Proc. of the 4th Pannonian Symp. on Math. Stat., Bad Tatzmannsdorf, Austria 1983
F. Konecny, J. Mogyoródi, W. Wertz, eds

LOCALLY MOST POWERFUL TESTS FOR TWO-SIDED HYPOTHESES

U. MÜLLER-FUNK, F. PUKELSHEIM AND H. WITTING

Freiburg

SUMMARY. The theory of two-sided tests is much less developed than that of their one-sided counterparts. This paper comprises several aspects of two-sided testing problems that are largely independent of each other but are linked by the idea to make them amenable to optimization techniques by linearizations around the null-hypothesis. More precisely we strive to prove the existence of locally most powerful (LMP) unbiased two-sided tests and to exhibit their form.

Section 1 derives an extension of the Neyman-Pearson fundamental lemma for certain non-linear objective functions under linear side conditions. In particular necessary and sufficient conditions for local optimality are given. This generalizes a result of Isaacson (1951). Here we heavily rely on the methods of convex programming.

In section 2 the simplest two-sided sequential testing problem is discussed. But unlike the one-sided case, cf. Berk (1975), it becomes difficult to determine explicit expressions.

Section 3 shows that for the most important two-sided nonparametric problems the LMP unbiased rank test is asymptotically

equivalent to the intuitively obvious rank test, namely to that one based on the squared one-sided rank statistic. To this end several results on one-sided linear rank tests, cf. Hájek (1962), Hájek-Sidák (1967), are extended to the two-sided case.

In Section 0 basic facts of the theory of LMP two-sided tests, cf. Schmetterer (1966), are reviewed. This justifies the introduction of $\mathbb{L}_1$-differentiable classes of distributions, which seem for these kinds of problems more adequate than the $\mathbb{L}_2$-differentiable classes, discussed, among others, by Hájek (1962) and Le Cam (1966).

Notation: $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ denotes a k-parametric class of distributions on the sample space $(\mathfrak{X}, \mathfrak{B})$, $\dot{L}_{\theta_0}$ and $\ddot{L}_{\theta_0}$ the first and second $\mathbb{L}_1(\theta_0)$-derivative of $\mathfrak{P}$ as defined below. The point θ_0 is always assumed to be an interior point of Θ and $U(\theta_0)$ is taken to be an open and convex neighbourhood of θ_0.

$\mathbb{L}_1(\theta_0)$ denotes the set of all P_{θ_0}-integrable real valued functions on the sample space $(\mathfrak{X}, \mathfrak{B})$; here, functions which are P_{θ_0}-a.e. identical, are identified. $\mathbb{L}_1^k(\theta_0)$ and $\mathbb{L}_1^{k\times k}(\theta_0)$ are the set of k-vectors or k×k-matrices, whose elements are members of $\mathbb{L}_1(\theta_0)$. Derivatives with respect to the parameter θ at θ_0 are denoted by ∇, f.i. $\nabla f(x,\theta_0)$ or $\nabla E_{\theta_0}\varphi_i$; correspondingly, derivatives with respect to the i-th component of θ are designated by ∇_i. k-vectors are considered to be column vectors; the transposed vector of $\nabla f(x,\theta_0)$ is denoted as $\nabla^T f(x,\theta_0)$ instead of $(\nabla f(x,\theta_0))^T$. The determinant and the trace of a k×k-matrix H are abbreviated as det H and tr H , respectively.

0. Introduction: Locally optimal tests and $\mathbb{L}_1$-differentiable classes of distributions.

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a one-parameter class of distributions and θ_0 be an interior point of Θ. Suppose the power function of any test $\varphi \in \Phi$ be twice differentiable at θ_0 and let $\Phi_{\alpha,o}$ be the set of all tests $\varphi \in \Phi$ which are locally unbiased and of level α at θ_0, i.e.

$$\Phi_{\alpha,o} := \{\varphi \in \Phi : E_{\theta_o}\varphi = \alpha,\ \nabla E_{\theta_o}\varphi = 0\} \quad .$$

Then a test $\varphi^* \in \Phi_{\alpha,o}$ is called a LMP unbiased level α test for the hypotheses $\mathbb{H}: \theta = \theta_0$ against $\mathbb{K}: \theta \neq \theta_0$, if φ^* maximizes the curvature of the power function at θ_0 among all tests $\varphi \in \Phi_{\alpha,o}$, i.e.

$$\varphi^* \in \Phi_{\alpha,o} : \nabla\nabla E_{\theta_o}\varphi^* = \sup_{\varphi \in \Phi_{\alpha,o}} \nabla\nabla E_{\theta_o}\varphi \quad . \tag{0.1}$$

An important feature for the theory of LMP tests is the property that the derivative of an arbitrary power function can be taken under the integral sign, i.e. that there exists functions $\dot{L}_{\theta_o}(x)$ and $\ddot{L}_{\theta_o}(x)$ such that

$$\nabla E_{\theta_o}\varphi = E_{\theta_o}(\varphi \dot{L}_{\theta_o}), \qquad \nabla\nabla E_{\theta_o}\varphi = E_{\theta_o}(\varphi \ddot{L}_{\theta_o}) \quad .$$

In this case (0.1) is equivalent to

$$\varphi^* \in \Phi_{\alpha,o} : E_{\theta_o}(\varphi^* \ddot{L}_{\theta_o}) = \sup_{\varphi \in \Phi_{\alpha,o}} E_{\theta_o}(\varphi \ddot{L}_{\theta_o}) \ . \tag{0.2}$$

For this optimization problem we are going to use the following short-hand notation

$$E_{\theta_o}\varphi = \alpha \tag{0.3}$$

$$E_{\theta_o}(\varphi \dot{L}_{\theta_o}) = 0 \tag{0.4}$$

$$E_{\theta_o}(\varphi \ddot{L}_{\theta_o}) = \sup. \tag{0.5}$$

A solution always exists and can be given by the fundamental lemma with constants c_0, $c_1 \in \mathbb{R}$ in the form

$$\varphi^*(x) = \begin{cases} 1 \text{ if } \ddot{L}_{\theta_0}(x) > c_1 L_{\theta_0}(x) + c_0 \\ 0 \text{ if } \ddot{L}_{\theta_0}(x) < c_1 L_{\theta_0}(x) + c_0 \end{cases} [P_{\theta_0}], \qquad (0.6)$$

provided $\alpha \in (0,1)$.

The notion of a LMP test and its programming approach along the lines of (0.3-6) is not restricted to one-parameter classes, and therefore we shall develope from the very beginning the following theory for k-parameter classes.

