Z. Wahrscheinlichkeitstheorie verw. Gebiete
57, 73-86 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Efficient Robust Tests in Parametric Models*

Rudolf Beran
Department of Statistics, University of California, Berkeley, CA 94720, USA

Summary. Let $\{P_\theta : \theta \in \Theta\}$, Θ an open subset of R^k, be a regular parametric model for a sample of n independent, identically distributed observations. Formulated and solved in this paper is a robust version of the classical multi-sided hypothesis testing problem concerning θ, or a subvector of θ. In the robust testing problem, the usual parametric null hypothesis and alternatives are both replaced with larger, more realistic, sets of possible distributions for each observation. These sets, defined in terms of a Hellinger metric projection of the actual distribution onto a subspace associated with the parametric null hypothesis, are required to shrink as sample size increases, so as to avoid trivial asymptotics. One construction of an asymptotically minimax test for the robust testing problem is based upon the robust estimate of θ developed in Beran (1979); another construction amounts to an adaptively modified $C(\alpha)$ test.

1. Introduction

Robust testing, in parametric models, of simple or composite hypotheses versus composite alternatives is the theme of this paper. Classical procedures for such testing problems include generalized likelihood ratio tests, $C(\alpha)$ tests, and tests based on quadratic forms in parameter estimates. Large sample optimality within the parametric model is the justification for these tests.

Fundamental goals in robust testing are controlling level under small, arbitrary departures from the parametric null hypothesis, and retaining good power under small, arbitrary departures from specified parametric alternatives. The classical tests need not be robust in this sense. Both the power and, contrary to common belief, the level of the one-sample two-sided t-test can be severely distorted by small departures from normality. On the other hand, the chi-square goodness-of-fit test in a multinomial model is robust, within the

* Research supported by National Science Foundation Grant MCS 75-10376

0044-3719/81/0057/0073/$02.80

framework of this paper, provided the critical value of the test is adjusted to control level under small departures from the null hypothesis. Both of these assertions are justified in Sect. 4 of this paper.

The theory of robust testing developed here parallels the asymptotic minimax approach to robust estimation taken in Beran (1979). As in earlier work in robust testing (such as Huber (1965), Huber-Carol (1970), Huber and Strassen (1973), Rieder (1978), Wang (1979)), the customary parametric hypotheses are replaced by larger, more realistic sets of possible distributions for each observation. To avoid trivial limiting power, these sets are required to shrink as sample size increases. Unlike the authors cited above, we specify the sets in terms of a projection of the actual distribution onto a subspace associated with the parametric null hypothesis. This formulation proves advantageous, intuitively and technically, in treating testing problems that involve nuisance parameters. Plausible modifications of $C(\alpha)$ tests and of the classical estimate-based tests turn out to be asymptotically minimax for the enlarged robust testing problem.

Throughout the paper, distances between probabilities are measured in the Hellinger metric and the projection mentioned in the previous paragraph is performed in a related Hilbert space. Other choices of Hilbert space might be made in formulating and solving the robust testing problem; the answers obtained would depend on this choice (cf. Millar (1979), who explores some alternative Hilbert spaces for robust estimation). Even for finite sample spaces, different metrics on probabilities determine different contamination models. A distinctive and attractive consequence of the Hellinger metric formulation: least favorable distributions for the enlarged robust testing problem may be found within the original parametric family; thus, the asymptotic minimax criterion is not too pessimistic. Other issues in the choice of contamination model for robustness studies have been discussed by Hampel (1971), Bickel (1978), Beran (1981), and Millar (1981). None of the discussions seems definitive to us.

Section 2 establishes an asymptotic upper bound on the power of every level α test in the formal robust testing problem. Attainability of the upper bound is demonstrated in Sect. 3 through two constructions of asymptotically minimax robust tests. Several examples of robust and non-robust tests in familiar models are considered in Sect. 4.

2. Asymptotic Minimax Bounds

As in the estimation paper, Beran (1981), let H be the set of elements $\xi(dP)^{1/2}$, where P is a probability on a Euclidean space $\mathscr{X}$ with Borel sets $\mathscr{B}$, and ξ is a random variable in $L_2(P)$. Suppose $\xi(dP)^{1/2}$ and $\eta(dQ)^{1/2}$ are elements of H and that $\mu=2^{-1}(P+Q)$. Define the inner product

$$\langle \xi(dP)^{1/2}, \eta(dQ)^{1/2}\rangle = \int \xi\eta \left(\frac{dP}{d\mu}\right)^{1/2} \left(\frac{dQ}{d\mu}\right)^{1/2} d\mu \tag{2.1}$$

and, for arbitrary real a, b, the linear combination

$$a\xi(dP)^{1/2}+b\eta(dQ)^{1/2}=\left[a\xi\left(\frac{dP}{d\mu}\right)^{1/2}+b\eta\left(\frac{dQ}{d\mu}\right)^{1/2}\right](d\mu)^{1/2}. \tag{2.2}$$

The corresponding norm $\|\cdot\|$ on H is given by

$$\begin{aligned}\|\xi(dP)^{1/2}\|^2&=\langle\xi(dP)^{1/2},\xi(dP)^{1/2}\rangle\\&=\int\xi^2\,dP.\end{aligned} \tag{2.3}$$

In particular, if $(dP)^{1/2}$ denotes the element $1(dP)^{1/2}$, the Hellinger distance between the probabilities P and Q may be defined as $\|(dQ)^{1/2}-(dP)^{1/2}\|$. The elements $\xi(dP)^{1/2}$ and $\eta(dQ)^{1/2}$ are called equivalent if $\|\xi(dP)^{1/2}-\eta(dQ)^{1/2}\|=0$. The set $\bar{H}$ of equivalence classes in H forms a Hilbert space with the above inner product and norm.

Suppose $\zeta=(\zeta_1,\zeta_2,\ldots,\zeta_k)'$ is a random vector whose components lie in $L_2(P)$. Then $\zeta(dP)^{1/2}$ designates the vector $(\zeta_1(dP)^{1/2}, \zeta_2(dP)^{1/2}, \ldots, \zeta_k(dP)^{1/2})'$ and $\|\zeta(dP)^{1/2}\|^2$ means $\int|\zeta|^2dP$. Similarly, if $\eta(dQ)^{1/2}$ belongs to H, $\langle\zeta(dP)^{1/2}, \eta(dQ)^{1/2}\rangle$ denotes the column vector of componentwise inner products.

