Kybernetika

Asunción Rubio; Jan Ámos Víšek

Discriminability of robust test under heavy contamination

*Kybernetika*, Vol. 29 (1993), No. 4, 379--390

Persistent URL: http://dml.cz/dmlcz/125633

KYBERNETIKA — VOLUME 29 (1993), NUMBER 4, PAGES 377–388

# DISCRIMINABILITY OF ROBUST TEST UNDER HEAVY CONTAMINATION

ASUNCIÓN MARIA RUBIO AND JAN ÁMOS VÍŠEK[1]

The paper studies the upper limit of possible contamination level which still allows to discriminate by a robust (likelihood ratio) test a hypothesis from an alternative. Having found this limit the test of a hypothesis against an alternative, both being increasingly contaminated with the increasing number of observations, are considered. The level of contamination is adjusted so that it allows discrimination with an a priori fixed risk. The asymptotic distribution of the test statistic is found and the tightness of approximation to the power of test based on this asymptotic distribution is illustrated by a small numerical example.

## 1. INTRODUCTION

The notion of the contamination of data is one of the basic ones in robust statistics. Although the majority of methods constructed in robust statistics assume, at least implicitely, some level of contamination and accomodate the whole approach to it (see e.g. [6] or [7]), the estimation of the contamination level deserves relatively very small attention (see [12], [14] or [15]). However, having estimated the contamination level of data, in a similar way as we estimate other characteristics of data, as the location or the scatter, we may select a procedure with an appropriate "level of robustness", avoiding the procedures with (extremely) high breakdown point. It may allow us to reach directly a good efficiency by relatively simple methods. The benefit of it, besides others, may be e.g. a decrease of probability of the computational error.

On the other hand the question "How high the contamination level can be still allowing a reliable estimation?" is (leaving aside its practical impact) very attractive and the results achieved in study of breakdown point may be viewed as an answer to it. Similarly for the robust testing one may ask: "How heavy could a contamination be to allow still the discrimination of a hypothesis from an alternative (on a corresponding significance level and a power of test)?" In the present paper this question is studied in a framework of the model of contamination with the general neighbourhoods.

The setup of the paper will be as follows. Let us have a simple hypothesis and a simple alternative, which are assumed to be fixed (which corresponds to the fact

[1]The authors has obtained a support from the Department of Education and Science of the Spanish Government.

that in many cases they are "given" by (physical) circumstances). Independent observations, generated either by the hypothesis or by the alternative, are available but it is assumed that they are contaminated. Then, it is easy to see that there is an upper bound of the contamination level, and for the contamination level higher than it the problem of testing has no reasonable solution, because of the nonempty intersection of the contamination-generated neighbourhoods of the hypothesis and of the alternative. Hence it is clear that this bound does not depend on the number of observations. However assuming e. g. the value of (minimal) sum of the error probabilities as a characterization of test risk, one may expect that with the increasing number of observations a level of contamination, permitting to construct a test still being able to discriminate the hypothesis from the alternative with an a priori given risk, would increase, too. Therefore finding the mentioned upper bound of the admissible contamination we will try to build up a model in which the level of the contamination would also increase with the increasing number of observations and moreover the distributions (under the hypothesis and under the alternative) of the test statistics would converge to asymptotic ones.

The model is proposed in Section 3 and the desired property is established in Theorem 2. Numerical examples are presented in Section 4. Now, let us give some notations.

## 2. NOTATIONS

Let us denote by $R$ the real line and by $\mathcal{N}$ the set of all positive integers. Let $(\Omega, \mathcal{B})$ be a measurable space and let $\mathcal{M}$ stand for the set of all probability measures on it. Let $P_0$ and $P_1$ be distinct probability measures. For real numbers $\varepsilon_i$ and $\delta_i$ fulfilling

$$0 \le \varepsilon_i, 0 \le \delta_i, 0 < \varepsilon_i + \delta_i < 1, \quad i = 0, 1 \tag{1}$$

let us define

$$\begin{aligned}
\mathcal{P}_i(\varepsilon_i, \delta_i) &= \{Q \in \mathcal{M} : Q(B) \ge (1-\varepsilon_i)P_i(B) - \delta_i \text{ for all } B \in \mathcal{B}\} \\
\mathcal{P}_i^{\otimes n}(\varepsilon_i, \delta_i) &= \left\{ \bigotimes_{j=1}^{n} Q_j : Q_j \in \mathcal{P}_i(\varepsilon_i, \delta_i) \text{ for } j = 1, \dots, n \right\} \\
H_i(\varepsilon_i, \delta_i) &= \left\{ w_n : w_n \in \mathcal{P}_i^{\otimes n}(\varepsilon_i, \delta_i) \text{ for all } n \in \mathcal{N} \right\}
\end{aligned}$$

where "$\bigotimes_{j=1}^{n}$" denotes $n$th convolution.

Let us recall a definition of the least favorable pair (LFP) for $(\mathcal{P}_0, \mathcal{P}_1)$. We say that the pair of probability measures $(Q_0, Q_1)$ is the least favourable pair for $(\mathcal{P}_0, \mathcal{P}_1)$ if it satisfies

$$\begin{aligned}
Q_0(\{\pi > t\}) &= \sup\{Q'(\{\pi > t\}) : Q' \in \mathcal{P}_0(\varepsilon_0, \delta_0)\}, \\
Q_1(\{\pi > t\}) &= \inf\{Q''(\{\pi > t\}) : Q'' \in \mathcal{P}_1(\varepsilon_1, \delta_1)\}
\end{aligned}$$

for all positive $t$ and $\pi \in \mathrm{d}Q_1/\mathrm{d}Q_0$.

