KYBERNETIKA — VOLUME 25 (1988), NUMBER 4

ESTIMATION OF CONTAMINATION LEVEL IN MODEL OF CONTAMINACY WITH GENERAL NEIGHBOURHOODS

JAN ÁMOS VÍŠEK

Strongly consistent estimator of contamination level in the model of contaminacy combining the Huber model and the model with total-variation-neighbourhoods is proposed. Order of convergence is established, too. Numerical example is included.

1. INTRODUCTION

It is intuitively clear that any study aiming to produce a robust statistical procedure assumes more or less explicitly presence of a *contamination* of data. However what is even more important is the fact that the user is forced at any practical application of robust procedure to express, at least in a vague way, his or her idea about the level of contamination (for some examples see Víšek [3]). And at this moment not having any estimator of contamination level one may "overestimate" and obtain less efficient procedures than is an optimal one or to "underestimate" it. Then the probabilistic characteristics of the procedure could be (and usually are) rather different from the assumed ones. Let us give examples of the both cases.

Let us consider simple situation when testing, in the framework of i.i.d. model, a simple hypothesis $\mathbb{H}_0$: $P = P_0$ against an alternative $\mathbb{H}_1$: $P = P_1$ under presence of contaminacy, i.e. we test

$$\mathscr{H}_{0AS} = \{Q\colon Q \in \mathbb{M},\ \|Q - (1 - \varepsilon_{0AS})\, P_0 - \varepsilon_{0AS} H_0\| \leqq \delta_{0AS};\ H_0 \in \mathbb{M}\}$$

against

$$\mathscr{H}_{1AS} = \{Q\colon Q \in \mathbb{M},\ \|Q - (1 - \varepsilon_{1AS})\, P_1 - \varepsilon_{1AS} H_1\| \leqq \delta_{1AS};\ H_1 \in \mathbb{M}\}$$

where $\mathbb{M}$ is the set of all probability measures over an appropriate measurable space, say $(\mathscr{X}, \mathscr{A})$, and $\varepsilon_{0AS} \geqq 0$, $\varepsilon_{1AS} \geqq 0$, $\delta_{0AS} \geqq 0$, $\delta_{1AS} \geqq 0$, $0 < \varepsilon_{0AS} + \delta_{0AS} < 1$, $0 < \varepsilon_{1AS} + \delta_{1AS} < 1$. (Index "$AS$" points out that the hypotheses were established under our *assumption* about the level of contamination which we had expressed by assigning values ε_{0AS}, ε_{1AS}, δ_{0AS} and δ_{1AS} to contamination parameters.) Let us assume in what follows that the sample size is fixed and denote by $\beta(\alpha)$ the second kind error probabilities of the most powerful test of the level α ($\alpha \in (0, 1)$). Then we have

$$\beta_{AS}(\alpha) = \inf_{C \in \mathscr{C}_{AS}(\alpha)} \sup_{Q \in H_{1AS}} (1 - Q^{(n)}(C)),$$

where

$$\mathscr{C}_{AS}(\alpha) = \{C \in \mathscr{A}^{(n)}: \sup_{Q \in H_{0AS}} Q^{(n)}(C) \leqq \alpha\}.$$

Under mild conditions (see Rieder [2]) there is a pair $(Q_0, Q_1) \in H_{0AS} \times H_{1AS}$ such that for any $\alpha \in (0, 1)$ there is a $C_\alpha \in \mathscr{A}$ such that $Q_0(C_\alpha) = \alpha$ and $1 - Q_1(C_\alpha) = \beta_{AS}(\alpha)$. Let us denote the class of these sets for all $\alpha \in (0, 1)$ by $\mathscr{E}$, i.e.

$$\mathscr{E} = \{C_\alpha \in \mathscr{C}_{AS}(\alpha): Q_0(C_\alpha) = \alpha, 1 - Q_1(C_\alpha) = \beta_{AS}(\alpha), \quad \alpha \in (0, 1)\}.$$

Now let us suppose that our assumption about the contamination level was wrong and that the true level of contamination is given by the parameters $\varepsilon_{0AC} = \beta_{0AS} + \tau_0$, $\varepsilon_{1AC} = \varepsilon_{1AS} + \tau_1$, $\delta_{0AC} = \delta_{0AS} + \xi_0$ and $\delta_{1AC} = \delta_{1AS} + \xi_1$. Denoting for $i = 0$ and 1

$$\mathscr{H}_{iAC} = \{Q: \|Q - (1 - \varepsilon_{iAC}) P_i - \varepsilon_{iAC} H_i\| \leqq \delta_{iAC}; H_i \in \mathbb{M}\}$$

("*AC*" again emphasizes that this level is *actual*) and for any $\alpha \in (0, 1)$

$$\mathscr{E}_\alpha = \{C \in \mathscr{E}: \sup_{Q \in H_{0AC}} Q^{(n)}(C) \leqq \alpha\},$$

we have for the actual second kind error probability of our tests (constructed for the assumed values ε_{iAS} and δ_{iAS})

$$\beta_{AC}(\alpha) = \inf_{C \in \mathscr{E}_\alpha} \sup_{Q \in H_1} (1 - Q^{(n)}(C)).$$

But if we had known this actual level we would have constructed an optimal test for it and we would have attained the second kind error probability

$$\beta_{AT}(\alpha) = \inf_{C \in \mathscr{C}_{AT}(\alpha)} \sup_{Q \in H_{1AC}} (1 - Q^{(n)}(C))$$

where again

$$\mathscr{C}_{AT}(\alpha) = \{C \in \mathscr{A}^{(n)}: \sup_{Q \in H_{0AC}} Q^{(n)}(C) \leqq \alpha\}.$$

Examples below offer possibility to create an idea about relations among these three curves, namely $\beta_{AS}(\alpha)$, $\beta_{AC}(\alpha)$ and $\beta_{AT}(\alpha)$ (as functions of α). In all examples the sample size was 40. The pairs of probability measures P_0 and P_1 and the values of contamination level parameters for which the corresponding values of β's were evaluated are described under every figure.

It seems the just presented examples hint that if we can estimate the contamination level it might be a help for practical application of robust procedures. The rest of this paper is a modest attempt to make a very first step in this area. In the second section we give notation, in the third one a definition of contamination level is proposed, the fourth section contains a characterization of this level and the last one brings a (strongly) consistent estimator of it.

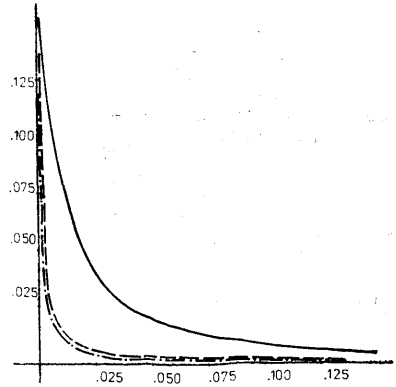

Fig. 1. The assumed (———), the actual (— — —) and the attainable (— . — . —) dependence of the second kind error probabilities on the first kind errors for $P_0 = \mathcal{N}(0, 1)$, $P_1 = \mathcal{N}(1, 1)$, $\varepsilon_{iAS} = 2\delta_{iAS} = \cdot 05$ and $\varepsilon_{iAC} = 2\delta_{iAC} = \cdot 03$.

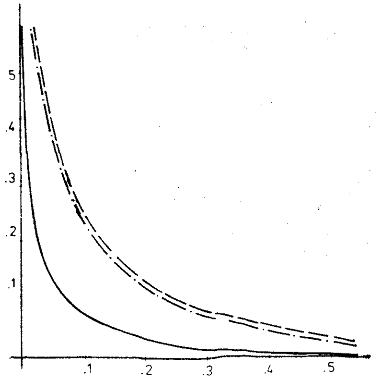

Fig. 2. The assumed (———), the actual (— — —) and the attainable (— . — . —) dependence of the second kind error probabilities on the first kind errors for $P_0 = \mathcal{N}(0, 1)$, $P_1 = \mathcal{N}(1, 1)$, $\varepsilon_{iAS} = 2\delta_{iAS} = \cdot 03$ and $\varepsilon_{iAC} = 2\delta_{iAC} = \cdot 05$.

