PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 14, Fasc. 2 (1993), pp. 173–187

ON THE ROLE OF CONTAMINATION LEVEL AND THE LEAST FAVOURABLE BEHAVIOUR OF GROSS-ERROR SENSITIVITY

BY

J. Á. VÍŠEK (PRAGUE)

Abstract. The notion of contamination level is introduced and its characterization for any pair of distribution functions is given. A possibility of reformulation of some basic problems of the robust statistics based on this notion is discussed. Finally, the behaviour of the gross-error sensitivity under the least favourable distribution is studied and the result is illustrated by a numerical example.

Introduction. The problems connected with the optimality of robust procedure have been studied by many authors (see, e.g., [3]). Presumably one of the best known studies is that one given by Hampel et al. [4], based on the influence function. In fact, the problems have been already formulated in Huber's pioneering paper [5]. In the present paper we shall show that the constraint under which Huber gave his famous minimax solution may be reformulated by means of the notion of contamination level. We shall also demonstrate that the bounds imposed on the gross-error sensitivity in the well-known Hampel's extremal problem is a one-to-one function of the contamination level (under general conditions). This implies that the notion of contamination level appears to be one of the basic notions or, in other words, the notion of contamination level may be used as the fundamental one for robust statistics.

The gross-error sensitivity introduced by Hampel in [3] is one of basic characteristics of the robust procedures. This characteristic has been in detail studied in [4]. One thing which is not explicitly emphasized in [4] is the fact that the results yielded by the approach via the influence function are given under the central model. In difference, Huber's minimax solution is given under the least favourable distribution. It occurred from the practical experiences that the data are usually better fitted by a model with heavy tails, like the Student one with a small number of degrees of freedom (for a large, exhaustive and very nice discussion see [4]). In such a case, Huber's approach may better reflect the real situation. Generally, we can say that in the case where the data are distributed according to a distribution which is (rather) near to the central model the approach via the influence function may give acceptable approximations for the behaviour of the robust procedures. It is true that we may already select as the central model a distribution with heavy tails, however in this case the corresponding formulas may be more complicated and we may get into some, not only computational, troubles. In other cases, where we prefer a simple central model although the data may be distributed according to a (complicated, unknown) distribution which may differ from it, it may be better to consider Huber's approach and to study the behaviour of the statistics under the least favourable distribution.

Moreover, a large attention which was devoted to the procedures with high breakdown point indicates that in some cases our idea about the character of randomness may be so vague that we should rely on the results describing the behaviour of the robust procedure under the least favourable distribution. Then it may appear that the behaviour of some characteristics of the robust procedure is different from their behaviour in the neighbourhood of the central model. We will demonstrate this phenomenon by the example of the behaviour of the gross-error sensitivity.

To make the paper easy understandable even for a reader which is not in an everyday touch with the robust statistics we will repeat in the next section some basic notions. The reader familiar with them may freely skip this part of the paper.

Notation and preliminaries. Let $\boldsymbol{N}$ denote the set of all positive integers. $\mathscr{B}$ is assumed to be the Borel σ-algebra of the subsets of the real line $\boldsymbol{R}$, and $\mathscr{F}$ the set of all one-dimensional distribution functions. Huber's result [5] was derived under the following condition:

CONDITION A. Assume that $F_0 \in \mathscr{F}$ has a density $f_0(t)$ with a convex support. Moreover, let f_0 be twice continuously differentiable with $-\log f_0(t)$ strictly convex on the support of $f_0(t)$. ∎

Without any loss of generality let us suppose that $\sup\{t \in \boldsymbol{R}; f_0'(t) > 0\} = 0$. For any $\varepsilon \in [0, 1]$ define the contamination model of data

$$(1) \qquad \mathscr{P}_{F_0}(\varepsilon) = \{G \in \mathscr{F}:\ G = (1-\varepsilon) F_0 + \varepsilon H;\ H \in \mathscr{F}\}.$$