Let $\{P_\theta^n=P_\theta\times P_\theta\times\ldots\times P_\theta$ n-times: $\theta\in\Theta\}$ be the parametric model for a sample of size n. The $\{P_\theta:\theta\in\Theta\}$ are assumed to satisfy the following regularity conditions: The parameter space Θ is an open subset of R^k. The mapping $\theta\to P_\theta$ has the property that

(i) for every $\theta\in\Theta$, there exists $\eta_\theta\in L_2^k(P_\theta)$ such that

$$\lim_{t\to 0}|t|^{-1}\,\|(dP_{\theta+t})^{1/2}-(dP_\theta)^{1/2}-t'\eta_\theta(dP_\theta)^{1/2}\|=0; \tag{2.4}$$

(ii) for every $\theta\in\Theta$, the Fisher information matrix

$$I(\theta)=4\int\eta_\theta\eta_\theta'\,dP_\theta \tag{2.5}$$

is non-singular.

It is evident that the quadratic mean derivative η_θ is unique, up to equivalence in $L_2^k(P_\theta)$, and that $\int\eta_\theta\,dP_\theta=0$.

Fix $\theta_0\in\Theta$. Let M be the subspace of $\bar{H}$ which is spanned by the components of $\eta_{\theta_0}(dP_{\theta_0})^{1/2}$ and let π_M be the orthogonal projection which maps $\bar{H}$ into M. The subspace M is the tangent space to the parametric model (regarded as the set of elements $\{(dP_\theta)^{1/2}:\theta\in\Theta\}$ at the point $(dP_{\theta_0})^{1/2}$. Suppose the column vectors θ and η_θ are partitioned into two subvectors of dimensions k_1 and k_2 respectively: $\theta'=(\theta_1',\theta_2')$ and $\eta_\theta=(\eta_{\theta,1}',\eta_{\theta,2}')$. Let M_2 be the subspace of $\bar{H}$ which is spanned by the components of $\eta_{\theta_0,2}(dP_{\theta_0})^{1/2}$ and let M_1 be the orthocomplement in M of M_2.

Consider the following testing problem: The distribution of the sample of size is $Q^n=Q\times Q\times\ldots\times Q$ n-times. The null hypothesis regarding Q is

$$\begin{aligned}H_n:\ &\|\pi_{M_1}((dQ)^{1/2}-(dP_{\theta_0})^{1/2})\|\leqq n^{-1/2}a,\\&\|(dQ)^{1/2}-(dP_{\theta_0})^{1/2}\|\leqq n^{-1/2}c;\end{aligned} \tag{2.6}$$

the alternative is

$$K_n: \ \|\pi_{M_1}((dQ)^{1/2}-(dP_{\theta_0})^{1/2})\| \geqq n^{-1/2} b,$$
$$\|(dQ)^{1/2}-(dP_{\theta_0})^{1/2}\| \leqq n^{-1/2} c \tag{2.7}$$

where $0 \leqq a < b < c < \infty$. This problem is a robust version of the classical testing problem: $\theta_1 = \theta_{0,1}$ versus $\theta_1 \neq \theta_{0,1}$, the subvector θ_2 being regarded as nuisance parameters. If no nuisance parameters exist, M_1 coincides with M.

The common requirement in H_n and K_n, that $(dQ)^{1/2}$ lie within a Hellinger ball of radius $n^{-1/2}c$ about $(dP_{\theta_0})^{1/2}$ serves two technical purposes: keeping the product measures Q^n and $P_{\theta_0}^n$ from separating in the Hellinger metric, so as to allow non-trivial asymptotic power calculations; and providing sufficient compactness for minimax considerations. The hypothesis H_n further asserts that the portion of Q which resembles a member of the parametric family $\{P_\theta: \theta \in \Theta\}$ is near P_{θ_0}; distance is measured in the subspace M_1, which ignores variations in the nuisance parameter θ_2. The alternative K_n asserts that the portion of Q which resembles a member of the parametric family is bounded away from P_{θ_0}, distance still being measured in M_1.

The power of every level α test for H_n versus K_n is subject to the asymptotic upper bound established in Theorem 1 below. Let $d_\alpha(k;r)$ be the upper α point of the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter r. Let $\beta_\alpha(k;r,s)$ denote the probability that a random variable with noncentral chi-square distribution, k degrees of freedom and noncentrality parameter s, exceeds $d_\alpha(k;r)$.

Theorem 1. *If $\{\psi_n; n \geqq 1\}$ is any sequence of tests such that*

$$\limsup_n \sup_{Q \in H_n} E_{Q^n}[\psi_n(x)] \leqq \alpha \tag{2.8}$$

then

$$\limsup_n \inf_{Q \in K_n} E_{Q^n}[\psi_n(x)] \leqq \beta_\alpha(k_1; 4a^2, 4b^2). \tag{2.9}$$

To prove Theorem 1, consider first the following simpler parametric testing problem. The distribution of the sample of size n is $P^n_{\theta_0+n^{-1/2}h}$, where $h \in R^k$. The null hypothesis regarding h is

$$H_n^*: \ 4\|\pi_{M_1}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2})\|^2 \leqq r,$$
$$4\|\pi_{M_2}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2})\|^2 = u; \tag{2.10}$$

the alternative is

$$K_n^*: \ 4\|\pi_{M_1}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2})\|^2 = s,$$
$$4\|\pi_{M_2}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2})\|^2 = u \tag{2.11}$$

where $0 \leqq r \leqq s < \infty$ and $u \geqq 0$. The following asymptotic minimax bound is well-known for the special case $M_1 = M$ (c.f. Hájek and Šidák (1967), LeCam (1972)).