To introduce an appropriate concept of differentiation let $L_{\theta_0}(\cdot,\theta)$ denote the likelihood ratio of P_θ with respect to P_{θ_0}, i.e. $L_{\theta_0}(\cdot,\theta)<\infty$ P_{θ_0}-a.e. and $L_{\theta_0}(\cdot,\theta)$ solves the Lebesgue-decomposition

$$P_\theta(B) = \int_B L_{\theta_0}(x,\theta)dP_{\theta_0} + P_\theta(B \cap \{L_{\theta_0}(\cdot,\theta) = \infty\}) \forall B \in \mathfrak{B} \quad .$$

Such a function $L_{\theta_0}(x,\theta)$ always exists, is $P_{\theta_0}+P_\theta$-unique and an element of $\mathbb{L}_1(\theta_0)$. In the special case of a dominated class $\mathfrak{P}$ with densities $f(x,\theta)$, $\theta \in \Theta$, the likelihood ratio can be chosen as

$$L_{\theta_0}(x,\theta) = \frac{f(x,\theta)}{f(x,\theta_0)} 1_{\{f(\cdot,\theta_0)>0\}}(x) + \infty 1_{\{f(\cdot,\theta_0)=0\}}(x) \quad [P_{\theta_0}+P_\theta] \ .$$

<u>Definition</u>. Let $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ be a k-parameter class of distributions, θ_0 an interior point of Θ and $L_{\theta_0}(\cdot,\theta) \in \mathbb{L}_1(\theta_0)$ the likelihood ratio of P_θ relative to P_{θ_0}.

1) $\mathfrak{P}$ is said to be $\mathbb{L}_1(\theta_0)$-differentiable with derivative $\dot{L}_{\theta_0}(\cdot)$ if $\dot{L}_{\theta_0}(\cdot) \in \mathbb{L}_1^k(\theta_0)$ such that

$$||L_{\theta_0}(\cdot,\theta) - 1 - (\theta-\theta_0)^T \dot{L}_{\theta_0}(\cdot)||_{\mathbb{L}_1(\theta_0)} = o(|\theta-\theta_0|) \text{ for } \theta \to \theta_0. \quad (0.7)$$

2) $\mathfrak{P}$ is said to be twice $\mathbb{L}_1(\theta_0)$-differentiable with second derivative $\ddot{L}_{\theta_0}(\cdot)$, if $\mathfrak{P}$ is $\mathbb{L}_1(\theta)$-differentiable for all $\theta \in U(\theta_0)$ and if $\ddot{L}_{\theta_0}(\cdot) \in \mathbb{L}_1^{k\times k}(\theta_0)$ such that

$$||\dot{L}_\theta(\cdot)L_{\theta_0}(\cdot,\theta)-\dot{L}_{\theta_0}(\cdot)-(\theta-\theta_0)^T\ddot{L}_{\theta_0}(\cdot)||_{\mathbb{L}_1^k(\theta_0)} = o(|\theta-\theta_0|) \text{ for } \theta\to\theta_0 \quad (0.8)$$

and

$$\int_{\{L_{\theta_0}(\cdot,\theta)=\infty\}} |\dot{L}_\theta(x)|dP_\theta = o(|\theta-\theta_0|) \quad \text{for } \theta\to\theta_0 \quad (0.9)$$

Remark.

1) It can be shown that (0.7) implies the following behavior of the singular parts

$$P_\theta(\{L_{\theta_0}(\cdot,\theta) = \infty\}) = o(|\theta-\theta_0|) \qquad \text{for } \theta\to\theta_0 \quad (0.10)$$

On the contrary the behavior of the singular parts is demanded separately in condition (0.9).

2) Notice that we did not assume that P_{θ_0} dominates P_θ nor even that P_{θ_0} and P_θ are absolutely continuous.

We now formulate some basic facts on $\mathbb{L}_1(\theta_0)$-differentiable classes, the proofs of which can be found in Witting (1985), section 1.8.

Theorem 1: Let $\mathfrak{P} = \{P_\theta : \theta\in\Theta\}$ be a k-parameter class of distributions.

a) If $\mathfrak{P}$ is $\mathbb{L}_1(\theta_0)$-differentiable with derivative $\dot{L}_{\theta_0}$, then

$$\forall\varphi \in \Phi: \quad \nabla E_{\theta_0}\varphi = E_{\theta_0}(\varphi\dot{L}_{\theta_0}) .$$

b) If $\mathfrak{P}$ is twice $\mathbb{L}_1(\theta_0)$-differentiable with second derivative $\ddot{L}_{\theta_0}$, then

$$\forall\varphi \in \Phi: \quad \nabla\nabla^T E_{\theta_0}\varphi = E_{\theta_0}(\varphi\ddot{L}_{\theta_0}) .$$

Theorem 2: Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be dominated by a σ-finite measure μ with μ-densities $f(x,\theta)$

a) If there exists a μ-null set $N \in \mathcal{B}$ such that for all $\theta \in U(\theta_o)$

1) $f(x,\theta)>0$ and $\nabla f(x,\theta)$ exists and is continuous in θ for all $x \in N^c$,

2) $\nabla f(\cdot,\theta) \in \mathbb{L}_1^k(\mu)$,

3) $\int |\nabla_i f(x,\theta)| d\mu \to \int |\nabla_i f(x,\theta_o)| d\mu$ for $\theta \to \theta_o$ and $i=1,\dots,k$,

then $\mathcal{P}$ is $\mathbb{L}_1(\theta_o)$-differentiable with derivative

$$\dot{L}_{\theta_o}(x) = \nabla f(x,\theta_o)/f(x,\theta_o) = \nabla \log f(x,\theta_o) \quad [P_{\theta_o}] \quad .$$

b) If the conditions of a) are fulfilled for all $\theta \in U(\theta_o)$ and if in addition for all $\theta \in U(\theta_o)$

4) $\nabla\nabla^T f(x,\theta)$ exists and is continuous in θ for all $x \in N^c$,

5) $\nabla\nabla^T f(\cdot,\theta) \in \mathbb{L}_1^{k\times k}(\mu)$,

6) $\int |\nabla_i \nabla_j f(x,\theta)| d\mu \to \int |\nabla_i \nabla_j f(x,\theta_o)| d\mu$ for $\theta \to \theta_o$ and $i,j=1,\dots,k$,

then $\mathcal{P}$ is twice $\mathbb{L}_1(\theta_o)$-differentiable with second derivative

$$\ddot{L}_{\theta_o}(x) = \nabla\nabla^T f(x,\theta_o)/f(x,\theta_o) = \nabla\nabla^T \log f(x,\theta_o) +$$

$$+ (\nabla \log f(x,\theta_o))\ (\nabla^T \log f(x,\theta_o)) \quad [P_{\theta_o}]$$

Theorem 3: For $i=1,\dots,n$ let $\mathcal{P}_i = \{P_{i,\theta} : \theta \in \Theta\}$ be k-parameter classes of distributions with the same parameter set Θ. Then $\mathcal{P} = \{\bigotimes_{i=1}^n P_{i,\theta} : \theta \in \Theta\}$ is twice differentiable with first and second derivative

$$\dot{L}_{\theta_o}(x) = \sum_{i=1}^n \dot{L}_{i,\theta_o}(x_i) \quad [P_{\theta_o}], \tag{0.11}$$

$$\ddot{L}_{\theta_o}(x) = \sum_{i=1}^n \ddot{L}_{i,\theta_o}(x_i) + \sum\sum_{1\le i\ne j\le n} \dot{L}_{i,\theta_o}(x_i)\dot{L}_{j,\theta_o}(x_j)[P_{\theta_o}], \tag{0.12}$$

provided the factor classes are once or twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives $\dot{L}_{i,\theta_o}(x_i)$ or $\ddot{L}_{i,\theta_o}(x_i)$, respectively.