Now, let the random variables $X_1, X_2, \ldots, X_n$ be independent and identically distributed according to a distribution function $G^*(t - \Delta)$, $\Delta \in \boldsymbol{R}$. ($G^*$ may be any distribution from $\mathscr{P}_{F_0}(\varepsilon)$ for some fix $\varepsilon \in [0, 1]$.) Let $\hat{\Delta}_\psi$ be the M-estimator of Δ given as a solution of the equation (for a moment let us assume that there is a solution)

$$(2) \qquad \sum_{i=1}^{n} \psi(X_i - t) = 0,$$

where $\psi\colon \boldsymbol{R}\to\boldsymbol{R}$ is a function which will be specified later. Finally, for any $G\in\mathscr{P}_{F_0}(\varepsilon)$ and any ψ let us denote by

(3) $$V(\psi, G)$$

the asymptotic variance of the M-estimator $\hat{\Delta}_\psi$ under G (if the variance exists). Huber's result then says:

ASSERTION 1 (Huber [5]). *Let Condition* A *be fulfilled. Then*

(4) $$\inf_{\Psi(F_\varepsilon)} V(\psi, F_\varepsilon) = V(\psi_\varepsilon, F_\varepsilon) = \sup_{\mathscr{G}(\psi_\varepsilon)} V(\psi_\varepsilon, G),$$

where

(5) $$\psi_\varepsilon(t) = \begin{cases} -k(\varepsilon), & \{t\in\boldsymbol{R}\colon f_0'(t)/f_0(t) > k(\varepsilon)\},\\ -f'(t)/f(t), & \{t\in\boldsymbol{R}\colon |f_0'(t)/f_0(t)| \leqslant k(\varepsilon)\},\\ k(\varepsilon), & \{t\in\boldsymbol{R}\colon f_0'(t)/f_0(t) < -k(\varepsilon)\},\end{cases}$$

and $k(\varepsilon)$ *is related to* ε *by the equation*

(6) $$(1-\varepsilon)^{-1} = \int_{t_0(\varepsilon)}^{t_1(\varepsilon)} f_0(t)\,dt + \frac{f_0\big(t_0(\varepsilon)\big)+f_0\big(t_1(\varepsilon)\big)}{k(\varepsilon)}$$

with $t_0(\varepsilon) < t_1(\varepsilon)$ *being the end points of the interval* $\{t\in\boldsymbol{R}\colon |f_0'(t)/f_0(t)| \leqslant k(\varepsilon)\}$, *which means that for finite interval we have*

(7) $$f_0'\big(t_i(\varepsilon)\big)/f_0\big(t_i(\varepsilon)\big) = (-1)^i k(\varepsilon), \quad i = 0, 1$$

(see [5]). *The distribution* F_ε *is such that* $f_\varepsilon'(t)/f_\varepsilon(t) = -\psi_\varepsilon(t)$, *i.e.*, $\hat{\Delta}_{\psi_\varepsilon}$ *is the maximum likelihood estimator for the* F_ε *with density*

(8) $$f_\varepsilon(t) = \begin{cases} (1-\varepsilon)\, f_0\big(t_0(\varepsilon)\big)\exp\{k(\varepsilon)\big(t-t_0(\varepsilon)\big)\} & \text{for } t < t_0(\varepsilon),\\ (1-\varepsilon)\, f_0(t) & \text{for } t_0(\varepsilon) \leqslant t \leqslant t_1(\varepsilon),\\ (1-\varepsilon)\, f_0\big(t_1(\varepsilon)\big)\exp\{-k(\varepsilon)\big(t-t_0(\varepsilon)\big)\} & \text{for } t > t_1(\varepsilon).\end{cases}$$

Finally,

$$\Psi(F_\varepsilon) = \{\psi\colon \boldsymbol{R}\to\boldsymbol{R};\ \mathrm{E}_{F_\varepsilon}\psi = 0,\ V(\psi, F_\varepsilon) \text{ exists}\}$$

and

$$\mathscr{G}(\psi_\varepsilon) = \{G\in\mathscr{P}_{F_0}(\varepsilon);\ \mathrm{E}_G\psi_\varepsilon = 0,\ V(\psi_\varepsilon, G) \text{ exists}\}.$$

For the proof see [5]. ∎

Contamination level. The parameter $\varepsilon\in[0, 1]$ in (1) may be interpreted as a level of contamination of data, however some care is necessary.

