality and efficiency of certain non-parametric test statistics. *Annals of Mathematical Statistics*, 29:972–994, 1958.

I. Chervoneva and B. Iglewicz. Orthogonal basis approach for comparing nonnormal continuous distributions. *Biometrika*, 92:679–690, 2005.

Y. Cheung and J. Klotz. The Mann Whitney Wilcoxon distribution using linked lists. *Statistica Sinica*, 7:805–813, 1997.

G. Claeskens and N. Hjort. Goodness of fit via non-parametric likelihood ratios. *Scandinavian Journal of Statistics*, 31:487–513, 2004.

A. Cohen and H. Sackrowitz. Unbiasedness of the chi-squared, likelihood ratio, and other goodness of fit tests for the equal cell case. *Annals of Statistics*, 3:959–964, 1975.

H. Cramér. On the composition of elementary errors. *Skandinavisk Aktuarietidskrift*, 11: 13–74, 141–180, 1928.

N. Cressie and T. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, 46:440–464, 1984.

M. Csörgö. *Quantile Processes with Statistical Applications*. SIAM, Philadelphia, USA, 1983.

M. Csörgö and L. Horváth. *Weighted Approximations in Probability and Statistics*. Wiley, New York, USA, 1993.

M. Csörgö and P. Révész. Strong approximations of the quantile process. *The Annals of Statistics*, 6:822–894, 1978.

M. Csörgö, S. Csörgö, L. Horváth, and D. Mason. Weighted empirical and quantile process. *Annals of Probability*, 14:31–85, 1986.

M. Csörgö, L. Horváth, and Q. Shao. Convergence of integrals of uniform empirical and quantile processes. *Stochastic Processes and Their Applications*, 45:283–294, 1993.

S. Csörgö. Weighted correlation tests for scale families. *test*, 11:219–248, 2002.

S. Csörgö and J. Faraway. The exact and asymptotic distribuitons of Cramér-von Mises statistics. *JRSSB*, 58:221–234, 1996.

C. Cunnane. Unbiased plotting positions - a review. *Journal of Hydrology*, 37:205–222, 1978.

J. Cwik and J. Mielniczuk. Data-dependent bandwidth choice for a grade kernel estimate. *Statistics and Probability Letters*, 16:397–405, 1993.

R. D'Agostino and M. Stephens. *Goodness-of-Fit Techniques*. Marcel Dekker, New York, USA, 1986.

G. Dallal and L. Wilkinson. An analytic approximation to the distribution of lilliefors' test for normality. *The American Statistician*, 40:294–296, 1986.

T. de Wet. Goodness-of-fit tests for location and scale families based on a weighted l_2-Wasserstein distance measure. *Test*, 11:89–107, 2002.

E. del Bario, J. Cuesta-Albertos, and C. Matrán. Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test*, 9:1–96, 2000.

E. del Barrio, J. Cuesta Albertos, and C. Matrán Y J. Rodríguez Rodríguez. Tests of goodness of fit based on the L2-Wasserstein distance. *Annals of Statistics*, 27:1230–1239, 1999.

E. Del Barrio, E. Giné, and F. Utzet. Asymptotics for l_2 functionals of the empirical quantile process, with applications to tests of fit based on the weighted Wasserstein distances. *Bernoulli*, 11:131–189, 2005.

K. Doksum. Empirical probability plots and statistical inference for nonlinear models in the two sample case. *Annals of Statistics*, 2:267–277, 1974.

H. Doss and R. Gill. An elementary approach to weak convergence for quantile processes, with applications to censored survival data. *JASA*, 87:869–877, 1992.

F. Drost. Asymptotics for generalized chi-square goodness-of-fit tests. In *CWI Tract*. Centrum voor Wiskunde en Informatica, Amsterdam, The Netherlands, 1988.

J. Durbin. Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics*, 1:279–290, 1973.

J. Durbin and M. Knott. Components of Cramér - von Mises statistics. *Journal of the Royal Statistical Society, Series B*, 34:290–307, 1972.

J. Durbin, M. Knott, and C. Taylor. Components of Cramér - von Mises statistics: II. *Journal of the Royal Statistical Society, Series B*, 37:216–237, 1975.

M. Dwass. Some k-sample rank-order tests. In *Contributions to Probabilitiy and Statistics, Essays in Honor of H. Hotelling*, pages 198–202. Stanford University Press, Stanford, USA, 1960.

B. Efron and R. Tibshirani. Using specially designed exponential families for density estimation. *Annals of Statistics*, 24:2431–2461, 1996.

J. Einmahl and D. Mason. Strong limit theorems for weighted quantile processes. *Annals of Probability*, 16:1623–1643, 1988.