## 3. RESULTS

Let $\varepsilon_1, \delta_0$ and $\delta_1$ be real numbers such that there exists $\varepsilon \in R$ and $B_\varepsilon \in \mathcal{B}$ such that (1) is fulfilled (with $\varepsilon_0 = \varepsilon$) and

$$(1-\varepsilon_1)P_1(B_\varepsilon) + \varepsilon_1 + \delta_1 < (1-\varepsilon)P_0(B_\varepsilon) - \delta_0. \tag{2}$$

Let $\mathcal{E}$ be a set of all such $\varepsilon \in (0,1)$ for which (1) and (2) hold and put

$$\varepsilon_0 = \sup \mathcal{E}.$$

To simplify further notation put

$$\nu_1 = \frac{\varepsilon_1 + \delta_1}{1-\varepsilon_1}, \quad \omega_1 = \frac{\delta_1}{1-\varepsilon_1}$$

and

$$\nu_\varepsilon = \frac{\varepsilon + \delta_0}{1-\varepsilon}, \quad \omega_\varepsilon = \frac{\delta_0}{1-\varepsilon}$$

and define for any $t > 0$ and $\Delta \in \mathrm{d}P_1/\mathrm{d}P_0$

$$\psi_{0\varepsilon}(t) = (\nu_1 + \omega_\varepsilon t)^{-1}\left[t \cdot P_0(\Delta < t) \; - \; P_1(\Delta < t)\right]$$

and

$$\psi_{1\varepsilon}(t) = (\nu_\varepsilon t + \omega_1)^{-1}\left[P_1(\Delta > t) \; - \; t \cdot P_0(\Delta > t)\right]$$

**Assertion 1.** The function $\psi_{0\varepsilon}(t)$ is strictly increasing on $\{t \in R,\ \psi_{0\varepsilon}(t) > 0\}$ and $\psi_{1\varepsilon}(t)$ is strictly decreasing on $\{t \in R,\ \psi_{1\varepsilon}(t) > 0\}$.

Proof. The proof is essentially contained in [6] but since it is short we will present it for the convenience of the reader. Let us assume at first the function

$$\varphi_\varepsilon(t) = P_0(\Delta < t) \; - \; t^{-1} \cdot P_1(\Delta < t)).$$

Let $0 < t_1 < t_2$, $t_1, t_2 \in \{\, t \in R, \psi_{0\varepsilon}(t) > 0\}$. Since for any $\omega \in \{t_1 \le \Delta < t_2\}$ we have $\mathrm{d}P_0(\omega) > t_2^{-1} \cdot \mathrm{d}P_1(\omega)$, we obtain also

$$P_0(t_1 \le \Delta < t_2) > t_2^{-1} \cdot P_1(t_1 \le \Delta < t_2) > t_2^{-1} \cdot P_1(\Delta < t_2) - t_1^{-1} \cdot P_1(\Delta < t_1)$$

and hence

$$\varphi_\varepsilon(t_2) \; - \; \varphi_\varepsilon(t_1) = P_0(t_1 \le \Delta < t_2) - t_2^{-1} \cdot P_1(\Delta < t_2) + t_1^{-1} \cdot P_1(\Delta < t_1) > 0.$$

On the other hand

$$\psi_{0\varepsilon}(t) \; = \; \frac{t}{\nu_1 + \omega_\varepsilon \cdot t} \cdot \varphi_\varepsilon(t),$$

and since $t\,(\nu_1 \, + \, \omega_\varepsilon \cdot t)^{-1}$ is also increasing on $\{t > 0\}$, the proof follows. □

Since for $t \searrow 0$ the function $\psi_{0\varepsilon}(t) \searrow 0$ and for $t \nearrow \infty$ it converges to $\omega_\varepsilon^{-1} = \frac{1-\varepsilon}{\delta_0} > 1$ (the last inequality follows from (1)), there is a uniquely given number $\Delta_{0\varepsilon}$ such that $\psi_{0\varepsilon}(\Delta_{0\varepsilon}) = 1$ or

$$\Delta_{0\varepsilon} P_0(\Delta < \Delta_{0\varepsilon}) - P_1(\Delta < \Delta_{0\varepsilon}) = \nu_1 + \omega_\varepsilon \Delta_{0\varepsilon} \tag{3}$$

and similarly for $\psi_{1\varepsilon}$ there is also uniquely given $\Delta_{1\varepsilon}$ so that

$$P_1(\Delta_{1\varepsilon} < \Delta) - \Delta_{1\varepsilon} P_0(\Delta_{1\varepsilon} < \Delta) = \nu_\varepsilon \Delta_{1\varepsilon} + \omega_1.$$

