Marie Hušková

Asymptotic properties of some tests under almost regular assumptions

Commentationes Mathematicae Universitatis Carolinae, Vol. 23 (1982), No. 1, 173--182

Persistent URL: http://dml.cz/dmlcz/106142

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

23,1 (1982)

ASYMPTOTIC PROPERTIES OF SOME TESTS UNDER ALMOST REGULAR ASSUMPTIONS

M. HUŠKOVÁ

Abstract: Under "regular" assumptions (density absolutely continuous, Fisher's information finite) the asymptotic properties of tests based either on the loglikelihood statistic or on simple linear rank statistics were studied by many authors (e.g. [2]). The aim of this paper is to investigate the properties of such tests under "almost regular" assumptions (density absolutely continuous, Fisher's information infinite).

Key words: Asymptotically optimal tests, nonregular case, rank tests.

Classification: 62G10, 62E20

1. Introduction. Let $(X_{n1},\dots,X_{nn})$, $n=1,2,\dots$; be a random vector with density $\prod_{i=1}^{n} f(x_i-\theta_{ni})$ (with respect to Lebesgue measure), where $\theta_{n1},\dots,\theta_{nn}$ are regression constants.

Consider the sequence of the testing problems $\{H_n,A_n\}_{n=1}^{\infty}$, where $H_n = \{(X_{n1},\dots,X_{nn})$ has the density $\prod_{i=1}^{n} f(x_i)\}$, $A_n = \{(X_{n1},\dots,X_{nn})$ has the density $\prod_{i=1}^{n} f(x_i-\theta_{ni})\}$. It is known that under the "regularity" conditions (f absolutely continuous, finite Fisher's information, $\{\theta_{n1},\dots,\theta_{nn}\}_{n=1}^{\infty}$ fulfils Noether's condition) the asymptotically most powerful test can be based on either of the following statistics

$$(1.1)\qquad L_n = \sum_{i=1}^{n} \ln(f(X_i-\theta_{ni})/f(X_i)), \quad n=1,2,\dots$$

(1.2) $$S_n(f) = \sum_{i=1}^{n} \theta_{ni}\, a_n(R_{ni}, t), \quad n=1,2,\dots$$

(1.3) $$S_n^o(f) = \sum_{i=1}^{n} \theta_{ni}\, a_n^o(R_{ni}, f), \quad n=1,2,\dots .$$

Here R_{ni} denotes the rank of X_{ni} in the sequence $X_{n1},\dots,X_{nn}$,

(1.4) $$a_n(i,f) = -E\, f'(X_{(i)})/f(X_{(i)}) \quad i=1,\dots,n,$$

(1.5) $$a_n^o(i,f) = -f'(F^{-1}(\tfrac{i}{n+1}))/f(F^{-1}(\tfrac{i}{n+1})) \quad i=1,\dots,n,$$

where $X_{(i)}$ denotes the i-th order statistic from the sample of size n from the distribution with the density f, F^{-1} denotes the quantile function corresponding to f.

The critical regions corresponding to the asymptotic most powerful test (with level α) have the following form:

$$L_n + 1/2\, I(f) \geq \Phi^{-1}(1-\alpha)(I(f))^{1/2},$$

$$S_n(f) \geq \Phi^{-1}(1-\alpha)(I(f))^{1/2},$$

$$S_n^o(f) \geq \Phi^{-1}(1-\alpha)(I(f))^{1/2},$$

when $I(f)$ is Fisher's information, Φ is the distribution function $N(0,1)$ and the asymptotic maximum power equals to $1-\Phi(\Phi^{-1}(1-\alpha)-I(f))$.

In the present paper, analogous results are established under "almost regular" assumptions (f absolutely continuous and Fisher's information infinite). We show among others that the asymptotic most powerful test can be based on the same rank statistics as under "the regular" assumptions. The results concerning estimation theory under "almost regular" assumptions are published in [4], [5].

In the following, the probability measures induced by

$\prod_{i=1}^{n} f(x_i)$ and $\prod_{i=1}^{n} f(x_i-\theta_{ni})$ will be denoted by P_n and Q_n resp. for the expectations with respect to P_n and Q_n we shall write E_{P_n}, E_{Q_n} (similarly var_{P_n}, var_{Q_n} etc.).