J. Einmahl and I. McKeague. Empirical likelihood based hypothesis testing. *Bernoulli*, 9:267–290, 2003.

K. Entacher and H. Leeb. Inversive pseudorandom number generators: Empirical results. In *Proceedings of the Conference Parallel Numerics 95*, pages 15–27, Sorrento, Italy, 1995.

T. Epps and L. Pulley. A test for normality based on the empirical characteristic function. *Biometrika*, 70:723–726, 1983.

R. Eubank and V. LaRiccia. Components of pearson's phi-squared distance measure for the k-sample problem. *Journal of the American Statistical Association*, 85:441–445, 1990.

R. Eubank, V. LaRiccia, and R. Rosenstein. Test statistics derived as components of Pearson's phi-squared distance measure. *Journal of the American Statistical Association*, 82:816–825, 1987.

J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*. Chapman and Hall, London, UK, 1996.

R. Farrell. On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Annals of Mathematical Statistics*, 43:170–180, 1972.

P. Feigin and C. Heathcore. The empirical characteristic function and the Cramér-von Mises statistic. *Sankhya A*, 38:309–325, 1977.

J. Filliben. The probability plot coefficient test for normality. *Technometrics*, 17:111–117, 1975.

M. Fisz. Some non-parametric tests for the k-sample problem. *Colloquium Math.*, 7:289–296, 1960.

M. Fligner and T. Killeen. Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, 71:210–213, 1976.

M. Fligner and G. Policello. Robust rank procedures for the Behrens-Fisher problem. *JASA*, 76:162–168, 1981.

G. Gajek. On improving density estimators which are not bona fide functions. *Annals of Statistics*, 14:1612–1618, 1986.

R. Gentleman, R. Irizarry, V. Carey, S. Dutoit, and W. E. Huber. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, New York, USA, 2005.

D. Gillen and S. Emerson. Nontransitivity in a class of weighted logrank statistics under nonproportional hazards. *Statistics and Probability Letters*, 77:123–130, 2007.

I. Glad, N. Hjort, and N. Ushakov. Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, 30:415–427, 2003.

P. Good. *Resampling Methods: a Practical Guide to Data Analysis*. Birkhauser, Boston, USA, 3rd edition, 2005.

P. Greenwood and M. Nikulin. *A Guide to Chi-Squared Testing*. Wiley, New York, USA, 1996.

I. Gringorten. A plotting rule for extreme probability paper. *Journal of Geophysical Research*, 68:813–814, 1963.

N. Gürtler and N. Henze. Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Annals of the Institute of Statistical Mathematics*, 52:267–286, 2000.

J. Hájek, Z. Šidák, and P. Sen. *Theory of Rank Tests*. Academic Press, San Diego, USA, 2nd edition, 1999.

P. Hall. Orthogonal series distribution function estimation, with applications. *Journal of the Royal Statistical Society, Series B*, 45:81–88, 1983.

P. Hall. On the rate of convergence of orthogonal series density estimators. *Journal of the Royal Statistical Society, Series B*, 48:115–122, 1986.

P. Hall. On Kullback-Leibler loss and density estimation. *Annals of Statistics*, 15: 1491–1519, 1987.

P. Hall and R. Murison. Correcting the negativity of high-order kernel density estimators. *Journal of Multivariate Analysis*, 47:103–122, 1993.

W. Hall and D. Mathiason. On large-sample estimation and testing in parametric models. *International Statistical Review*, 58:77–97, 1990.

M. Halperin, P. Gilbert, and J. Lachin. Distribution-free confidence intervals for $pr(x_1 < x_2)$. *Biometrics*, 43:71–80, 1987.

F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel. *Robust Statistics: The Approach Based on the Influence Function*. Springer-Verlag, New York, USA, 1986.

D. Hand, F. Daly, A. Lunn, K. McConway, and E. Ostrowsky. *A Handbook of Small Data Sets*. Chapman and Hall, London, UK, 1994.

M. Handcock and M. Morris. *Relative Distribution Methods in Social Siences*. Springer-Verlag, New York, USA, 1999.

J. Hart. *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, Berlin, Germany, 1997.

A. Hazen. *Flood Flows*. Wiley, New York, USA, 1930.

C. Heathcore. A test of goodness of fit for symmetric random variables. *Australian Journal of Statistics*, 14:172–181, 1972.

N. Henze. A new flexible class of tests for exponentiality. *Communications in Statistics - Theory and Methods*, 22:115–133, 1993.

N. Henze. Do components of smooth tests of fit have diagnostic properties? *Metrika*, 45:121–130, 1997.

N. Henze and B. Klar. Properly rescaled components of smooth tests of fit are diagnostic. *Australian Journal of Statistics*, 38:61–74, 1996.