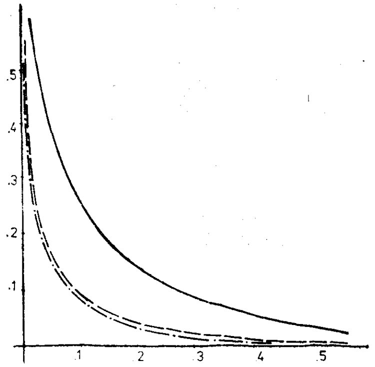

Fig. 3. The assumed, the actual and the attainable dependence of the second kind error probabilities on the first kind errors for $P_0 = \mathcal{N}(0, 1)$, $P_1 = \mathcal{N}(1, 1)$, $\varepsilon_{iAS} \approx 2\delta_{iAS} = \cdot 05$, $\varepsilon_{iAC} = 2\delta_{iAC} = \cdot 02$.

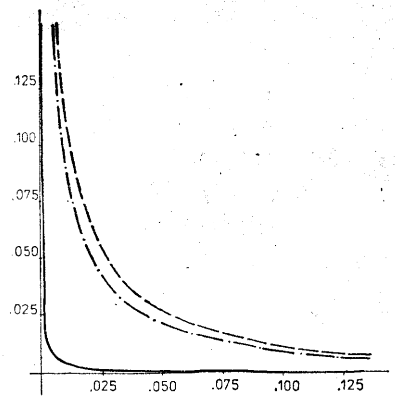

Fig. 4. The assumed, the actual and the attainable dependence of the second kind error probabilities on the first kind errors for $P_0 = \mathcal{N}(0, 1)$, $P_1 = \mathcal{N}(1, 1)$, $\varepsilon_{iAS} = 2\delta_{iAS} = 0{\cdot}2$, $\varepsilon_{iAC} = 2\delta_{iAC} = {\cdot}05$.

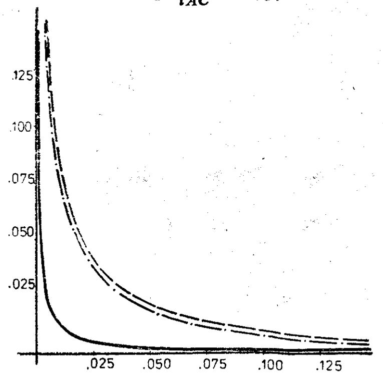

Fig. 5. The assumed, the actual and the attainable dependence of the second kind error probabilities on the first kind errors for P_0 equal to Weibull distribution with $c = 1$ and $p = 3$, P_1 — Weibull with $c = 2$, $p = 3$, $\varepsilon_{iAS} = 2\delta_{iAS} = {\cdot}05$ and $\varepsilon_{iAC} = 2\delta_{iAC} = {\cdot}03$.

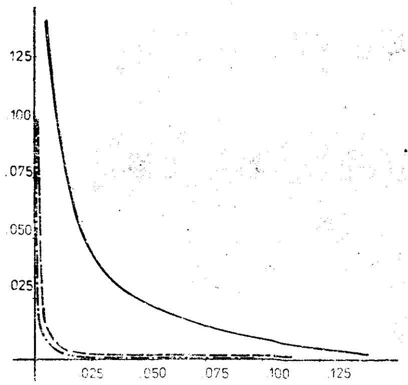

Fig. 6. The assumed, the actual and the attainable dependence of the second kind error probabilities on the first kind errors for P_0 equal to Weibull distribution with $c = 1$ and $p = 3$, P_1 — Weibull with $c = 2$, $p = 3$, $\varepsilon_{iAS} = 2\delta_{iAS} = {\cdot}03$ and $\varepsilon_{iAC} = 2\delta_{iAC} = {\cdot}05$.

2. NOTATION

Let $\boldsymbol{N}$ be the set of all positive integers, $\mathbf{R}$ the real line, $\mathbf{R}^+$ its positive part and $(\mathscr{X}, \mathscr{A})$ a measurable space. Let us denote by $\mathbb{M}$ the set of all probability measures on $(\mathscr{X}, \mathscr{A})$, and for any σ-finite measure μ on $(\mathscr{X}, \mathscr{A})$ let $\mathscr{P}_\mu$ denote the set of all probability measures on $(\mathscr{X}, \mathscr{A})$ which are absolutely continuous with respect to μ. Then for any probability measure from $\mathscr{P}_\mu$ – denoted by a capital letter, say H – denote by a small letter with index μ – in our case h_μ – a version of its density with respect to μ.

In the rest of paper we shall consider a fixed pair Q and P from $\mathbb{M}$ and we shall choose and fix a finite measure ν such that Q and P are absolutely continuous with respect to it. For the densities q_ν and p_ν we write only q and p since it cannot cause any confusions. It will be convenient to have also a notation for the following sets:

For any $\varepsilon \in [0, 1]$ put

$$S_\varepsilon = \{x \in \mathscr{X}: (1 - \varepsilon)\, p(x) > q(x)\}$$

and for any μ such that $\nu \ll \mu$ and $H \in \mathscr{P}_\mu$

$$S_\varepsilon(h_\mu) = \{x \in \mathscr{X}: (1 - \varepsilon)\, p_\mu(x) + \varepsilon\, h_\mu(x) > q_\mu(x)\}\,.$$

(The sets $S_\varepsilon(h_\mu)$ may differ according to the given versions of p_μ, q_μ and h_μ. As they will be used as an integration region, it cannot cause any difficulty.)

In accordance with commonly used notation, for any $A \in \mathscr{X}$ let us denote by A^c the complement of A with respect to $\mathscr{X}$. Finally, denote by $\|H - K\|$ the total variation of any pair of measures $H, K \in \mathbb{M}$.

3. DEFINITION OF CONTAMINATION LEVEL

To be able to give a meaningful definition of contamination level we have to impose conditions on the function which will be used to reach a balance between ε and δ.

Definition 1. Let $w(x, y)$: $[0, 1]^2 \to \mathbf{R}$ be a continuous mapping such that for any $x_0 \in (0, 1]$ and $y_0 \in (0, 1]$ the following conditions are fulfilled.

(i) The function $w(\lambda x_0, (1 - \lambda)\, y_0)$ is a convex function of $\lambda \in [0, 1]$ and there is a unique $\lambda_0 \in (0, 1)$ such that

$$w(\lambda_0 x_0, (1 - \lambda_0)\, y_0) = \min_{\lambda \in [0,1]} w(\lambda x_0, (1 - \lambda)\, y_0)\,.$$

(ii) The function $w(\lambda x_0, \lambda y_0)$ is nondecreasing in $\lambda \in [0, \min\{1/x_0, 1/y_0\}]$.

Then the mapping $w(x, y)$ will be called the *regular weight function.*

To see that the next definition is not empty we shall need the following lemmas.

Lemma 1. For any $H, K \in \mathscr{P}_\mu$ we have

$$\|H - K\| = \int_{\{x\in\mathscr{X}: k_\mu(x) < h_\mu(x)\}} [h_\mu(x) - k_\mu(x)] \,\mathrm{d}\mu \,.$$

The proof is transparent and will be omitted.

Lemma 2. For any $\varepsilon \in [0, 1]$ and μ such that $P \ll \mu$ and $Q \ll \mu$ there is a probability measure $H^\varepsilon \in \mathscr{P}_\mu$ such that

$$h^\varepsilon_\mu(x) \leqq 1/\varepsilon \,\{q_\mu(x) - (1 - \varepsilon)\, p_\mu(x)\} \quad \text{for} \quad x \in \{q_\mu(x) \geqq (1 - \varepsilon)\, p_\mu(x)\}$$

and

$$h^\varepsilon_\mu(x) = 0 \quad \text{elsewhere}\,.$$

Proof. Since

$$\int_{\{q_\mu \geqq (1-\varepsilon)p_\mu\}} \{q_\mu(x) - (1 - \varepsilon)\, p_\mu(x)\} \,\mathrm{d}\mu =$$
$$= \varepsilon + \int_{\{q_\mu < (1-\varepsilon)p_\mu\}} \{(1 - \varepsilon)\, p_\mu(x) - q_\mu(x)\} \,\mathrm{d}\mu$$

we have

$$1/\varepsilon \int_{\{q_\mu \geqq (1-\varepsilon)p_\mu\}} \{q_\mu(x) - (1 - \varepsilon)\, p_\mu(x)\} \,\mathrm{d}\mu \geqq 1$$

and the existence of H^ε follows. □

Lemma 3. We have for any $\varepsilon \in [0, 1]$

$$\inf_H \|(1 - \varepsilon)\, P + \varepsilon H - Q\| = \int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\} \,\mathrm{d}\nu\,,$$

where the inf is taken over $H \in \mathbb{M}$.