Proposition 1. *If* $\{\psi_n; n \geqq 1\}$ *is any sequence of tests such that*

$$\limsup_n \sup_{Q \in H_n^*} E_{Q^n}[\psi_n(x)] \leqq \alpha \tag{2.12}$$

then

$$\limsup_n \inf_{Q \in K_n^*} E_{Q^n}[\psi_n(x)] \leqq \beta_\alpha(k_1; r, s). \tag{2.13}$$

Proof. Suppose not. Then, there exists a test sequence $\{\psi_n\}$ and $\varepsilon > 0$ such that

$$\begin{aligned} &\limsup_n \sup_{Q \in H_n^*} E_{Q^n}[\psi_n(x)] \leqq \alpha \\ &\limsup_n \inf_{Q \in K_n^*} E_{Q^n}[\psi_n(x)] = \beta_\alpha(k_1; r, s) + 2\varepsilon. \end{aligned} \tag{2.14}$$

By going to a subsequence, we may assume without loss of generality that

$$\begin{aligned} &\limsup_n \sup_{Q \in H_n^*} E_{Q^n}[\psi_n(x)] \leqq \alpha \\ &\inf_{Q \in K_n^*} E_{Q^n}[\psi_n(x)] \geqq \beta_\alpha(k_1; r, s) + \varepsilon \end{aligned} \tag{2.15}$$

for every $n \geqq 1$.

For $j = 1, 2$, choose ρ_j so that the components of $\rho_j(dP_{\theta_0})^{1/2}$ are orthonormal and span the subspace M_j. Then, there exists a k_j-dimensional column vector t_j such that $\pi_{M_j}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2}) = 2^{-1} t_j' \rho_j (dP_{\theta_0})^{1/2}$; thus the length of this projection is $2^{-1}|t_j|$. Let $\rho = (\rho_1', \rho_2')'$, $t = (t_1', t_2')$ and write θ_n for $\theta_0 + n^{-1/2} h$. The log-likelihood ratio of $P_{\theta_n}^n$ relative to $P_{\theta_0}^n$ is defined, up to a $P_{\theta_0}^n$-null set, by

$$L_n(\theta_n, \theta_0) = \sum_{i=1}^n \log\left[\frac{dP_{\theta_n, c}}{dP_{\theta_0}}(x_i)\right], \tag{2.16}$$

where $P_{\theta_n, c}$ denotes the part of P_{θ_n} which is absolutely continuous with respect to P_{θ_0}. Under $\{P_{\theta_0}^n\}$, the following expansion holds for $L_n(\theta_n, \theta_0)$ because of assumption (2.4) on the parametric model (see LeCam (1969)):

$$\begin{aligned} L_n(\theta_n, \theta_0) &= 2n^{-1/2} \sum_{i=1}^n h'\eta_{\theta_0}(x_i) - 2\|h'\eta_{\theta_0}(dP_{\theta_0})^{1/2}\|^2 + o_p(1) \\ &= t'Z_n - 2^{-1}|t|^2 + o_p(1), \end{aligned} \tag{2.17}$$

where $Z_n = n^{-1/2} \sum_{i=1}^n \rho(x_i)$.

It is evident that the sequence $\{(\psi_n, Z_n); n \geqq 1\}$ is tight under $\{P_{\theta_0}^n\}$. By going to a subsequence, we can assume that $\{(\psi_n, Z_n)\}$ converges weakly to (ψ, Z), where Z has a standard normal distribution on R^k and ψ has range $[0, 1]$. For every $h \in K_n^*$,

$$\begin{aligned} 1 - \beta_\alpha(k_1; r, s) - \varepsilon &\geqq \liminf_n E_{\theta_n}[1 - \psi_n] \\ &\geqq \liminf_n E_{\theta_0}[(1 - \psi_n) \exp(L_n)] \\ &\geqq E_{\psi, Z}[(1 - \psi) \exp(t'Z - 2^{-1}|t|^2)] \\ &= E_Z[(1 - \psi_0(Z)) \exp(t'Z - 2^{-1}|t|^2)] \\ &= 1 - E[\psi_0(Z + t)], \end{aligned} \tag{2.18}$$

where $\psi_0(Z)=E[\psi|Z]$ is also a test. A similar argument shows that, for every $h\in H_n^*$,

$$\begin{aligned}\alpha &\geqq \limsup_n E_{\theta_n}[\psi_n(x)]\\ &\geqq \liminf_n E_{\theta_n}[\psi_n(x)]\\ &\geqq E_Z[\psi_0(Z)\exp(t'Z-2^{-1}|t|^2)]\\ &=E[\psi_0(Z+t)].\end{aligned} \tag{2.19}$$

The inequalities (2.18) and (2.19) imply the following conclusion: there exists a test ψ_0 such that $E[\psi_0(Z+t)]\leqq\alpha$ for every $t\in\{t: |t_1|^2\leqq r, |t_2|^2=u\}$ and $E[\psi_0(Z+t)]\geqq\beta_\alpha(k_1; r, s)+\varepsilon$ for every $t\in\{t: |t_1|^2=s, |t_2|^2=u\}$. But this is false, because the infimum power of every level α test for this problem cannot exceed $\beta_\alpha(k_1; r, s)$. (This fact is a consequence of the Hunt-Stein theorem. A closely related example appears on p. 338 of Lehmann (1959)). The contradiction completes the proof of Proposition 1.

Proof of Theorem 1. Choose $\varepsilon>0$ small enough that $\varepsilon<\min(4a^2, 2c^2-2b^2)$. Set $r=4a^2-\varepsilon$, $s=4b^2+\varepsilon$ and $u=4c^2-4b^2-2\varepsilon$. The quadratic mean differentiability assumption (2.4) implies that, for every compact set C in R^k,

$$\begin{aligned}&\lim_{n\to\infty}\sup_{h\in C} n\|\pi_{M_j}((dP_{\theta_n})^{1/2}-(dP_{\theta_0})^{1/2})\|^2\\ &\quad=\sup_{h\in C}\|\pi_{M_j}(h'\eta_{\theta_0}(dP_{\theta_0})^{1/2})\|^2.\end{aligned} \tag{2.20}$$

Thus, there exists $n_0(\varepsilon)$ such that for every $n>n_0(\varepsilon)$, $\{P_{\theta_n}: h\in H_n^*\}\subset H_n$ and $\{P_{\theta_n}: h\subset K_n^*\}\subset K_n$.