N. Henze and S. Meintanis. Goodness-of-fit tests based on a new characterization of the exponential distribution. *Communications in Statistics - Theory and Methods*, 31:1479–1497, 2002.

R. Hilgers. On the Wilcoxon-Mann-Whitney-test as nonparametric analogue and extension of t-test. *Biometrical Journal*, 24:1–15, 2007.

N. Hjort and I. Glad. Nonparametric density estimation with a parametric start. *Annals of Statistics*, 23:882–904, 1995.

J. Hodges and E. Lehmann. Some problems in minimax point estimation. *Annals of Mathematical Statistics*, 21:182–197, 1956.

J. Hodges and E. Lehmann. *Hodges-Lehmann Estimators*. In S. Kotz, L. Johnson, and C. Read, editors, *Encyclopedia of Statistical Sciences, Volume 3*. Wiley, New York, USA, 1983.

P. Holland. A variation on the minimum chi-square test. *Journal of Mathematical Psychology*, 4:377–413, 1967.

M. Hollander and D. Wolfe. *Nonparametric Statistical Methods*. Wiley, New York, USA, 1999.

E. Holmgren. The P-P plot as a method for comparing treatment effects. *JASA*, 90:360–365, 1995.

T. Hothorn, K. Hornik, M. van de Wiel, and A. Zeileis. A lego system for conditional inference. *The American Statistician*, 60:257–263, 2006.

F. Hsieh. The empirical process approach for semiparametric two-sample models with heterogeneous treatment effect. *JRSSB*, 57:735–748, 1995.

F. Hsieh and B. Turnbull. Non- and semi-parametric estimation of the receiver operating characteristic curve. Technical Report 1026, school of operations research, Cornell University, 1992.

F. Hsieh and B. Turnbull. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics*, 24:25–40, 1996.

P. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the 5th Berkeley Symposium*, 1:221–233, 1967.

P. Huber. *Robust Statistics*. Wiley, New York, USA, 1974.

R. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50:361–365, 1996.

T. Inglot and A. Janic-Wróblewska. Data driven chi-square test for uniformity with unequal cells. *Journal of Statistical Computation and Simulation*, 73:545–561, 2003.

T. Inglot and T. Ledwina. Towards data driven selection of a penalty function for data driven Neyman tests. *Linear Algebra and its Applications*, 417:579–590, 2006.

T. Inglot, W. Kallenberg, and T. Ledwina. Data driven smooth tests for composite hypotheses. *Annals of Statistics*, 25:1222–1250, 1997.

A. Janic-Wróblewska. Data-driven smooth test for a location-scale family. *Statistics*, 38: 337–355, 2004.

A. Janic-Wróblewska and T. Ledwina. Data driven rank test for two-sample problem. *Scandinavian Journal of Statistics*, 27:281–297, 2000.

A. Janic-Wróblewska and T. Ledwina. Data-driven smooth tests for a location-scale family revisited. *Jourmal of Statistical Theory and Practice*, to appear, 2009.

A. Janssen. Global power functions of goodness of fit tests. *Annals of Statistics*, 29:239–253, 2000.

A. Janssen. Principal component decomposition of non-parametric tests. *Probability Theory and Related Fields*, 101:193–209, 1995.

A. Janssen and T. Pauls. A monte carlo comparison of studentized permutation and bootstrap for heteroscedastic two-sample problems. *Computational Statistics*, 20:369–383, 2005.

H. Javitz. *Generalized Smooth Tests of Goodness of Fit, Independence and Equality of Distributions*. PhD thesis, unpublished thesis, Univ. of Calif., Berkely, USA, 1975.

M. Kac and A. Siegert. An explicit representation of a stationary Gausian process. *Annals of Mathematical Statistics*, 18:438–442, 1947.

M. Kac, J. Kiefer, and J. Wolfowitz. On tests of normality and other tests of goodness-of-fit based on distance methods. *Annals of Mathematical Statistics*, 26:189–211, 1955.

W. Kaigh. EDF and EQF orthogonal component decompositions and tests of uniformity. *Nonparametric Statistics*, 1:313–334, 1992.

W. Kallenberg and T. Ledwina. Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Annals of Statistics*, 23:1594–1608, 1995a.

W. Kallenberg and T. Ledwina. On data-driven Neyman's tests. *Probability and Mathematical Statistics*, 15:409–426, 1995b.

W. Kallenberg and T. Ledwina. Data-driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association*, 92:1094–1104, 1997.