For any distribution function we denote its density by the corresponding lower case letter (it will be clear from the context with respect to which measure the density is understood).

For any pair of distribution functions G and F there is an $\varepsilon\in[0, 1]$ and a distribution function H so that

(9) $$G(t) = (1-\varepsilon)\,F(t) + \varepsilon H(t).$$

(Indeed, we have at least $G(t)=(1-1)F(t)+1\cdot G(t)$.) If, however, there is $\varepsilon\in[0, 1)$ for which (9) holds, then for any positive $\varepsilon^*\in(\varepsilon, 1]$ we have also

$$G(t)=(1-\varepsilon^*)F(t)+\varepsilon^* H^*(t) \qquad \text{for } H^*(t)=\frac{1}{\varepsilon^*}\{(\varepsilon^*-\varepsilon)F(t)+\varepsilon H\}.$$

Hence, we need the following definition.

Definition 1. For any pair of distribution functions G and F, the *contamination level* of G with respect to F will be given by the value

$$\varepsilon_{G,F}=\inf\{\varepsilon:\ G(t)=(1-\varepsilon)F(t)+\varepsilon H(t),\ H(t)\ \text{d.f.}\}.$$

Example 1. By considering a mixture of two normal laws

$$g(x)=\frac{1}{\sqrt{2\pi}}\left\{(1-\varepsilon)\exp\left\{-\frac{t^2}{2}\right\}+\varepsilon\sigma^{-1}\exp\left\{-\frac{t^2}{2\sigma^2}\right\}\right\},$$

a straightforward computation gives $\varepsilon_{G,F}=(1-\sigma^{-1})\varepsilon$ for $\sigma>1$ while $\varepsilon_{G,F}=\varepsilon$ for $\sigma<1$.

Lemma 1 (Characterization of $\varepsilon_{G,F}$). *Let G and F be distribution functions and ν a σ-finite measure such that $G\ll\nu$ as well as $F\ll\nu$ and put*

$$\mathscr{D}=\{g^*(t), f^*(t):\ \nu\{g^*(t)\neq g(t)\}=0,\ \nu\{f^*(t)\neq f(t)\}=0\}.$$

Then

$$\varepsilon_{G,F}=\inf_{\mathscr{D}}\ \sup_{\{t:f^*(t)>0\}}\frac{f^*(t)-g^*(t)}{f^*(t)}.$$

Proof. Let $\varepsilon_{G,F}<1$. Then, by Definition 1, for any $\varepsilon\in(\varepsilon_{G,F}, 1]$ there is $h_\varepsilon(t)$ so that

$$g(t)=(1-\varepsilon)f(t)+\varepsilon h_\varepsilon(t),$$

which yields $\{f(t)-g(t)\}f^{-1}(t)\leqslant\varepsilon$ (for any $\varepsilon\in(\varepsilon_{G,F}, 1]$ and $t\in\{z: f(z)>0\}$). Hence

$$\sup_{\{t:f(t)>0\}}\{f(t)-g(t)\}f^{-1}(t)\leqslant\varepsilon_{G,F}.$$

Since

$$\sup_{\{t:f(t)>0\}}\{f(t)-g(t)\}f^{-1}(t)\leqslant 1,$$

the inequality $\sup_{\{t:f(t)>0\}}\{f(t)-g(t)\}f^{-1}(t)\leqslant\varepsilon_{G,F}$ holds also for $\varepsilon_{G,F}=1$. Assume that