2. Main results. We start this section by formulation of "almost regular" assumptions:

(AR) 1) f is absolutely continuous, there exist real numbers $y_1,\dots,y_k$ and $\delta > 0$ such that f is expressible as

$f(x)=a_j\,(y_j-x)+\psi_j(x)$ for $y_j-\delta< x\leq y_j$, $j=1,\dots,k$

$=b_j(-y_1+x)+\psi_j(x)$ for $y_j\leq x$ $y_j+\delta$, $j=1,\dots,k$,

$a_j\geq 0$, $b_j\geq 0$, $\sum_{j=1}^{k}(a_j+b_j)>0, \psi_j(y_j)=\psi_j'(y_j)=0$, ψ_j'' is continuous on $(y_j-\delta, y_j+\delta)$ and

$$\int_{x\notin\bigcup_{j=1}^{n}(y_j-\delta,\, y_j+\delta)} (f'(x))^2/f(x)\,dx<+\infty$$

2) $$\lim_{n\to\infty}\max_{1\leq i\leq n}|\theta_{ni}|=0,\ \sum_{i=1}^{n}\theta_{ni}=0,\ \sum_{\substack{i=1\\ \theta_{ni}\neq 0}}^{n}\theta_{ni}^2\ \ln|\theta_{ni}^{-1}|=1.$$

First we state the assertions on the asymptotic properties of the loglikelihood statistic L_n:

Theorem 2.1. If the assumptions AR are satisfied, then

1) $\{Q_n\}$ is contiguous to $\{P_n\}$ and conversely;

2) $$L_n=-\sum_{i=1}^{n}\sum_{j=1}^{k}\theta_{ni}|y_j-X_i|^{-1}\ I\{|X_i-y_j|<\gamma\}$$

$$-\ 1/2\ I^*(f)+o_{P_n}(1)\qquad n\longrightarrow\infty\ .$$

where $0<\gamma\leq\delta$, $I\{A\}$ denotes the indicator of a set A,

$I^*(f)=\sum_{j=1}^{k}(a_1+b_j)$;

3) $\mathcal{L}(L_n|P_n)\longrightarrow_w N(-\ 2\ I^*(f))^{-1},\ I^*(f))$,

$\mathcal{L}(L_n|Q_n)\longrightarrow_w N(\ 2\ I^*(f))^{-1},\ I^*(f))$;

4) the asymptotic maximum power for $\{H_n, A_n\}_{n=1}^{\infty}$ is equaled to $1-\Phi(\Phi^{-1}(1-\alpha) - I^*(f))$ and is reached by the test with the critical region

$$L_n + (2\, I^*(f))^{-1} \geq \Phi^{-1}(1-\alpha)(I^*(f))^{1/2}.$$

Proof: Assertions 1,3,4, follow directly from Theorem 4.3 in [3].

As for 2, using Lemma 4.1 in [3] one can get similarly as in the proof of Lemma VI.2.1.a and Lemma VI.2.1.b in [2] the following relations:

$$\mathrm{var}_{P_n}\{\sum_{i=1}^{n} \ln(f(X_{ni}-\theta_{ni})/f(X_{ni}))I\{X_{ni} \notin \bigcup_{j=1}^{k} (y_j-\gamma, y_j+\gamma)\}\} =$$

$$= O(\sum_{i=1}^{n} \theta_{ni}^2), \quad n \to \infty, \quad 0<\gamma<\delta, \tag{2.1}$$

$$E_{P_n}\{\sum_{i=1}^{n} \ln(f(X_{ni}-\theta_{ni})/f(X_{ni}))\} = -2\sum_{i=1}^{n} \int (\sqrt{f(x-\theta_{ni})} - \sqrt{f(x)})^2 dx +$$

$$+ o(1) = -1/2 \sum_{i=1}^{n} \theta_{ni}^2 \ln|\theta_{ni}|^{-1} \; I^*(f) + o(1), \; n \to \infty. \tag{2.2}$$

Since the assumption AR we have

$$\sum_{i=1}^{n} \{\ln(f(X_{ni}-\theta_{ni})/f(X_{ni})) - |X_{ni}-y_j|^{-1}\theta_{ni}\} . I\{|X_{ni}-y_j| < \gamma\}\} \to 0$$

for $n \to \infty$. The last relation together with (2.1) and (2.2) imply 2). Q.E.D.