W. Kallenberg, J. Oosterhoff, and B. Schriever. The number of classes in chi-squared goodness-of-fit tests. *Journal of the American Statistical Association*, 80:959–968, 1985.

M. Kaluszka. On the Devroye-Gyorfi methods of correcting density estimators. *Statistics and Probability Letters*, 37:249–257, 1998.

R. Kanwal. *Linear Integral Equations, Theory and Technique*. Academic Press, New York, USA, 1971.

M. Karpenstein-Machan and R. Maschka. Investigations on yield structure and local adaptibility. *Agrobiological Research*, 49:130–143, 1996.

M. Karpenstein-Machen, B. Honermeier, and F. Hartmann. *Produktion Aktuell, Triticale*. DLG Verlag, Frankfurt, Germany, 1994.

J. Kiefer. Deviations between the sample quantile process and the sample DF. In M. Puri, editor, *Proceedings of the Conference on Nonparametric Techniques in Statistical Inference*, pages 299–319, Cambridge, UK, 1970. Cambridge University Press.

B. Kimball. On the choice of plotting positions on probability paper. *JASA*, 55:546–560, 1960.

B. Klar. Diagnostic smooth tests of fit. *Metrika*, 52:237–252, 2000.

M. Knot. The distribution of the Cramér-von Mises statistic for small sample sizes. *Journal of the Royal Statistical Society, Series B*, 36:430–438, 1974.

D. Knuth. *The Art of Computer Programming, Volume 2*. Addison-Wesley, Reading, MA, USA, 1969.

A. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Gior. Ist. Ital. Attuari*, 4:83–91, 1933.

M. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, USA, 2008.

H. Lancaster. *The Chi-Squared Distribution*. Wiley, London, UK, 1969.

R. Larsen, T. Curran, and W. J. Hunt. An air quality data analysis system for interrelating effects, standards, and needed source reductions: Part 6. calculating concentration reductions needed to achieve the new national ozone standard. *Journal of Air Pollution Control Association*, 30:662–669, 1980.

T. Ledwina. Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, 89:1000–1005, 1994.

A. Lee. *U-Statistics*. Marcel Dekker, New York, USA, 1990.

J. Lee and N. Tu. A versatile one-dimensional distribution plot: The BLiP plot. *The American Statistician*, 51:353–358, 1997.

E. Lehmann. Consistency and unbiasedness of certain nonparametric tests. *Annals of Mathematical Statistics*, 22:165–179, 1951.

E. Lehmann. The power of rank tests. *Annals of Mathematical Statistics*, 24:23–43, 1953.

E. Lehmann. *Nonparametrics. Statistical Methods Based on Ranks*. Prentice Hall, Upper Saddle River, NJ, USA, 1998.

E. Lehmann. *Elements of Large-Sample Theory*. Springer, New York, USA, 1999.

E. Lehmann and J. Romano. *Testing Statistical Hypotheses (3rd Ed.)*. Springer, New York, USA, 2005.

Y. Lepage. A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika*, 58: 213–217, 1971.

P. Lewis. Distribution of the Anderson-Darling statistic. *Annals of Mathematical Statistics*, 32:1118–1124, 1961.

G. Li, R. Tiwari, and M. Wells. Quantile comparison functions in two-sample problems with applications to comparisons of diagnostic markers. *JASA*, 91:689–698, 1996.

W. Lidicker and F. McCollum. Allozymic variation in California sea otters. *Journal of Mammalogy*, 78:417–425, 1997.

H. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62:399–402, 1967.

C. Lin and S. Sukhatme. Hoeffding type theorem and power comparisons of some two-sample rank tests. *Journal of the Indian Statistical Association*, 31:71–83, 1993.

T. Lumley. Non-transitivity of the Wilcoxon rank sum test. Personal communication, 2009.

D. Mage. An objective graphical method for testing normal distributional assumptions using probability plots. *The American Statistician*, 36:116–120, 1982.

H. Mann and A. Wald. On the choice of the number of class intervals in the application of the chi-square test. *Annals of Mathematical Statistics*, 13:306–317, 1942.

H. Mann and D. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60, 1947.

D. Mason. Weak convergence of the weighted empirical quantile process in $l^2(0,1)$. *Annals of Probability*, 12:243–255, 1984.

F. Massey. The distribution of the maximum deviation between two sample cumulative step functions. *Annals of Mathematical Statistics*, 22:125–128, 1951.

M. Matsui and A. Takemura. Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Annals of the Institute of Statistical Mathematics*, 57:183–199, 2005.

R. Mee. Confidence intervals for probabilities and tolerance regions based on a generalisation of the Mann-Whitney statistic. *Journal of the American Statistical Association*, 85:793–800, 1990.

