The Annals of Statistics
1978, Vol. 6, No. 5, 1080–1094

A ROBUST ASYMPTOTIC TESTING MODEL

BY HELMUT RIEDER

University of Freiburg and
University of California, Berkeley[1]

In order to obtain quantitative results about the influence of outliers on tests, their maximum size and minimum power over certain neighborhoods are evaluated asymptotically. The neighborhoods are defined in terms of ε-contamination and total variation, the tests considered are based on statistics $n^{-\frac{1}{2}} \sum_{i=1}^{n} IC(x_i)$, IC a rather arbitrary function.

Furthermore, the unique IC^* is determined that leads to a maximin test with respect to this subclass of tests. A comparison with the likelihood ratio of least favorable pairs shows that the test based on $n^{-\frac{1}{2}} \sum_{i=1}^{n} IC^*(x_i)$ is in fact maximin among all tests at a given level.

Tests based on (M)-statistics are also considered.

1. Introduction. Let $(\Omega, \mathscr{B})$ denote a measurable space, $\mathscr{M}$ the set of probability measures on $(\Omega, \mathscr{B})$, and let P_0, P_1 be two distinct elements of $\mathscr{M}$.

Consider the problem to decide between P_0 and P_1 looking at n independent random variables $x_1, \cdots, x_n$. Assume, however, that their distributions need not exactly coincide with either P_0 or P_1 but are only known to lie in some neighborhood $\mathscr{P}_0$ of P_0 or $\mathscr{P}_1$ of P_1.

Let $\alpha \in (0, 1)$ denote a given level, and let ϕ_n be a test based on a statistic T_n, such that $E_{P_0^{\otimes n}} \phi_n \leqq \alpha$ ($P_0^{\otimes n}$ denotes the n-fold product of P_0). Because deviations from P_0 may increase the size of ϕ_n, the critical value of T_n must be adjusted. Similarly, deviations from P_1 may decrease the power of ϕ_n, and the adjustment of the critical value will add to this decrease.

To be more precise, consider the neighborhoods

$$\mathscr{P}_j = \{Q \in \mathscr{M}\colon Q(B) \geqq (1 - \varepsilon_j)P_j(B) - \delta_j \quad \text{for all} \quad B \in \mathscr{B}\}$$

for given parameters $\varepsilon_j, \delta_j \in [0, 1]$, $0 < \varepsilon_j + \delta_j < 1$, $j = 0, 1$, and consider tests based on statistics of the form

$$T_n(IC) = n^{-\frac{1}{2}} \sum_{i=1}^{n} IC(x_i)\,,$$

IC satisfying some integrability assumptions.

After calculating the maximum size of ϕ_n, the appropriate critical value can be determined, and then the minimum power can be evaluated. This is done asymptotically, as the sample size tends to infinity (Theorem 3.4).

In the underlying asymptotic model, the centers of the neighborhoods belong

Received October 1976; revised October 1977.

[1] Now at University of California, Berkeley. Research partially supported by the Deutsche Forschungsgemeinschaft.

AMS 1970 *subject classifications.* Primary 62G35; Secondary 62E20.

Key words and phrases. Asymptotic normality, asymptotic relative efficiency, contiguity, asymptotic maximin test, least favorable pair, ε-contamination, total variation.

to a smoothly parametrized family. They approach each other, and the neighborhoods shrink at the same rate (Section 2).

If IC is unbounded, then the limit minimum power is 0 (Theorem 3.6). A criterion for unbiasedness involving bounds on IC is derived. For unbiased tests a measure of asymptotic relative efficiency is considered (Section 5).

The unique IC^* is determined that leads to a maximin test with respect to this subclass of tests (Theorem 3.7). A comparison with the likelihood ratio of least favorable pairs shows that the test based on $T_n(IC^*)$ is in fact a maximin test among all tests at level α (Theorem 4.4).

The results are extended to (M)-statistics, which do not coincide with, but can be approximated by $(T_n(IC))$ for some IC.

This paper was partly inspired by the work of C. Huber-Carol (1970), who studied the likelihood ratio of least favorable pairs in a similar asymptotic setup. Actually, she considered the special case, where $\varepsilon_j = 0$, $\delta_0 = \delta_1$, and required Ω to be the real line and the parametric family to have monotone likelihood ratio. By making use of the contiguity of least favorable pairs these restrictions can be dispensed with (Theorem 4.1).

2. The model. Let $(\Omega, \mathscr{B})$ be a measurable space and $\mathscr{M}$ the set of probability measures on $(\Omega, \mathscr{B})$. For some $\tau > 0$ let $\{P_\theta : |\theta| \leqq \tau\}$ denote a one real parameter family in $\mathscr{M}$. Parameters $\varepsilon_j, \delta_j \in [0, \infty)$, $0 < \varepsilon_j + \delta_j$, $j = 0, 1$, are given.

ASSUMPTIONS. The following regularity properties are assumed to hold throughout the paper.

(2.1) $P_\theta \ll P_0$ for all $|\theta| \leqq \tau$. Let p_θ denote a suitable version of $\dfrac{dP_\theta}{dP_0}$.

(2.2) $\theta \to p_\theta(x)$ is twice differentiable for all $x \in \Omega$.

Put $\Lambda(x) = \dfrac{\partial}{\partial\theta}\Big|_{\theta=0} \log p_\theta(x)$.

$$0 < \int \Lambda^2 \, dP_0 < \infty . \tag{2.3}$$

$$\lim_{\theta\to 0} \int \left(\frac{p_\theta^{\frac{1}{2}} - 1}{\theta}\right)^2 dP_0 = \tfrac{1}{4} \int \Lambda^2 \, dP_0 . \tag{2.4}$$

(2.5) There exists a function h in $L^1(P_0)$ such that

$$\sup_{|\theta| \leqq \tau} \left|\frac{\partial^2}{\partial\theta^2} p_\theta(x)\right| \leqq h(x) \qquad \text{for all} \quad x \in \Omega .$$

$$\frac{\varepsilon_0 + \delta_0 + \delta_1}{2\tau} < \int \left(\Lambda - \frac{\varepsilon_1 - \varepsilon_0}{2\tau}\right)^+ dP_0 . \tag{2.6}$$

NOTATION AND DEFINITIONS. Let $\mathbb{R}$ denote the real line, $\mathbb{N}$ the set of nonnegative integers. For $Q', Q'' \in \mathscr{M}$ put $d_{\mathrm{Var}}(Q', Q'') = \sup\{|Q'(B) - Q''(B)| : B \in \mathscr{B}\}$.