Let

$$\Delta_0 = \sup\{\Delta_{0\varepsilon} : \varepsilon \in \mathcal{E}\}$$

and

$$\Delta_1 = \inf\{\Delta_{1\varepsilon} : \varepsilon \in \mathcal{E}\}.$$

**Lemma 1.** The set $\mathcal{E}$ has a form $(0, \varepsilon_0)$ and it holds:

$$\frac{1-\varepsilon_0}{1-\varepsilon_1} P_0\left(\Delta < \frac{1-\varepsilon_0}{1-\varepsilon_1}\right) - P_1\left(\Delta < \frac{1-\varepsilon_0}{1-\varepsilon_1}\right) = \nu_1 + \frac{\delta_0}{1-\varepsilon_1}, \tag{4}$$

$$\Delta_0 = \Delta_1 = \frac{1-\varepsilon_0}{1-\varepsilon_1}.$$

Proof. Let $\varepsilon \in \mathcal{E}$ and $\tilde{\varepsilon} \in R$, $0 < \tilde{\varepsilon} < \varepsilon$. Then evidently (1) holds for $\tilde{\varepsilon}$. Since for some $B_\varepsilon \in \mathcal{B}$ the relation (2) is fulfilled, we have also

$$(1-\varepsilon_1)P_1(B_\varepsilon) + \varepsilon_1 + \delta_1 < (1-\tilde{\varepsilon})P_0(B_\varepsilon) - \delta_0.$$

So $\tilde{\varepsilon} \in \mathcal{E}$, too. Moreover, let $t > 0$. Then $\omega_{\tilde{\varepsilon}} \cdot t < \omega_\varepsilon \cdot t$ and hence $\psi_{0\varepsilon}(t) < \psi_{0\tilde{\varepsilon}}(t)$ which implies

$$\Delta_{0\tilde{\varepsilon}} < \Delta_{0\varepsilon}. \tag{5}$$

Let us assume that $\psi_{0\varepsilon}(\frac{1-\varepsilon}{1-\varepsilon_1}) \leq 1$. Then

$$(1-\varepsilon)\, P_0\left(\Delta < \frac{1-\varepsilon}{1-\varepsilon_1}\right) - \delta_0 \leq (1-\varepsilon_1)\, P_1\left(\Delta < \frac{1-\varepsilon}{1-\varepsilon_1}\right) + \varepsilon_1 + \delta_1. \tag{6}$$

Let $B \in \mathcal{B}$ and let us consider the set $C = \left\{\Delta \leq \frac{1-\varepsilon}{1-\varepsilon_1}\right\} \cap B^c$ where the superindex "$c$" stands for the complement. Due to the fact that at any point $\omega \in C$ we have $(1-\varepsilon_1)\, \mathrm{d}P_1(\omega) \leq (1-\varepsilon)\, \mathrm{d}P_0(\omega)$ we obtain

$$(1-\varepsilon_1)\, P_1(C) \leq (1-\varepsilon)\, P_0(C)$$

and hence (using (6))

$$(1-\varepsilon)\, P_0\left(\left\{\Delta < \frac{1-\varepsilon}{1-\varepsilon_1}\right\} \cap B\right) - \delta_0 \leq (1-\varepsilon_1)\, P_1\left(\left\{\Delta < \frac{1-\varepsilon}{1-\varepsilon_1}\right\} \cap B\right) + \varepsilon_1 + \delta_1. \tag{7}$$

Since similarly for any $\omega \in \left\{\Delta > \frac{1-\varepsilon}{1-\varepsilon_1}\right\}$ we have

$$(1-\varepsilon)\,P_0\left(\left\{\Delta > \frac{1-\varepsilon}{1-\varepsilon_1}\right\} \cap B\right) < (1-\varepsilon_1)\,P_1\left(\left\{\Delta > \frac{1-\varepsilon}{1-\varepsilon_1}\right\} \cap B\right),$$

taking into account (7) we obtain for any $B \in \mathcal{B}$

$$(1-\varepsilon)\,P_0\,(B)\;-\;\delta_0 \le (1-\varepsilon_1)\,P_1\,(B)\;+\;\varepsilon_1\;+\;\delta_1$$

but it is in a contradiction with the assumption that $\varepsilon \in \mathcal{E}$ (see (2)). So we have found that $\psi_{0\varepsilon}(\frac{1-\varepsilon}{1-\varepsilon_1}) > 1$ and then from the Assertion 1 follows that $\Delta_{0\varepsilon} < \frac{1-\varepsilon}{1-\varepsilon_1}$ for any $\varepsilon \in \mathcal{E}$. Let us assume that $\Delta_0 < \frac{1-\varepsilon_0}{1-\varepsilon_1}$ and for any $\varepsilon \in \mathcal{E}$ denote by $C_\varepsilon$ the set $\{\omega \in \Omega; \Delta < \Delta_{0\varepsilon}\}$. Consider a monotone sequence $\{\varepsilon_n\}_{n=1}^{\infty} \nearrow \varepsilon_0$. From the continuity of probability measure in a nondecreasing sequence $\{C_{\varepsilon_n}\}_{n=1}^{\infty}$ (see (5)) we obtain

$$\psi_{0\varepsilon_0}(\Delta_0) = 1. \tag{8}$$

But then from the assumption that $\Delta_0 < \frac{1-\varepsilon_0}{1-\varepsilon_1}$ we obtain

$$\psi_{0\varepsilon_0}\left(\frac{1-\varepsilon_0}{1-\varepsilon_1}\right) > 1.$$

Let us define $\tilde{\psi}(\varepsilon) = \psi_{0\varepsilon}\left(\frac{1-\varepsilon}{1-\varepsilon_1}\right)$. Then we have $\tilde{\psi}(\varepsilon_0) > 1$ which implies that there exists $\varepsilon^* \in R$, $\varepsilon^* > \varepsilon_0$ and $\tilde{\psi}(\varepsilon^*) > 1$, i.e.