Clearly, the asymptotically optimal test for $\{H_n, A_n\}_{n=1}^{\infty}$ depends only on X_{ni} lying in the neighbourhoods of $y_1, \dots$ $\dots, y_k$. Consider general simple linear rank statistics

$$S_n = \sum_{i=1}^{n} \theta_{ni}\, a_n(R_{in}), \tag{2.3}$$

where scores $a_n(1), \dots, a_n(n)$ satisfy:

$$\lim_{n\to\infty} \int_0^1 (a_n([un]+1) - \varphi(u))^2\, du = 0 \tag{2.4}$$

$$0 < \int_0^1 \varphi^2(u) du < +\infty, \tag{2.5}$$

with $[un]$ denoting the largest integer not exceeding un.

It is well known that if $X_{n1},\dots,X_{nn}$ are i.i.d. random variables with continuous distribution function and if

$$(2.6)\quad \sum_{i=1}^{n} \theta_{ni} = 0, \quad \lim_{n\to\infty} \max_{1\le i\le n} |\theta_{ni}|^2 \Big(\sum_{j=1}^{n} \theta_{nj}^2\Big)^{-1} = 0$$

then the asymptotic distribution of S_n is normal with parameters $(0, \sum_{i=1}^{n} \theta_{ni}^2 \int_0^1 \varphi^2(u)du)$.

Since the contiguity of $\{Q_n\}$ to $\{P_n\}$ and the proof of Theorem V.1.5 a in [2] one can assert

$$(2.7)\quad S_n - \sum_{i=1}^{n} \theta_{ni}\, \varphi(F(X_i)) = o_{Q_n}(1), \text{ as } n \to \infty.$$

By Theorem in [6] we have

$$(2.8)\quad \mathcal{L}\Big(\Big(\sum_{i=1}^{n} \theta_{ni}^2\Big)^{-1/2}\Big(\sum_{i=1}^{n} \theta_{ni}\, \varphi(F(X_i)) - a_n\Big)\Big) \to_w$$
$$\to_w N\Big(0, \int_0^1 \varphi^2(u)du\Big) \text{ as } n \to \infty,$$

where $a_n = \Big(\sum_{j=1}^{n} \theta_{nj}^2\Big)^{-1/2} \sum_{i=1}^{n} \theta_{ni} \int \varphi(F(x)) f(x-\theta_{ni})dx$.

With respect to the assumption AR 1 we can write for $0<\varepsilon<\delta$

$$(2.9)\quad b_n = b_{n1}(\varepsilon) + b_{n2}(\varepsilon),$$

$$b_{n1}(\varepsilon) = \Big(\sum_{j=1}^{n} \theta_{nj}^2\Big)^{-1/2} \sum_{v=1}^{k} \sum_{i=1}^{n} \theta_{ni} \int \varphi(F(x))(f(x-\theta_{ni}) - f(x)) \cdot$$
$$\cdot I\{|\theta_{ni}|\, \ln^{1/4}|\theta_{ni}|^{-1} < |x-y_v| < \varepsilon\}\, dx,$$

$$b_{n2}(\varepsilon) = \Big(\sum_{j=1}^{n} \theta_{nj}^2\Big)^{-1/2} \sum_{i=1}^{n} \theta_{ni} \int \varphi(F(x)) f(x-\theta_{ni}) I\{|x-y_v| \ge \varepsilon\}.$$

Similarly as in the regular case we get

$$(2.10)\quad b_{n2}(\varepsilon) = O\Big(\Big(\sum_{i=1}^{n} \theta_{ni}^2\Big)^{1/2}\Big) \text{ as } n \to \infty \text{ uniformly in } \varepsilon,$$
$0 < \varepsilon < \delta$.