Define for each $n \in \mathbb{N}$ and for $j = 0, 1$

$$\tau_n = n^{-\frac{1}{2}}\tau\,, \qquad \varepsilon_{jn} = n^{-\frac{1}{2}}\varepsilon_j \qquad \text{and} \qquad \delta_{jn} = n^{-\frac{1}{2}}\delta_j \tag{2.7}$$

$$P_{0n} = P_{-\tau_n}\,, \qquad P_{1n} = P_{\tau_n} \tag{2.8}$$

$$\mathscr{P}_{jn} = \{Q \in \mathscr{M}\colon Q(B) \geqq (1 - \varepsilon_{jn})P_{jn}(B) - \delta_{jn} \ \text{ for all } \ B \in \mathscr{B}\} \tag{2.9}$$

$$\mathscr{P}_{jn}^{\otimes n} = \{\textstyle\bigotimes_{i=1}^n Q_i : Q_i \in \mathscr{P}_{jn} \ \text{ for } \ i = 1, \cdots, n\} \tag{2.10}$$

$$H_j = \{(w_n) : w_n \in \mathscr{P}_{jn}^{\otimes n} \ \text{ for all } \ n \in \mathbb{N}\}\,. \tag{2.11}$$

Let Ω^n be endowed with the product σ-field $\mathscr{B}^{\otimes n}$. For a sequence (ϕ_n) of tests $\phi_n : \Omega^n \to [0, 1]$ put

$$\begin{aligned} \alpha_n(\phi_n) &= \sup\{E_{u_n}\phi_n : u_n \in \mathscr{P}_{0n}^{\otimes n}\} \\ \beta_n(\phi_n) &= \inf\{E_{v_n}\phi_n : v_n \in \mathscr{P}_{1n}^{\otimes n}\}\,. \end{aligned} \tag{2.12}$$

Given $\alpha \in (0, 1)$. Then (ϕ_n) is called an asymptotic test for H_0 at level α iff

$$\limsup\nolimits_n \alpha_n(\phi_n) \leqq \alpha\,. \tag{2.13}$$

Denote the set of asymptotic tests for H_0 at level α by Φ_α. Then (ϕ_n^*) is called an asymptotic maximin test for H_0 versus H_1 at level α iff $(\phi_n^*) \in \Phi_\alpha$ and

$$\liminf\nolimits_n (\beta_n(\phi_n^*) - \beta_n(\phi_n)) \geqq 0 \qquad \text{for all} \quad (\phi_n) \in \Phi_\alpha\,. \tag{2.14}$$

REMARKS.

1. From (2.2) and (2.5) it follows that

$$\int \Lambda\, dP_0 = 0\,.$$

2. By Corollary 2.25 of Witting and Noelle (1970), page 66, the sequences

$$(P_{-\tau_n}^{\otimes n})\,, \qquad (P_0^{\otimes n}) \qquad \text{and} \qquad (P_{\tau_n}^{\otimes n})$$

are mutually contiguous.

3. Condition (2.4) is adapted to this corollary. It means interchangeability of the limit and the integral. It is fulfilled, for instance, if $p_\theta(x) > 0$ for all $|\theta| \leqq \tau$, $x \in \Omega$ and if the Fisher information is continuous in 0.

4. Condition (2.6) is fulfilled for small parameters ε_j, δ_j. In view of Lemma 4.3 of [8] it is equivalent to

$$\mathscr{P}_{0n} \cap \mathscr{P}_{1n} = \emptyset \qquad \text{for large} \quad n\,.$$

5. Note that

$$\mathscr{P}_{jn} = \textstyle\bigcup_{Q'} \{Q'' \in \mathscr{M}\colon d_{\mathrm{Var}}(Q', Q'') \leqq \delta_{jn}\}\,,$$

where Q' runs through $\{(1 - \varepsilon_{jn})P_{jn} + \varepsilon_{jn}H : H \in \mathscr{M}\}$. Hence the considered neighborhoods are a natural generalization of ε-contamination and total variation neighborhoods.

6. For reasons of symmetry, the parameter sequence of the null hypothesis has been chosen to be $(-\tau_n)$. The pair (0), (τ_n) could have been considered just as well, the modifications being obvious.

EXAMPLE. Exponential families satisfy the above regularity assumptions. The densities may be assumed to be of the following form:

$$p_\theta = c(\theta)e^{\theta\chi},$$

where $\int \chi \, dP_0 = 0$, $0 < \int \chi^2 \, dP_0 < \infty$.

Because $\theta \to \int e^{\theta\chi} \, dP_0$ is analytic on the interior Z^0 of the natural parameter space Z, the function $c(\theta)$ and in particular the expectation and variance of χ with respect to P_θ are continuous and bounded on $[-\tau, \tau]$, provided that $[-\tau, \tau] \subset Z^0$.