$$\frac{1-\varepsilon^*}{1-\varepsilon_1}P_0\left(\Delta < \frac{1-\varepsilon^*}{1-\varepsilon_1}\right) - P_1\left(\Delta < \frac{1-\varepsilon^*}{1-\varepsilon_1}\right) > \nu_1 + \omega_{\varepsilon^*}\frac{1-\varepsilon^*}{1-\varepsilon_1}.$$

Putting $B_{\varepsilon^*} = \left\{\Delta < \frac{1-\varepsilon^*}{1-\varepsilon_1}\right\}$ one obtains

$$(1-\varepsilon^*)P_0(B_{\varepsilon^*}) - \delta_0 > (1-\varepsilon_1)P_1(B_{\varepsilon^*}) + \varepsilon_1 + \delta_1,$$

i.e. (2) holds for $\varepsilon^*$. Moreover

$$0 < \varepsilon_1 + \delta_1 + (1-\varepsilon_1)P_1(B_{\varepsilon^*}) < (1-\varepsilon^*)P_0(B_{\varepsilon^*}) - \delta_0 \le 1 - (\varepsilon^* + \delta_0).$$

i.e. (1) is fulfilled for $\varepsilon^*$, too. So we have found that the assumption $\Delta_0 < \frac{1-\varepsilon_0}{1-\varepsilon_1}$ implies existence of $\varepsilon^* > \varepsilon_0, \varepsilon^* \in \mathcal{E}$ and therefore $\Delta_0 = \frac{1-\varepsilon_0}{1-\varepsilon_1}$. Similarly it is possible to show that $\Delta_1 = \frac{1-\varepsilon_0}{1-\varepsilon_1}$. □

**Remark 1.** The assertion of Lemma 1 may be written also as

$$\Delta_0 P_0(\Delta < \Delta_0) - P_1(\Delta < \Delta_0) = \nu_1 + \omega_0\Delta_0$$

from which follows

$$P_0(\Delta < \Delta_0) - \omega_0 > 0.$$

We will need this inequality several times in the sequal.

For any $\varepsilon \in (0, \varepsilon_0)$ write $\varepsilon = \varepsilon(\tau) = \varepsilon_0 - \tau$ (for some $\tau \in (0, \varepsilon_0)$ ).

**Lemma 2.** Let us write

$$\Delta_{0\varepsilon} = \Delta_0 - \tau R_0(\tau) \quad \text{and} \quad \Delta_{1\varepsilon} = \Delta_0 + \tau R_1(\tau).$$

Then

$$\lim_{\tau \searrow 0} R_0(\tau) = \frac{\omega_0}{(1-\varepsilon_1)[P_0(\Delta < \Delta_0) - \omega_0]}$$

and

$$\lim_{\tau \searrow 0} R_1(\tau) = \frac{1+\nu_0}{(1-\varepsilon_1)\,[P_0(\Delta_0 < \Delta) + \nu_0]}.$$

Proof. It follows from (5) that the mapping $\Delta_{0\varepsilon} : \mathcal{E} \to (0, \Delta_0)$ is nondecreasing, and since $\Delta_0 = \sup\{\Delta_{0\varepsilon} : \varepsilon \in \mathcal{E}\}$ we have

$$\lim_{\tau \searrow 0} \Delta_{0\varepsilon(\tau)} = \Delta_0.$$

Let $\varepsilon \in \mathcal{E}$. Substracting (3) from an analogous relation written for $\varepsilon_0$ (and $\Delta_0$, naturally — see (4)) one obtains

$$\begin{aligned}&\Delta_0\,(P_0\,(\Delta < \Delta_0) - P_0\,(\Delta < \Delta_{0\varepsilon})) + (\Delta_0 - \Delta_{0\varepsilon})P_0(\Delta < \Delta_{0\varepsilon}) - P_1(\Delta_{0\varepsilon} \le \Delta < \Delta_0)\\ &= (\omega_0 - \omega_\varepsilon)\Delta_0 + \omega_\varepsilon(\Delta_0 - \Delta_{0\varepsilon}),\end{aligned}$$

i.e.

$$\begin{aligned}&\Delta_0 P_0(\Delta_{0\varepsilon} \le \Delta < \Delta_0) - P_1(\Delta_{0\varepsilon} \le \Delta < \Delta_0)\\ =\;&(\omega_0 - \omega_\varepsilon)\Delta_0 + (\Delta_0 - \Delta_{0\varepsilon})(\omega_\varepsilon - P_0(\Delta < \Delta_{0\varepsilon})).\end{aligned} \tag{9}$$

From (3) follows that

$$P_0(\Delta < \Delta_{0\varepsilon}) - \omega_\varepsilon > 0 \tag{10}$$

because of $\nu_1 > 0$ (and hence also $\Delta_{0\varepsilon} > 0$ — see also Remark 1). A straightforward computation gives

$$P_1(\Delta_{0\varepsilon} \le \Delta < \Delta_0) \le \Delta_0 P_0(\Delta_{0\varepsilon} \le \Delta < \Delta_0) \tag{11}$$

which together with (9) results in

$$(\Delta_0 - \Delta_{0\varepsilon})(\omega_\varepsilon - P_0(\Delta < \Delta_{0\varepsilon})) \ge (\omega_\varepsilon - \omega_0)\Delta_0$$

and finally (see (10))

$$\Delta_0 - \Delta_{0\varepsilon} \le [P_0(\Delta < \Delta_{0\varepsilon}) - \omega_\varepsilon]^{-1} \cdot \frac{\tau \cdot \delta_0}{(1 - \varepsilon_0 + \tau)(1 - \varepsilon_1)}.$$