A prominent example is a k-parameter exponential family in θ and $T(x)$ with derivatives

$$\dot{L}_{\theta_o}(x) = T(x) - E_{\theta_o}T \quad [P_{\theta_o}],$$

$$\ddot{L}_{\theta_o}(x) = (T(x) - E_{\theta_o}T)(T(x)-E_{\theta_o}T)^T - Cov_{\theta_o}T \quad [P_{\theta_o}],$$

where $Cov_{\theta_o}T$ denotes the covariance-matrix of T under P_{θ_o}.

Similarly, a translation family $\mathfrak{P}$ is twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives

$$\dot{L}_{\theta_o}(x) = -\frac{f'(x-\theta_o)}{f(x-\theta_o)}[P_{\theta_o}],\ \ddot{L}_{\theta_o}(x) = \frac{f''(x-\theta_o)}{f(x-\theta_o)}[P_{\theta_o}],$$

if $\mathfrak{P}$ is generated by a Lebesgue-density f such that $f>0$, f twice differentiable with first and second derivative f' and f'', respectively, and $\int|f'(x)|d\lambda<\infty$ and $\int|f''(x)|d\lambda<\infty$.

For the nonparametric testproblems we are particularly interested in regression families $\mathfrak{P} = \{\bigotimes_{i=1}^{n} F_{\theta_o+\eta d_i} : \eta\in U(0)\subset\mathbb{R}\}$, where $\mathfrak{F} = \{F_\theta : \theta\in\Theta\}$ is a one-parameter class and $d_1,\dots,d_n$ are given regression coefficients. If $\mathfrak{F}$ is twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives $\dot{L}_{\theta_o}(x)$ and $\ddot{L}_{\theta_o}(x)$, then $\mathfrak{P}$ is also twice $\mathbb{L}_1(0)$-differentiable with derivatives

$$\dot{L}_o(x) = \sum_{i=1}^{n} d_i\dot{L}_{\theta_o}(x_i) \quad [P_{\theta_o}], \tag{0.13}$$

$$\ddot{L}_o(x) = \sum_{i=1}^{n} d_i^2\ddot{L}_{\theta_o}(x_i) + \sum\sum_{1\le i\ne j\le n} d_i d_j \dot{L}_{\theta_o}(x_i)\dot{L}_{\theta_o}(x_j) \ [P_{\theta_o}].$$

For the asymptotic theory the following transformation of $\ddot{L}_0(x)$ turns out to be useful

$$\ddot{L}_0(x) = \sum_{i=1}^{n} d_i^2\left[\ddot{L}_0(x_i) - \dot{L}_0^2(x_i)\right] + \left(\sum_{i=1}^{n} d_i\dot{L}_0(x_i)\right)^2 \quad [P_{\theta_0}] \tag{0.14}$$

Under Noether-type conditions on the regression coefficients the second term in (0.14) asymptotically dominates the first. This is the reason why the two-sided test (0.6) has a test statistic which is asymptotically χ^2-distributed, provided the Fisher information $E_{\theta_0}(\dot{L}_{\theta_0})^2$ is positive and finite.

1. Locally optimal test in k-parameter classes for two-sided hypotheses. Let $\mathfrak{P} = \{P_\theta : \theta \in \Theta\}$ be a k-parameter class of distributions, which is twice $\mathbb{L}_1(\theta_0)$-differentiable with first and second derivative $\dot{L}_{\theta_0}$ and $\ddot{L}_{\theta_0}$ respectively. Then an arbitrary powerfunction is twice differentiable, and so an LMP unbiased level α test φ^* for $\mathbb{H}: \theta=\theta_0$ against $\mathbb{K}: \theta\neq\theta_0$ is defined as a solution of

$$E_{\theta_0}\varphi = \alpha, \tag{1.1}$$

$$E_{\theta_0}(\varphi\dot{L}_{\theta_0}) = 0, \tag{1.2}$$

$$H(\varphi) := E_{\theta_0}(\varphi\ddot{L}_{\theta_0}) \geq 0, \tag{1.3}$$

$$\det H(\varphi) = \sup. \tag{1.4}$$

The determinant here serves as a scalar measure of curvature and is related to the Gauß curvature of the powerfunction, cf Isaacson (1951). The condition (1.3) means that $H(\varphi)$ is non--negative definite and secures that the curvature "goes upwards".

Theorem 4: There always exists a LMP unbiased level α test φ^*.

Proof: This is the usual weak compactness argument. □

Theorem 4 also follows from the following typical two step argument. The set of matrices

$$\mathfrak{H} := H(\Phi_{\alpha,o,\geq}) ,$$

$$\Phi_{\alpha,o,\geq} := \{\varphi\in\Phi : E_{\theta_o}\varphi=\alpha,\ E_{\theta_o}(\varphi\dot{L}_{\theta_o})=0,\ H(\varphi)\geq 0\}$$

can easily seen to be convex and compact. Hence the continuous function $H\to\det H$ assumes a maximum on $\mathfrak{H}$ at H^*, say, and so we can find a test φ^* with $H(\varphi^*) = H^*$.

We now turn to the problem of determining the form of a LMP unbiased test. The idea is to replace the given optimization problem with the non-linear objective function $H\to\det H$ by a quasi-linear optimization problem.

The matrix part of our original problem is of the form

$$\det{}^{1/k} H^* = \sup_{H\in\mathfrak{H}} \det{}^{1/k} H, \qquad (1.5)$$

where $\mathfrak{H}$ is a convex and compact set of non-negative definite $k\times k$-matrices. The dual problem turns out to be of the form

$$\frac{1}{k\,\det^{1/k} G^*} = \inf_{G\in\mathfrak{G}} \frac{1}{k\,\det^{1/k} G} , \qquad (1.6)$$

where $\mathfrak{G}$ is the set of all non-negative definite $k\times k$-matrices G satisfying

$$\operatorname{tr}(GH) \leq 1 \qquad \forall H \in \mathfrak{H} . \qquad (1.7)$$

Theorem 5: If $H \in \mathfrak{h}$ and $G \in \mathfrak{g}$, then

$$\det{}^{1/k} H \le \frac{1}{k \det^{1/k} G} \tag{1.8}$$

and equality holds if and only if

$$H \text{ positive definite} \quad \text{and} \quad G = H^{-1}/k. \tag{1.9}$$

Proof: a) $\det^{1/k}(GH)$ is the geometric mean of the eigenvalues of GH, and therefore less than or equal to the arithmetic mean

$$\det{}^{1/k}(GH) \le \frac{1}{k} \operatorname{tr}(GH) \le \frac{1}{k} \quad . \tag{1.10}$$

The last inequality follows from the definition of $\mathfrak{g}$. The conditions on equality are based on the fact that all eigenvalues be equal. □

It is a simple consequence from (1.8) that

$$\max_{H \in \mathfrak{h}} \det{}^{1/k} H \le \min_{G \in \mathfrak{g}} \frac{1}{k \det^{1/k} G} \quad .$$

However, we even have equality.

Theorem 6: (Duality) If there exists a test φ with $E_{\theta_o} \varphi = \alpha$, $E_{\theta_o}(\varphi \dot{L}_{\theta_o}) = 0$, $E_{\theta_o}(\varphi \ddot{L}_{\theta_o}) > 0$, then

$$\max_{H \in \mathfrak{h}} \det{}^{1/k} H = \min_{G \in \mathfrak{g}} \frac{1}{k \det^{1/k} G} \quad . \tag{1.11}$$

This is a special case of Theorem 4 in Pukelsheim (1980). Indeed, as in the experimental design context, we can choose functions j on $\mathfrak{h}$ other than det as a scalar measure of curvature. The duality theorem stated above will carry over provided

a) j is positive on the set of positive definite $k \times k$-matrices and upper semicontinuous on the set of non-negative definite $k \times k$-matrices,

b) j is positive homogeneous,

c) j is concave.