Then the supremum b of $c(\theta) + |E_\theta \chi|^2 + \mathrm{Var}_\theta \chi$ over $|\theta| \leqq \tau$ is finite, and h may be chosen to be

$$h = b(e^{-\tau\chi} + e^{\tau\chi})(2\chi^2 + 2b^2 + b). \qquad \square$$

3. General framework. We consider the subclass of asymptotic tests (ϕ_n) that are based on a special kind of test statistics:

$$\begin{aligned} \phi_n &= 1 && \text{if} \quad T_n(IC) > k_n \\ &= \gamma_n && \text{if} \quad T_n(IC) = k_n \\ &= 0 && \text{if} \quad T_n(IC) < k_n , \end{aligned} \tag{3.1}$$

where $\gamma_n \in [0, 1]$, $k_n \in \mathbb{R}$ and

$$T_n(IC) = n^{-\frac{1}{2}} \sum_{i=1}^n IC(x_i) \qquad x_1, \cdots, x_n \in \Omega . \tag{3.2}$$

We shall refer to (ϕ_n) as an asymptotic test based on $(T_n(IC))$. The function $IC \colon \Omega \to \mathbb{R}$ is assumed to be measurable, bounded and to satisfy

$$\int IC \, dP_0 = 0 , \qquad \int IC^2 \, dP_0 \in (0, \infty) . \tag{3.3}$$

First, for an appropriate choice of the critical values k_n, the limits of the maximum size α_n and the minimum power β_n will be evaluated. Then the maximization problem expressed in (2.14) under the side condition (2.13) will be solved within this subclass.

To begin with, observe the following asymptotic normality.

PROPOSITION 3.1. *For all $(w_n) \in H_0 \cup H_1$ it holds that*

$$\mathscr{L}_{w_n}(T_n(IC) - E_{w_n} T_n(IC)) \Rightarrow \mathscr{N}(0, E_{P_0} IC^2) .$$

PROOF. Note that

$$\limsup_n n^{\frac{1}{2}} \sup \{d_{\mathrm{Var}}(Q_n, P_0) \colon Q_n \in \mathscr{P}_{0n} \cap \mathscr{P}_{1n}\} < \infty .$$

Because of the boundedness of IC we have

$$\lim_n \mathrm{Var}_{Q_n} IC = E_{P_0} IC^2$$

uniformly in $Q_n \in \mathscr{P}_{0n} \cup \mathscr{P}_{1n}$.

The Lindeberg condition is trivially fulfilled. $\square$

By means of the next two lemmas the stochastically extreme limits of $(T_n(IC))$ can be determined. Denote by P^{IC} the image law of IC under P, and by $(P^{IC})^{-1}$

the left-continuous pseudoinverse of the distribution function of P^{IC}. Extend it to 0 and 1 by taking the one-sided limits. Then $(P^{IC})^{-1}(0) = \inf_{[P]} IC$, the essential infimum of IC with respect to P, and $(P^{IC})^{-1}(1) = \sup_{[P]} IC$, the essential supremum of IC with respect to P.

Define

$$
\begin{aligned}
a_{0n} &= (P_{0n}^{IC})^{-1}\left(\frac{\delta_{0n}}{1-\varepsilon_{0n}}\right) \\
a_{1n} &= (P_{1n}^{IC})^{-1}\left(\frac{1-(\varepsilon_{1n}+\delta_{1n})}{1-\varepsilon_{1n}}\right) \qquad (3.4)\\
\tilde{v}_{0n}(IC) &= (1-\varepsilon_{0n}) \int IC \vee a_{0n}\, dP_{0n} - \delta_{0n} a_{0n} + (\varepsilon_{0n}+\delta_{0n}) \sup IC \\
\tilde{u}_{1n}(IC) &= (1-\varepsilon_{1n}) \int IC \wedge a_{1n}\, dP_{1n} + (\varepsilon_{1n}+\delta_{1n}) \inf IC - \delta_{1n} a_{1n}\,,
\end{aligned}
$$

where $x \vee y = \max\{x, y\}$, $x \wedge y = \min\{x, y\}$, $\sup IC = \sup\{IC(x) : x \in \Omega\}$, $\inf IC = \inf\{IC(x) : x \in \Omega\}$.

The following lemma is related to Lemma 2.4 of Huber and Strassen (1973) and generalizes Lemma 4.2 of [8].

LEMMA 3.2. *It holds that*

$$
\begin{aligned}
\sup\{\textstyle\int IC\, dQ' : Q' \in \mathscr{P}_{0n}\} &= \tilde{v}_{0n}(IC) \\
\inf\{\textstyle\int IC\, dQ'' : Q'' \in \mathscr{P}_{1n}\} &= \tilde{u}_{1n}(IC)\,.
\end{aligned}
$$

PROOF. The indices may be dropped from the notation. Then let $v : \mathscr{B} \to [0, 1]$ be defined by

$$
\begin{aligned}
v(B) &= ((1-\varepsilon)P(B) + \varepsilon + \delta) \wedge 1 \qquad &&\text{if } B \neq \emptyset \\
&= 0 &&\text{if } B = \emptyset\,.
\end{aligned}
$$

In this proof it is no restriction to assume that IC is nonnegative. Then the integral

$$\int_0^\infty v(IC > t)\, dt$$

is an upper bound for $\int IC\, dQ'$, $Q' \in \mathscr{P}$, and, by a straightforward computation, can be shown to equal $\tilde{v}(IC)$, defined by (3.4). It remains to prove that this upper bound is approximated by integrals $\int IC\, dQ'$, $Q' \in \mathscr{P}$.

Define

$$
\begin{aligned}
\gamma &= (P(IC = a))^{-1}\left(P(IC \leq a) - \frac{\delta}{1-\varepsilon}\right) \qquad &&\text{if } P(IC = a) > 0 \\
&= 0 &&\text{if } P(IC = a) = 0\,.
\end{aligned}
$$

Choose $x_0 \in \Omega$ such that $IC(x_0) > a$, and let I_{x_0} denote the one point mass in x_0. Consider the probability measure R defined by

$$R(B) = \gamma(1-\varepsilon)P(B \cap \{IC = a\}) + (1-\varepsilon)P(B \cap \{IC > a\}) + (\varepsilon+\delta)I_{x_0}(B) \qquad B \in \mathscr{B}\,.$$

Then $R \in \mathscr{P}$ and

$$\int IC\, dR = (1-\varepsilon) \int IC \vee a\, dP - \delta a + (\varepsilon + \delta) IC(x_0)\,.$$

Letting $IC(x_0)$ tend to sup IC proves the first equality. The other one is obtained similarly. □