Every sequence of tests $\{\psi_n\}$ which satisfies (2.8) necessarily satisfies (2.12) as well. By Proposition 1,

$$\begin{aligned}\limsup_n \inf_{Q\in K_n} E_{Q^n}[\psi_n(x)] &\leqq \limsup_n \inf_{Q\in K_n^*} E_{Q^n}[\psi_n(x)]\\ &\leqq \beta_\alpha(k_1; 4a^2-\varepsilon, 4b^2+\varepsilon).\end{aligned} \tag{2.21}$$

Letting ε tend to zero completes the argument.

3. Constructing Asymptotic Minimax Tests

It is not immediately clear, from the proof of Theorem 1, whether the asymptotic bound on infimum power is actually attained by some test sequences. Under slightly stronger assumptions on the parametric model, the answer is affirmative.

Theorem 2. *In addition to the assumptions for Theorem 1, suppose that the mapping $\theta\to P_\theta$ is one-to-one. Then, for every $\alpha\in(0,1)$ and every $\theta_0\in\Theta$, there exists a sequence of robust tests $\{\varphi_n; n\geqq 1\}$ such that*

$$\lim_{n\to\infty}\sup_{Q\in H_n} E_{Q^n}[\varphi_n(x)]=\alpha \tag{3.1}$$

and

$$\lim_{n\to\infty} \inf_{Q\in K_n} E_{Q^n}[\varphi_n(x)] = \beta_\alpha(k_1; 4a^2, 4b^2). \tag{3.2}$$

Theorem 2 will be proved by giving two possible constructions for the asymptotically minimax robust tests $\{\varphi_n\}$. To be avoided here are the trivial non-robust tests which satisfy (3.1) and (3.2) by confounding sample size with the distance between $Q\in K_n$ and P_{θ_0}. A similar phenomenon in robust estimation has been discussed in Beran (1981).

Suppose $\{Q_n; n\geqq 1\}$ is any sequence of probabilities on $(\mathscr{X}, \mathscr{B})$ such that $\{n^{1/2}((dQ_n)^{1/2}-(dP_\theta)^{1/2}); n\geqq 1\}$ converges weakly to an element in H. Under the assumptions of Theorem 2, there exist estimates $\{\theta_n^*; n\geqq 1\}$ such that, for every $\theta\in\Theta$, the distributions of $\{n^{1/2}(\theta_n^*-\theta); n\geqq 1\}$ under $\{Q_n^n\}$ are tight. A discretized version $\hat\theta_n$ of θ_n^* is defined as follows: Let d be an arbitrary positive constant. Cover the parameter space $\Theta\subset R^k$ with disjoint semi-closed hypercubes of side length $n^{-1/2}d$. Set $\hat\theta_n$ equal to the center of the hypercube which contains θ_n^*. Evidently, $\{n^{1/2}(\hat\theta_n-\theta); n\geqq 1\}$ is also tight under $\{Q_n^n\}$ for every $\theta\in\Theta$. For a general construction of θ_n^* with the desired properties, see LeCam (1969), pp. 104–107.

Separability of Θ and quadratic mean continuity of P_θ as a function of θ ensure the existence of a probability μ on $(\mathscr{X}, \mathrm{B})$ such that $P_\theta \ll \mu$ for every $\theta\in\Theta$. Let p_θ be the density of P_θ with respect to μ. Let $\{e_i; 1\leqq i\leqq k\}$ be the usual orthonormal basis for R^k; e_j is a k-dimensional column vector whose jth component is 1 and whose other components are 0. For $x\in\mathscr{X}$ and $\theta\in\Theta$, define

$$\eta_n(x,\theta)=\begin{cases}\sum_{j=1}^{n} n^{1/2}(p_\theta^{-1/2}(x)\,p_{\theta+n^{-1/2}e_j}^{1/2}(x)-1)\,e_j & \text{if } p_\theta(x)>0\\ 0 & \text{if } p_\theta(x)=0\end{cases} \tag{3.3}$$

If the model $\{P_\theta\colon \theta\in\Theta\}$ is such that

$$\lim_{t\to 0} \|\eta_{\theta+t}(dP_{\theta+t})^{1/2}-\eta_\theta(dP_\theta)^{1/2}\|=0, \tag{3.4}$$

it is possible, in what follows, to use η_θ rather than $\eta_n(\cdot,\theta)$ (c.f. Beran (1981)).

Let $\{c_n^*; n\geqq 1\}$ be positive, real-valued statistics such that c_n^* is a function of the sample $(x_1, x_2, \ldots, x_n)$ and $\{n^{1/4-\delta}c_n^*\}$ is tight under $\{Q_n^n\}$ for some $\delta\in(0, 1/4)$. One possible nontrivial choice is

$$c_n^* = \lambda[\sup_x |\hat F_n(x) - F_{\hat\theta_n}(x)|]^{1/2-2\delta}, \tag{3.5}$$

where $\lambda>0$, $\hat F_n$ is the empirical c.d.f. of the sample, and F_θ is the c.d.f. of P_θ. Discretize c_n^* as follows. Let v be an arbitrary positive constant. Starting with the origin as an endpoint, cover the positive real axis with disjoint, semiclosed intervals of length $n^{-1/4+\delta}v$. Set $\hat c_n$ equal to the center of the interval which contains c_n^*. Tightness of $\{n^{1/4-\delta}c_n^*\}$ under $\{Q_n^n\}$ implies tightness of $\{n^{1/4-\delta}\hat c_n\}$. Note that $\hat c_n \geqq 2^{-1}n^{-1/4+\delta}v$.