$$\sup_{\{t:f(t)>0\}}\{f(t)-g(t)\}f^{-1}(t)<\varepsilon_{G,F}.$$

Then there is $\delta>0$ such that: (i) $\varepsilon_{G,F}-\delta>0$, and (ii) for all $t\in\{z: f(z)>0\}$

$$f(t)-g(t)\leqslant(\varepsilon_{G,F}-\delta)f(t).$$

Since the last inequality holds also for $t \notin \{z: f(z) > 0\}$, the function

$$h(t) = (\varepsilon_{G,F} - \delta)^{-1} \{g(t) - [1 - (\varepsilon_{G,F} - \delta)] f(t)\}$$

is a density and we have

$$g(t) = [1 - (\varepsilon_{G,F} - \delta)] f(t) + (\varepsilon_{G,F} - \delta) h(t),$$

which contradicts the definition of $\varepsilon_{G,F}$. ∎

Remark 1. Since $\varepsilon_{G,F}$ is given uniquely (by Definition 1), it is clear that the characterization does not depend on ν. ∎

Remark 2. It is clear that the set $\mathscr{D}$ had to be included in Lemma 1 because changing $f(t)$ (to be positive) and $g(t)$ (to be zero) at one point t such that $\nu(\{t\}) = 0$, we obtain $\varepsilon_{G,F} = 1$ for any pair G, F. ∎

Remark 3. Arguing as at the end of the proof of Lemma 1 we may show that there is $h_{\varepsilon_{G,F},\nu}(t)$ such that

$$g(t) = (1 - \varepsilon_{G,F}) f(t) + \varepsilon_{G,F} h_{\varepsilon_{G,F},\nu}(t).$$

Integrating we obtain

$$G(t) = (1 - \varepsilon_{G,F}) F(t) + \varepsilon_{G,F} H_{\varepsilon_{G,F}}(t),$$

so that the infimum from Definition 1 is attained. ∎

LEMMA 2. *Let Condition* A *be fulfilled. Then*

$$\max_{G \in \mathscr{P}_{F_0}(\varepsilon)} \varepsilon_{G,F_0} = \varepsilon_{F_\varepsilon,F_0} = \varepsilon.$$

Proof. It is clear from Definition 1 that

$$\sup_{G \in \mathscr{P}_F(\varepsilon)} \varepsilon_{G,F_0} \leqslant \varepsilon.$$

So it is sufficient to show that $\varepsilon_{F_\varepsilon,F_0} = \varepsilon$. We are going to use Lemma 1 with $\nu = \lambda$ (Lebesgue measure).

Let $t \in [t_0(\varepsilon), t_1(\varepsilon)]$. Then

$$\frac{f_0(t) - f_\varepsilon(t)}{f_0(t)} = \varepsilon,$$

and the proof follows. ∎

COROLLARY 1. *The contamination model* $\mathscr{P}_{F_0}(\varepsilon)$ (see (1)) *coincides with the set*

$$\mathscr{C}_{F_0}(\varepsilon) = \{G \in \mathscr{F}: \varepsilon_{G,F_0} \leqslant \varepsilon\}.$$

Proof. The previous lemma proves that $\mathscr{P}_{F_0}(\varepsilon) \subset \mathscr{C}_{F_0}(\varepsilon)$. The opposite inclusion follows from Definition 1 and Remark 3. ∎

Gross-error sensitivity. Corollary 1 brought the result which has been promised in the Introduction that Huber's result may be equivalently formulated using the notion of contamination level. Let us keep it in mind in what follows.

In 1968 Hampel [3] introduced the notion of gross-error sensitivity and studied the extremal problem of finding the ψ-function which generates the M-estimator with the minimal variance in the family of all M-estimators with the gross-error sensitivity bounded by a given limit. Let us briefly remind necessary notions and then the results.