In particular, the generalized means of order p, given by $j_p(H) := [\frac{1}{k} \operatorname{tr}(H^p)]^{1/p}$, lead to the dual objective function $1/[kj_q(G)]$, where $p,q \in (-\infty,1)$ and $pq = p + q$. If $H \in \mathfrak{H}$ and $G \in \mathfrak{G}$, then

$$j_p(H) \leq \frac{1}{kj_q(G)} ,$$

with equality if and only if

$$H \text{ positive definite} \quad \text{and} \quad G = H^{p-1}/k . \tag{1.12}$$

For the sake of clarity we continue with the simple case of D-optimality, i.e. with $j(H) = \det^{1/k} H$, which is the limiting case of $j_p(H)$ for $p \to 0$. The duality theorem 6 implies the following linearization of our original problem:

<u>Theorem 7</u>: (Equivalence) If there exists a test $\tilde{\varphi}$ with $E_{\theta_o}\tilde{\varphi} = \alpha$, $E_{\theta_o}(\tilde{\varphi}\dot{L}_{\theta_o}) = 0$, $E_{\theta_o}(\tilde{\varphi}\ddot{L}_{\theta_o}) > 0$, then:

a) The following two statements are equivalent:

   1) $H^* \in \mathfrak{H}$ maximizes $H \to \det^{1/k} H$ for $H \in \mathfrak{H}$.

   2) $H^* \in \mathfrak{H}$ is positive definite and H^* maximizes $H \to \operatorname{tr}(H\, H^{*-1})$ for $H \in \mathfrak{H}$.

b) The following three statements are equivalent:

   1) $\varphi \in \Phi_{\alpha,o,\geq}$ maximizes $\varphi \to \det^{1/k} H(\varphi)$ for $\varphi \in \Phi_{\alpha,o,\geq}$.

   2) $H(\varphi^*)$ is positive definite and φ^* maximizes $\varphi \to E_{\theta_o}[\varphi \operatorname{tr}(\ddot{L}_{\theta_o} H(\varphi^*)^{-1})]$.

   3) $H(\varphi^*)$ is positive definite and there exist $c_o \in \mathbb{R}$, $c \in \mathbb{R}^k$ such that

$$\varphi^*(x) = \begin{cases} 1 & \text{if } \operatorname{tr}[\ddot{L}_{\theta_o}(x)H(\varphi^*)^{-1}] > c^T\dot{L}_{\theta_o}(x) + c_o \\ 0 & \text{if } \operatorname{tr}[\ddot{L}_{\theta_o}(x)H(\varphi^*)^{-1}] < c^T\dot{L}_{\theta_o}(x) + c_o \end{cases} \quad [P_{\theta_o}],$$

$$E_{\theta_o}\varphi^* = \alpha, \quad E_{\theta_o}[\varphi^*\dot{L}_{\theta_o}] = 0.$$

Proof: a) $H^* \in \mathfrak{H}$ is optimal if and only if there exists some matrix $G^* \in \mathfrak{G}$ such that $\det^{1/k}H^* = \{k \det^{1/k}G^*\}^{-1}$. By Theorem 5 this is equivalent to the positive definiteness of H^* and $G^* = H^{*-1}/k$. According to the definition of $\mathfrak{G}$ thus $\operatorname{tr}(H\,H^{*-1}) \leq k \quad \forall H \in \mathfrak{H}$.

b) The equivalence of 1) and 2) follows from part a). Since 2) involves the maximization of a linear functional an appropriate form of the Neyman Pearson lemma yields the equivalence of 2) and 3). Notice, that initially we have Lagrange multipliers $c_o \in \mathbb{R}$, $c \in \mathbb{R}^k$ and a non-negative definite $k\times k$-matrix C; however, it turns out that for an optimal test φ^* we can choose $C=0$.□

Sufficiency of Condition b3) in families with differentiable densities is due to Isaacson (1951).

As an example consider a k-parameter exponential family in θ and $T(x)$. Then φ^* is LMP unbiased level α if and only if there exist $c_o \in \mathbb{R}$ and $a \in \mathbb{R}^k$ such that

$$\varphi^*(x) = \begin{cases} 1 & \text{if } (T(x)-a)^T H(\varphi^*)^{-1}(T(x)-a) > c \\ 0 & \text{if } (T(x)-a)^T H(\varphi^*)^{-1}(T(x)-a) < c \end{cases} \quad [P_{\theta_o}],$$

$$E_{\theta_o}\varphi^* = \alpha, \quad E_{\theta_o}(\varphi^*\dot{L}_{\theta_o}) = 0.$$

2. Locally optimal two-sided sequential tests in one-parametric classes.

The by far most important test in sequential analysis is the Sequential Probability Ratio Test (SPRT). Though basically a test for two simple hypotheses, it is usually applied to composite hypotheses. To treat this type of testing problems right from the beginning one can take recourse to a local approach. For one-parameter families and for one-sided hypotheses Berk (1975) has shown that a LMP-test exists and takes on the form of a SPRT. Therefore the question arises whether also a two-sided LMP unbiased test exists and to what extent it can be made explicit. To be more precise let us assume an i.i.d. sequence of basic observations the common distribution of which belongs to a one-parametric family $\{P_\theta : \theta \in \Theta \subset \mathbb{R}\}$. Suppose that this class is twice $\mathbb{L}_1(\theta_0)$-differentiable and that $0 < I := E_{\theta_0}(\dot{L}^2_{\theta_0}) < \infty$, $J := E_{\theta_0}|\ddot{L}_{\theta_0}| < \infty$. We are going to look for a solution of the following problem

$$E_{\theta_0}(\varphi_N \ddot{L}_N) = \sup, \tag{2.1}$$

$$E_{\theta_0}(\varphi_N \dot{L}_N) = 0, \tag{2.2}$$

$$E_{\theta_0}(\varphi_N) = \alpha, \tag{2.3}$$

$$E_{\theta_0}(N^\kappa) \leq \gamma, \tag{2.4}$$

where $\kappa \geq 1$ is some fixed constant. As before the local character of the problem is indicated by the fact that all conditions only involve the parameter point θ_0. Display (2.1-3) are exactly those of the corresponding non-sequential problem (0.3-5). (We shall not enter into a discussion of the regularity conditions required for the twice differentiability of the power curve of any sequential test under the expectation.) As for (2.4) it is intuitively clear that some restriction on the size of the stopping time N is needed in order to prevent the optimal procedure from sampling infinitely long. Note, that Wald's equations show that the supremum appearing in (2.1) lies between 0 and $\gamma(2I+J)$.