S. Meintanis. Goodness-of-fit tests for the logistic distribution based on empirical transforms. *Sankhya, Series B*, 66:306–326, 2004a.

S. Meintanis. A class of omnibus tests for the Laplace distribution based on the empirical characteristic function. *Communications in Statistics - Theory and Methods*, 33:925–948, 2004b.

S. Meintanis. Consistent tests for symmetry stability with finite mean based on the empirical characteristic function. *Journal of Statistical Planning and Inference*, 128:373–380, 2005.

J. Michael. The stabilized probability plot. *Biometrika*, 70:11–17, 1983.

P. Mielke and K. Berry. *Permutation Tests: A Distance Function Approach*. Springer, New-York, USA, 2001.

J. Mielniczuk. Grade estimation of Kullback-Leibler information number. *Probability and Mathematical Statistics*, 13:139–147, 1992.

A. Mood. On the asymptotic efficiency of certain nonparametric two-sample tests. *Annals of Mathematical Statistics*, 25:514–522, 1954.

D. Moore. *Tests of Chi-Squared Type.* In R. D'Agostino and M. Stephens, editors, *Goodness-of-Fit Techniques*, chapter 3, pages 63–95. Marcel Dekker, New York, USA, 1986.

L. Moses. Rank tests for dispersion. *The Annals of Mathematical Statitstics*, 34:973–983, 1963.

G. Neuhaus. Local asymptotics for linear rank statistics with estimated score functions. *Annals of Statistics*, 15:491–512, 1987.

R. Newcombe. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. part 1: General issues and tail-area-based methods. *Statistics in Medicine*, 25: 543–557, 2006a.

R. Newcombe. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. part 2: Asymptotic methods and evaluation. *Statistics in Medicine*, 25:559–573, 2006b.

J. Neyman. Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift*, 20:149–199, 1937.

J. Neyman. Contribution to the theory of the χ^2 test. In *Proceedings of the First Berkeley Symposium of Mathematical Statistics and Probability*, pages 239–273, 1949.

J. Oosterhoff. The choice of cells in chi-square tests. *Statistica Neerlandica*, 39:115–128, 1985.

E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

E. Parzen. Nonparametric statistical data modeling (with discussion). *JASA*, 74:105–131, 1979.

E. Parzen. FUN.STAT: Quantile approach to two sample statistical data analysis. Technical report, Texas A& M University, College Station, Texas, USA, 1983.

E. Parzen. Statistical methods, mining, tow sample data analysis, comparison distributions, and quantile limit theorems. In *Proceedings of the International Conference on Asymptotic Methods in Probability and Statistics, 8-13 July 1997, Carleton University, Canada*, 1997.

E. Parzen. *Statistical Methods, Mining, Two-Sample Data Analysis, Comparison Distributions, and Quantile Limit Theorems.* In *Asymptotic Methods in Probability and Statistics.* Elsevier, Amsterdam, The Netherlands, 1999.

E. Pearson and H. Hartley. *Biometrika Tables for Statisticians, Vol. 2.* Cambridge University Press, Cambridge, UK, 1972.

E. Pearson and M. Stephens. The goodness-of-fit tests based on W_N^2 and U_N^2. *Biometrika*, 49:397–402, 1962.

K. Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50:157–175, 1900.

A. Pettitt. A two-sample Anderson-Darling rank statistic. *Biometrika*, 63:161–168, 1976.

H. Piepho. Exact confidence limits for covariate-dependent risk in cultivar trials. *Journal of Agricultural, Biological and Environmental Statistics*, 5:202–213, 2000.

E. Pitman. *Notes on Non-parametric Statistical Inference*. Columbia University, New York, USA, 1948.

R. Potthof. Use of the wilcoxon statistic for a generalized Behrens-Fisher problem. *Annals of Mathematical Statistics*, 34:1596–1599, 1963.

N. Pya. Goodness-of-fit tets for the logistic distribution. *Mathematical Journal*, 4:68–75, 2004.

R. Pyke and G. Shorack. Weak convergences of a two-sample empirical process and a new approach to Chernoff-Savage theorems. *Annals of Mathematical Statistics*, 39:755–771, 1968.

A. Qu, B. Lindsay, and B. Li. Improving generalised estimating functions using quadratic inference functions. *Biometrika*, 87:823–836, 2000.

M. Quine and J. Robinson. Efficiencies of chi-square and likelihood ratio goodness-of-fit tests. *Annals of Statistics*, 13:727–742, 1985.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. URL http://www.R-project.org. ISBN 3-900051-07-0.