Define

$$s_0(IC) = -\frac{\delta_0}{\tau}\inf_{[P_0]} IC + \frac{\varepsilon_0 + \delta_0}{\tau}\sup IC \tag{3.5}$$

$$s_1(IC) = -\frac{\varepsilon_1 + \delta_1}{\tau}\inf IC + \frac{\delta_1}{\tau}\sup_{[P_0]} IC\,.$$

LEMMA 3.3. *It holds that*

$$\lim_n n^{\frac{1}{2}}\tilde{v}_{0n}(IC) = -\tau(\textstyle\int IC\Lambda\, dP_0 - s_0(IC))$$
$$\lim_n n^{\frac{1}{2}}\tilde{u}_{1n}(IC) = \tau(\textstyle\int IC\Lambda\, dP_0 - s_1(IC))\,.$$

PROOF. We may confine ourselves to the proof of the first equality. The crucial point is to show that

$$\lim_n a_{0n} = (P_0^{IC})^{-1}(0)$$

and

$$\lim_n n^{\frac{1}{2}} \textstyle\int_{\{IC<a_{0n}\}} (a_{0n} - IC)\, dP_{0n} = 0\,.$$

Put $\eta_{0n} = d_{\mathrm{Var}}(P_{0n}, P_0)$ and $\gamma_{0n} = \delta_{0n}/(1-\varepsilon_{0n})$. Without restriction $\gamma_{0n} > 0$. Then also $P_0(IC \leqq a_{0n}) > 0$, hence $a_{0n} \geqq (P_0^{IC})^{-1}(0)$. On the other hand, a_{0n} is not greater than $(P_0^{IC})^{-1}(\gamma_{0n} + \eta_{0n})$, which converges to $(P_0^{IC})^{-1}(0)$.

To prove the second convergence it is sufficient to consider

$$n^{\frac{1}{2}} \textstyle\int_{\{IC<a_{0n}\}} (a_{0n} - IC)\, dP_0\,.$$

An upper bound is

$$n^{\frac{1}{2}}(\gamma_{0n} + \eta_{0n})(a_{0n} - (P_0^{IC})^{-1}(0))\,,$$

which tends to 0. □

Define

$$s(IC) = \tau(E_{P_0} IC^2)^{-\frac{1}{2}}(2\textstyle\int IC\Lambda\, dP_0 - (s_0(IC) + s_1(IC)))\,, \tag{3.6}$$

and for $\alpha \in (0, 1)$ define

$$k_\alpha(IC) = (E_{P_0} IC^2)^{\frac{1}{2}} u_\alpha - \tau(\textstyle\int IC\Lambda\, dP_0 - s_0(IC))\,, \tag{3.7}$$

where u_α is the upper α-point of the standard normal Φ.

Then, if (ϕ_n) is the asymptotic test (3.1) based on $(T_n(IC))$, the preceding results imply the following theorem.

THEOREM 3.4. *Provided that the critical values k_n tend to $k_\alpha(IC)$, we have*

$$\lim_n \alpha_n(\phi_n) = \alpha$$
$$\lim_n \beta_n(\phi_n) = 1 - \Phi(u_\alpha - s(IC))\,.$$

So far, IC has been assumed to be bounded. One might try to extend this result to unbounded IC.

Assume that

$$\textstyle\int IC^2(|\Lambda| + h)\, dP_0 < \infty\,. \tag{3.8}$$

THEOREM 3.5. *If* $\sup IC = \infty$, *there exists a sequence* $(R_{0n}^{\otimes n})$ *in* H_0 *such that*

$$\mathscr{L}_{R_{0n}^{\otimes n}}(T_n(IC)) \Rightarrow I_\infty \quad (\textit{one point mass in } \infty)\,.$$

If $\inf IC = -\infty$, *there exists a sequence* $(R_{1n}^{\otimes n})$ *in* H_1 *such that*

$$\mathscr{L}_{R_{1n}^{\otimes n}}(T_n(IC)) \Rightarrow I_{-\infty} \quad (\textit{one point mass in } -\infty)\,.$$

PROOF. If $\sup IC = \infty$, choose x_{0n} in Ω such that

$$|IC(x_{0n})|^2 = o(n^{\frac{1}{2}})\,, \qquad IC(x_{0n}) \to \infty \quad \text{as} \quad n \to \infty\,.$$

Define

$$R_{0n} = (1 - (\varepsilon_{0n} + \delta_{0n}))P_{0n} + (\varepsilon_{0n} + \delta_{0n})I_{x_{0n}}\,.$$

Observe that

$$\lim_n \operatorname{Var}_{R_{0n}} IC = E_{P_0} IC^2\,,$$

whereas

$$\lim_n n^{\frac{1}{2}} E_{R_{0n}} IC = \infty\,.$$

Furthermore, the Lindeberg condition is fulfilled. The proof of the second statement is similar. □

As a consequence, we obtain the following result, which complements Theorem 3.4, since $s(IC) = -\infty$ if IC is unbounded.

THEOREM 3.6. *Let* IC *be unbounded* (*satisfying* (3.3) *and* (3.8)). *Let* (ϕ_n) *be the asymptotic test based on* $(T_n(IC))$ *satisfying* $\limsup_n \alpha_n(\phi_n) < 1$.

Then we have

$$\lim_n \beta_n(\phi_n) = 0\,.$$

PROOF. Obvious, since the sequences $\{\mathscr{L}_{P_{jn}^{\otimes n}}(T_n(IC)) : n \in \mathbb{N}\}$, $j = 0, 1$, are tight. □

EXAMPLE. For quite a few one parameter exponential families on the real line, including the normal location model, the gamma and the Poisson model, the logarithmic derivative Λ is unbounded.

Hence in view of Theorem 3.6 the asymptotic test based on $(T_n(\Lambda))$, which is optimal for $(P_{0n}^{\otimes n})$ versus $(P_{1n}^{\otimes n})$, ends up with error probability 1. □

We arrive at the problem of maximizing $s(IC)$ under the side conditions (3.3) and (3.8).