If we allow for randomized stopping rules then we can ensure the mere existence of an optimal solution by a simple compactness argument. In order to get a clue of what an optimal test might look like, it is near at hand to turn the above problem into an unconstrained optimal stopping problem and to try standard techniques. The natural way to do this is to write down a corresponding dual problem. Formally this can be done by introducing Lagrange multipliers c_1, $c_2 \in \mathbb{R}$, $c_3 > 0$. (The case $c_3=0$ can be immediately excluded from our considerations.) According to that, we proceed from the primal to the dual objective function by the following chain of inequalities:

$$\begin{aligned} E_{\theta_o}(\varphi_N \ddot{L}_N) &\leq E_{\theta_o}(\varphi_N \{\ddot{L}_N - c_1 \dot{L}_N - c_2\}) - c_3 E_{\theta_o} N^K + c_2\alpha + c_3\gamma \\ &\leq E_{\theta_o}[\{\ddot{L}_N - c_1 \dot{L}_N - c_2\}^+ - c_3 N^K] + c_2\alpha + c_3\gamma \qquad (2.5) \\ &\leq \sup_M E_{\theta_o} Z_M + c_2\alpha + c_3\gamma , \end{aligned}$$

where for all $n \in \mathbb{N}$

$$Z_n = \{\ddot{L}_n - c_1 \dot{L}_n - c_2\}^+ - c_3 n^K = (S_{n1}^2 + S_{n2} - c_1 S_{n1} - c_2)^+ - c_3 n^K,$$

$$S_{n1} = \dot{L}_n = \sum_{i=1}^n \dot{L}_{\theta_o}(x_i), \quad S_{n2} = \sum_{i=1}^n \{\ddot{L}_{\theta_o}(x_i) - \dot{L}_{\theta_o}^2(x_i)\} .$$

Therefore the new objective function essentially consists of the value v of the optimal stopping problem

$$\sup_M E_{\theta_o} Z_M = v,$$

where the supremum is taken over the class $\mathfrak{M}$ of all stopping variables M such that $P_{\theta_o}(M<\infty) = 1$ and $E_{\theta_o}(Z_M^-)<\infty$.

It is implicit in our approach that we can restrict attention to the terminal decision rules

$$\varphi_n = \begin{cases} 1 & \text{if } \ddot{L}_n - c_1\dot{L}_n - c_2 > 0 \\ 0 & \text{if } \ddot{L}_n - c_1\dot{L}_n - c_2 < 0 \end{cases} \quad [P_{\theta_o}],\ n\in\mathbb{N}. \qquad (2.6)$$

Now, if there is an optimal $M^*=M_c^*$ whith $E_{\theta_o} M_c^* < \infty$, $c = (c_1,c_2,c_3)$, we can try to determine c in such a way that the side conditions (2.2-4) are fulfilled. The ensueing sequential test $(M_c^*, (\varphi_{mc}^*))$ will solve the original problem (2.1-4). The reason is that once we fix a stopping rule M_c^* and restrict attention to the σ-algebra of events prior to M_c^*, then we are essentially back in the non-sequential case and can discuss the equality of the objective functions in the usual way (cf Witting (1984), section 2.4). (It is common practice in sequential optimization problems, however, not to bother about the way in which c is to be determined but rather go the other way round, i.e. to make α etc. depend on c.)

Next let us enter into the above stopping problem.

Lemma 8. a) If $\kappa = 1$ then $v = \infty$ irrespective of $c\in\mathbb{R}^2\times(0,\infty)$.
b) Suppose that $E|(\dot{L}_{\theta_o}(x))|^{2\kappa/(\kappa-1)} < \infty$, $\kappa>1$. Then

$$v \le E_{\theta_o}(\sup_{m\ge 1} Z_m^+) < \infty \quad \forall c\in\mathbb{R}^2\times(0,\infty). \qquad (2.7)$$

Proof: a) It can be drawn from Rootzén (1976) that $m^{-1}Z_m$ lies dense in $[-I-c_3,\infty)$ and hence that $M(a) = \inf\{m\ge1 : Z_m\ge a\}$ is a finite stopping time for every $a>0$. As $E_{\theta_o} Z_{M(a)}^- = 0$ and $E_{\theta_o} Z_{M(a)} \ge a$ it follows that $v=\infty$.

b) This assertion is a consequence of well known results. In fact for every $0<b\leq c_3$

$$E_{\theta_o}(\sup_{m\geq 1} Z_m^+) \geq E_{\theta_o}(\sup_{m\geq 1}(S_{m1}^2 - bm^\kappa)^+) + E_{\theta_o}(\sup_{m\geq 1}(S_{m2} - c_1 S_{m1})^+) + |c_2|.$$

The second term on the RHS is finite, cf Chow-Robbins-Siegmund (1971), p. 92. As for the first one we apply Lemma 1 in Chow-Teicher (1978), p. 364, with $p = (\kappa+1)/\kappa$, which yields

$$E_{\theta_o}(\sup_{m\geq 1}(S_{m1}^2-bm^\kappa)^+)\leq b\kappa 2^\kappa \sum_{n\geq 1} n^{\kappa-1}\{P_{\theta_o}(\max_{1\leq m\leq n} S_{m1}^+ > \frac{b}{4} n^{\kappa/2}) +$$

$$+ P_{\theta_o}(\max_{1\leq m\leq n}(-S_{m1})^+ > \frac{b}{4} n^{\kappa/2})\} .$$

Employing a Baum-Katz-type inequality, cf Chow-Teicher (1978), p. 362, we immediately arrive at the result. □

(2.7) is the basic condition underlying the theory of optimal stopping. We shall indicate below how to go around the difficulties in case $\kappa = 1$, where this assumption is apparently violated.

To get near the optimal solution it is fortunate that the reward sequence $(Z_m)_{m\geq 1}$ allows for a stationary Markov representation. To realize this put $d = c_1^2/4 + c_2$ and

$$f(s_1,s_2,s_3) = (s_1^2 + s_2 - d)^+ - c_3 s_3^\kappa, \quad (s_1,s_2,s_3)\in\mathbb{R}^2\times\mathbb{N}.$$

The sequence $(S_{n1}, S_{n2}, n)_{n\in\mathbb{N}}$ together with some starting point (s_1, s_2, s_3) forms a homogeneous Markov chain and

$$f(S_{n1} - c_1/2, S_{n2} + 0, n + 0) = Z_n, \quad n\in\mathbb{N} .$$

To solve the stopping problem we have to compute the smallest regular excessive majorant g of f and to look at the stopping variable

$$M^* = \inf\{m\geq 1: f(S_{m1}+s_1,\ S_{m2}+s_2,\ m+s_3) = $$
$$= g(S_{m1}+ s_1,\ S_{m2} + s_2,\ m+s_3)\} \ .$$

Because the reward sequence tends to $-\infty$ it is well known that M^* is finite P_{θ_o}-a.s. and hence optimal for $\kappa>1$, cf Chow-Robbins-Siegmund (1971), theorem 4.5., p. 70 and chapter 5.1 or Shiryayev (1978), theorem 8, p. 74. Moreover, $E_{\theta_o} M^*$ is finite, too, as can be seen with the help of Lemma 8. Even in the case of an exponential family, however, it seems very unlikely that one can obtain a concise formula describing the stopping region.