R. Randles and R. Hogg. Adaptive distribution-free tests. *Communications in Statistics*, 2:337–356, 1973.

K. Rao and D. Robson. A chi-square statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics*, 3:1139–1153, 1974.

L. Rayleigh. On the problems of random vibrations and flights in one,two, or three dimensions. *Philosophical Magazine*, 37:321–347, 1919.

J.C.W. Rayner and D.J. Best. *A Contingency Table Approach to Nonparametric Testing*. Chapman and Hall, New York, USA, 2001.

J.C.W. Rayner and D.J. Best. Neyman-type smooth tests for location-scale families. *Biometrika*, 73:437–446, 1986.

J.C.W. Rayner and D.J. Best. *Smooth Tests of Goodness-of-Fit*. Oxford University Press, New York, USA, 1989.

J.C.W. Rayner, O. Thas, and B. De Boeck. A generalised Emerson recurrence relation. *Australian and New Zealand Journal of Statistics*, 50:235240, 2008.

J.C.W. Rayner, O. Thas, and D.J. Best. *Smooth Tests of Goodness of Fit: Using R*. Wiley, Singapore, 2009.

T. Read and N. Cressie. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer-Verlag, New York, 1988.

R. Risebrough. Effects of environmental pollutants upon animals other than man. In *Proceedings of the 6th Berkeley Symposium on Mathematics and Statistics*, pages 443–463, Berkeley, 1972. University of California University Press.

J. Romano. A bootstrap revival of some nonparametric distance tests. *Journal of the American Statistical Association*, 83:698–708, 1988.

M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

F. Scholz and M. Stephens. k-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82:918–924, 1987.

G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

D. Scott. *Multivariate Density Estimation*. Wiley, New York, USA, 1992.

P. Sen. A note on asymptotically distribution-free condence intervals for $\Pr[x < y]$ based on two independent samples. *Sankhya, A.*, 29:95–102, 1967.

J. Shao. Jackknife variance estimators for two sample linear rank statistics. Technical Report 88-61, Department of Statistics, Purdue University, USA, 1988.

J. Shao. Differentiability of statistical functionals and consistency of the jackknife. *The Annals of Statistics*, 21:61–75, 1993.

G. Shorack. *Probability for Statisticians*. Springer-Verlag, New York, USA, 2000.

G. Shorack and J. Wellner. *Empirical Processes with Applications to Statistics*. Wiley, New York, USA, 1986.

S. Siegel and J. Tukey. A nonparametric sum of rank procedure for relative spread in unpaired samples. *Journal of the American Statistical Association*, 55:429–444, 1960.

B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, UK, 1986.

J. Simonoff. *Smoothing Methods in Statistics*. Springer, New York, USA, 1996.

N. Smirnov. Sur les ecarts de la courbe de distribution empirique (in Russian). *Rec. Math.*, 6:3–26, 1939.

T. E. Speed. *Statistical Analysis of Gene Expression Microarray Data*. Chapman and Hall, Boca Raton, Florida, USA, 2003.

M. Stephens. Use of the Kolmogorov-Smirnov, Cramér-von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society, Series B*, 32:115–122, 1970.

M. Stephens. Asymptotic results for goodness-of-fit statistics when parameters must be estimated. Technical Report 159 and 180, Department of Statistics, Stanford University, 1971.

M. Stephens. EDF statistics for goodness-of-fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

M. Stephens. Asymptotic results for goodness-of-fit statistics with unknown parameters. *Annals of Statistics*, 4:357–369, 1976.

B. Sukhatme. On certain two-sample nonparametric tests for variances. *Annals of Mathematical Statistics*, 28:188–194, 1957.

P. Switzer. Confidence procedures for two-sample problems. *Biometrika*, 63:13–25, 1976.

O. Thas. *Nonparametrical Tests Based on Sample Space Partitions*. PhD thesis, Ghent University, 2001.

O. Thas and J. Ottoy. Goodness-of-fit tests based on sample space partitions: An unifying overview. *Journal of Applied Mathematics and Decision Sciences*, 6:203–212, 2002.

O. Thas and J. Ottoy. Some generalization of the Anderson-Darling statistic. *Statistics and Probability Letters*, 64:255–261, 2003.

O. Thas and J. Ottoy. An extension of the Anderson-Darling k-sample test to arbitrary sample space partition sizes. *Journal of Statistical Computation and Simulation*, 74: 561–666, 2004.

O. Thas and J.C.W. Rayner. Informative statistical analyses using smooth goodness-of-fit tests. *Journmal of Statistical Theory and Practice*, to appear, 2009.