DEFINITION 2. For a convex subset $\mathscr{F}_1$ of $\mathscr{F}$ let $T(F)$: $\mathscr{F}_1 \to F$ be a real functional and Δ_t the distribution function putting the mass 1 at the point $t \in \boldsymbol{R}$. The *influence function* $\mathrm{IF}(t, T, F)$ of the functional T at the distribution $F \in \mathscr{F}_1$ is given by

$$\lim_{\tau \to 0+} \frac{T((1-\tau)F + \tau\Delta_t) - T(F)}{\tau} \tag{10}$$

at those points t at which the limit exists. ■

EXAMPLE 2. Let $F_{\exp}(v)$ be the double exponential distribution and $T_{\mathrm{med}}(F)$ the median. Keep in mind that for any $F \in \mathscr{F}$ such that there is a point $m \in \boldsymbol{R}$ such that $F(m) = \frac{1}{2}$ we have $T_{\mathrm{med}}(F) = m$. Let $t > 0$. Then we have

$$(1-\tau)F_{\exp}(z) + \tau\Delta_t(z) > \tfrac{1}{2} \quad \text{for any } z \geqslant t.$$

On the other hand,

$$(1-\tau)F_{\exp}(z) < \tfrac{1}{2} \quad \text{for } z \leqslant 0.$$

So to establish $T_{\mathrm{med}}((1-\tau)F_{\exp} + \tau\Delta_t)$ we obviously need to find a point $m \in (0, t)$ such that $(1-\tau)F_{\exp}(m) = \frac{1}{2}$. Using Taylor's expansion one easily verifies that m is given as a solution of the equation

$$(1-\tau)\cdot\tfrac{1}{2} + \tfrac{1}{2}\cdot m + o(m) = \tfrac{1}{2},$$

and (10) then yields

$$\mathrm{IF}(t, T_{\mathrm{med}}, F_{\exp}) = \begin{cases} -1 & \text{for } t < 0, \\ 0 & \text{for } t = 0, \\ 1 & \text{for } t > 0. \end{cases} \quad ■$$

DEFINITION 3. The *gross-error sensitivity* of the functional T at the distribution $F \in \mathscr{F}_1$ is given by

$$\gamma^*(T, F) = \sup_t |\mathrm{IF}(t, T, F)|, \tag{11}$$

where the supremum is taken over all the points at which the influence function exists. ■

EXAMPLE 3. It follows from Example 2 that $\gamma^*(T_{\text{med}}, F_{\text{exp}}) = 1$. Similarly, we may find that $\gamma^*(T_{\text{med}}, \Phi) = \sqrt{\pi/2}$ in the standard normal law or $\gamma^*(T_{\text{med}}, F_{\text{L}}) = 2$ in the logistic distribution (see [4], Example 2.5.4). ∎

In what follows we shall write $\gamma^*(\psi, F)$ for the gross-error sensitivity of the M-estimator generated by the function ψ.

We shall consider the following two results given in [4].

ASSERTION 2 (Hampel [4]). *Let Θ be an open convex subset of $\mathbf{R}$ and $\{F_\theta\}_{\theta\in\Theta}$ a family of distribution functions with strictly positive densities $f_\theta(x)$ which are assumed to be absolutely continuous. Let us put*

$$s(t, F_\theta) = \left\{\frac{\partial}{\partial\theta} f_\theta(t)\right\} / f_\theta(t)$$

and fix some $\theta_0 \in \Theta$. We shall write briefly F, f and $s(t)$ instead of $F_{\theta_0}, f_{\theta_0}$ and $s(t, F_{\theta_0})$, respectively. Let $\int s(t)\,dF(t) = 0$ and the Fisher information $J_F = \int s^2(t)\,dF(t) \in (0, \infty)$. Finally, let $b > 0$ be a constant. Then there is a real number a such that