Proof. Fix some $H \in \mathbb{M}$ and put $\mu = \nu + H$. Then $\nu \ll \mu$ and hence there is a Radon-Nikodym density $d(x)$ of ν with respect to μ. Then for appropriate version of densities we may write

$$p_\mu(x) = d(x) \,.\, p(x) \quad \text{and} \quad q_\mu(x) = d(x) \,.\, q(x)$$

having $p_\mu(x)$ positive iff $p(x)$ is positive.

Then for any $\varepsilon \in [0, 1]$ and $x_0 \in S_\varepsilon$ we have

$$(1 - \varepsilon)\, p_\mu(x_0) > q_\mu(x_0)$$

and finally

$$(1 - \varepsilon)\, p_\mu(x_0) + \varepsilon\, h_\mu(x_0) > q_\mu(x_0)\,,$$

i.e. $S_\varepsilon \subset S_\varepsilon(h_\mu)$. According to Lemma 1 we may write

$$\|(1 - \varepsilon)\, P + \varepsilon H - Q\| = \int_{S_\varepsilon(h_\mu)} \{(1 - \varepsilon)\, p_\mu(x) + \varepsilon\, h_\mu(x) - q_\mu(x)\} \,\mathrm{d}\mu \geqq$$
$$\geqq \int_{S_\varepsilon} \{(1 - \varepsilon)\, p_\mu(x) - q_\mu(x)\} \,\mathrm{d}\mu = \int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\} \,\mathrm{d}\nu\,.$$

Due to Lemma 2 we may find $H^\varepsilon \in \mathscr{P}_\nu$ such that $h^\varepsilon_\nu(x) = 0$ for $x \in S_\varepsilon$. Then we have

$$\int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\} \,\mathrm{d}\nu = \int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) + \varepsilon\, h^\varepsilon_\nu(x) - q(x)\} \,\mathrm{d}\nu =$$
$$= \int_{S_\varepsilon(h^\varepsilon_\nu)} \{(1 - \varepsilon)\, p(x) + \varepsilon\, h^\varepsilon_\nu(x) - q(x)\} \,\mathrm{d}\nu = \|(1 - \varepsilon)\, P + \varepsilon H^\varepsilon - Q\|\,.$$

(Last but one equality holds due to $S_\varepsilon = S_\varepsilon(h^\varepsilon_\nu)$.) So we obtain also

$$\int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\} \,\mathrm{d}\nu \geqq \inf_H \|(1 - \varepsilon)\, P + \varepsilon H - Q\|\,. \quad □$$

Lemma 4. For any $\varepsilon \in [0, 1]$ denote by $\delta(\varepsilon)$ the integral

$$\int_{S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\}\, \mathrm{d}\nu$$

and put $\mathbb{D} = \{(\varepsilon, \delta(\varepsilon)), \varepsilon \in [0, 1]\}$. Then $\mathbb{D}$ is closed and hence the $\min\limits_{(\varepsilon,\delta)} w(\varepsilon, \delta)$, where min is taken over $(\varepsilon, \delta) \in \mathbb{D}$, exists.

Proof. Let us have any sequence $\{\varepsilon_n, \delta_n\}_{n=1}^{\infty} \subset \mathbb{D}$ which converges to a point $(\varepsilon_0, \delta_0)$. It implies that

$$\lim_{n\to\infty} \varepsilon_n = \varepsilon_0 \quad \text{and} \quad \lim_{n\to\infty} \delta_n = \delta_0 .$$

Since on the other hand

$$-q(x) \leqq \{(1 - \varepsilon)\, p(x) - q(x)\} \,.\, I_{S_\varepsilon}(x) \leqq p(x) \quad \text{for all} \quad 0 < \varepsilon < 1 ,$$

the Lebesgue convergence theorem gives

$$\lim_{n\to\infty} \int_{S_{\varepsilon_n}} \{(1 - \varepsilon_n)\, p(x) - q(x)\}\, \mathrm{d}\nu = \int_{S_{\varepsilon_0}} \{(1 - \varepsilon_0)\, p(x) - q(x)\}\, \mathrm{d}\nu ,$$

i.e.

$$\delta_0 = \int_{S_{\varepsilon_0}} \{(1 - \varepsilon_0)\, p(x) - q(x)\}\, \mathrm{d}\nu = \delta(\varepsilon_0)$$

and it says that $(\varepsilon_0, \delta_0) \in \mathbb{D}$. □

Definition 2. For $Q, P \in \mathbb{M}$ and a regular weight function w a pair $(\varepsilon, \delta) \in [0, 1]^2$ such that

(1) $$(\varepsilon, \delta) = \arg\min w(x, y)$$

over the set

(2) $$\{\inf_H \|(1 - \varepsilon)\, P + \varepsilon H - Q\| = \delta\, ;\; (\varepsilon, \delta) \in [0, 1]^2\} ,$$

the inf is over all $H \in \mathbb{M}$, will be called the *contamination level* of Q with respect to P and denoted by $(\varepsilon^w_{Q,P}, \delta^w_{Q,P})$.

Remark 1. Let us note that the definition of contamination level was inspired by the model of contaminacy with general neighbourhoods (which includes as a special case Huber's model of contaminacy) but the sense of $\varepsilon^w_{Q,P}$ is a little different from the parameter of Huber's model of contaminacy. For more details see Víšek [3]. Moreover, it is clear that generally we may have a whole set of pairs satisfying Definition 2. Nevertheless it is clear that we are able to characterize $(\varepsilon^w_{Q,P}, \delta^w_{Q,P})$ in a more convenient way which enables us to show uniqueness of contamination level. We shall do it in the next section. Let us remind that due to Lemma 3 we have $\delta^W_{Q,P} = \delta(\varepsilon^W_{Q,P})$ and the set over which the minimum is taken is the set $\mathbb{D}$ of Lemma 4.

4. CHARACTERIZATION AND UNIQUENESS OF CONTAMINATION LEVEL

Since in the rest of paper we shall assume the weight function to be fixed we shall omit the index w in $\varepsilon^w_{Q,P}$ and $\delta^w_{Q,P}$. Although Q and P were already fixed we shall write

$\varepsilon_{Q,P}$ and $\delta_{Q,P}$ with indexes because ε and δ will be used also for another purposes. It is clear from Lemma 3 and 4 that the following characterization theorem is true.

Theorem 1. A pair $(\varepsilon^*, \delta^*)$ is equal to $(\varepsilon_{Q,P}, \delta_{Q,P})$ iff

(3) $$(\varepsilon^*, \delta^*) = \arg\min w(x, y)$$

and

(4) $$\delta^* = \int_{S_{\varepsilon^*}} \{(1 - \varepsilon^*)\, p(x) - q(x)\}\, \mathrm{d}\nu\,.$$

Remark 2. Since the definition of $\varepsilon_{Q,P}$, $\delta_{Q,P}$ does not depend on the measure ν with respect to which densities p and q are taken, the values of ε^* and δ^* also do not depend on the chosen version p and q. A formal way showing this directly may be based on the idea used in the proof of Lemma 3.

Remark 3. Notice that the existence of $(\varepsilon_{Q,P}, \delta_{Q,P})$ was proved in Lemmas 2, 3 and 4.

Lemma 5. For any $\varepsilon \in [0, 1]$ the function $\delta(\varepsilon)$ (see Lemma 4) has a continuous derivative

(5) $$\delta'(\varepsilon) = -\int_{S_\varepsilon} p\, \mathrm{d}\nu$$

(i.e. this derivative exists and is equal to the right-hand side of (5) – at the end points of the interval $[0, 1]$ the derivative is meant from one side). Moreover, $\delta'(\varepsilon)$ is nondecreasing in ε and hence $\delta(\varepsilon)$ is convex.