H. Thode. *Testing for Normality*. Marcel Dekker, New York, USA, 2002.

M. Tiku. Chi-square approximations for the distributions of goodness-of-fit statistics U_N^2 and W_N^2. *Biometrika*, 52:630–633, 1965.

A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, USA, 2006.

J. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, USA, 1977.

V. Tuscher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences*, 98: 5115–5121, 2001.

S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability Applications*, 18:785–786, 1973.

M. van de Wiel. The split-up algorithm: A fast symbolic method for computing p values of rank statistics. *Computational Statistics*, 16:519–538, 2001.

A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK, 1998.

A. Van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, USA, 2nd edition, 2000.

B. van der Waerden. Order tests for the two-sample problem and their power. *Indagationes Mathematicae*, 14:453–458, 1952.

B. van der Waerden. Order tests for the two-sample problem and their power. *Indagationes Mathematicae*, 15:303–310, 1953.

C. van Eeden. Note on the consistency of some distribution-free tests for dispersion. *Journal of the American Statistical Association*, 59:105–119, 1964.

R. von Mises. *Wahrscheinlichkeitsrechnung*. Deuticke, Vienna, Austria, 1931.

R. von Mises. On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, 18:309–348, 1947.

G. Wahba. Data-based optimal smoothing of orthogonal series density estimates. *Annals of Statistics*, 9:146–156, 1958.

L. Wasserstein and J. J. Boyer. Bounds on the power of linear rank tests for scale parameters. *American Statistician*, 45:10–13, 1991.

G. Watson. Goodness-of-fit tests on a circle. *Biometrika*, 48:109–114, 1961.

G. Watson. Density estimation by orthogonal series. *Annals of Mathematical Statistics*, 40:1496–1498, 1969.

F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1:80–83, 1945.

H. Wouters, O. Thas, and J. Ottoy. Data driven smooth tests and a diagnostic tool for lack-of-fit for circular data. *Australian and New Zealand Journal of Statistics*, to appear.

S. Zaremba. A generalisation of Wilcoxon's test. *Monatshefte f ur Mathematik*, 66:359–370, 1962.

J. Zhang. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society, Series B*, 64:281–294, 2002.

J. Zhang and Y. Wu. A family of simple distribution functions to approximate complicated distributions. *Journal of Statistical Computing and Simulation*, 70:257–266, 2001.

Index

K-sample problem, 165
2 sample poblem, 164

adaptive test, 98, 190, 269, 288, 306
Akaike's Information Criterion (AIC), 104
Anderson–Darling test, 129, 299
ANOVA, 268
Ansari–Bradley (AB) test, 259
asymptotic efficiency, 46
asymptotic relative efficiency (ARE), 46
asymptotically maximin test, 46
asymptotically most powerful test (AMPT), 45
asymptotically uniformly most powerful test (AUMPT), 45

Bahadur representation, 28
BAN, 12
Barton smooth model, 79, 272, 283
Bayesian information criterion (BIC), 99
best asymptotically normal, 12
bootstrap, 128, 336
boxplot, 52
Brownian bridge, 24

Capon test, 256
characteristic function, 145
Chernoff-Lehmann test, 17
circular data, 142
comparison density, 108
comparison density function, 284
comparison distribution, 29, 62, 191, 213, 273
comparison distribution for discrete data, 73
comparison distribution function, 29
comparison distribution process, 298
composite null hypothesis, 88
conditional test, 177
consistent test, 43
contingency table approach, 311, 315
continuous mapping theorem, 24
contrast process, 192, 297
covariance function, 23
Cramér-von Mises test, 130
cultivars data, 6

data-driven test, 99, 288
decomposition of the comparison density, 63
diagnostic property, 84, 86, 113, 116, 243, 245, 278
dilution effect, 97
discrete comparison density, 74
distribution free, 125
distribution function (CDF), 181
Dvoretzky–Kiefer–Wolfowitz inequality, 21

effective order, 98
efficiency, 36
empirical characteristic function (ECF), 151
empirical difference process, 28
empirical distribution function, 19
empirical distribution function (EDF), 182
empirical process, 22
empirical quantile function, 145
empirical quantile process (EQP), 146
envelope power function, 44
estimated empirical process, 127
estimated empirical quantile process, 149
estimation equation, 35
exact null distribution, 171

O. Thas, *Comparing Distributions*,
DOI 10.1007/978-0-387-92710-7,

exchangeability, 175
exponential distribution, 153

Fisher–Yates–Terry–Hoeffding test, 253
Fligner–Killeen test, 264
Fourier basis, 33
Freeman-Tukey statistic, 13
functional central limit theorem, 23

Gaussian process, 23
gene expression data, 166
Glivenko–Cantelli Theorem, 21
grade transformation, 29
Gram–Schmidt orthogonalisation, 33