Let m be an absolutely continuous function mapping R^+ into $[0,1]$ such that $m(0)=1$, $\sup_x[xm(x)]<\infty$ and the derivative m' is bounded. For instance,

$m(x)=\min\{1, x^{-1}\}$; or $m(x)=x^{-1}\sin(x)$; or $m(x)=\max\{(1-x^2), 0\}$. Define the random window w_n by

$$w_n(x; x_1, x_2, \ldots, x_n)=m(\hat{c}_n|\eta_n(x, \hat{\theta}_n)|). \tag{3.6}$$

For brevity, we will write $w_n(x)$, suppressing the dependence upon the sample. Let $\xi_n(\cdot, \hat{\theta}_n)$ be a weighted and recentered version of $\eta_n(\cdot, \hat{\theta}_n)$, defined as follows:

$$\xi_n(x, \hat{\theta}_n)=\{\eta_n(x, \hat{\theta}_n)-[\int w_n(t)\, dP_{\hat{\theta}_n}]^{-1}\int \eta_n(t, \hat{\theta}_n) w_n(t)\, dP_{\hat{\theta}_n}\} w_n(x). \tag{3.7}$$

Partition k-vectors and $k\times k$ matrices according to the dimensions k_1, k_2 of the subspaces M_1, M_2. In particular, $\theta=(\theta_1', \theta_2')'$, where θ_i is $k_i\times 1$, and $I(\theta)=\{I_{ij}(\theta)\}$, where $I_{ij}(\theta)$ is $k_i\times k_j$. Let

$$I_{11.2}(\theta)=I_{11}(\theta)-I_{12}(\theta)I_{22}^{-1}(\theta)I_{21}(\theta) \tag{3.8}$$

and let

$$\eta_{\theta,1.2}=\eta_{\theta,1}-I_{12}(\theta)I_{22}^{-1}(\theta)\eta_{\theta,2}. \tag{3.9}$$

Estimates Test. Suppose $\{Q_n;\ n\geqq 1\}$ is any sequence of probabilities on $(\mathscr{X}, \mathscr{B})$ such that $\{n^{1/2}((dQ)^{1/2}-(dP_{\theta_0})^{1/2})\}$ converges weakly to an element $\zeta(dR)^{1/2}$ in H. There exist asymptotically minimax estimates $\{T_n\}$ such that the limiting distribution of $\{n^{1/2}(T_n-\theta_0);\ n\geqq 1\}$ under $\{Q_n^n\}$ is normal with mean $4I^{-1}(\theta_0)\langle\eta_{\theta_0}(dP_{\theta_0})^{1/2}, \zeta(dR)^{1/2}\rangle$ and covariance matrix $I^{-1}(\theta_0)$, for every possible choice of $\zeta(dR)^{1/2}$ in H. One construction of such $\{T_n\}$ is

$$T_n=\hat{\theta}_n+2n^{-1}I_n^{-1}(\hat{\theta}_n)\sum_{i=1}^{n}\xi_n(x_i, \hat{\theta}_n), \tag{3.10}$$

where $I_n(\hat{\theta}_n)=\int \xi_n(x, \hat{\theta}_n)\xi_n'(x, \hat{\theta}_n)\, dP_{\hat{\theta}_n}$ (Proposition 1 in Beran (1981)). Define φ_n as the test which rejects H_n if

$$n(T_{n,1}-\theta_{0,1})' I_{n,11.2}(\hat{\theta}_n)(T_{n,1}-\theta_{0,1})>d_\alpha(k_1; 4a^2). \tag{3.11}$$

Here $I_{n,11.2}(\hat{\theta}_n)$ is defined by analogy with (3.8) and $T_{n,1}$ is the first k_1 components in T_n.

Scores Test. Let $\bar{\theta}_n$ be the random vector obtained by replacing the first k_1 components of $\hat{\theta}_n$ with the first k_1 components of θ_0. Let

$$Z_n(\bar{\theta}_n)=2n^{-1/2}\sum_{i=1}^{n}\xi_n(x_i, \bar{\theta}_n)$$

and let

$$Z_{n,1.2}(\bar{\theta}_n)=Z_{n,1}(\bar{\theta}_n)-I_{n,12}(\bar{\theta}_n)I_{n,22}^{-1}(\bar{\theta}_n)Z_{n,2}(\bar{\theta}_n). \tag{3.12}$$

Define φ_n as the test which rejects H_n if

$$Z_{n,1.2}'(\bar{\theta}_n)I_{n,11.2}^{-1}(\bar{\theta}_n)Z_{n,1.2}(\bar{\theta}_n)>d_\alpha(k_1; 4a^2). \tag{3.13}$$

These two tests are robust modifications of, respectively, the usual estimates test and Neyman's (1959) $C(\alpha)$-test. As is proved below, both tests satisfy (3.1) and (3.2) under the assumptions of Theorem 2. In place of $I_n(\hat{\theta}_n)$ or $I_n(\bar{\theta}_n)$, it is possible to use $J_n(\hat{\theta}_n)=n^{-1}\sum_{i=1}^{n}\xi_n(x_i,\hat{\theta}_n)\xi_n'(x_i,\hat{\theta}_n)$ or the analogously defined $J_n(\bar{\theta}_n)$. If $I(\theta)$ is continuous in θ, both $I(\hat{\theta}_n)$ or $I(\bar{\theta}_n)$ will also serve. The local asymptotic analysis carried out in this paper is unable to discriminate among these matrix estimates.

Proof of Theorem 2. Consider first the scores test. Let $\{Q_n;\ n\geqq 1\}$ be any sequence of probabilities on $(\mathscr{X},\mathscr{B})$ such that $\{n^{1/2}((dQ_n)^{1/2}-(dP_{\theta_0})^{1/2});\ n\geqq 1\}$ converges weakly to an element $\zeta(dR)^{1/2}$ in H. By the argument for Proposition 1 in Beran (1981), the limiting distribution of $\{Z_n(\bar{\theta}_n)+I(\theta_0)n^{1/2}(\bar{\theta}_n-\theta_0);\ n\geqq 1\}$ under $\{Q_n^n\}$ is $N(4\langle\eta_{\theta_0}(dP_{\theta_0})^{1/2},\ \zeta(dR)^{1/2}\rangle,\ I(\theta_0))$. Moreover, the matrix $I_n(\bar{\theta}_n)$ and its surrogates all converge in Q_n^n-probability to $I(\theta_0)$. Thus, the limiting distribution of $\{Z_{n,1.2}(\bar{\theta}_n);\ n\geqq 1\}$ is $N(4\langle\eta_{\theta_0,1.2}(dP_{\theta_0})^{1/2},\ \zeta(dR)^{1/2}\rangle,\ I_{11.2}(\theta_0))$. It follows that the limiting distribution under $\{Q_n^n\}$ of the test statistic in (3.13) is noncentral chi-square with k_1 degrees of freedom and noncentrality parameter

$$\gamma^2=4\,\|\pi_{M_1}(\zeta(dR)^{1/2})\|^2. \tag{3.14}$$

To verify (3.14), recall that the subspace M_2 is spanned by the components of $\eta_{\theta_0,2}(dP_{\theta_0})^{1/2}$; hence, the components of $\eta_{\theta_0,1.2}(dP_{\theta_0})^{1/2}$ form a basis for the subspace M_1.