$$\psi_b(t) = \max\{-b, \min\{s(t)-a, b\}\} \tag{12}$$

satisfies

$$\int \psi_b(t)\,dF(t) = 0 \quad \text{and} \quad d = \int \psi_b(t)\,s(t)\,dF(t) > 0,$$

and ψ_b minimizes

$$\int \psi^2(t)\,dF(t)\left\{\int \psi(u)\,s(u)\,dF(u)\right\}^{-2} \tag{13}$$

among all mappings that satisfy

$$\int \psi(t)\,dF(t) = 0, \quad \int \psi(t)\,s(t)\,dF(t) \neq 0 \tag{14}$$

and

$$\sup_t \left|\psi(t)\left\{\int \psi(u)\,s(u)\,dF(u)\right\}^{-1}\right| \leqslant b/d. \tag{15}$$

Any other solution of this extremal problem coincides with a non-zero multiple of ψ_b almost everywhere with respect to F.

For the proof see [4], Theorem 2.4.1. ∎

Remark 4. Let us observe that ψ_b coincides with ψ_ε for $b = k(\varepsilon)$ and $F = F_0$. Having forgotten, for a while, that ψ_ε was found as a solution of the location problem, we know also from the Huber result that if Condition A is fulfilled for the distribution function F_0, then in the set of all distribution functions having the contamination level ε_{G,F_0} not greater than ε there is a distribution function F_ε (with $\varepsilon_{F_\varepsilon,F_0} = \varepsilon$) such that for $b = k(\varepsilon)$ we have $\psi_b = \psi_\varepsilon$, where $k(\varepsilon)$ and ψ_ε are given by (6) and (5), respectively. The ψ-function which is the solution of Huber's problem coincides with the ψ-function which is the solution of Hampel's problem. (Hampel has introduced for these estimators the name the *optimal B-robust estimators* (see [4], 2.4b).) Sometimes this coincidence is explained by the fact that *B*- and *V*-robustness, in this case, also coincide and that *V*-robustness takes into account the change of variance, and hence it is related to Huber's approach. However, the arguments are not completely correct, because although in both cases the constraint was the same (namely, as we have demonstrated, un upper bound on the contamination level), the models under which the results were obtained were different. In one case the central model, in the other the least favourable one. It may be also of interest that for $b \to 0$ we obtain as a limit case of ψ_b just ψ_{med}. Obviously, as we have seen in Example 2, we may obtain ψ_{med} also as the influence function of the maximum likelihood estimator of location for the double exponential distribution. As we shall see later on, if Condition A is fulfilled, $k(\varepsilon)$ is a decreasing continuous function,

$$k(\varepsilon)\colon [0, 1) \to \left(0, \sup_t |s(t, F)|\right).$$

This means that $b \to 0$ corresponds to $\varepsilon \to 1$, so that the function ψ_{med}, according to the former "definition", is obtained as the limit case for $\varepsilon \to 1$, i.e., for the situation when the portion of the contamination tends to 100%. Since sometimes, e.g., when we study the robust procedures with high breakdown point, we argue that the contamination higher than 50% is senseless, the interpretation of ψ_{med} as the influence function of the maximum likelihood estimator of location for the double exponential distribution seems to be more satisfactory. ∎

Remark 5. Let us put

$$J_F(\varepsilon) = \int \psi_\varepsilon(t)\, s(t)\, dF(t) \tag{16}$$

(the notation reflects the fact that for $\varepsilon = 0$ we obtain the Fisher information J_F). Therefore, as (13) represents the variance $V(\psi, F)$ (see (3)), Hampel's result says that ψ_b minimizes the variance among all the *M*-estimators having bounded the gross-error sensitivity by $k(\varepsilon)/J_F(\varepsilon)$. Observe again that the variances are computed with respect to the central model in $\mathscr{P}_F(\varepsilon)$ (see (1)). ∎

ASSERTION 3 (Rousseeuw [6]). *Let the assumptions of Assertion 2 be fulfilled. Moreover, let f be symmetric. Then the mapping $b \to \gamma^*(\psi_b, F)$ is a strictly increasing continuous bijection from $(0, \sup_t |s(t)|)$ onto $(\gamma^*(\psi_{\text{med}}, F), \gamma^*(s(t), f))$.*

For the proof see [6] or [4], Lemma 2.5.1.