Using Lemma 4.1 in [5] and the Schwarz equality we obtain

(2.11) $$b_{n1}(\varepsilon) \le D\left(\sum_{i=1}^{n} \theta_{ni}^2\right)^{-1/2} \sum_{j=1}^{m} \sum_{v=1}^{k} \theta_j^2 \left(\int_{|x-y_v|<\varepsilon} \varphi^2(F(x))f(x)dx \cdot \int_{|x-y_v|<\varepsilon} (\sqrt{f(x-\theta_{ni})} - \sqrt{f(x)})^2 dx \right)^{1/2}.$$

This relation together with (2.9 - 2.10) yields (ε can be chosen small enough):

$$b_n = o(1), \text{ as } n \to \infty .$$

The derived results can be summarized in the following theorem:

Theorem 2.2. If (2.4)-(2.6) and AR are satisfied then the asymptotic distribution of S_n given by (2.3) is normal with parameters 0 and $\sum_{i=1}^{n} \theta_{ni}^2 \int_0^1 \varphi^2(u)du$ both under $\{P_n\}$ and $\{Q_n\}$.

From the assertion of this theorem one can see that no simple linear rank statistic S_n generated by a square-integrable function performs suitable test statistics for testing problems $\{H_n, A_n\}_{n=1}^{\infty}$.

Now we shall formulate the assertion on the asymptotic properties of $S_n(f)$ and $S_n^0(f)$ given by (1.2) and (1.3), resp.

Theorem 2.3. If AR assumption is satisfied then

1) $$S_n(f) = \sum_{i=1}^{n} \sum_{j=1}^{k} \theta_{ni} \, a_n(R_{ni}, f) I\left\{ F(y_j - \gamma) < \frac{R_{ni}}{n+1} < F(y_j + \gamma) \right\} + o_{P_n}(1), \; n \to \infty, \; 0 < \gamma < \tilde{\delta};$$

2) $$\mathcal{L}(S_n(f) \mid P_n) \to_w N(0, I^*(f)), \quad n \to \infty,$$
$$\mathcal{L}(S_n(f) \mid Q_n) \to_w N(I^*(f), I^*(f)), \quad n \to \infty;$$

3) the asymptotic maximum power for $\{H_n, A_n\}_{n=1}^{\infty}$ is reached by the rank test with the critical region (with level α):

$$S_n(f) \geq \Phi^{-1}(1-\alpha)(I^*(f))^{1/2}.$$

If, moreover, the function $\varphi_f(u)=f'(F^{-1}(u))/f(F^{-1}(u))$, $u \in (0,1)$ is expressible as a sum of monotone functions then 1,2,3 remain true if we replace $S_n(f)$ by $S_n^o(f)$.

The proof is postponed to Section 3.

In other words, the assertion 1 of Theorem 2.3 gives that the test statistics $S_n(f)$ and $S_n^*(f)$ are asymptotically equivalent to the statistics depending only on the ranks lying in the neighbourhoods of $F(y_1),\ldots,F(y_k)$.

As an application consider the two-sample case. Let $(X_{N1},\ldots,X_{Nm})$ and $(X_{Nm+1},\ldots,X_{Nm+n})$ be independent random samples of size m and n, from the distribution $F(x-(N \ln N)^{-1/2})$ and $F(x)$, resp., $N=n+m$, i.e. $\theta_{Ni}=(N \ln N)^{-1/2}$, $i=1,\ldots,m$; $\theta_{Ni}=0$, $i=1+m,\ldots,N$. Consequently, for $\min(m,n)\longrightarrow\infty$ and $n/N\longrightarrow\lambda$, $\lambda \in (0,1)$: $N^{-1}\sum_{i=1}^{N} a_N(i,f) \longrightarrow 0$,

$$\sum_{i=1}^{N}(\theta_{Ni}-\bar{\theta}_N)^2 \longrightarrow \lambda(1-\lambda)/2, \quad \bar{\theta}_N= N^{-1}\sum_{i=1}^{N} \theta_{Ni},$$

which together with Theorem 2.3 implies (under assumption AR 1)

$$\sum_{i=1}^{m} a_N(R_{Ni},f)(N \ln N)^{-1/2} = \sum_{i=1}^{m} a_N(R_{Ni},f)\ I\{F(y_j-\gamma) < R_{Ni}(N+1)^{-1} < F(y_j+\gamma)\} + o_{P_n}(1),\ n\longrightarrow\infty,\ 0<\gamma<\delta$$

$$\mathcal{L}\Big(\sum_{i=1}^{m} a_N(R_{Ni},f)(N \ln N)^{-1/2} \mid P_N\Big) \longrightarrow_w N(0,1/2\lambda(1-\lambda)\ I^*(f)),\quad n\longrightarrow\infty,$$

$$\mathcal{L}\Big(\sum_{i=1}^{m} a_N(R_{Ni},f)(N \ln N)^{-1/2} \mid Q_N\Big) \longrightarrow_w N\big(\sqrt{1/2\lambda(1-\lambda)}\,I^*(f),\ 1/2\lambda(1-\lambda)\ I^*(f)\big),\quad n\longrightarrow\infty.$$