Let $S_n(c)=\{Q\colon \|(dQ)^{1/2}-(dP_{\theta_0})^{1/2}\|\leqq n^{-1/2}c\}$. We will show that

$$\begin{aligned}\lim_{n\to\infty}\ \sup_{Q\in S_n(c)}\ &|E_{Q^n}[\varphi_n(x)]\\ &-\beta_\alpha(k_1;\,4a^2,4n\|\pi_{M_1}((dQ)^{1/2}-(dP_{\theta_0})^{1/2})\|^2)|=0.\end{aligned} \tag{3.15}$$

Suppose (3.15) is false. Then there exists a sequence of probabilities $\{Q_n\in S_n(c);\ n\geqq 1\}$ such that the absolute difference on the left side of (3.15) stays bounded away from zero. By considering subsequences, we may assume without loss of generality that $\{n^{1/2}((dQ_n)^{1/2}-(dP_{\theta_0})^{1/2})\}$ converges weakly to an element $\zeta(dR)^{1/2}$ in H. Then, by the result of the preceding paragraph,

$$\begin{aligned}\lim_{n\to\infty}E_{Q_n^n}[\varphi_n(x)]&=\beta_\alpha(k_1;\,4a^2,\gamma^2)\\ &=\lim_{n\to\infty}\beta_\alpha(k_1;\,4a^2,4n\|\pi_{M_1}((dQ_n)^{1/2}-(dP_{\theta_0})^{1/2})\|^2).\end{aligned} \tag{3.16}$$

The contradiction proves (3.15).

Both $\{Q\in H_n\}$ and $\{Q\in K_n\}$ are subsets of $S_n(c)$. Moreover, β_α is continuous, strictly monotone increasing in its third argument and

$$\lim_{n\to\infty}\ \sup_{Q\in H_n}\ n\|\pi_{M_1}((dQ)^{1/2}-(dP_{\theta_0})^{1/2})\|^2=a^2. \tag{3.17}$$

It follows, therefore, from (3.15) that

$$\begin{aligned}\lim_{n\to\infty}\ \sup_{Q\in H_n}\ E_{Q^n}[\varphi_n(x)]&=\beta_\alpha(k_1;\,4a^2,4a^2)\\ &=\alpha.\end{aligned} \tag{3.18}$$

The verification of (3.2) is analogous. This completes the proof of Theorem 2 by way of the scores test.

Consider now the estimates test. Let $\{Q_n;\, n \geqq 1\}$ be any sequence of probabilities on $(\mathscr{X}, \mathscr{B})$ such that $\{n^{1/2}((dQ_n)^{1/2} - (dP_{\theta_0})^{1/2});\, n \geqq 1\}$ converges weakly to an element $\zeta(dR)^{1/2}$ in H. By assumption, the limiting distribution of $\{n^{1/2}(T_n - \theta_0);\, n \geqq 1\}$ under $\{Q_n^n\}$ is normal with mean $4I^{-1}(\theta_0)\langle \eta_{\theta_0}(dP_{\theta_0})^{1/2}, \zeta(dR)^{1/2}\rangle$ and covariance matrix $I^{-1}(\theta_0)$. Hence, the limiting distribution of the test statistic in (3.11) is noncentral chi-square with k_1 degrees of freedom and noncentrality parameter γ^2, defined in (3.14). The rest of the argument is as before.

4. Examples

The four examples worked in this section illustrate the application of Theorems 1 and 2 to the normal location-scale model, the multinomial model, and the canonical exponential model.

4.1 Testing Location in the Normal Model

Suppose Θ is $R \times R^+$, $\theta = (\mu, \sigma^2)$, and P_θ is $N(\mu, \sigma^2)$. The quadratic mean differentiability assumption (2.4) is satisfied with

$$\eta_\theta(x) = \begin{pmatrix} (2\sigma^2)^{-1}(x-\mu) \\ (4\sigma^2)^{-1}[-1 + \sigma^{-2}(x-\mu)^2] \end{pmatrix}. \tag{4.1}$$

The Fisher information matrix is

$$I(\theta) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 2^{-1}\sigma^{-4} \end{pmatrix}. \tag{4.2}$$

Let $(\bar{x}_n, s_n^2)$ be the usual unbiased estimates of (μ, σ^2). The two-sided t-test for the classical hypothesis $\mu = 0$ rejects if $n^{1/2} s_n^{-1} |\bar{x}_n|$ is too large. A robust version of this classical testing problem is given by (2.6) and (2.7), with $\theta_0 = (0, 1)$, say. In this case, M_1 is the subspace spanned by $2^{-1}x(dP_{\theta_0})^{1/2}$ while M_2 is the subspace spanned by $4^{-1}(-1 + x^2)(dP_{\theta_0})^{1/2}$. We will show that ψ_n, the two-sided t-test of level α, is nonrobust in both level and power because

$$\begin{aligned} &\lim_{c \to \infty} \liminf_n \sup_{Q \in H_n} E_{Q^n}[\psi_n(x)] = 1 \\ &\lim_{c \to \infty} \limsup_n \inf_{Q \in K_n} E_{Q^n}[\psi_n(x)] \leqq \alpha. \end{aligned} \tag{4.3}$$

To prove the first limit in (4.3), consider the sequence of probabilities $\{Q_n;\, n \geqq 1\}$ defined by

$$Q_n = (1 - (nu)^{-1}h)P_{\theta_0} + (nu)^{-1}hA(n^{1/2}u), \tag{4.4}$$

where h, u are positive constants and $A(t)$ is the unit probability supported by the point t. Since,

$$(dQ_n)^{1/2} = (1-(nu)^{-1}h)^{1/2}(dP_{\theta_0})^{1/2} + (nu)^{-1/2}h^{1/2}(dA(n^{1/2}u))^{1/2} \tag{4.5}$$

by mutual singularity of A and P_{θ_0}, it follows that

$$\pi_{M_1}(dQ_n)^{1/2} = \pi_{M_1}(dP_{\theta_0})^{1/2} = 0$$

and

$$\lim_{n\to\infty} n\,\|(dQ_n)^{1/2} - (dP_{\theta_0})^{1/2}\|^2 = hu^{-1}. \tag{4.6}$$

Thus, $Q_n \in H_n$ for all sufficiently large n, provided c^2 exceeds hu^{-1}.