In the same way as (11) one may derive

$$\Delta_{0\varepsilon} P_0(\Delta_{0\varepsilon} \le \Delta < \Delta_0) \le P_1(\Delta_{0\varepsilon} \le \Delta < \Delta_0)$$

which together with (4) gives

$$(\Delta_0 - \Delta_{0\varepsilon})\,\omega_\varepsilon - P_0(\Delta < \Delta_{0\varepsilon}) \le (\Delta_0 - \Delta_{0\varepsilon})P_0(\Delta_{0\varepsilon} \le \Delta < \Delta_0) - (\omega_0 - \omega_\varepsilon)\Delta_0,$$

i.e.

$$(\Delta_0 - \Delta_{0\varepsilon})(\omega_\varepsilon - P_0(\Delta < \Delta_0)) \le (\omega_\varepsilon - \omega_0)\Delta_0. \tag{12}$$

Taking into account Remark 1 and the inequality $\omega_0 = \frac{\delta_0}{1-\varepsilon_0} \ge \frac{\delta_0}{1-\varepsilon_0+\tau} = \omega_\varepsilon$ we get

$$P_0(\Delta < \Delta_0) - \omega_\varepsilon > 0. \tag{13}$$

Now from (12) and (13) we obtain

$$\Delta_0 - \Delta_{0\varepsilon} \ge [\,P_0(\Delta < \Delta_0)\, - \,\omega_\varepsilon\,]^{-1} \cdot \frac{\tau \cdot \delta_0}{(1-\varepsilon_0+\tau)(1-\varepsilon_1)}$$

and the proof of the first assertion follows from the monotone convergence of $\Delta_{0\varepsilon}$ to $\Delta_0$ and continuity of the probability. A proof of the second assertion of Lemma 2 is similar. □

Now for a fix $\varepsilon_1$ and for any $\varepsilon \in \mathcal{E}$ we may find LFP $(Q_{0\varepsilon}, Q_{1\varepsilon})$ for $\mathcal{P}_0(\varepsilon, \delta_0), \mathcal{P}_1(\varepsilon_1, \delta_1)$, the likelihood ratio of which is given by

$$\frac{q_{1\varepsilon}}{q_{0\varepsilon}} = \frac{1-\varepsilon_1}{1-\varepsilon} \text{ median } \{\Delta_{0\varepsilon}, \Delta, \Delta_{1\varepsilon}\} \tag{14}$$

where densities $q_{0\varepsilon}$ and $q_{1\varepsilon}$ may be taken with respect to $P_0 + P_1$ or $Q_{0\varepsilon} + Q_{1\varepsilon}$, see [9]. Keep in mind that although $\varepsilon_1$ and $\delta_1$ are fix (and hence also $\mathcal{P}_1(\varepsilon_1)$ is fix), $q_{1\varepsilon}$ depends on $\varepsilon$ — see (6.4) of [9]. From the relation (6.4) of [9] and from Lemma 1 it also follows that there is a probability measure $Q_0$ such that

$$\lim_{\varepsilon \nearrow \varepsilon_0} \|Q_{0\varepsilon} - Q_0\| = 0 \quad \text{ and } \quad \lim_{\varepsilon \nearrow \varepsilon_0} \|Q_{1\varepsilon} - Q_0\| = 0.$$

Recalling that for any $\varepsilon \in (0, \varepsilon_0)$ we have written $\varepsilon = \varepsilon_0(\tau) = \varepsilon_0 - \tau$ for some $\tau \in (0, \varepsilon_0)$, put for $h > 0$ $\tau_n = h.n^{-\frac{1}{2}}$ and $\varepsilon_n = \max\{0, \varepsilon_0 - \tau_n\}$.

**Assertion 2.** If $\|P_n - P_n'\| \to 0$ as $n \to \infty$, then the sequences $\{P_n\}_{n=1}^\infty$ and $\{P_n'\}_{n=1}^\infty$ are mutually contiguous.

For the proof see [10], Lemma 2.1 of Chapter 1. □

**Theorem 1.** Sequences of probability measures $\{Q_{j\varepsilon_n}^{\otimes n}\}_{n=1}^\infty$, $j = 0, 1$ are (mutually) contiguous.

Proof. The proof follows immediately from Assertion 2 due to

$$\lim_{\varepsilon \nearrow \varepsilon_0} \|Q_{0\varepsilon} - Q_{1\varepsilon}\| = 0.$$ □

Keeping the notation of [10] let us write $\Lambda_{\varepsilon_0,h,n}$ for the logarithm of the likelihood ratio $\mathrm{d}Q_{1\varepsilon_n}^{\otimes n} / \mathrm{d}Q_{0\varepsilon_n}^{\otimes n}$. Then we have

$$\Lambda_{\varepsilon_0,h,n}(\mathbf{x}) = \sum_{i=1}^n \log \left[\frac{1-\varepsilon_1}{1-\varepsilon_0+\tau_n} \text{ median } \{\Delta_{0\varepsilon_n}, \Delta(x_i), \Delta_{1\varepsilon_n}\}\right]. \tag{15}$$