Consider the equations

$$\int (d_0 - \Lambda)^+ \, dP_0 = \frac{\varepsilon_1 + \delta_0 + \delta_1}{2\tau} \tag{3.9}$$

$$\int (\Lambda - d_1)^+ \, dP_0 = \frac{\varepsilon_0 + \delta_0 + \delta_1}{2\tau}\,.$$

Note that the solutions d_j, $j = 0, 1$, exist and are unique. Define

$$IC^* = (d_0 \vee \Lambda \wedge d_1) - \frac{\varepsilon_1 - \varepsilon_0}{2\tau}\,. \tag{3.10}$$

Note also that IC^* satisfies the side conditions.

THEOREM 3.7. *For each function IC satisfying* (3.3) *and* (3.8) *it holds that*

$$s(IC) \leqq s(IC^*)\,.$$

Equality implies that

$$IC = \gamma IC^* \quad P_0\text{-a.e.}$$

for some constant $\gamma > 0$.

REMARK. There seems to be a relation between this theorem and Lemma 5 of Hampel (1968), page 51, if IC denotes the influence curve of an (M)-estimate.

PROOF. Write

$$\int IC\Lambda\, dP_0 = \int IC IC^*\, dP_0 + \int IC(\Lambda - IC^*)\, dP_0\,.$$

Apply the Cauchy–Schwarz inequality to the first integral on the right-hand side. Observe that

$$\int IC(\Lambda - IC^*)\, dP_0 \leqq \tfrac{1}{2}(s_0(IC) + s_1(IC)) \tag{3.11}$$

with equality holding for $IC = IC^*$, such that in particular

$$s^2(IC^*) = 4\tau^2 E_{P_0}(IC^*)^2\,. \tag{3.12}$$

□

4. The likelihood ratio of least favorable pairs. According to Huber (1965), (1968) there exists a least favorable pair $(Q_{0n}, Q_{1n} \mid \pi_n)$ for $(\mathscr{P}_{0n}, \mathscr{P}_{1n})$ (notation adapted to [8]).

Using the log likelihood

$$\textstyle\sum_{i=1}^n \log \pi_n(x_i) \qquad x_1, \cdots, x_n \in \Omega$$

one obtains an asymptotic maximin test for H_0 versus H_1 at level α if the critical values are chosen appropriately.

Let us recall that the special version π_n of the likelihood ratio dQ_{1n}/dQ_{0n} is given by

$$\pi_n = \frac{1-\varepsilon_{1n}}{1-\varepsilon_{0n}}\left(\Delta_{0n} \vee \frac{dP_{1n}}{dP_{0n}} \wedge \Delta_{1n}\right).$$

The truncation points Δ_{jn} are the unique solutions of the equations

$$\Delta_{0n} P_{0n}\left(\frac{dP_{1n}}{dP_{0n}} < \Delta_{0n}\right) - P_{1n}\left(\frac{dP_{1n}}{dP_{0n}} < \Delta_{0n}\right) = \frac{\varepsilon_{1n} + \delta_{1n}}{1-\varepsilon_{1n}} + \frac{\delta_{0n}}{1-\varepsilon_{0n}}\Delta_{0n}$$

$$P_{1n}\left(\frac{dP_{1n}}{dP_{0n}} > \Delta_{1n}\right) - \Delta_{1n} P_{0n}\left(\frac{dP_{1n}}{dP_{0n}} > \Delta_{1n}\right) = \frac{\varepsilon_{0n} + \delta_{0n}}{1-\varepsilon_{0n}}\Delta_{1n} + \frac{\delta_{1n}}{1-\varepsilon_{1n}}\,.$$

Put

$$L_n = \frac{1}{2\tau}\textstyle\sum_{i=1}^n \log \pi_n(x_i) \qquad x_1, \cdots, x_n \in \Omega\,.$$

Recall the definition (3.10) of IC^*; put $\sigma_*^2 = E_{P_0}(IC^*)^2$.

THEOREM 4.1. *The following asymptotic normality holds*:

$$\mathscr{L}_{Q_{0n}^{\otimes n}}(L_n) \Longrightarrow \mathscr{N}(-\tau\sigma_*^2, \sigma_*^2)$$
$$\mathscr{L}_{Q_{1n}^{\otimes n}}(L_n) \Longrightarrow \mathscr{N}(+\tau\sigma_*^2, \sigma_*^2)\,.$$

The proof is based on two lemmas.

LEMMA 4.2. *The sequences* $(Q_{jn}^{\otimes n})$, $j = 0, 1$, *are mutually contiguous.*

PROOF. If, say, $(Q_{1n}^{\otimes n})$ were not contiguous to $(Q_{0n}^{\otimes n})$, there would exist some sequence (B_n), $B_n \in \mathscr{B}^{\otimes n}$, such that

$$\lim Q_{0n}^{\otimes n}(B_n) = 0$$
$$\limsup_n Q_{1n}^{\otimes n}(B_n) > 0 .$$

Let ϕ_n denote the Neyman–Pearson test for $Q_{0n}^{\otimes n}$ versus $Q_{1n}^{\otimes n}$, based on L_n, at level $Q_{0n}^{\otimes n}(B_n)$. Then, because of the inequalities

$$\int \phi_n \, dP_{0n}^{\otimes n} \leqq \int \phi_n \, dQ_{0n}^{\otimes n} = Q_{0n}^{\otimes n}(B_n)$$
$$\int \phi_n \, dP_{1n}^{\otimes n} \geqq \int \phi_n \, dQ_{1n}^{\otimes n} \geqq Q_{1n}^{\otimes n}(B_n) ,$$

it follows that ϕ_n tends to zero in $P_{0n}^{\otimes n}$- but not in $P_{1n}^{\otimes n}$-probability—contradicting the contiguity of $(P_{1n}^{\otimes n})$ to $(P_{0n}^{\otimes n})$. □