Proof. Let $\varepsilon' > \varepsilon$, $\varepsilon \in [0, 1)$, $\varepsilon' \in (0, 1]$. Then $1 - \varepsilon' < 1 - \varepsilon$ and hence for any $x \in S_{\varepsilon'}$ we have $x \in S_\varepsilon$, i.e. $S_{\varepsilon'} \subset S_\varepsilon$. Therefore

$$\int_{S_{\varepsilon'}} [(1 - \varepsilon')\, p - q]\, \mathrm{d}\nu - \int_{S_\varepsilon} [(1 - \varepsilon)\, p - q]\, \mathrm{d}\nu =$$
$$= -(\varepsilon' - \varepsilon) \int_{S_{\varepsilon'}} p\, \mathrm{d}\nu - \int_{S_\varepsilon - S_{\varepsilon'}} [(1 - \varepsilon)\, p - q]\, \mathrm{d}\nu \leqq -(\varepsilon' - \varepsilon) \int_{S_{\varepsilon'}} p\, \mathrm{d}\nu$$

and finally

$$\limsup_{\varepsilon' \to \varepsilon+} \frac{\delta(\varepsilon') - \delta(\varepsilon)}{\varepsilon' - \varepsilon} \leqq -\int_{S_\varepsilon} p\, \mathrm{d}\nu\,.$$

On the other hand

$$-(\varepsilon' - \varepsilon) \int_{S_{\varepsilon'}} p\, \mathrm{d}\nu - \int_{S_\varepsilon - S_{\varepsilon'}} \{(1 - \varepsilon)\, p - q\}\, \mathrm{d}\nu \geqq$$
$$\geqq -(\varepsilon' - \varepsilon) \int_{S_{\varepsilon'}} p\, \mathrm{d}\nu - \int_{S_\varepsilon - S_{\varepsilon'}} \{(1 - \varepsilon)\, p - (1 - \varepsilon')\, p\}\, \mathrm{d}\nu = -(\varepsilon' - \varepsilon) \int_{S_\varepsilon} p\, \mathrm{d}\nu$$

and hence

$$\liminf_{\varepsilon' \to \varepsilon+} \frac{\delta(\varepsilon') - \delta(\varepsilon)}{\varepsilon' - \varepsilon} \geqq -\int_{S_\varepsilon} p\, \mathrm{d}\nu\,.$$

Similarly

$$\lim_{\varepsilon' \to \varepsilon-} \frac{\delta(\varepsilon') - \delta(\varepsilon)}{\varepsilon' - \varepsilon} = -\int_{S_\varepsilon} p\, \mathrm{d}\nu\,.$$

Continuity of the derivative follows from regularity of the probability measure.

Monotonicity of the derivative is implied by monotonicity of the sets S_ε which was shown above. □

Theorem 2. The pair $(\varepsilon_{Q,P}, \delta_{Q,P})$ is given uniquely.

Proof. Let us assume that $P \neq Q$ and that there are two pairs satisfying (1) and (2), say $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$. Without any restriction on generality let us assume $\varepsilon_1 < \varepsilon_2$ – due to Theorem 1 we know that it is not possible to have simultaneously $\varepsilon_1 = \varepsilon_2$ and $\delta_1 \neq \delta_2$. Similarly for $\varepsilon_1 \neq \varepsilon_2$ we have, under assumption $P \neq Q$, $\delta(\varepsilon_1) \neq \delta(\varepsilon_2)$. But then we have

$$\delta_i = \delta(\varepsilon_i)\,, \quad i = 1, 2$$

and

$$w(\varepsilon_1, \delta_1) = w(\varepsilon_2, \delta_2)\,.$$

Due to Lemma 5 we know that the limits of $\delta(\varepsilon)$ for ε tending to zero from right and to one from left exist and surely

$$(6) \qquad \lim_{\varepsilon \to 0_+} \delta(\varepsilon) \leqq 1 \quad \text{and} \quad \lim_{\varepsilon \to 1_-} \delta(\varepsilon) = 0\,.$$

Together with convexity of $\delta(\varepsilon)$ it implies that there are $x_0 \in (0, 1]$ and $y_0 \in (0, 1]$ such that the line going through the points $(0, y_0)$ and $(x_0, 0)$ contains also the points $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$. Applying requirement (i) of Definition 1 we may find a point $(\varepsilon_3, \delta_3)$ of this line such that

$$w(\varepsilon_3, \delta_3) < w(\lambda x_0, (1 - \lambda)\, y_0) \quad \text{for all} \quad \lambda \in (0, 1)$$

with exception of $\lambda = \varepsilon_3/x_0$. It is easy to see that

$$\varepsilon_1 < \varepsilon_3 < \varepsilon_2 \quad \text{and} \quad \delta_2 < \delta_3 < \delta_1\,,$$

i.e. the point $(\varepsilon_3, \delta_3)$ is an inner point of the abscissa with the end points $(\varepsilon_1, \delta_1)$ and $(\varepsilon_2, \delta_2)$ since the opposite would distort convexity assumption in (i) of Definition 1. Now let us find an intersection of the line going through the origin and the point $(\varepsilon_3, \delta_3)$ with the curve $\{(\varepsilon, \delta(\varepsilon): \varepsilon \in [0, 1]\}$ and denote it $(\varepsilon_4, \delta_4)$. Then we have – see (ii) of Definition 1 –

$$(7) \qquad w(\varepsilon_4, \delta(\varepsilon_4)) = w(\varepsilon_4, \delta_4) \leqq w(\varepsilon_3, \delta_3) < w(\varepsilon_1, \delta_1) = w(\varepsilon_2, \delta_2)\,.$$

But at the start of the proof we have assumed that $w(\varepsilon_1, \delta_1)$ and $w(\varepsilon_2, \delta_2)$ represent minimum of the function $w(\varepsilon, \delta(\varepsilon))$ which contradicts with (7).

To finish the proof let us assume that $P = Q$. Then we have for any $\varepsilon \in [0, 1]$

$$\inf_H \|(1 - \varepsilon)\, P + \varepsilon H - Q\| = \|(1 - \varepsilon)\, P + \varepsilon P - Q\| = 0\,,$$

where the inf is taken over $H \in \mathbb{M}$.

But taking into account Definition 1 a straightforward implication yields the fact that the function $w(\varepsilon, 0)$ is strictly increasing in ε and hence

$$\arg\min w(\varepsilon, 0) = (0, 0)\,.$$ □

5. ESTIMATION OF CONTAMINATION LEVEL

The characterizing theorem for contamination level parameters gives us a hint for estimation of them. Since p is known (it is chosen by a statistician as a model by which he or she would like to explain data allowing an "inaccuracy", in the form of contamination) it seems that if we estimate q (q is unknown "true" density which generated data) by a density estimator $\hat{q}$ we may use (3) and (4) (in (4) we substitute $\hat{q}$ for q) to establish a pair $(\hat{\varepsilon}_{Q,P}, \hat{\delta}_{Q,P})$ which we hope may serve as an estimator of $(\varepsilon_{Q,P}, \delta_{Q,P})$. Hence we need to introduce some additional notation necessary for density estimation. Let us restrict ourselves in the rest of paper on $\mathscr{X} = \boldsymbol{R}$ and $\mathscr{A} = \mathscr{B}$ (σ-algebra of Borel sets).

Let $\{X_k(\omega)\}_{k=1}^{\infty}$ be a sequence of random variables (r.v.'s) (defined on a probability space $(\Omega, \mathscr{C}, Pr), X_k: \Omega \to \boldsymbol{R}$) which are independent and identically distributed according to a distribution function $\tilde{Q}$ which corresponds to the above fixed probability measure Q with q vanishing outside an interval (c, d) (c and d may be infinite). Further, let $\psi = \{\psi_n(x, y)\}_{n=1}^{\infty}$ be a sequence of functions defined on an interval $(a, b)^2$ with $-\infty \leqq a \leqq c < d \leqq b \leqq \infty$. (We may define $\psi_n(x, y) = 0$ outside $(a, b)^2$ – hence let us assume that $(a, b) = \boldsymbol{R}$.) Put

$$V_n = \sup_{x \in \boldsymbol{R}} \sup_{y \in \boldsymbol{R}} \psi_n(x, y) .$$

Finally, let us denote for any $y \in \boldsymbol{R}$ and $\omega \in \Omega$ by $Q_n(y, \omega)$ the empirical distribution function corresponding to the first n r.v.'s $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$, i.e.

$$Q_n(y, \omega) = 1/n \sum_{k=1}^{n} I_{\{X_k(\omega) \leqq y\}} ,$$

where I_A is the indicator of a set A. Then by a density estimator $\hat{q}_n$ of the density q we shall understand

$$(8) \qquad \hat{q}_n(x, \omega) = \int \psi_n(x, y) \, \mathrm{d}Q_n(y, \omega) = 1/n \sum_{k=1}^{n} \psi_n(x, X_k(\omega))$$

(see [1]). Moreover for any $P^* \in \mathscr{P}_\lambda$ let us denote by

$$\pi_{P^*}(n) = \sup_{x \in \boldsymbol{R}} |\mathsf{E}_{P^*} \hat{q}_n(x, \omega) - p^*(x)| .$$

As the next assertions will refer to Q^∞ (or P^∞) we shall write $\hat{q}_n(x, \mathbf{x})$ instead of $\hat{q}_n(x, \omega)$ where $\mathbf{x} = \{X_1(\omega), X_2(\omega), \ldots\}$. In what follows we shall need a result obtained by Csörgö and Revesz:

Theorem 3 ([1], Theorem 6.21). Let $\hat{q}_n(x, \mathbf{x})$ be the estimator of density q and for any $x \in \boldsymbol{R}$

$$\lim_{y \to -\infty} |\psi_n(x, y)| \{Q(y) . \log(-\log Q(y))\}^{1/2} = 0$$

and

$$\lim_{y \to \infty} |\psi_n(x, y)| \{(1 - Q(y)) \log(-\log(1 - Q(y))\}^{1/2} = 0 .$$

Moreover let

$$V_n' = o(n^{1/2}(\log \log n)^{-1/2})$$

and

$$\pi_Q(n) = o(1) .$$

Then

(9) $$\lim_{n\to\infty} \sup_{x\in \boldsymbol{R}} |\hat{q}_n(x, \mathbf{x}) - q(x)| = 0 \quad \text{a.s.} \quad Q^\infty .$$

Definition 3. We shall say that $\{p_n(x)\}_{n=1}^\infty$ is a monotone quantile sequence of order r if:

(10) $$\sup_{x\in \boldsymbol{R}} |p_n(x) - p(x)| = o(n^{-r})$$

and for any $n \in \boldsymbol{N}$ and $x \in \boldsymbol{R}$ either

$$0 \leqq p_n(x) \leqq p_{n+1}(x) \leqq p(x)$$

or

$$p(x) \leqq p_{n+1}(x) \leqq p_n(x) \quad \text{together with} \quad \int p_1(x)\,\mathrm{d}\nu < \infty .$$

Theorem 2′. Let $\{p_n(x)\}_{n=1}^\infty$ be a monotone quantile sequence of order O. Define for any $n \in \boldsymbol{N}$ and $\mathbf{x} \in \boldsymbol{R}^\infty$ $\hat{\varepsilon}_{Q,P}(n, \mathbf{x})$ and $\hat{\delta}_{Q,P}(n, \mathbf{x})$ as a solution of

$$(\varepsilon, \delta) = \arg\min w(x, y)$$

and

$$\delta = \int_{S_\varepsilon(n,\mathbf{x})}\{(1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x})\}\,\mathrm{d}\nu$$

where $S_\varepsilon(n, \mathbf{x}) = \{x \in \boldsymbol{R}: (1 - \varepsilon)\, p_n(x) > \hat{q}_n(x, \mathbf{x})\}$. Then the pair $(\hat{\varepsilon}_{Q,P}(n, \mathbf{x}), \hat{\delta}_{Q,P}(n, \mathbf{x}))$ is given uniquely.

The proof of the theorem may be carried out along the same lines as of Theorem 2 with the help of lemma:

Lemma 5′. Under assumption of Theorem 2′ fix some $\varepsilon \in [0, 1]$, $n \in \boldsymbol{N}$ and $\mathbf{x} \in \boldsymbol{R}^\infty$ and put

(11) $$\delta(\varepsilon, n, \mathbf{x}) = \int_{S_\varepsilon(n,\mathbf{x})} \{(1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x})\}\,\mathrm{d}\nu .$$

Then $\delta(\varepsilon, n, \mathbf{x})$ has a continuous derivative (with respect to ε)

(12) $$\delta'(\varepsilon, n, \mathbf{x}) = - \int_{S_\varepsilon(n,\mathbf{x})} p_n(x)\,\mathrm{d}\nu$$

(this again means that the derivative exists and is equal to the right-hand side of (12) – at the end points the derivative is understood from one side). Moreover $\delta'(\varepsilon, n, \mathbf{x})$ is nondecreasing and hence $\delta(\varepsilon, n, \mathbf{x})$ is (for fixed $n \in \boldsymbol{N}$ and $\mathbf{x} \in \boldsymbol{R}^\infty$) a convex function of $\varepsilon \in [0, 1]$.

Proof. Let us look at the proof of Lemma 5. What is important in this proof is the monotonicity of S_ε in ε and integrability of p and q. Since $p_n(x)$ is integrable by Definition 3 as well as $\hat{q}_n$ (for any fixed $\mathbf{x} \in \boldsymbol{R}^\infty$ and $n \in \boldsymbol{N}$) and $S_\varepsilon(n, \mathbf{x})$ is also monotone in ε (again for fixed $n \in \boldsymbol{N}$ and $\mathbf{x} \in \boldsymbol{R}^\infty$) the proof is analogous to the proof of Lemma 5.

Theorem 4. Under the assumption of Theorem 2′ and Theorem 3 the pair $(\hat{\varepsilon}_{Q,P}(n, \mathbf{x}), \hat{\delta}_{Q,P}(n, \mathbf{x}))$ is a (strongly) consistent (with respect to Q^∞) estimator of $(\varepsilon_{Q,P}, \delta_{Q,P})$.

Proof. One may verify that for $\delta(\varepsilon, n, \mathbf{x})$ introduced in (11) we have

$$\text{(13)} \qquad |\delta(\varepsilon, n, \mathbf{x}) - \delta(\varepsilon)| =$$

$$= \left|\int_{S_\varepsilon(n,\mathbf{x})} \{(1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x})\}\, \mathrm{d}\nu - \int_{S_\varepsilon}\{(1 - \varepsilon)\, p(x) - q(x)\}\, \mathrm{d}\nu\right| \leqq$$

$$\leqq \int_{S_\varepsilon(n,\mathbf{x}) \cap S_\varepsilon} |(1 - \varepsilon)\,(p_n(x) - p(x)) + q(x) - \hat{q}_n(x, \mathbf{x})|\, \mathrm{d}\nu +$$

$$+ \int_{S_\varepsilon(n,\mathbf{x}) \cap S_\varepsilon{}^c} \{(1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x})\}\, \mathrm{d}\nu +$$

$$+ \int_{S_\varepsilon{}^c(n,\mathbf{x}) \cap S_\varepsilon} \{(1 - \varepsilon)\, p(x) - q(x)\}\, \mathrm{d}\nu\,.$$

Now let $B \in \mathbf{R}^\infty$ be a set such that

$$Q^\infty(B^c) = 0$$

and for any $\mathbf{x} \in B$ the relation (9) is fulfilled. Let $\mathbf{x}_0 \in B$ and $x \in S_\varepsilon(n, \mathbf{x}_0) \cap S_\varepsilon^c$ (for some fixed $\varepsilon \in [0, 1]$ and $n \in \mathbf{N}$). Then we have

$$q(x) - (1 - \varepsilon)\, p(x) \geqq 0\,,$$

hence

$$0 \leqq (1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x}_0) \leqq (1 - \varepsilon)\,(p_n(x) - p(x)) + q(x) - \hat{q}_n(x, \mathbf{x}_0)$$

and finally

$$0 \leqq \int_{S_\varepsilon(n,\mathbf{x}_0) \cap S_\varepsilon{}^c} \{(1 - \varepsilon)\, p_n(x) - \hat{q}_n(x, \mathbf{x}_0)\}\, \mathrm{d}\nu \leqq$$

$$\leqq \int_{S_\varepsilon(n,\mathbf{x}_0) \cap S_\varepsilon{}^c} \{(1 - \varepsilon)\,(p_n(x) - p(x)) + q(x) - \hat{q}_n(x, \mathbf{x}_0)\}\, \mathrm{d}\nu\,.$$

Similar inequality can be obtained for the last integral in (13) and since ε and n was arbitrary it implies that

$$\text{(14)} \qquad |\hat{\delta}(\varepsilon, n, \mathbf{x}_0) - \delta(\varepsilon)| \leqq$$

$$\leqq \int |(1 - \varepsilon)\,(p_n(x) - p(x)) + q(x) - \hat{q}_n(x, \mathbf{x}_0)|\, \mathrm{d}\nu \leqq$$

$$\leqq \{\sup_{x \in \mathbf{R}} |p(x) - p(x)| + \sup_{x \in \mathbf{R}} |\hat{q}_n(x, \mathbf{x}_0) - q(x)|\}\, \nu(\mathbf{R})\,,$$

i.e. we have

$$\text{(15)} \qquad \delta(\varepsilon, n, \mathbf{x}_0) \to \delta(\varepsilon)$$

and due to the fact that the upper bound in (14) does not depend on ε, the convergence is uniform in $\varepsilon \in [0, 1]$ (for any fixed $\mathbf{x}_0 \in B$). Let us assume that $\{(\hat{\varepsilon}_{Q,P}(n, \mathbf{x}_0), \hat{\delta}_{Q,P}(n, \mathbf{x}_0)\}_{n=1}^\infty$ does not converge to $(\varepsilon_{Q,P}, \delta_{Q,P})$. Then there is a subsequence $\{n_k\}_{k=1}^\infty$ of $\mathbf{N}$ such that

$$\hat{\varepsilon}_{Q,P}(n_k, \mathbf{x}_0) \to \varepsilon^* \neq \varepsilon_{Q,P} \quad \text{as} \quad k \to \infty$$

and due to uniform convergence in (15) also

$$\hat{\delta}_{Q,P}(n_k, \mathbf{x}_0) = \delta(\hat{\varepsilon}_{Q,P}(n_k, \mathbf{x}_0), n_k, \mathbf{x}_0) \to \delta(\varepsilon^*) \neq \delta_{Q,P} \quad \text{as} \quad k \to \infty\,.$$