3. <u>Proof of Theorem 2.3</u>. Let us start with treating $S_n(f)$. Decompose $a_n(i,f)$ as follows (for n large enough):

$$(3.1)\quad a_n(i,f)= a_{n1\nu}(i,f)+ a_{n2\nu}(i,f)+ a_{n3\nu}(i,f),\quad \nu=1,\dots,n;\ i=1,\dots,n,$$

$$a_{n1\nu}(i,f)=\sum_{j=1}^{k} E'f\,(X_{(i)})/f(X_{(i)})\, I\{|X_{(i)}-y_j|\le |\theta_{n\nu}|\ln^{1/4}|\theta_{n\nu}|^{-1}\},$$

$$a_{n2\nu}(i,f)=\sum_{j=1}^{k} E'f\,(X_{(i)})/f(X_{(i)})\, I\{|\theta_n|\ \ln^{1/4}(|\theta_n|^{-1})< |X_{(i)}-y_j|\ \ln^{-1}|\theta_n|^{-1}\},$$

$$a_{n3\nu}(i,f)= E'f\,(X_{(i)})/f(X_{(i)})\, I\{|X_{(i)}-y_j|\ge \ln^{-1}|\theta_{n\nu}|^{-1},\ j=1,\dots,k\}.$$

By direct computation we get

$$(3.2)\quad E_{P_n}|\sum_{i=1}^{n} \theta_{ni}a_{n1i}(R_{ni},f) = O(\sum_{i=1}^{n}|\theta_{ni}|^2\ \ln^{1/4}|\theta_{ni}|^{-1})$$

$$(3.3)\quad \mathrm{var}_{P_n}\{\sum_{i=1}^{n}\theta_{ni}a_{n3i}(R_{ni},f)\} = O(\sum_{i=1}^{n}\theta_{ni}^2\ \ln|\theta_{ni}|^{-1}).$$

The assumptions ensure that the functions

$$g_j(x) = \frac{f'(x)}{f(x)} - \frac{1}{|y_j-x|}\quad \text{if } |x-y_j|<\delta,\quad j=1,\dots,k$$
$$= 0 \qquad\qquad \text{otherwise}\qquad j=1,\dots,k,$$

are bounded. Using this fact and taking into account (3.2), (3.3) we arrive at the following:

$$(3.4)\quad S_n(f) =\sum_{i=1}^{n}\theta_{ni}\ a^*_{n2i}(R_{ni},f) + o_{P_n}(1),\quad n\longrightarrow\infty,$$

where

$$a^*_{n2i}(\nu,f) =\sum_{j=1}^{k} E|X_{(\nu)}-y_j|^{-1}\ I\{|\theta_{ni}|\ln^{1/4}|\theta_{ni}|^{-1}< |X_{(\nu)}-y_j|\le \ln^{-1}|\theta_{ni}|^{-1}\}.$$

Thus it suffices to treat $\sum_{i=1}^{n}\theta_{ni}\ a^*_{n2i}(R_{ni},f)$.

By simple considerations we get that there exists a constant D (not depending on n,ν) such that

$$n^{-1}\sum_{i=1}^{n} a^{*2}_{n2\nu}(i,f)\le D\ \ln|\theta_{ni}|^{-1},\quad \nu=1,\dots,n,$$

$$\max_{1\le i\le N}|a^{*2}_{n2\nu}(i,f)|\le D\ n^{1/2}\ \ln^{-1}|\theta_{ni}|^{-1},\quad \nu=1,\dots,n.$$

This together with Lemma 2.1 in [1] yields

$$(3.5)\quad \mathrm{var}_{P_n}\{\sum_{i=1}^{n}\theta_{ni}(a^*_{n2i}(R_{ni},f)-a^*_{n2i}([U_in]+1,f))\} =$$
$$= O(\sum_{i=1}^{n}\theta^2_{ni}\max_{1\le\nu\le n}|a^*_{n2i}(\nu,f)\; n^{-1}(\sum_{s=1}^{n}a^*_{n2i}(s,f))^{1/2}) =$$
$$= O(\sum_{i=1}^{n}\theta^2_{ni}\; \ln^{-1/2}|\theta_{ni}|^{-1}),$$

where $U_1,\dots,U_n$ is the random sample of size n from the uniform distribution on (0,1).