LEMMA 4.3. *For all* $(w_n) \in H_0 \cup H_1$ *we have*

$$\lim_n \operatorname{Var}_{w_n} L_n = \sigma_*^{\,2} .$$

PROOF. Take $p_{\tau_n}/p_{-\tau_n}$ as a version of dP_{1n}/dP_{0n}. Put

$$d_{jn} = \frac{1}{2\tau_n} \log \Delta_{jn} .$$

It turns out to be sufficient to prove that

$$\lim_n d_{jn} = d_j . \tag{*}$$

For this implies that

$$\lim_n \frac{1}{2\tau_n} \log \pi_n(x) = IC^*(x) \qquad x \in \Omega$$

and

$$\limsup_n \sup \left\{ \frac{1}{2\tau_n} \log \pi_n(x) : x \in \Omega \right\} < \infty .$$

Because of

$$\limsup_n n^{\frac{1}{2}} \sup \{ d_{\mathrm{Var}}(Q_n, P_0) : Q_n \in \mathscr{P}_{0n} \cup \mathscr{P}_{1n} \} < \infty$$

we may then conclude that

$$\lim_n \operatorname{Var}_{Q_n} \frac{1}{2\tau_n} \log \pi_n = \sigma_*^{\,2} ,$$

uniformly in $Q_n \in \mathscr{P}_{0n} \cup \mathscr{P}_{1n}$.

Let us prove (*) in the case $j = 0$ (the case $j = 1$ can be treated similarly).

For $d \in \mathbb{R}$, $u > 0$, $x \in \Omega$ define the functions

$$f(d, u, x) = \frac{e^{2ud} p_{-u}(x) - p_u(x)}{u(1 + e^{2ud})}$$
$$F(d, u) = \int f(d, u, x)^+ \, dP_0(x)$$
$$G(d) = \int (d - \Lambda(x))^+ \, dP_0(x) .$$

By means of Taylor expansions, based on Assumptions (2.2) and (2.5), one obtains

$$(**)\qquad \lim_{u\to 0} F(d, u) = G(d) \qquad d \in \mathbb{R}\,.$$

Note furthermore that $d \to F(d, u)$ is isotone[2] and that G is strictly isotone in a neighborhood of d_0.

Note also that d_{0n} and d_0 are given by

$$F(d_{0n}, \tau_n) = \frac{\varepsilon_1 + \delta_1}{\tau(1 - \varepsilon_{1n})(1 + e^{2\tau_n d_{0n}})} + \frac{\delta_0 e^{2\tau_n d_{0n}}}{\tau(1 - \varepsilon_{0n})(1 + e^{2\tau_n d_{0n}})}$$

$$G(d_0) = \frac{\varepsilon_1 + \delta_0 + \delta_1}{2\tau}\,.$$

The sequence (d_{0n}) is bounded. It is bounded by above, since $\Delta_{0n} \leqq 1$ for large n, which is a consequence of $\mathscr{P}_{0n} \cap \mathscr{P}_{1n} = \emptyset$ and Lemma 4.3, Lemma 4.4 of [8]. On the other hand, convergence of a subsequence $(d_{0\nu})$ to $-\infty$ implies for each $r \in \mathbb{R}$ that

$$0 < \frac{\varepsilon_1 + \delta_1}{2\tau} \leqq \liminf_\nu F(d_{0\nu}, \tau_\nu) \leqq \lim_\nu F(r, \tau_\nu) = G(r)\,.$$

However $\lim_{r\to-\infty} G(r) = 0$.

Boundedness of (d_{0n}) implies that

$$\lim_n F(d_{0n}, \tau_n) = G(d_0)\,.$$

Then $\lim_n d_{0n} = d_0$ follows, because of $(**)$ and the isotony properties of F and G. □

Proof of Theorem 4.1. Because of uniformly bounded summands and Lemma 4.3 the Lindeberg condition is trivially fulfilled, and we obtain

$$\mathscr{L}_{w_n}(L_n - E_{w_n} L_n) \Rightarrow \mathscr{N}(0, \sigma_*^2)\,.$$

Then, if b denotes a cluster point of $(E_{Q_{0n}^{\otimes n}} L_n)$, we may conclude that

$$\mathscr{L}_{Q_{0n}^{\otimes n}}\left(\log \frac{dQ_{1n}^{\otimes n}}{dQ_{0n}^{\otimes n}}\right) \Rightarrow \mathscr{N}(2\tau b, 4\tau^2\sigma_*^2)\,.$$

Denote this limit normal by H_b.

Because $(Q_{1n}^{\otimes n})$ is contiguous to $(Q_{0n}^{\otimes n})$, the equality

$$\int e^t\, dH_b(t) = 1$$

must hold. This implies

$$b = -\tau\sigma_*^2\,.$$

Hence we have shown that

$$\lim_n E_{Q_{0n}^{\otimes n}} L_n = -\tau\sigma_*^2\,.$$

[2] Monotone increasing.

By symmetry, it follows that

$$\lim_n E_{Q_{1n}^{\otimes n}} L_n = \tau\sigma_*^2 ,$$

and the proof is complete. □

As a consequence of Theorem 4.1, Theorem 3.4 and the equality $s(IC^*) = 2\tau\sigma_*$, the following statement is true.

THEOREM 4.4. *Let* (ϕ_n^*) *be the asymptotic test based on* $(T_n(IC^*))$, *where* IC^* *is defined by* (3.10); *the critical values* k_n *are supposed to tend to* $k_\alpha(IC^*)$, *given by* (3.7).

Then (ϕ_n^*) *is an asymptotic maximin test for* H_0 *versus* H_1 *at level* α.

5. Unbiasedness and efficiency. Let IC_l satisfy (3.3), (3.8) and let $(\phi_{n,l})$ be based on $(T_n(IC_l))$ such that

$$\limsup \alpha_n(\phi_{n,l}) = \alpha , \qquad l = 1, 2 .$$

Then $(\phi_{n,1})$ is called unbiased iff

$$\liminf_n \beta_n(\phi_{n,1}) \geqq \alpha .$$

Recall the definitions (3.5), (3.6) of s_0, s_1, s.