Let B be any Borel set whose boundary is a Lebesgue null-set. A standard weak convergence argument establishes

$$\lim_{u\to 0}\lim_{n\to\infty} Q_n^n[n^{1/2}\bar{x}_n \in B] = P[Z+h \in B], \tag{4.7}$$

where Z is a $N(0,1)$ random variable. Applying Chebyshev's inequality to $n^{-1}\sum x_i^2$ and drawing on (4.7) yields

$$\lim_{u\to 0}\lim_{n\to\infty} Q_n^n[|s_n^2 - 1| > \varepsilon] = 0 \tag{4.8}$$

for every $\varepsilon > 0$. It follows from (4.7) and (4.8) that

$$\lim_{u\to 0}\lim_{n\to\infty} Q_n^n[n^{1/2}s_n^{-1}\bar{x}_n \in B] = P[Z+h \in B]. \tag{4.9}$$

The first part of (4.3) holds because $h>0$ is arbitrary in the argument above.

The reasoning for the second part of (4.3) is based on the sequence of probabilities

$$Q_n = (1-(nu)^{-1}h)P_{\theta_n} + (nu)^{-1}hA(n^{1/2}u), \tag{4.10}$$

where h, u are positive constants and $\theta_n = (-n^{-1/2}h, 1)$. In this case,

$$\begin{aligned}\lim_{n\to\infty} n\|\pi_{M_1}((dQ_n)^{1/2} - (dP_{\theta_0})^{1/2})\|^2 &= 4^{-1}h^2\\ \lim_{n\to\infty} n\|(dQ_n)^{1/2} - (dP_{\theta_0})^{1/2}\|^2 &= hu^{-1} + 4^{-1}h.\end{aligned} \tag{4.11}$$

Thus, $Q_n \in K_n$ for all sufficiently large n, provided c^2 exceeds $hu^{-1}+4^{-1}h$ and b^2 is less than $4^{-1}h^2$. Essentially the same argument as in the previous paragraph yields the result

$$\lim_{u\to 0}\lim_{n\to\infty} Q_n^n[n^{1/2}s_n^{-1}\bar{x}_n \in B] = P[Z \in B], \tag{4.12}$$

which implies the second part of (4.3).

Let $(\hat{\mu}_n, \hat{\sigma}_n^2)$ be a discretized robust initial estimate of (μ, σ^2) which has the tightness property required in the test constructions of Sect. 3. Many M-

estimates with bounded score function will serve. Note that η_θ for the normal model satisfies (3.4). Thus, an asymptotically minimax estimates test for location in the normal model is to reject H_n if

$$n\hat{\sigma}_n^{-2} T_{n,1}^2 > d_\alpha(1; 4a^2), \tag{4.13}$$

where

$$T_{n,1} = \hat{\mu}_n + n^{-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_n) w_n(x_i). \tag{4.14}$$

If the function m which enters into the definition of the window w_n happens to be $m(x) = \min\{1, x^{-1}\}$, or $m(x) = \max\{(1-x^2), 0\}$, or $m(x) = x^{-1}\sin(x)$, the corresponding location estimates $T_{n,1}$ are adaptive versions of estimates associated with the names Huber, Tukey, Andrews, respectively.

4.2 Testing Scale in the Normal Model

The statistical model $\{P_\theta\colon \theta \in \Theta\}$ is the same as in subsection 4.1. The two-sided chi-square test for the classical hypothesis $\sigma^2 = 1$ rejects if s_n^2 is too large or too small. A robust version of the problem is given by (2.6) and (2.7) with $\theta_0 = (0, 1)$, say. M_1 is now the subspace spanned by $4^{-1}(-1+x^2)(dP_{\theta_0})^{1/2}$ while M_2 is the subspace spanned by $2^{-1}x(dP_{\theta_0})^{1/2}$. The two-sided test based on s_n^2 has both of the nonrobustness properties in (4.3), for the appropriately defined H_n and K_n. To prove the first limit, consider the sequence of probabilities

$$Q_n = (1-(nu)^{-1}h) P_{\theta_0} + (2nu)^{-1} h [A(n^{1/4}u^{1/2}) + A(-n^{1/4}u^{1/2})]; \tag{4.15}$$

for the second limit, use the similar sequence obtained by replacing θ_0 with $\theta_n = (0, 1 - n^{-1/2}h)$.

An asymptotically minimax estimates test for scale in the normal model is to reject H_n if

$$2^{-1} n (T_{n,2} - 1)^2 > d_\alpha(1; 4a^2), \tag{4.16}$$

where

$$T_{n,2} = \hat{\sigma}_n^2 + n^{-1} \sum_{i=1}^{n} [(x_i - \hat{\mu}_n)^2 - \hat{\sigma}_n^2 - A_n] w_n(x_i) \tag{4.17}$$

and

$$A_n = [\int w_n(t) dP_{\hat{\theta}_n}]^{-1} \int [(t-\hat{\mu}_n)^2 - \hat{\sigma}_n^2] w_n(t) dP_{\hat{\theta}_n} \tag{4.18}$$

for $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$. The estimate $T_{n,2}$ is an adaptive one-step M-estimate of σ^2.