The fact that $\delta(\varepsilon^*) \neq \delta_{Q,P}$ follows from convexity of $\delta(\varepsilon)$ and from (6). From Theorem 2 we have for some $\tau > 0$

$$\text{(16)} \qquad 0 < \tau < -w(\varepsilon_{Q,P}, \delta_{Q,P}) + w(\varepsilon^*, \delta(\varepsilon^*))\,.$$

Making use of the uniform continuity of $w(x, y)$ on $[0, 1]^2$ we may find $k_0 \in \boldsymbol{N}$ such that for any $k \in \boldsymbol{N}$, $k > k_0$ we have

$$(17) \qquad |w(\hat{\varepsilon}_{Q,P}(n_k, \mathbf{x}_0), \hat{\delta}_{Q,P}(n_k, \mathbf{x}_0)) - w(\varepsilon^*, \delta(\varepsilon^*))| < \tau/2$$

and

$$(18) \qquad |w(\varepsilon_{Q,P}, \delta_{Q,P}) - w(\varepsilon_{Q,P}, \delta(\varepsilon_{Q,P}, n_k, \mathbf{x}_0))| < \tau/2 .$$

But from (16), (17) and (18) it follows for any $k > k_0$

$$w(\hat{\varepsilon}_{Q,P}(n_k, \mathbf{x}_0), \hat{\delta}_{Q,P}(n_k, \mathbf{x}_0)) > w(\varepsilon_{Q,P}, \delta(\varepsilon_{Q,P}, n, \mathbf{x}_0))$$

which contradicts with definition of $(\hat{\varepsilon}_{Q,P}(n, \mathbf{x}_0), \hat{\delta}_{Q,P}(n, \mathbf{x}_0))$. So we have $\hat{\varepsilon}_{Q,P}(n, \mathbf{x}_0) \to \to \varepsilon_{Q,P}$ and $\hat{\delta}_{Q,P}(n, \mathbf{x}_0) \to \delta_{Q,P}$ as $n \to \infty$, but it is nothing else then

$$(\hat{\varepsilon}_{Q,P}, \hat{\delta}_{Q,P}) \to (\varepsilon_{Q,P}, \delta_{Q,P}) \quad \text{a.s.} \quad Q^\infty$$

since $\mathbf{x}_0 \in B$ was arbitrary. □

For a heavy tailed contamination we may offer an idea about the order of consistency of the contamination level estimator. In what follows we would like to assume that the support of q is a subset of the support of p. From the practical point of view it seems to be a natural demand because the opposite would imply that we try to explain data by a density which – at least asymptotically – does not cover the range of them.

Since on the other hand $q = (1 - \varepsilon) p + \varepsilon h$, the support of q cannot be a proper subset of support of p. So we assume that we guess the "explaining" density p just to "cover" all (possible) data.

Definition 4. We shall say that Q is a heavy-tailed with respect to P if we have for the corresponding densities:

$$\lim_{x \to c+} \frac{q(x)}{p(x)} = \infty \quad \text{and} \quad \lim_{x \to d-} \frac{q(x)}{p(x)} = \infty ,$$

(i.e. the limits exist and have the required values. For the meaning of c and d see the beginning of Section 5.)

Further we shall restrict, ourselves to kernel estimators of density, i.e. we shall assume that $\psi_n(x, y) = h_n^{-1} \tilde{\lambda}((x - y) h_n^{-1})$ for some density $\tilde{\lambda}$ and a real sequence $\{h_n\}$ which properties will be given in the next theorem. We shall need the following theorem of Csörgö and Révész:

Theorem 5 ([1], Theorem 6.2.5). Suppose that

(i) q is vanishing outside the interval $[0, 1]$,
(ii) q is twice differentiable over $(0, 1)$ and $|q''| \leqq C < \infty$,
(iii) q is strictly positive on $(0, 1)$, say $q \geqq \alpha > 0$,
(iv) $\tilde{\lambda}(x) \leqq C$,

(v) $\tilde{\lambda}(-x) = \lambda(x)$,

(vi) $\lim_{x\to\infty} x^4\, \tilde{\lambda}(x) = 0$,

(vii) $\tilde{\lambda}$ is twice differentiable on an interval $-\infty \leqq -a < +a \leqq \infty$,

(viii) $h_n \searrow 0$, $nh_n \nearrow \infty$,

and

(ix)
$$\frac{\log^4 n}{nh_n \log h_n^{-1}} \to 0\,, \qquad \frac{nh_n^5}{\log h_n^{-1}} \to 0\,.$$

Then for any $\Delta > 0$ we have

$$\lim_{n\to\infty}\left[\frac{nh_n}{2\Lambda^2 \log h_n^{-1}}\right]^{1/2} \sup_{\Delta<x<1-\Delta}\left|\frac{\hat{q}_n(x, \mathbf{x}) - q(x)}{q^{1/2}(x)}\right| = 1 \quad \text{a.s.} \quad Q^\infty\,, \tag{19}$$

where $\Lambda^2 = \int_{-\infty}^{\infty} \tilde{\lambda}^2(x)\, dx$.

To be able to apply the recalled theorem we have to consider $c = 0$ and $d = 1$ which may seem at the first glance a little restricting. Due to assumption that the support of q is not larger than the support of p the remedy in many cases may be a transformation of data by means of

$$z = F_P(x)$$

where F_P is the distribution function corresponding to the probability measure P (through the identity-random variable).

Notice that due to the form of Definition 2 the values of $\varepsilon_{Q,P}$ and $\delta_{Q,P}$ do not depend on a transformation of random variable.

Theorem 6. Let the assumptions of Theorem 3 and Assumptions (i)–(vii) of Theorem 5 be fulfilled. Put $h_n = h\,.\,n^{-1/2}$ for some $h > 0$. Moreover let Q be heavy-tailed with respect to P and for some $\tau > 0$, $\{p_n\}_{n=1}^{\infty}$ be a monotone quantile sequence of order $-(\frac{1}{2} - \tau)$. Having defined for any $x_0 \in (0, 1]$ and $y_0 \in (0, 1]$ the function of λ by

$$\tilde{w}_{x_0,y_0}(\lambda) = w(\lambda x_0, (1 - \lambda)\, y_0)$$

let us assume that for the λ_0 – the point of the unique minimum of $\tilde{w}_{x_0,x_0}(\lambda)$ (see Definition 1) – we have

$$\sup_{0<\lambda_1<\lambda_2<\lambda_0} \frac{\tilde{w}_{x_0,y_0}(\lambda_1) - \tilde{w}_{x_0,y_0}(\lambda_2)}{\lambda_2 - \lambda_1} > K \tag{20}$$

and

$$\inf_{\lambda_0<\lambda_1<\lambda_2<1} \frac{\tilde{w}_{x_0,y_0}(\lambda_2) - \tilde{w}_{x_0,y_0}(\lambda_1)}{\lambda_2 - \lambda_1} > K \tag{21}$$

for some $K > 0$. Then for any $\tau > 0$

$$|\hat{\varepsilon}_{Q,P} - \varepsilon_{Q,P}| + |\hat{\delta}_{Q,P} - \delta_{Q,P}| = o(n^{-(1/2-\tau)}) \quad \text{a.s.} \quad Q^\infty\,.$$

Proof. Let $B \subset [0, 1]^\infty$ be a set such that $Q^\infty(B^c) = 0$ and for any $\mathbf{x} \in B$ the relation (9) and (19) are fulfilled. Let $\mathbf{x}_0 \in B$. We shall prove that there is a $\Delta \in (0, 1)$

and $n_0 \in \boldsymbol{N}$ such that for any $\varepsilon \in [0, 1]$ and $n \in \boldsymbol{N}$, $n \geqq n_0$

$$S_\varepsilon \cup S_\varepsilon(n, \mathbf{x}_0) \subset (\Delta, 1 - \Delta) .$$