THEOREM 5.1. $(\phi_{n,1})$ *is unbiased* iff

$$s_0(IC_1) + s_1(IC_1) \leqq 2 \int IC_1 \Lambda \, dP_0 .$$

PROOF. Obvious, in view of Theorem 3.4 and Theorem 3.6. □

Let $(\phi_{n,l})$ be unbiased, $l = 1, 2$.

DEFINITION 5.2. The asymptotic relative efficiency of $(\phi_{n,2})$ with respect to $(\phi_{n,1})$, denoted by $\mathrm{ARE}_{\mathrm{mx}}((\phi_{n,2}):(\phi_{n,1}))$ is defined by

$$\mathrm{ARE}_{\mathrm{mx}}((\phi_{n,2}):(\phi_{n,1})) = \frac{s^2(IC_2)}{s^2(IC_1)} .$$

REMARKS.

1. The usual conventions about dividing by 0 are presumed.
2. $\mathrm{ARE}_{\mathrm{mx}}$ is a generalization of the Pitman efficiency. It allows the same interpretation in terms of the sample size n and the minimum power β_n.

We are going to compare the asymptotic relative efficiencies of the asymptotic tests based on $(T_n(IC))$ and $(T_n(IC^*))$, respectively, when either deviations from the parametric model are allowed or not.

To be more precise, denote by (ϕ_n) the asymptotic test based on $(T_n(IC))$, such that $\lim_n \alpha_n(\phi_n) = \alpha$. Let (ϕ_n^*) denote the asymptotic test based on $(T_n(IC^*))$, such that $\lim_n \alpha_n(\phi_n^*) = \alpha$.

On the other hand, let $(\hat{\phi}_n)$ be the asymptotic test based on $(T_n(IC))$, such that $\lim_n E_{P_{0n}^{\otimes n}} \hat{\phi}_n = \alpha$, and let $(\hat{\phi}_n^*)$ be the asymptotic test based on $(T_n(IC^*))$, such that $\lim_n E_{P_{0n}^{\otimes n}} \hat{\phi}_n^* = \alpha$.

Note that the Pitman efficiency of $(\hat{\phi}_n^*)$ with respect to $(\hat{\phi}_n)$ under $(P_{1n}^{\otimes n})$, denoted by ARE $((\hat{\phi}_n^*) : (\hat{\phi}_n))$, is given by

$$\mathrm{ARE}\,((\hat{\phi}_n^*) : (\hat{\phi}_n)) = \frac{(\int IC^* IC \, dP_0)^2}{E_{P_0}(IC^*)^2 E_{P_0}(IC)^2} \,.$$

The following theorem shows that even in the case $s(IC) \geqq 0$ the asymptotic relative efficiency of (ϕ_n) with respect to (ϕ_n^*) is smaller than the Pitman efficiency of $(\hat{\phi}_n^*)$ with respect to $(\hat{\phi}_n)$ under $(P_{1n}^{\otimes n})$ and thus may shed some light on the stability of tests based on $(T_n(IC^*))$.

THEOREM 5.3. *Assume that* $s(IC) \geqq 0$. *Then we have*

$$\mathrm{ARE}_{\mathrm{mx}}\,((\phi_n) : (\phi_n^*)) \leqq \mathrm{ARE}\,((\hat{\phi}_n^*) : (\hat{\phi}_n)) \,.$$

PROOF. This follows from (3.11) and (3.12). □

EXAMPLE. Let P_0 denote a probability measure on $\mathbb{R}$ with continuous distribution function F_0. Consider the densities

$$p_\theta(x) = 1 + \theta(F_0(x) - \tfrac{1}{2}) \qquad x \in \mathbb{R},\ |\theta| \leqq 2 \,.$$

Put $\tau = 2$. Given $\delta \in (0, \frac{1}{4})$, put $\varepsilon_0 = \varepsilon_1 = 0$, $\delta_0 = \delta_1 = \delta$. Consider $IC = \Lambda$. Then (ϕ_n) is unbiased iff $\delta \leqq \frac{1}{6}$. Furthermore, we obtain

$$IC^*(x) = (-\tfrac{1}{2} + \delta^{\frac{1}{2}}) \vee (F_0(x) - \tfrac{1}{2}) \wedge (\tfrac{1}{2} - \delta^{\frac{1}{2}})$$

and

$$\mathrm{ARE}_{\mathrm{mx}}\,((\phi_n) : (\phi_n^*)) = 1 - 16\delta^{\frac{3}{2}} + o(\delta^{\frac{3}{2}})$$
$$\mathrm{ARE}\,((\hat{\phi}_n^*) : (\hat{\phi}_n)) = 1 - 8\delta^{\frac{3}{2}} + o(\delta^{\frac{3}{2}}) \qquad \text{as} \quad \delta \to 0 \,.$$

□

6. (M)**-statistics.** Test statistics of a more general kind can often be approximated in $(P_0^{\otimes n})$-probability by the sequence $(T_n(IC))$ for some suitable IC, which may be obtained by projection or differentiation. This approximation holds then also under other sequences of probability measures, provided that they are contiguous to $(P_0^{\otimes n})$.

However, under rather weak assumptions, there exist sequences in $H_0 \cup H_1$ which are not contiguous to $(P_0^{\otimes n})$.

PROPOSITION 6.1. *If there is an unbounded measurable function* $f : \Omega \to [0, \infty)$ *that satisfies*

$$\int f(1 + |\Lambda| + h) \, dP_0 < \infty$$

or if there is a $B_0 \in \mathscr{B}$, $B_0 \neq \emptyset$ *such that*

$$P_0(B_0) = 0 \,,$$

then there exists a sequence in $H_0 \cup H_1$ *which is not contiguous to* $(P_0^{\otimes n})$.

REMARK. For every infinite Ω the assumptions are fulfilled (provided that $\{x\} \in \mathscr{B}$ for every $x \in \Omega$).

PROOF. Under the first assumption, put

$$IC = f^{\frac{1}{2}} - E_{P_0} f^{\frac{1}{2}} \,.$$

Then the assertion follows in view of Theorem 3.5 and the tightness of $\{\mathscr{L}_{P_0^{\otimes n}}(T_n(IC)) : n \in \mathbb{N}\}$.