Remark: In the case $\kappa=1$ the above representation can be somewhat simplified. In fact it suffices to look at the chain (S_{n1}, S_{n2}) and the function $f_1(s_1,s_2)=(s_1^2+s_2-d)^+$, respectively. Here we get the representation

$$f_1(S_{n1} - c_1/2,\ S_{n2} + 0) - c_3 n = Z_n, \qquad n\in\mathbb{N} \ ,$$

and we have to get hold of the smallest regular c-excessive majorant g_1 instead. Of course the stopping rule takes on the form

$$M^* = \inf\{m\geq 1: f_1(S_{m1}+s_1,\ S_{m2} + s_2) = g_1(S_{m1} +s_1,\ S_{m2} + s_2)\}.$$

Clearly

$$\sup\{E_{\theta_o}(f_1(S_{M1}+s_1, S_{M2}+s_2)\wedge a - c_3 M): M\in\mathfrak{M}\} \to v \quad \text{if } a\uparrow\infty \ .$$

It can be shown that the family of smallest c-excessive majorants ${}_a g$ corresponding to $f_1\wedge a$ converges towards a function g_1 which is the smallest regular c-excessive majorant of f_1. Unfortunately we can not show that the stopping variables ${}_aM^*$ solving the truncated problem tend to M^*, but in any case $E_{\theta_o}({}_aM^*)\to v$.

3. Locally optimal two-sided rank-tests. Let $X_1,\ldots X_n$ be real valued independent random variables with continuous distribution functions $F_1,\ldots,F_n$ and let us consider the problem of testing randomness against non-randomness

$$\mathbb{H}: F_i = F_j \quad \forall i \neq j, \qquad \mathbb{K}: F_i \neq F_j \quad \exists i \neq j. \tag{3.1}$$

A rank test is defined as a test which factorizes with respect to the rank statistic $R_n(x)=(R_{n1}(x),\ldots,R_{nn}(x))$, where $x=(x_1,\ldots,x_n)$ and $R_{ni}(x)$ denotes the number of x_j, $j=1,\ldots,n$, which are less than or equal to x_i.

According to the theory of one-sided rank tests, cf Hájek (1962), Hájek-Sidák (1967), Witting-Nölle (1970), one is interested in a rank test $\varphi^*(x)=\psi^*(R_n(x))$, which is LMP unbiased along a one-parameter class of regression alternatives

$$\mathcal{P} = \left\{ \bigotimes_{i=1}^{n} F_{\theta_o+\eta d_{ni}} \; : \; \eta \in U(0) \subset \mathbb{R} \right\}$$

among all rank tests. Here $P_\eta = \bigotimes_{i=1}^{n} F_{\theta_o+\eta d_{ni}}$ is the joint distribution of $X_1,\ldots,X_n$ such that $P_\eta \in \mathbb{H}$ if and only if $\eta=0$. Hence, $P_o = F_{\theta_o}^{(n)}$ where F_{θ_o} is a continuous distribution function.

The determination of φ^* is equivalent with that of ψ^* as a LMP unbiased test for $\tilde{\mathbb{H}}: \eta=0$ against $\tilde{\mathbb{K}}: \eta\neq 0$ underlying $\mathcal{P}^{R_n} = \{P_\eta^{R_n}: \eta \in U(0) \subset \mathbb{R}\}$. Therefore, according to (0.1-3) we need the first and the second $\mathbb{L}_1(0)$-derivative of $\mathcal{P}^{R_n}$.

Theorem 9: Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a k-parameter class of distributions on $(\mathcal{X},\mathcal{B})$ which is twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives $\dot{L}_{\theta_o}(x)$ and $\ddot{L}_{\theta_o}(x)$, and let $R:(\mathcal{X},\mathcal{B}) \to (\mathcal{R},\mathcal{T})$ be an arbitrary statistic. Then, the class of distributions $\mathcal{P}^R=\{P^R_\theta:\theta\in\Theta\}$ is twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives

$$\dot{\ell}_{\theta_o}(r) = E_{\theta_o}(\dot{L}_{\theta_o}(X)\,|\,R = r) \qquad [P^R_{\theta_o}],$$

$$\ddot{\ell}_{\theta_o}(r) = E_{\theta_o}(\ddot{L}_{\theta_o}(X)\,|\,R = r) \qquad [P^R_{\theta_o}]. \tag{3.2}$$

The proof of this theorem can also be found in Witting (1985), section 1.8.

For deriving ψ^* by means of Theorem 9 it is convenient to replace $X_1,\ldots,X_n$ by $F^{-1}(U_1),\ldots,F^{-1}(U_n)$, where F^{-1} is the generalized inverse of $F = F_{\theta_o}$ and $U_1,\ldots,U_n$ are independent rectangular $(0,1)$-distributed random variables. Let $U_{n\uparrow 1},\ldots,U_{n\uparrow n}$ denote the corresponding ordered random variables,

$$a(u) := \ddot{L}_{\theta_o}(F^{-1}(u)), \quad b(u) := \dot{L}_{\theta_o}(F^{-1}(u)), \quad B(u,v) := b(u)b(v) \tag{3.3}$$

the score generating functions,

$$a_{ni} := Ea(U_{n\uparrow i}), \quad b_{ni} := Eb(U_{n\uparrow i}), \quad B_{nij} := Eb(U_{n\uparrow i}, U_{n\uparrow j}) \tag{3.4}$$

the corresponding scores and

$$a_n(u) := \sum_{i=1}^{n} a_{ni} 1_{\left[\frac{i-1}{n},\frac{i}{n}\right)}(u), \quad b_n(u) := \sum_{i=1}^{n} b_{ni} 1_{\left[\frac{i-1}{n},\frac{i}{n}\right)}(u)$$

$$B_n(u,v) := \sum_{i=1}^{n}\sum_{j=1}^{n} B_{nij} 1_{\left[\frac{i-1}{n},\frac{i}{n}\right)\times\left[\frac{j-1}{n},\frac{j}{n}\right)}(u,v) \tag{3.5}$$

the step functions defined by these scores.

Hence, all rank statistics, in particular those of the form

$$S_n(r)= \sum_{i=1}^{n} d_{ni}^2 a_{nr_{ni}} + \sum\sum_{1\le i\neq j\le n} d_{ni}d_{nj}B_{nr_{ni}r_{nj}} + c_1 \sum_{i=1}^{n} d_{ni}b_{nr_{ni}} + c_0 \quad (3.6)$$

with fixed constants c_0, $c_1 \in \mathbb{R}$, are distribution free on $\mathbb{H}$.

Theorem 10: Let $\mathfrak{F}=\{F_\theta:\theta\in\Theta\}$ be a one-parameter class of distributions, which is twice $\mathbb{L}_1(\theta_o)$-differentiable with derivatives $\dot{L}_{\theta_o}$ and $\ddot{L}_{\theta_o}$. Let $d_{n1},\dots,d_{nn}$ be given regression coefficients and $\alpha\in(0,1)$. Then there exists a level α rank test φ^* which is LMP unbiased along the one-parameter subclass

$\mathfrak{K}=\left\{\bigotimes_{i=1}^{n} F_{\theta_o+\eta d_{ni}}:\eta\in U(0)\subset\mathbb{R}\right\}$. It holds

$$\varphi^*(x)=\psi^*(R_n(x)),\ \psi^*(r)=\begin{cases}1 & \text{if}\ S_n(r)>0\\ 0 & \text{if}\ S_n(r)<0\end{cases} \quad (3.7)$$

with S_n according to (3.6), where $c_1,c_0\in\mathbb{R}$ are determined such that

$$E_{\theta_o}\psi^* = \alpha\ , \qquad E_{\theta_o}(\psi^*\dot{\ell}_{\theta_o}) = 0 \quad (3.8)$$

and $\dot{\ell}_{\theta_o}$ is given by (3.2).