Taking into account that Q is heavy-tailed with respect to P we may for any $K_1 > 1$ find a Δ such that for any $x \in [0, \Delta] \cup [1 - \Delta, 1]$ we have

$$\frac{q(x)}{p(x)} > K_1 .$$

Denote for any $0 < \beta < \alpha$ (see assumption (iii) of Theorem 5) by

$$A_1 = \{x \in [0, \Delta] \cup [1 - \Delta, 1] \colon p(x) < \beta\}$$

and

$$A_2 = \{x \in [0, \Delta] \cup [1 - \Delta, 1] \colon p(x) \geqq \beta\} .$$

Then for any $x \in A_1$ it holds

$$q(x) - p(x) > \alpha - \beta$$

and for any $x \in A_2$ analogously

$$q(x) - p(x) > (K_1 - 1)\, p(x) > (K_1 - 1) \,.\, \beta .$$

Hence there is a $\gamma > 0$ such that for any $x \in [0, \Delta] \cup [1 - \Delta, 1]$

$$q(x) - \gamma > p(x) .$$

It implies that $S_\varepsilon \subset (\Delta, 1 - \Delta)$. Let us find $n_0 \in \boldsymbol{N}$ such that for any $n \in \boldsymbol{N}$, $n \geqq n_0$

$$\sup_{x \in [0,1]} |\hat{q}_n(x, \mathbf{x}_0) - q(x)| < \gamma/2 \quad \text{and} \quad \sup_{x \in [0,1]} |p_n(x) - p(x)| < \gamma/2 .$$

But then for any $x \in [0, \Delta] \cup [1 - \Delta, 1]$ and $n \geqq n_0$

$$\hat{q}_n(x, \mathbf{x}_0) > p_n(x) ,$$

i.e. $S_\varepsilon(n, \mathbf{x}_0) \subset (\Delta, 1 - \Delta)$, too. Now we may proceed step by step as in the proof of Theorem 4. We obtain

$$(14') \qquad |\hat{\delta}(\varepsilon, n, \mathbf{x}_0) - \delta(\varepsilon)| \leqq$$
$$\leqq \{ \sup_{x \in (\Delta, 1 - \Delta)} |p_n(x) - p(x)| + \sup_{x \in (\Delta, 1 - \Delta)} |\hat{q}_n(x, \mathbf{x}_0) - q(x)|\}\, \nu(\mathbf{R}) ,$$

i.e. we have

$$\hat{\delta}(\varepsilon, n, \mathbf{x}_0) - \delta(\varepsilon) = o(n^{-(1/2 - \tau)})$$

uniformly in $\varepsilon \in [0, 1]$ (due to the fact that (14') does not depend on ε). Making use of (20) and (21) one may finish the proof in a similar way as the proof of the Theorem 4 was finished. □

Remark 4. The assumption of Theorem 6 may be considered to be rather restrictive, especially the assumption about the support of q not being out of interval $[0, 1]$. On the other hand since the values of $\varepsilon_{Q,P}$, $\delta_{Q,P}$ and $\hat{\varepsilon}_{Q,P}$ and $\hat{\delta}_{Q,P}$ do not depend on transformation of densities all of assumption of this kind are only of technical ones. Really, one may assume that data were transformed by an appropriate transformation then the values of $\hat{\varepsilon}_{Q,P}$ and $\hat{\delta}_{Q,P}$ were evaluated having desired (and above proved) properties and since not depending on this transformation the value of $\hat{\varepsilon}_{Q,P}$ and $\hat{\delta}_{Q,P}$ are the same for the original data and hence have also the desired properties if calculated directly from the original data.

6. NUMERICAL EXAMPLE

The main result of the paper – consistency of the contamination level estimator – is an asymptotic one and hence to reach a practical applicability needs to perform some numerical study. Such a study should show how to choose "free" parameters, namely the monotone quantile sequence, the type of density estimator etc. Here we restrict ourselves only on presenting a few basic results to offer the reader a possibility to create an idea how the estimator really works. But we are aware that the issue needs deeper, more complex study which should recognize valuability (or invaluability) of contamination level estimation.

For the just described study the kernel estimator was assumed in the form

$$\hat{q}_n(x, \mathbf{x}) = \frac{1}{n\, h_n \sqrt{(2\pi)}} \sum_{i=1}^{n} \exp\left\{-\frac{(x - x_i)^2}{2h_n^2}\right\}.$$

Throughout the whole study successively the samples containing always 40 standard normal number were generated (i.e. n was fixed being equal to 40). Any time the normality was checked by means of the χ^2-test and the Kolmogorov-Smirnov test (75% – quantile for χ^2-test and 80% – quantile for the Kolmogorov-Smirnov-test were used) together with 80% confidence intervals for the mean and variance. The study was divided into three parts. At the first one the "optimal" width h_n of window was found. Since the contamination level estimator is based on the integration of the function $(1 - \varepsilon)\, p_n - \hat{q}_n$ over the region of its positivity, as an "optimal" width of window was assumed such for which the integral

$$\int_{\{f(x) - \hat{q}(x,\mathbf{x}) > 0\}} [f(x) - \hat{q}_n(x, \mathbf{x})]\, dx$$

(where $f(x)$ is the standard normal density and $\mathbf{x}$ is the sample) was minimal. 50 samples were generated and at each of them the optimal value of h_n – the value minimizing the above given integral – was found. The mean of these 50 values (equal to 2·13203) was used as the width of window in the next two steps of the study. The quantile p_{40} we have considered in the form *const* . $f(x)$ and the goal of the second step was to find the "optimal" value of *const*. The framework was as follows. Again 50 samples was generated (and checked for normality) and contaminated in such a way that the 8 observations were multiplied by 3, i.e. we obtained 50 samples from the mixture

$$80\% \mathcal{N}(0, 1) + 20\% \mathcal{N}(0, 9)$$

(and consequently $Q = 0{\cdot}8P_1 + 0{\cdot}2P_2$ where P_1 and P_2 were probability measures

Table 1.

Case number	Mean	Variance	Chi square statistic before contamination	Chi square statistic after contamination	Kolmogorov-Smirnov statistic before contamination	Kolmogorov-Smirnov statistic after contamination	Estimated value of epsilon before contamination	Estimated value of epsilon after contamination
1	—0·0331	0·7414	7·2	3·2	0·082	0·100	0·006401	0·038343
2	—0·0660	1·1192	4·4	7·2	0·061	0·091	0·009111	0·046755
3	—0·1362	1·2904	4·8	8·8	0·139	0·147	0·047990	0·107097
4	0·1428	1·0749	1·6	5·2	0·127	0·127	0·026614	0·103544
5	—0·0360	0·9666	2·8	2·8	0·098	0·098	0·010155	0·040947
6	—0·1152	1·2560	5·6	9·2	0·163	0·163	0·069016	0·146373
7	—0·0398	0·9751	3·2	6·8	0·069	0·137	0·007071	0·078992
8	—0·1595	0·9724	6·8	10·8	0·126	0·151	0·003917	0·059766
9	0·1028	0·7861	4·0	3·2	0·132	0·132	0·027846	0·098341
10	—0·0367	0·7595	7·6	7·2	0·127	0·102	0·045531	0·061999
11	—0·0037	1·2730	8·4	12·0	0·111	0·161	0·069100	0·138265
12	—0·0961	0·9411	3·2	2·4	0·137	0·115	0·031319	0·040475
13	—0·1288	0·9126	5·2	5·6	0·160	0·135	0·048938	0·087377
14	—0·1087	0·7321	6·0	5·2	0·116	0·116	0·000753	0·038268
15	—0·0788	0·9888	8·8	6·4	0·100	0·136	0·008767	0·060098
16	0·1470	1·1524	4·4	6·8	0·083	0·095	0·025742	0·071244
17	0·0761	0·8924	3·6	4·0	0·076	0·076	0·028803	0·071697
18	0·1063	0·8080	6·4	4·4	0·139	0·139	0·054623	0·057006
19	—0·1197	0·8096	6·8	6·8	0·154	0·119	0·058225	0·068387
20	—0·0009	1·0582	1·2	5·6	0·060	0·093	0·020072	0·081555
21	0·1418	0·8744	6·4	4·0	0·076	0·101	0·005468	0·036011
22	0·0730	0·8241	6·4	11·2	0·096	0·127	0·016448	0·058796
23	—0·0099	1·0590	6·8	4·4	0·062	0·080	0·019964	0·068583
24	0·0016	1·0966	8·4	10·0	0·094	0·079	0·018142	0·053666
25	—0·1433	1·2489	4·4	12·0	0·092	0·155	0·000000	0·050949
26	—0·0658	0·8350	6·0	4·0	0·132	0·094	0·042856	0·055881
27	0·1906	0·9023	2·0	6·8	0·101	0·115	0·017113	0·037184
28	—0·1664	0·8284	6·8	9·6	0·145	0·198	0·031210	0·070357
29	—0·0278	1·1520	3·6	12·8	0·131	0·134	0·057530	0·157378
30	—0·0458	1·2811	6·4	15·6	0·128	0·175	0·085353	0·144683
31	—0·0277	0·7941	7·2	2·8	0·126	0·126	0·049226	0·062877
32	—0·1347	1·2078	4·8	4·0	0·120	0·131	0·021253	0·035756
33	0·0980	0·9867	6·0	6·4	0·125	0·125	0·036545	0·102035
34	0·1535	1·2594	6·8	12·8	0·146	0·200	0·117009	0·168553
35	—0·0347	1·0985	3·6	4·4	0·159	0·159	0·065334	0·077020
36	0·0682	1·1276	4·4	8·8	0·078	0·128	0·008612	0·066701
37	0·0153	0·7903	1·6	0·8	0·061	0·057	0·000300	0·009714
38	—0·0039	0·9284	1·6	1·6	0·058	0·105	0·001538	0·072945
39	0·0298	0·8049	8·8	8·8	0·104	0·081	0·006765	0·034500
40	—0·1851	1·0918	4·8	10·4	0·157	0·207	0·055808	0·111866