Hahn orthonormal polynomial vectors, 148
Hardy-Weinberg Equilibrium, 10
Hilbert Space, 30
Hilbert space, 79, 112
histogram, 49
Hodges–Lehmann Estimator, 234

implied hypothesis, 246
improved density estimator, 107
information divergence, 80
integrated squared error (ISE), 38
interquartile range, 27
interquartile range (IQR), 53

Kac and Siegert decomposition of Gaussian processes, 24
kernel density estimation, 42
Klotz test, 256
Kolmogorov–Smirnov test, 123, 297
Kruskal–Wallis (KW) test, 265
Kruskal-Wallis test, 286
ks.test, 126
ksboot.test, 129
Kullback-Leibler, 39
KURT statistic, 278

Laplace transform, 154
Lebesgue space, 30
Legendre polynomials, 34, 276
Lehmann test, 264
Lepage test, 270
likelihood ratio test, 13
likely ordering, 197, 226, 259
lillie.test, 129
Lilliefors test, 128
locally asymptotically linear, 34, 88
locally asymptotically most powerful test (LAMPT), 46
locally most powerful linear rank test, 187
locally most powerful test (LMPT), 45
location shift model, 202
location-scale distribution, 57
location-scale family, 67
location-scale invariant distribution, 128, 140, 148, 152

M-estimator, 35
Mann–Whitney statistic, 227
Mann–Whitney test, 225
maximin most powerful test, 44
mean integrated squared error (MISE), 38
Mercer's theorem, 24
method of moments estimator (MME), 35
midrank, 181
minimum chi-square estimator, 14
minimum discrepancy estimators, 15
minimum distance estimator, 148
minimum quadratic influence function estimator (MQIFE), 115
Mood test, 262, 276, 287
most powerful test (MPT), 44

natural hypothesis, 246
Neyman smooth model, 78, 272, 283
Neyman's modified X^2 statistic, 13
Neyman–Pearson lemma, 47
nonparametric Behrens–Fisher problem, 248
nonparametric density estimation, 37
nonparametric density estimator, 50
normal scores test, 253
nuisance parameters, 127

order statistics, 20, 180
ordinal dominance curve, 29
orthogonal complement, 32
orthogonal projection, 31
orthogonal series estimator, 39, 107
outliers, 55

Parseval's relation, 80
PCB concentration data, 5
Pearson χ^2 test, 109
Pearson chisquared test, 312
Pearson's ϕ^2 measure, 79, 96, 273
Pearson-Fisher test, 12
permutation null distribution, 177
permutation test, 171
Pitman efficiency, 46
placement, 248
plotting position, 57
population comparison distributions, 62
population probability plot, 29, 56

power divergence statistic, 13
power function, 43
PP plot, 57, 124, 203
principal component decomposition, 147
principal components of a Gaussian process, 25
probabilistic index, 197
probability integral transformation (PIT), 29, 101
probability plot, 56
Pseudo-random generator data, 4
pulse rate data, 5

QQ plot, 57, 201
Quadratic Inference Function (QIF), 115
quantile function, 27, 145
quantile process, 28
quartile, 27

randomisation hypothesis, 175
randomisation test, 175
rank generating function, 185
rank score process, 186
rank statistic, 227
rank tests, 179
ranks, 180
regression constants, 182
regression-based density estimation, 52
relative data, 68
relative distribution, 62
relative distribution function, 29

sample path, 22
sample quantile, 27
sample space partition test, 155, 313
score statistic, 82
score test, 13
scores, 183
semiparametric hypotheses, 111, 244
shift function, 203
Siegel–Tukey test, 260
simple linear rank statistic, 183, 250, 276
SKEW statistic, 277
smooth model, 77, 271
Smooth Test, 77, 82, 271
spacings, 147
SSP test, 155
stochastic equivalence, 264
stochastic ordening, 203
stochastic ordering, 124, 196, 298
strip chart, 55
Sukhatme's test, 261

test function, 43, 176
Theorem of Kac and Siegert, 25
ties, 20, 181
tightness, 23
transitivity, 197
travel times, 167
two-sample t-test, 222, 224

unbiased test, 43
uniform empirical process, 22
uniformly most powerful, 187
uniformly most powerful test (UMPT), 44
unlikely ordering, 197, 199

van der Waerden test, 254
vector space, 108

Wald test, 13, 317
Wasserstein distance, 146
Watson test, 142
weak convergence, 23
Wilcoxon rank sum statistic, 228, 276
Wilcoxon rank sum test, 225
Wilcoxon–Mann–Whitney test, 228, 276

Z-estimator, 35