Proof: According to Theorem 9 and (0.11) the class $\mathfrak{K}^{R_n}$ is twice $\mathbb{L}_1(0)$-differentiable with derivatives (3.2). We recall that for i.i.d. random variables $X_1,\dots,X_n$ with continuous distribution function F it holds true that:

1) $X_i = X_{n\uparrow R_{ni}(X)}$ $F^{(n)}$-a.s., $i=1,\dots,n$.

2) The order statistic $M_n(X) = (X_{n\uparrow 1},\dots,X_{n\uparrow n})$ and the rank statistic $R_n(X) = (R_{n1}(X),\dots,R_{nn}(X))$ are independent,

3) $X_{n\uparrow 1},\dots,X_{n\uparrow n}$ are distributed as $F^{-1}(U_{n\uparrow 1}),\dots,F^{-1}(U_{n\uparrow n})$.

Therefore we can simplify the derivatives (3.2) in the following way

$$\dot{\ell}_{\theta_o}(r) = \sum_{i=1}^{n} d_{ni}E_{\theta_o}(\dot{L}_{\theta_o}(X_{n\uparrow r_{ni}})|R_n=r) = \sum_{i=1}^{n} d_{ni}b_{nr_{ni}},$$

$$\ddot{\ell}_{\theta_o}(r) = \sum_{i=1}^{n} d_{ni}^2 E_{\theta_o}(\ddot{L}_{\theta_o}(X_{n\uparrow r_{ni}})|R_n=r) +$$

$$+ \sum\sum_{1\leq i\neq j\leq n} d_{ni}d_{nj}E_{\theta_o}(\dot{L}_{\theta_o}(X_{n\uparrow r_{ni}})\dot{L}_{\theta_o}(X_{n\uparrow r_{nj}})|R_n = r) =$$

$$= \sum_{i=1}^{n} d_{ni}^2 a_{nr_{ni}} + \sum\sum_{1\leq i\neq j\leq n} d_{ni}d_{nj}B_{nr_{ni}r_{nj}} .$$

Thus according to the fundamental lemma (0.6) the optimal test is of the form (3.6-8).□

According to Theorem 10 there always exists a rank test which is LMP unbiased level α for any one-parameter subclass

$$\mathfrak{F} = \{ \bigotimes_{i=1}^{n} F_{\theta_o+\eta d_{ni}} : \eta \in U(0) \subset \mathbb{R}\} .$$

However, this test is distinct from the test which is based on the square of the corresponding linear rank statistik $\sum_{i=1}^{n} d_{ni}b_{nr_{ni}}$.

On the other hand both are asymptotically equivalent under sequences of distributions which are contiguous to the hypothesis. To prove this we need some further regularity assumption on the class $\mathfrak{F}$ as well as on the regression coefficients.

<u>Theorem 11</u>: Let a, $b:(0,1)\to\mathbb{R}$ be real-valued functions with $\int a^2(u)d\lambda<\infty$ and $\int b^4(u)d\lambda<\infty$ and let $B:(0,1)\times(0,1)\to\mathbb{R}$ be defined by $B(u,v)=b(u)b(v)$. Then with a_n, b_n and B_n according to (3.4-5) and $R^*_{ni}:=R_{ni}/(n+1)$, $i=1,\dots,n$,

a) $\quad a_n \xrightarrow[\mathbb{L}_2]{} a,\quad b_n \xrightarrow[\mathbb{L}_2]{} b,\quad B_n \xrightarrow[\mathbb{L}_2]{} B,$ (3.9)

b) $\quad a_n(R^*_{n1}) \xrightarrow[\mathbb{L}_2]{} a(U_1),\quad b_n(R^*_{n1}) \xrightarrow[\mathbb{L}_2]{} b(U_1),$ (3.10)

$\quad B_n(R^*_{n1},R^*_{n2}) \xrightarrow[\mathbb{L}_2]{} B(U_1,U_2).$

Proof: a) The first two statements, both in a) and in b), are well known from the theory of one-sided rank tests. The third statement in case a) follows similarly to the second one using that $\int B^2(u,u)d\lambda = \int b^4(u)d\lambda<\infty$; in case b) it can be proved by a standard (martingale-) argument. □

Theorem 12: Let the assumptions of Theorem 11 be fulfilled and let d_{ni}, $i=1,\dots,n$, $n\in\mathbb{N}$, be a system of regression coefficients which satisfies the Noether type condition

$$\sum_{i=1}^{n} d_{ni} = 0,\quad \max_{1\le i\le n} d_{ni}^2\to 0,\quad \sum_{i=1}^{n} d_{ni}^2\to 1 . \tag{3.11}$$

Then under the hypothesis $\mathbb{H}$

a) $$\sum_{i=1}^{n} d_{ni}^2 a_{nR_{ni}} \xrightarrow[\mathbb{L}_2]{} \int a^2 d\lambda,\quad \sum_{i=1}^{n} d_{ni}(b_{nR_{ni}}-b(U_i)) \xrightarrow[\mathbb{L}_2]{} 0,$$

$$\sum\sum_{1\le i\ne j\le n} d_{ni}d_{nj}(B_{nR_{ni}R_{nj}}-B(U_i,U_j))^2 \xrightarrow[\mathbb{L}_2]{} 0, \tag{3.12}$$

b) $$\mathcal{L}\Big(\sum_{i=1}^{n} d_{ni}b_{nR_{ni}}\Big) \to \mathfrak{N}(0,\int b^2 d\lambda),$$

$$\mathcal{L}\Big(\sum\sum_{1\le i\ne j\le n} d_{ni}d_{nj}B_{nR_{ni}R_{nj}}\Big) \xrightarrow[\mathcal{L}]{} \int b^2 d\lambda[\chi^2-1],$$

c) The test (3.6-8) is asymptotically equivalent to the asymptotically unbiased level α rank test

$$\tilde{\varphi}(x) = \tilde{\psi}(R_n(x)),\ \tilde{\psi}(r) = \begin{cases} 1 & \text{for } \left|\sum_{i=1}^{n} d_{ni}b_{nr_{ni}}\right| > (\int b^2\, d\lambda)^{1/2} u_{\alpha/2}, \\ 0 & \text{for } \left|\sum_{i=1}^{n} d_{ni}b_{nr_{ni}}\right| < (\int b^2\, d\lambda)^{1/2} u_{\alpha/2}. \end{cases}$$

Proof: a) Obviously, $Ea^2_{nR_{ni}}$ is independent of $i=1,\dots,n$; because of Theorem 11a,

$$Ea^2_{nR_{n1}} = \frac{1}{n}\sum_{i=1}^{n} a^2_{ni} = \int a_n^2 d\lambda \to \int a^2 d\lambda .$$

Analogously, $Ea_{nR_{ni}}a_{nR_{nj}}$ is independent of $1\le i\ne j\le n$

$$Ea_{nR_{n1}}a_{nR_{n2}} = \frac{1}{n(n-1)}\sum\sum_{1\le i\ne j\le n} a_{ni}a_{nj}$$

$$= \frac{-1}{n(n-1)}\sum_{i=1}^{n} a^2_{ni} + \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} a_{ni}\right)^2 \to$$

$$(\int a d\lambda)^2 \quad \text{with} \quad \int a d\lambda = \int \ddot{L} dF = 0.$$

Therefore, because of

$$\sum_{i=1}^{n} d^4_{ni} \le \max_{1\le j\le n} d^2_{nj} \sum_{i=1}^{n} d^2_{ni} \to 0,$$

$$E\left(\sum_{i=1}^{n} d^2_{ni} a_{nR_{ni}}\right)^2 = \sum_{i=1}^{n} d^4_{ni} Ea^2_{nR_{n1}} + \left[(\sum d^2_{ni})^2 - \sum d^4_{ni}\right] Ea_{nR_{n1}} a_{nR_{n2}} \to \int a^2 d\lambda .$$

The second statement again follows from the theory of one-sided rank tests.