(Tab. 1 contin.)

Case number	Mean	Variance	Chi square statistic		Kolmogorov-Smirnov statistic		Estimated value of epsilon	
			before contamination	after contamination	before contamination	after contamination	before contamination	after contamination
41	0·0553	1·0466	2·8	6·4	0·070	0·098	0·022936	0·083156
42	−0·0620	0·7245	4·8	3·2	0·120	0·070	0·024150	0·016618
43	−0·1806	1·0460	3·2	4·8	0·102	0·143	0·005437	0·046194
44	0·0430	1·1590	3·6	6·0	0·086	0·105	0·035179	0·076975
45	−0·0408	1·1078	3·6	4·8	0·100	0·125	0·047953	0·089803
46	0·0622	1·0244	3·2	6·0	0·081	0·106	0·029828	0·095459
47	−0·2003	0·9477	6·4	8·4	0·152	0·196	0·041174	0·111036
48	−0·0745	1·2449	4·4	6·8	0·116	0·141	0·064426	0·084307
49	0·0134	0·7373	3·6	1·6	0·063	0·092	0·003920	0·022636
50	0·0379	0·7970	2·0	1·2	0·065	0·067	0·000247	0·033998

generated by the distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, 9)$, respectively). Finally $P = P_1$ and

$$w(x, y) = \max\left[1000x - 1998y, 1000y - 499x\right]$$

were selected. It implies that the theoretical values of $\varepsilon_{Q,P}$ and $\delta_{Q,P}$ are equal to 0·0723 and 0·03615, respectively (see Definition 2). And now the values of const. was selected so that the mean values of $\hat{\varepsilon}_{Q,P}$ and $\hat{\delta}_{Q,P}$ over the above mentioned 50 contaminated samples were approximately equal to the theoretical values $\varepsilon_{Q,P}$ and $\delta_{Q,P}$, respectively.

The corresponding value of *const* was 0·870. (The results produced for this value of *const* are presented in Table 1 which is included mainly for the further purposes. In this table only $\hat{\varepsilon}_{Q,P}$ was presented because the precision of evaluation of $\hat{\delta}_{Q,P}$ is such that we obtain always $\hat{\delta}_{Q,P} = \frac{1}{2}\hat{\varepsilon}_{Q,P}$.)
This value 0·870 was used in the third, the last step of study. At this step for a few contamination levels the sets of 50 samples, contaminated in a corresponding way, (each again containing 40 "observations") were generated. In the following table the means over the above mentioned sets (for given values of contamination level) are presented (denoting them mean $\hat{\varepsilon}_{Q,P}$ and mean $\hat{\delta}_{Q,P}$, respectively). At its first column the percentage of contaminated units is given. It means that e.g. for the value 0·15 the samples from 85% $\mathcal{N}(0, 1)$ + 15% $\mathcal{N}(0, 9)$ were generated (again 50 of such samples) by means of multiplying by 3 six units in every "pure" sample from $\mathcal{N}(0, 1)$. The number of "contaminated" units, in this case six units, is given in the second column of Table 2.

For this mixture (85% $\mathcal{N}(0, 1)$ + 15% $\mathcal{N}(0, 9)$) the theoretical values of $\varepsilon_{Q,P}$ and $\delta_{Q,P}$ are 0·05422 and 0·02711. Since the function w is strictly convex we obtain for any mixture $\delta_{Q,P} = \frac{1}{2}\,\varepsilon_{Q,P}$. The same is true for estimated values $\hat{\varepsilon}_{Q,P}$ and $\hat{\delta}_{Q,P}$

Table 2.

Percentage of contamination	Number of "contaminated" units	$\varepsilon_{Q,P}$	Mean value of $\hat{\varepsilon}_{Q,P}$	Mean value of $\hat{\varepsilon}_{P,P}$
15%	6	0·05422	0·05994	0·0306
17·5%	7	0·06326	0·06445	0·0239
20%	8	0·07230	0·07264	0·0303
22·5%	9	0·08129	0·078055	0·0306
25%	10	0·09033	0·077936	0·0265

for all samples, i.e. $\hat{\varepsilon}_{Q,P} = 2\hat{\delta}_{Q,P}$. That is why in the following tables only values $\varepsilon_{Q,P}$ and $\hat{\varepsilon}_{Q,P}$ – in fact mean value of $\hat{\varepsilon}_{Q,P}$ over mentioned 50 samples – are presented (the former in the third, the latter in the fourth column). To offer the idea how the estimator works for noncontaminated samples at the last column the means over 50 noncontaminated samples are given.

As it follows from Table 1 in many cases values of χ^2 and Kolmogorov-Smirnov statistics for contaminated samples were still under selected quantiles (9·037 for χ^2 and 0·1655 for Kolmogorov-Smirnov tests – the opposite is true only for 12 samples from 50). But looking on this situation from practical point of view we must admit that the suspicion that sample is contaminated would usually arise when the sample is rejected being normal at least by one of the above mentioned tests. Naturally, to describe the behaviour of an estimator of contamination level for this framework (i.e. when we restrict ourselves on that part of sampling space at which χ^2 or the Kolmogorov-Smirnov statistics exceed some given level) would be much more difficult than in the nonrestricted case. Table 3 offers results of numerical study for such case (from the technical reasons the σ of contaminating $\mathcal{N}(O, \sigma^2)$ distribution was chosen to be 9). The value of *const* chosen for this case was 0·865.

Table 3.

Percentage of contamination	Number of "contaminated" units	$\varepsilon_{Q,P}$	Mean value of $\hat{\varepsilon}_{Q,P}$	Mean value of $\hat{\varepsilon}_{P,P}$
15%	6	0·08091	0·08701	0·03184
17·5%	7	0·09439	0·97928	0·03489
20%	8	0·10788	0·107403	0·03266
22·5%	9	0·12136	0·11289	0·03081
25%	10	0·13485	0·119484	0·02582

Two conclusions follow from this table (together with Table 1): At first the mean values of $\hat{\varepsilon}_{Q,P}$ are not considerably better estimation of $\varepsilon_{Q,P}$ then in the nonrestricted case. Secondly, estimation of zero-contamination level is biased but stable. It may lead to two conjectures. Firstly, it is probably worthless to built up the theory for restricted case. Secondly, may be that it would be possible to propose an adaptive estimator (with *const* adapting to the "true" contamination level) which would estimate better the contamination level.

(Received February 16, 1988.)

REFERENCES

[1] M. Csörgö and P. Révész: Strong Approximations in Probability and Statistics. Akadémiai Kiadó, Budapest 1981.

[2] H. Rieder: Least favourable pairs for special capacities. Ann. Statist. *5* (1977), 909—921.

[3] J. Á. Víšek: Estimating contamination level. In: Proc. of the Fifth Pannonian Symposium on Mathematical Statistics, Visegrád 1985. Akadémiai Kiadó, Budapest 1987, pp. 401—414.

RNDr. Jan Ámos Víšek, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.