Let us find a function $\varphi_{(-\tau_n,\tau_n)}(x)$ such that

$$\Lambda_{\varepsilon_0,h,n}(\mathbf{x}) = 2\sum_{i=1}^{n} \log \varphi_{(-\tau_n,\tau_n)}(x_i), \tag{16}$$

i.e.

$$\varphi(-\tau_n,\tau_n)(x_i) = \left[\frac{1-\varepsilon_1}{1-\varepsilon_0+\tau_n}\ \text{median}\{\Delta_{0\varepsilon_n},\, \Delta(x_i),\, \Delta_{1\varepsilon_n}\}\right]^{\frac{1}{2}}.$$

Let us evaluate $\varphi'(0) = \lim_{\tau\searrow 0} \frac{1}{\tau}\left\{\varphi_{(-\tau,\tau)}(x) - \varphi_{(0,0)}(x)\right\}$ for a fix $x$. Let $x \in \{\Delta < \Delta_0\}$. Then there is a $\tilde{\tau} = \tilde{\tau}(x) > 0$ such that for all $\tau \in (0,\tilde{\tau})$

$$\varphi_{(-\tau,\tau)} = \left(\frac{1-\varepsilon_1}{1-\varepsilon_0+\tau}\Delta_{0\varepsilon}\right)^{\frac{1}{2}}$$

and hence

$$\begin{aligned} \varphi'(0) &= \lim_{\tau\searrow 0}\frac{1}{\tau}\left\{\left(\frac{1-\varepsilon_1}{1-\varepsilon_0+\tau}\Delta_{0\varepsilon}\right)^{\frac{1}{2}} - \left(\frac{1-\varepsilon_1}{1-\varepsilon_0}\Delta_0\right)^{\frac{1}{2}}\right\} \\ &= \lim_{\tau\searrow 0}\frac{1}{\tau}\left\{\left(\frac{1-\varepsilon_1}{1-\varepsilon_0+\tau}\Delta_{0\varepsilon}\right)^{\frac{1}{2}} - \left(\frac{1-\varepsilon_1}{1-\varepsilon_0}\Delta_{0\varepsilon}\right)^{\frac{1}{2}}\right\} \\ &+ \lim_{\tau\searrow 0}\frac{1}{\tau}\left\{\left(\frac{1-\varepsilon_1}{1-\varepsilon_0}\right)^{\frac{1}{2}}\left(\Delta_{0\varepsilon}^{\frac{1}{2}} - \Delta_0^{\frac{1}{2}}\right)\right\} \end{aligned} \tag{17}$$

(keep in mind that $\varepsilon = \varepsilon(\tau)$). Now we easy find that the first limit is equal to $-\frac{(1-\varepsilon_1)^{\frac{1}{2}}}{2(1-\varepsilon_0)^{\frac{3}{2}}\Delta_0^{\frac{1}{2}}} = -\frac{1}{2}(1-\varepsilon_0)$. Making use of Lemma 2 we also easy compute the value of the second limit, namely $-\frac{\omega_0}{2(1-\varepsilon_0)[P(\Delta<\Delta_0)-\omega_0]}$ that finally gives

$$\varphi'(0) = \frac{P_0(\Delta < \Delta_0)}{2(1-\varepsilon_0)[\omega_0 - P_0(\Delta < \Delta_0)]}$$

for any $x \in \{\Delta < \Delta_0\}$. Similarly for any $x \in \{\Delta_0 < \Delta\}$ we find

$$\varphi'(0) = \frac{P_0(\Delta < \Delta_0)}{2(1-\varepsilon_0)\,[P_0(\Delta_0 < \Delta) + \nu_0]}$$

and moreover, due to the fact that for the case when $\tau > \tilde{\tau}$ we have $\varphi_{(-\tau,\tau)}(x) \in \left(\frac{1-\varepsilon_1}{1-\varepsilon}\Delta_{0\varepsilon}, \frac{1-\varepsilon_1}{1-\varepsilon}\Delta_{1\varepsilon}\right)$ (see (14)), convergence

$$\frac{1}{\tau}\left\{\varphi_{(-\tau,\tau)}(x) - \varphi_{(0,0)}(x)\right\} \longrightarrow_{\tau\searrow 0} \varphi'(0)$$

is uniform. Since all consideration were made under the assumption that the dominating measure is a probability, we have

$$\dot{\varphi}(0) = \varphi'(0) \qquad \text{a.s. } [Q_0]$$

where $\dot{\varphi}(0)$ denotes the derivative of $\varphi_{(-\tau,\tau)}(x)$ with respect to $\tau$ in the quadratic mean. Finally, defining $\Gamma = 4 \cdot E_{Q_0}\dot{\varphi}^2$ we may compute that

$$\Gamma = \left\{(1-\varepsilon_0)^2[P_0(\Delta < \Delta_0) - \omega_0][P_0(\Delta_0 < \Delta) + \nu_0]\right\}^{-1} \cdot P_0^2(\Delta < \Delta_0).$$

Denote by $\mathcal{L}[\mathcal{Y}|P]$ the distribution of a random variable $\mathcal{Y}$ under a probability measure $P$ and by $N(\mu, \sigma^2)$ the normal distribution with the mean $\mu$ and the variance $\sigma^2$. Finally let us put

$$D_n = 2n^{-\frac{1}{2}} \sum_{j=1}^{n} \dot{\varphi}_j(0) \tag{18}$$

where $\dot{\varphi}_j(0)$ denotes value of $\dot{\varphi}(0)$ (or $\varphi'(0)$, it is the same) at the point $x_j$.