Let us give now an assertion that allows us to connect the Rousseeuw result with the contamination level.

ASSERTION 4. *Let Condition A be fulfilled. Then*

$$\frac{dk(\varepsilon)}{d\varepsilon} = -\frac{k^2(\varepsilon)}{(1-\varepsilon)^2}\left\{f_0(t_0(\varepsilon)) + f_0(t_1(\varepsilon))\right\}^{-1}.$$

For the proof see [9], Lemma 1.

Remark 6. As follows from Assertions 3 and 4 under Condition A the mapping $\varepsilon \to \gamma^*(\psi_\varepsilon, F_0)$ is a strictly decreasing function of ε. Observe however that the behaviour of the gross-error sensitivity of ψ_ε has been studied in Assertion 3 again under the central model F_0. ■

Remark 7. Assertion 4 enables us to see that Hampel's extremal problem as well as Rousseeuw's result about monotonicity of the gross-error sensitivity can be reformulated by using the notion of contamination level. This reformulation may offer for some readers a better understanding and more natural description of the problems because in applications one may have some feelings about the contamination level of data but probably (only) a vague idea how much the gross-error sensitivity should be limited for such a contamination level. ■

Intensive studies of the procedures with the high breakdown point clearly prove that there are situations when the contamination level of data is not near zero. And these situations are presumably much more frequent than it is commonly recognized and accepted (for a nice discussion see [4], 1.2c and 1.2d, and also [2] and [1]). Then the results derived by the infinitesimal approach may be (however not necessarily) of limited use. This directly inspires the following questions: *What is the behaviour of the robust procedures and their characteristics, e.g., the behaviour of the gross-error sensitivity of ψ_ε under the least favourable distribution F_ε? Is it still monotone as under the central model F_0?*

As we shall see later the answer is negative. But then we ask probably immediately: *For what values of the contamination level is the gross-error sensitivity $\gamma^*(\psi_\varepsilon, F_\varepsilon)$ already increasing?* The answer to both the questions will be given in Lemma 4. We need however at first to prove an assertion and a lemma.

ASSERTION 5. *Let Condition A be fulfilled. Define for any fix $z < 0$ and $t > 0$ a function $r_z(t)$ as follows:*

$$r_z(t) = \int_z^t f_0(y)\,dy + \frac{f_0^2(z)}{f_0'(z)} - \frac{f_0^2(t)}{f_0'(t)}.$$

Then $r_z(t)$ is continuously differentiable and strictly decreasing.

Proof. A straightforward computation gives

$$\frac{dr_z(t)}{dt} = \frac{f_0^2(t)\,f_0''(t) - f_0(t)\,[f_0'(t)]^2}{[f_0'(t)]^2}. \tag{17}$$

Now, the requirement of the strict convexity of $-\log f_0(t)$ implies

$$f_0(t)\,f_0''(t) - [f_0'(t)]^2 < 0$$

for t in the support of f_0, and the proof follows. ■

In what follows let us write $r'_z(t)$ for $dr_z(t)/dt$.