Now, $\sum_{i=1}^{n}\theta_{ni}\; a^*_{n2i}([U_in]+1,f)$ is the sum of independent random variables. The asymptotic normality follows now in the classical way. It remains to show

$$(3.6)\quad \lim_{n\to\infty}\mathrm{var}_{P_n}\{\sum_{i=1}^{n}\theta_{ni}\; a^*_{n2i}([U_in]+1,f)\} = I^*(f).$$

Put for $i=1,2,\dots,n$

$$\psi^*_i(u) = |F^{-1}(u)-y_j|^{-1} \text{ if } |\theta_{ni}|\ \ln^{1/4}|\theta_{ni}|^{-1} < |F^{-1}(u)-y_j| \le$$
$$\le \ln^{-1}|\theta_{ni}|^{-1} \qquad j=1,\dots,k$$
$$= 0 \quad \text{otherwise.}$$

By a careful investigation we obtain

$$(3.7)\quad \overline{\lim_{n\to\infty}}\{\sum_{i=1}^{n}\theta^2_{ni}\,E(a^*_{n2i}([U_in]+1,f)-\psi^*_i(([U_in]+1)/(n+1))^2\}\le$$
$$\le \overline{\lim_{n\to\infty}}\{\sum_{i=1}^{n}\theta^2_{ni}\; n^{-1}\sum_{\nu=1}^{n}\int_0^1(\psi^*_i(u)-\psi^*_i(\tfrac{\nu}{n+1}))^2\frac{n!}{(\nu-1)!(n-\nu)!}\cdot$$
$$\cdot\, u^{\nu-1}(1-u)^{n-\nu}du\} = 0.$$

Further, using the same arguments as in the proof of Lemma V.I.6.a in [2] it can be shown that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\theta^2_{ni}\{n^{-1}\sum_{\nu=1}^{n}\psi^{*2}_i(\tfrac{\nu}{n+1}) - \int_0^1\psi^{*2}_i(u)du\} = 0.$$

The last relation together with (3.7) implies (3.6). Combining (3.4 - 3.6) together with the asymptotic normality of $\sum_{i=1}^{n}\theta_{ni}a^*_{n2i}([U_in]+1,f)$ we have

$$\mathcal{L}(S_n(f)|P_n)\longrightarrow_w N(0,I^*(f)),\qquad n\longrightarrow\infty$$

This fact together with the contiguity of $\{Q_n\}$ to $\{P_n\}$ implies

$$\mathcal{L}(S_n(f)\,|\,P_n) \longrightarrow_w N(I^*(f), I^*(f)), \qquad n \to \infty .$$

Assertion 3) is an immediate consequence of 2). Assertion 1) follows from (3.4) and (3.6).

The proof of the results on $S_n^o(f)$ is very similar (a little bit simpler) to that on $S_n(f)$ so that it is omitted. Q.E.D.

References

[1] HÁJEK J.: Some extensions of the Wald-Wolfowitz-Noether theorem, Ann. Math. Statist. 32(1961), 506-523.

[2] HÁJEK J., ŠIDÁK Zb.: Theory of rank tests, Academia, Praha, 1967.

[3] HUŠKOVÁ M., RATINGER T.: Contiguity in some nonregular cases, Proceedings of the Second Panonian conference 1981 (to appear).

[4] IBRAGIMOV I.A., HASMINSKIJ R.Z.: Asimptotičeskaja teorija ocenivanija, Nauka, Moskva, 1979.

[5] PFLUG G.: The limiting likelihood process for discontinuous densities, submitted 1981.

[6] BEHNEN K., NEUHAUS G.: A central limit theorem under contiguous alternatives, Annals of Statist. 2 (1975), 1349-1353.

Matematicko-fyzikální fakulta, Universita Karlova, Sokolovská 83, 18600 Praha 8, Czechoslovakia

(Oblatum 22.10. 1981)