**Theorem 2.** It holds

$$\mathcal{L}\left[D_n \,|\, Q_{j\varepsilon_n}^{\otimes n}\right] \longrightarrow N((-1)^{j+1}h \cdot \Gamma, \Gamma)$$

and

$$\mathcal{L}\left[\Lambda_{\varepsilon_0,h,n} \,|\, Q_{j\varepsilon_n}^{\otimes n}\right] \longrightarrow N\left(\frac{(-1)^{j+1}}{2}h^2\Gamma,\, h^2\Gamma\right). \tag{19}$$

The proof follows directly from Theorems 4.6, 4.5 and 4.3 of Chapter 2 of [10]. (In fact, specifying Theorems 4.3, 4.5 and 4.6 for our setup, i.e. for the setup in which we assume except of a system of the shifting alternatives also a system of the shifting hypotheses, we obtain Theorem 2.)

## 4. NUMERICAL EXAMPLES AND DISCUSSION

The numerical studies performed in [6] have proved the reliability of approximations, based on Edgeworth expansion, to error probabilities of the robust likelihood-ratio test based on $\Lambda_{\varepsilon_0,h,n}$ (defined in (15)). So we may use them to create an idea about two questions:

i) How does convergence described in Theorem 2 work?

ii) In which situations the result given in Theorem 2 is useful for an approximation to error probabilities of the robust test?

Both answers are presented in the form of tables. Let us explain a setup and the values gathered in them. $P_0$ and $P_1$ were assumed to be $N(0,1)$ and $N(\mu, 1)$, respectively.

The Tables 1a, 1b and 1c offer a possibility to make an idea about the convergence in (19), i.e. $\varepsilon_0$ and $h$ are fixed. Their values together with other parameters are referred on the upper margin of the tables. An approximation obtained from (19) is denoted by $\alpha^*$.

The setup of Table 1a, 1b and 1c was selected so that it shows how convergence given in (19) works. However this setup is not appropriate for building up an idea of a practical possibility to use (19) as an approximation to size or power of test.

**Table 1a.**

$\mu = .5 \quad \varepsilon_1 = .070 \quad \varepsilon_0 = .097 \quad \delta_0 = \delta_1 = .025$

$h = .027 \quad \Gamma = 3.520 \quad \alpha^* = .4899$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| EDG.APPROX. | .488824 | .488882 | .488903 | .488914 | .488920 |
| $n$ | 60 | 70 | 80 | 90 | 100 |
| EDG.APPROX. | .488925 | .488928 | .488931 | .488932 | .488934 |
| $n$ | 110 | 120 | 130 | 140 | 150 |
| EDG.APPROX. | .488935 | .488936 | .488937 | .488938 | .488938 |

**Table 1b.**

$\mu = .75 \quad \varepsilon_1 = .080 \quad \varepsilon_0 = .219 \quad \delta_0 = \delta_1 = .025$

$h = .139 \quad \Gamma = 2.684 \quad \alpha^* = .4546$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| EDG.APPROX. | .454125 | .454460 | .454585 | .454653 | .454697 |
| $n$ | 60 | 70 | 80 | 90 | 100 |
| EDG.APPROX. | .454728 | .454751 | .454769 | .454783 | .454795 |
| $n$ | 110 | 120 | 130 | 140 | 150 |
| EDG.APPROX. | .454805 | .454814 | .454822 | .454828 | .454834 |

**Table 1c.**

$\mu = 1.0 \quad \varepsilon_1 = .100 \quad \varepsilon_0 = .343 \quad \delta_0 = \delta_1 = .025$

$h = .243 \quad \Gamma = 2.359 \quad \alpha^* = .4260$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| EDG.APPROX. | .421214 | .421853 | .422093 | .422224 | .422307 |
| $n$ | 60 | 70 | 80 | 90 | 100 |
| EDG.APPROX. | .422366 | .422410 | .422445 | .422473 | .422496 |
| $n$ | 110 | 120 | 130 | 140 | 150 |
| EDG.APPROX. | .422515 | .422532 | .422546 | .422559 | .422570 |

Hence the Tables 2a, 2b and 2c collect the values of the approximation, yielded by (19), and the Edgeworth one for a situation when $\varepsilon$ is assumed to be fixed (value of which we have estimated (or guessed) from the character of given data) and the parameter $h$ of the asymptotic setting of Section 3 is taken $h = h(n)$ so that for every $n \in \mathcal{N}$ we have

$$\varepsilon = \varepsilon_0 - h(n)/\sqrt{n},$$

i.e.

$$h(n) = \sqrt{n}(\varepsilon_0 - \varepsilon).$$

It may seem strange that we have considered in the previous section the parameter $h$ to be fixed and now we select $h = h(n)$. But it is quite consistent. In the previous section we have for some fixed $h$ derived some asymptotic result. Now we try to use this result for a given situation in which we assume that bulk of data is distributed either according to probability model $P_0$ or according to $P_1$, but they are contaminated with fixed contamination level. Naturally, we have some fixed number of observations. After estimating values of parameters describing contamination level (namely $\varepsilon_0, \varepsilon_1, \delta_0$ and $\delta_1$), we have to select $h$ so that "shrinking setup includes" our concrete case (or in other words, asymptotic setup "runs" through our fixed case). Then we do the same for other sample size (but the same $\varepsilon_0, \varepsilon_1, \delta_0$ and $\delta_1$) and hence the parameter $h$ changes because in fact we approximate corresponding error probabilities in a different asymptotic model (i. e. for every column in the next tables we have to have different asymptotic model). In the examples described by the following tables $\varepsilon$'s are the same as in the Tables 1a, 1b and 1c. They corresponds to (the upper bound of) the usual level of contamination, (see [5]). Their values together with the other parameters are given again on the upper margin of the tables and $\alpha^*$ again denotes the approximation evaluated from (19) as above.