4.3 Testing Goodness-of-fit in the Multinomial Model

Suppose P_θ is a discrete distribution supported on $k+1$ points with probabilities $\{\theta_i;\ 1 \leq i \leq k+1\}$; thus $\theta_{k+1} = 1 - \sum_{j=1}^{k} \theta_i$. Let $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ and let the parameter space be the interior of the k-dimensional unit simplex. If $x=(x^{(1)}, x^{(2)}, \ldots, x^{(k+1)})$ denotes a typical observation, then one of the components equals one and all the others are zero. The quadratic mean differentiability requirement (2.4) is fulfilled with $\eta_\theta(x)=(\eta_\theta^{(1)}(x), \eta_\theta^{(2)}(x), \ldots, \eta_\theta^{(k)}(x))'$ and

$$\eta_\theta^{(j)}(x)=\theta_j^{-1} x^{(j)}-\theta_{k+1}^{-1} x^{(k+1)}. \tag{4.19}$$

The Fisher information is the matrix

$$I(\theta)=\operatorname{diag}\{\theta_j^{-1};\ 1\leqq j\leqq k\}+\theta_{k+1}^{-1} e e', \tag{4.20}$$

where e is the $k\times 1$ vector whose components all equal 1.

For the robust testing problem induced by the classical simple hypothesis $\theta=\theta_0$, the subspace M_1 coincides with M, which is spanned by the components of $\eta_{\theta_0}(dP_{\theta_0})^{1/2}$. Note that $I(\theta_0)$ can be used as the matrix in the robust test statistic (3.11) and that the maximum likelihood estimate of θ can serve as T_n. Thus, $T_n=(T_{n,1}, T_{n,2}, \ldots, T_{n,k})'$ with

$$T_{n,j}=n^{-1}\sum_{i=1}^{n} x_i^{(j)}. \tag{4.21}$$

That this choice of T_n has the asymptotic behavior required in the proof of Theorem 2 can be verified directly in this case, because the estimate is bounded. Define $T_{n,k+1}=1-\sum_1^k T_{n,j}$. The asymptotically minimax estimates test (3.11) reduces, in the present situation, to the following procedure: reject H_n if

$$n\sum_{j=1}^{k+1} \theta_{0,j}^{-1}(T_{n,j}-\theta_{0,j})^2>d_\alpha(k;\, 4a^2). \tag{4.22}$$

Only the choice of the critical value distinguishes this robust test from the usual chi-square goodness-of-fit test. The purpose of the larger critical value is to control level under small departures from the classical null hypothesis.

4.4 Testing in the Canonical Exponential Family

Let μ be a σ-finite measure on $(\mathscr{X},\mathscr{B})$ and let $h(x)=(h_1(x), h_2(x), \ldots, h_k(x))'$ be a Borel measurable function mapping $\mathscr{X}$ into R^k. Let $r(x)=(h_1(x), h_1(x), \ldots, h_k(x), 1)'$ and suppose the functions $\{h_j\}$ are such that $\int (a'r(x))^2\, d\mu>0$ for every nonzero column vector $a\in R^k$. The canonical exponential family $\{P_\theta\colon \theta\in\Theta\}$ generated by μ and h has densities

$$\frac{dP_\theta}{d\mu}(x)=\exp[\theta' h(x)-c(\theta)], \tag{4.23}$$

where θ is a column vector in R^k, $c(\theta)$ is the density normalizing constant, and $\Theta=\operatorname{int}\{\theta\in\mathrm{R}^k\colon \exp(c(\theta))<\infty\}$. This parametric model satisfies (2.4) with $\eta_\theta(x)=2^{-1}[h(x)-E_\theta h(x)]$ and information matrix $I(\theta)=\operatorname{Cov}_\theta h(x)$. The mapping

$\theta \to P_\theta$ is one-to one (c.f. Berk (1972)). Thus, asymptotically minimax robust tests for simple and composite classical hypotheses regarding θ can be constructed as in Theorem 2. Some simplifications occur because η_θ satisfies (3.4) and $I(\theta)$ is continuous on Θ. If the exponential family basis $h(x)$ is bounded over $\mathscr{X}$, the maximum likelihood estimate of θ will serve as T_n or as θ_n^* in the two test constructions. When $h(x)$ is not bounded, there exist robust modifications of the maximum likelihood estimate which can be used as θ_n^*. For justification of the last two assertions, see Sect. 4 of Beran (1981).

References

Beran, R.: Efficient robust estimates in parametric models. Z. Wahrscheinlichkeitstheorie verw. Gebiete **55**, 91–108 (1981)

Berk, R.H.: Consistency and asymptotic normality of MLE's for exponential models. Ann. Math. Statist. **43**, 193–204 (1972)

Bickel, P.J.: Some recent developments in robust statistics. Unpublished (1978)

Hájek, J., Šidák, Z.: Theory of Rank Tests. New York-London: Academic Press 1967

Hampel, F.R.: A general qualitative definition of robustness. Ann. Math. Statist. **42**, 1887–1896 (1971)

Huber, P.J.: A robust version of the probability ratio test. Ann. Math. Statist. **36**, 1753–1758 (1965)

Huber-Carol, C.: Etude asymptotique de tests robustes. Thesis, E.T.H. Zurich (1970)

Huber, P.J., Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Statist. **1**, 251–263 (1973)

LeCam, L.: Théorie Asymptotique de la Décision Statistique. Les Presses de l'Université de Montréal (1969)

LeCam, L.: Limits of experiments. Proc. Sixth Berkeley Sympos. Math. Statist. Prob. **1**, 245–261. University of California Press (1972)

Lehmann, E.L.: Testing Statistical Hypotheses. New York: Wiley 1959

Millar, P.W.: Robust estimation via minimum distance methods. Z. Wahrscheinlichkeitstheorie verw. Gebiete **55**, 73–89 (1981)

Neyman, J.: Optimal asymptotic tests of composite statistical hypotheses. Probability and Statistics. U. Grenander, ed. New York: Wiley 1959

Rieder, H.: A robust asymptotic testing model. Ann. Statist. **6**, 1080–1094 (1978)

Wang, P.: Asymptotic robust tests in the presence of nuisance parameters. Ph. D. dissertation, University of California, Berkeley (1979)

Received October 1, 1979; in final form December 12, 1980