The third one can be proved by similar arguments: For $i<j$ and $k<\ell$

$$\rho_{i,j;k,\ell} := E(B_{nR_{ni}R_{nj}} - B(U_i,U_j))(B_{nR_{nk}R_{n\ell}} - B(U_k,U_\ell))$$

can take three possible values only, namely $\rho_{1,2;3,4}$, $\rho_{1,2;2,3}$ and $\rho_{1,2;1,2}$. Since, according to the Cauchy-Schwarz inequality and Theorem 10b, for all $i<j$ and $k<\ell$

$$|\rho_{i,j;k,\ell}| \le E(B_{nR_{n1}R_{n2}} - B(U_1,U_2))^2 \to 0,$$

for proving (3.12) it is sufficient to verify that the coefficients of these three possible values are $0(1)$ for $n\to\infty$. This can be done by an elementary but somewhat lengthy computation.

b) Since $\mathcal{L}(\sum_{i=1}^{n} d_{ni} b(U_{ni})) \xrightarrow[\mathcal{L}]{} \mathcal{N}(0, \int b^2 d\lambda)$ follows from the Lindeberg-Feller theorem, the first assertion is an immediate consequence of a) and the Slutsky theorem. Similarly, the second one can be proved by taking into account that $E \sum_{i=1}^{n} d_{ni}^2 b^2(U_i) \to \int b^2 d\lambda$ according to the Noether condition (3.12).

c) Let $W_n := \sum_{i=1}^{n} d_{ni} b_{nR_{ni}}$. Then from a) and b) follows

$$S_n = W_n^2 + c_1 W_n + \int a^2 d\lambda - \int b^2 d\lambda + c_o + o_P(1) .$$

Therefore, the test (3.6-8) is asymptotically equivalent with the test

$$\tilde{\varphi}_n = \begin{cases} 1 & \text{for } (W_n + \frac{c_1}{2})^2 > c_2 , \\ 0 & \text{for } (W_n + \frac{c_1}{2})^2 < c_2 , \end{cases} \qquad c_2 := \int b^2 d\lambda - \int a^2 d\lambda + \frac{c_1^2}{4} - c_0 .$$

But it can easily be proved that the only test of this form, which is an asymptotically unbiased level α test for $\mathbb{H}$ against $\mathbb{K}$, is the test (3.12). ⊔

Notice, that the tests (3.6-8) and (3.12) are asymptotically unbiased in all directions and not only in the direction of $\not{k}$.

The above discussion is not restricted to testing problem of randomness but also extends to the problems of testing symmetry and of independence.

If one is interested in sequential tests that are asymptotically locally optimal one has to guarantee the validity of the functional limit theorem, too. For that purpose it is convenient that the rank statistic $S_n(r)$ together with the σ-algebra generated by the ranks form a martingale in the two-sample case with equal sample sizes (as well as in the other cases mentioned above). In fact

$$\begin{aligned}
E_{\theta_o}(S_{n+1}(R_{n+1})|R_n) &= E_{\theta_o}(E_{\theta_o}(\ddot{L}_{n+1}|R_{n+1})|R_n) = E_{\theta_o}(\ddot{L}_{n+1}|R_n) = \\
&= E_{\theta_o}(\ddot{L}_n|R_n) + 2E_{\theta_o}(\dot{L}_{\theta_o}(x_{n+1})\dot{L}_n|R_n) - \\
&- 2E_{\theta_o}(\dot{L}_{\theta_o}(y_{n+1})\dot{L}_n|R_n) + E_{\theta_o}(\dot{L}^2_{\theta_o}(x_{n+1})|R_n) + \\
&+ E_{\theta_o}(\dot{L}^2_{\theta_o}(y_{n+1})|R_n) - 2E_{\theta_o}(\dot{L}_{\theta_o}(x_{n+1})\dot{L}_{\theta_o}(y_{n+1})|R_n) + \\
&+ E_{\theta_o}(\ddot{L}_{\theta_o}(x_{n+1}) - \dot{L}^2_{\theta_o}(x_{n+1})|R_n) + \\
&+ E_{\theta_o}(\ddot{L}_{\theta_o}(y_{n+1}) - \dot{L}^2_{\theta_o}(y_{n+1})|R_n) = \\
&= E_{\theta_o}(\ddot{L}_n|R_n) + 2E_{\theta_o}(\dot{L}^2_{\theta_o}(x_1)) - \\
&- 2E_{\theta_o}(\dot{L}^2_{\theta_o}(x_1)) = S_n(R_n).
\end{aligned}$$

In the presence of the martingale property it suffices under $\mathbb{H}$ as well as under contiguous alternatives to establish the convergence of the finite dimensional marginals.

REFERENCES

BERK, R.M. (1975). Locally most powerful sequential tests. Ann. Stat. 3, 373-381.

CHOW, Y.S., ROBBINS, H. and SIEGMUND, D. (1971). Great expectations: The theory of optimal stopping. Houghton Mifflin, New York.

CHOW, Y.S. and TEICHER, H. (1978). Probability Theory. Springer, New York.

HÁJEK, J. (1962). Asymptotically most powerful rank-order tests. Ann. Math. Stat. 33, 1124-1147.

HÁJEK, J. and SIDÁK, Z. (1967). Theory of Rank Tests. Academic Press, New York.

ISAACSON, S.L. (1951). On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters. Ann. Math. Stat. 22, 217-234.

LE CAM, L. (1966). Likelihood functions for large numbers of independent observations. Research papers in statistics, ed. F.N. David, 167-187.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. J. of Stat. Planning and Inference 4, 339-364.

ROOTZÉN, H. (1976). Fluctuations of sequences which converge in distribution. Ann. Prob. 4, 456-463.

SCHMETTERER, L. (1966). Mathematische Statistik, 2. Aufl., Springer, Wien.

SHIRYAYEV, A.N. (1978). Optimal Stopping Rules. Springer, New York.

WITTING, H. (1985). Mathematische Statistik I, Parametrische Verfahren bei festem Stichprobenumfang. Teubner, Stuttgart.

WITTING, H. and NÖLLE, G. (1970). Angewandte Mathematische Statistik. Teubner, Stuttgart.

Ulrich Müller-Funk, Friedrich Pukelsheim, Hermann Witting
Institut für Mathematische Stochastik
der Albert-Ludwigs-Universität
Hebelstraße 27.
D-7800. Freiburg/Breisgau