**Table 2a.**

$\mu = .5 \quad \varepsilon_1 = .070 \quad \varepsilon_0 = .097 \quad \delta_0 = \delta_1 = .025$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| $\alpha^*$ | .4849 | .4787 | .4740 | .4699 | 4664 |
| EDG.APPROX | .4644 | .4499 | .4388 | .4295 | .4213 |

**Table 2b.**

$\mu = .75 \quad \varepsilon_1 = .080 \quad \varepsilon_0 = .219 \quad \delta_0 = \delta_1 = .025$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| $\alpha^*$ | .4415 | .4175 | .3993 | .3842 | 3710 |
| EDG.APPROX | .3945 | .2982 | .2586 | .2274 | .2017 |

**Table 2c.**

$\mu = 1.0 \quad \varepsilon_1 = .100 \quad \varepsilon_0 = .348 \quad \delta_0 = \delta_1 = .025$

| $n$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| $\alpha^*$ | .4033 | .3646 | .3358 | 3123 | .2921 |
| EDG.APPROX | .2552 | .1767 | .1281 | .0951 | .0716 |

**Remark 2.** It follows from the Tables 1a, 1b and 1c the Edgeworth approximations of error probabilities are very stable which is in the accordance with Theorem 2. The differences among $\alpha^*$ and the values given in the Tables 1a, 1b and 1c are due to the fact that the approximation to the standard normal distribution (used for evaluation of $P_0(\Delta < \Delta_0)$ and $P_1(\Delta_< \Delta_0)$) is not very tight. (A polynomial approximation from [1] with accuracy $10^{-5}$ which is usually sufficiently good was used.) However one finds out that the small deviations in approximation cause really not negligible changes in solution of (4). Maybe, a normalization of the Edgeworth expansion could bring a little better accuracy (see [2], [3] or [11]).

On the other hand a practical application of the results of Theorem 2 is possible only for a rather "small" size of sample (as Table 2a, 2b and 2c show) and for a heavy contamination i. e. for contamination not very far from the maximal possible one. At the first glance it may seem strange why with increasing $n$ the accuracy of the approximations decreases. The explanation is simple and follows from (18). In other words, due to the fact that our $Q_{0\varepsilon_n}$ and $Q_{1\varepsilon_n}$ are fixed, namely equal to

$Q_{0\varepsilon}$ and $Q_{1\varepsilon}$, the summands in (16) are the same, independently of $n$. However the asymptotic model, in which we embed our situation, assumes that they are equal to $\dot{\varphi}_j(0)$ (see (18)). Asymptotic distributions of such two sums of independent and identically distributed random variables are naturally disjoint. Hence the increasing inaccuracy. This is the reason why models with shrinking neighbourhoods while for the theoretical purposes very appealing and clarifying limiting situations, are of limited importance for practical applications. It does not mean that they should not be used at all. They can be used in situations when size of sample just crosses boundary above which the discrimination (under heavy contamination) is already possible with a given risk.

(Received April 16, 1992.)

## REFERENCES

[1] M. Abramowitz and I. A. Stegun: Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Dover Publications, 1964.

[2] O. Barndorff-Nielsen and D. R. Cox: Edgeworth and saddle-point approximations with statistical applications. J. Roy. Statist. Soc. Ser. B *1* (1979), 3, 279–312.

[3] C. A. Field and E. M. Ronchetti: Small Sample Asymptotics. A manuscript of a book, 1988.

[4] F. R. Hampel: The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. *69* (1974), 364, 383–393.

[5] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti and W. A. Stahel: Robust Statistics – The Approach Based on Influence Function. J. Wiley, New York 1986.

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

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

[8] D. Lambert: Influence functions for testing. J. Amer. Statist. Assoc. *76* (1981), 375, 649–657.

[9] H. Rieder: Least favorable pairs for special capacities. Ann. Statist. *5* (1977), 5, 909–921.

[10] G. G. Roussas: Contiguity of Probability Measures. Cambridge Univ. Press, Cambridge 1972. Russian translation: Izdatelstvo Mir, Moscow 1975.

[11] A. Rubio, L. Aguilar and J. Á. Víšek: Testing difference of models. Computational Statistics *8* (1993), 57–70.

[12] A. Rubio and J. Á. Víšek: Estimating the contamination level of data in the framework of linear regression analysis. Qüesttió 1993 (to appear).

[13] J. Á. Víšek: On second order efficiency of a robust test and approximations of its error probabilities. Kybernetika *19* (1983), 3, 387–407.

[14] J. Á. Víšek: Estimating contamination level. In: Proc. Fifth Pannonian Symposium on Mathematical Statistics, Visegrád, Hungary 1985, pp. 401–414.

[15] J. Á. Víšek: Estimation of contamination level in the model of contaminacy with general neighbourhoods. Kybernetika *25* (1989), 278–297.

*Prof. Asunción Maria Rubio, Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, 10071 Cáceres. Spain.*

*RNDr. Jan Ámos Víšek, CSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha. Czech Republic.*