Under the second assumption, choose $x_0 \in B_0$ and consider

$$R_{0n} = (1 - (\varepsilon_{0n} + \delta_{0n}))P_{0n} + (\varepsilon_{0n} + \delta_{0n})I_{x_0}$$
$$B_n = (\Omega \backslash B_0)^n .$$

Note that

$$\lim_n R_{0n}^{\otimes n}(B_n) = 0 ,$$

whereas

$$P_0^{\otimes n}(B_n) = 1 \qquad \text{for all} \quad n \in \mathbb{N} . \qquad \square$$

Nevertheless, the results of Section 3 can be extended to special other test statistics.

We confine ourselves here to (M)-statistics. Given a function $\psi: \mathbb{R} \to \mathbb{R}$, consider a measurable solution M_n of the equation

$$\sum_{i=1}^n \psi(\Lambda(x_i) - M_n) = 0 \qquad x_1, \cdots, x_n \in \Omega .$$

The sequence (M_n) is termed (M)-statistic induced by ψ.

By the arguments of Huber (1964), M_n may be thought of as a robustified version of $n^{-1} \sum_{i=1}^n \Lambda(x_i)$.

To be more precise, assume that

ψ is isotone, bounded and uniformly continuous;

$\int \psi \circ \Lambda \, dP_0 = 0$, $\quad \int \psi^2 \circ \Lambda \, dP_0 \in (0, \infty)$;

the function $\xi \to \int \psi(\Lambda(x) + \xi) \, dP_0(x)$ has a strictly positive derivative in 0, denoted by λ.

Define the asymptotic test (ϕ_n) based on (M_n) to be of the form (3.1) with $T_n(IC)$ replaced by $\lambda n^{\frac{1}{2}} M_n$. Recall the definitions (3.6) and (3.7) of $s(IC)$ and $k_\alpha(IC)$.

THEOREM 6.2. *Put $IC = \psi \circ \Lambda$. Then, if the critical values k_n tend to $k_\alpha(IC)$, we have*

$$\lim_n \alpha_n(\phi_n) = \alpha$$
$$\lim_n \beta_n(\phi_n) = 1 - \Phi(u_\alpha - s(IC)) .$$

PROOF. Because of the set inclusions

$$\{\sum_{i=1}^n \psi(\Lambda(x_i) - t) < 0\} \subset \{M_n < t\} \subset \{\sum_{i=1}^n \psi(\Lambda(x_i) - t) \leqq 0\}$$

the study of $(n^{\frac{1}{2}} M_n)$ can be reduced to the investigation of $(T_n(IC_n))$, where for fixed $t \in \mathbb{R}$

$$IC_n(x) = \psi(\Lambda(x) - n^{-\frac{1}{2}} t) \qquad x \in \Omega .$$

By the Lindeberg–Feller theorem we obtain

$$\mathscr{L}_{w_n}(T_n(IC_n) - E_{w_n} T_n(IC_n)) \Rightarrow \mathscr{N}(0, E_{P_0} IC^2)$$

for all $(w_n) \in H_0 \cup H_1$.

Recall the definition (3.4) of $\tilde{v}_{0n}(IC_n)$ and $\tilde{u}_{1n}(IC_n)$. Then by the arguments of the proof to Lemma 3.3 and, furthermore, by exploiting the uniform approximation of IC by IC_n, hence in particular of the pseudoinverse $(Q^{IC})^{-1}$ by $(Q^{IC_n})^{-1}$ for each $Q \in \mathscr{M}$, we obtain

$$\lim_n n^{\frac{1}{2}}\tilde{v}_{0n}(IC_n) = -\lambda t - \tau(\textstyle\int IC\Lambda\, dP_0 - s_0(IC))$$
$$\lim_n n^{\frac{1}{2}}\tilde{u}_{1n}(IC_n) = -\lambda t + \tau(\textstyle\int IC\Lambda\, dP_0 - s_1(IC))\,.$$

From this the theorem follows. □

Assume in addition that

$$P_0(\Lambda = d_0) = 0 = P_0(\Lambda = d_1)$$
$$P_0(d_0 < \Lambda < d_1) > 0\,.$$

Then the function

$$\psi^*(s) = (d_0 \vee s \wedge d_1) - \frac{\varepsilon_1 - \varepsilon_0}{2\tau} \qquad s \in \mathbb{R}$$

has all required properties. Hence the (M)-statistic induced by ψ^* leads to an asymptotic maximin test.

Via influence curves there corresponds to this (M)-statistic the sequence of trimmed means

$$n^{-1} \sum \left\{(\Lambda x)_{(i)} : F_0^{\Lambda}(d_0) \leqq \frac{i}{n+1} \leqq F_0^{\Lambda}(d_1)\right\}$$

where the $(\Lambda x)_{(i)}$ are the ordered values of $\Lambda(x_1), \cdots, \Lambda(x_n)$ and F_0^{Λ} is the distribution function of P_0^{Λ}.

It may be expected that they also lead to an asymptotic maximin test, however the author has not succeeded in treating them directly.

Acknowledgment. This paper arose from a portion of my doctoral dissertation written at the University of Freiburg. I would like to thank Professor H. Witting for his generous guidance and encouragement. I am also thankful to the referees for their valuable comments and suggestions.

REFERENCES

[1] HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Unpublished dissertation, Univ. California, Berkeley.

[2] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

[3] HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

[4] HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 269–278.

[5] HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

[6] HUBER-CAROL, C. (1970). Etude asymptotique de tests robustes. Thèse, ETH Zürich.

[7] RIEDER, H. (1976). Zur finiten und asymptotischen Theorie robuster Tests. Dissertation, Univ. Freiburg.

[8] RIEDER, H. (1977). Least favorable pairs for special capacities. *Ann. Statist.* **5** 909–922.
[9] WITTING, H. and NOELLE, G. (1970). *Angewandte Mathematische Statistik*. Teubner, Stuttgart.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720