LEMMA 3. *Under Condition A for any γ_1, γ_2 such that $0 < \gamma_1 < \gamma_2 < 1$ and $-\infty < t_0(\gamma_1) < t_1(\gamma_1) < \infty$ there are positive and finite constants $M_1(\gamma_1, \gamma_2)$ and $M_2(\gamma_1, \gamma_2)$ such that for any pair $\varepsilon_1, \varepsilon_2, \gamma_1 < \varepsilon_1 < \varepsilon_2 < \gamma_2$, we have*

$$(18)\qquad 0 < M_1(\gamma_1, \gamma_2)\{t_0(\varepsilon_2)-t_0(\varepsilon_1)+t_1(\varepsilon_1)-t_1(\varepsilon_2)\} \leqslant \varepsilon_2-\varepsilon_1 \leqslant M_2(\gamma_1, \gamma_2)\{t_0(\varepsilon_2)-t_0(\varepsilon_1)+t_1(\varepsilon_1)-t_1(\varepsilon_2)\}.$$

Proof. From Assertion 4 it follows that $t_0(\varepsilon_1) < t_0(\varepsilon_2)$ and $t_1(\varepsilon_2) < t_1(\varepsilon_1)$. Taking into account also (6), we conclude that

$$(19)\qquad -\infty < t_0(\gamma_1) \leqslant t_0(\varepsilon_1) \leqslant t_0(\varepsilon_2) \leqslant t_0(\gamma_2) < t_1(\gamma_2) \leqslant t_1(\varepsilon_2) \leqslant t_1(\varepsilon_1) \leqslant t_1(\gamma_1) < \infty.$$

Making use of (6) once again we obtain

$$(20)\qquad 0 < (1-\varepsilon_2)^{-1}-(1-\varepsilon_1)^{-1} = \int_{t_0(\varepsilon_2)}^{t_1(\varepsilon_2)} f_0(t)\,dt - \frac{f_0(t_0(\varepsilon_2))+f_0(t_1(\varepsilon_2))}{k(\varepsilon_2)} - \int_{t_0(\varepsilon_1)}^{t_1(\varepsilon_1)} f_0(t)\,dt - \frac{f_0(t_0(\varepsilon_1))+f_0(t_1(\varepsilon_1))}{k(\varepsilon_1)}.$$

Observing that $r'_z(t)$ does not depend on z (see (17)) and using (7) and (19) we may rewrite (20) in the form

$$(21)\qquad \frac{\varepsilon_2-\varepsilon_1}{(1-\varepsilon_1)(1-\varepsilon_2)} = -\left\{\int_{t_0(\varepsilon_1)}^{t_0(\varepsilon_2)} r'_z(t)\,dt + \int_{t_1(\varepsilon_2)}^{t_1(\varepsilon_1)} r'_z(t)\,dt\right\}.$$

Putting $\Gamma = [t_0(\gamma_1), t_0(\gamma_2)] \cup [t_1(\gamma_2), t_1(\gamma_1)]$, let us define

$$M_1^*(\gamma_1, \gamma_2) = -\sup_{t\in\Gamma} r'_z(t) \quad \text{and} \quad M_2^*(\gamma_1, \gamma_2) = -\inf_{t\in\Gamma} r'_z(t).$$

Since $r'_z(t)$ is continuous and Γ is compact, $M_1^*(\gamma_1, \gamma_2)$ as well as $M_2^*(\gamma_1, \gamma_2)$ is positive and finite, and we obtain

$$0 < M_1^*(\gamma_1, \gamma_2)\{t_0(\varepsilon_2)-t_0(\varepsilon_1)+t_1(\varepsilon_1)-t_1(\varepsilon_2)\} \leqslant \frac{\varepsilon_2-\varepsilon_1}{(1-\varepsilon_1)(1-\varepsilon_2)} \leqslant M_2^*(\gamma_1, \gamma_2)\{t_0(\varepsilon_2)-t_0(\varepsilon_1)+t_1(\varepsilon_1)-t_1(\varepsilon_1)\}.$$

Since $0 < (1-\gamma_2)^2 \leqslant (1-\varepsilon_1)(1-\varepsilon_2) \leqslant (1-\gamma_1)^2$, we have verified (18). ∎

Remark 8. It is clear that to cope with the situation when $t_0(\gamma_1) = -\infty$ or $t_1(\gamma_1) = \infty$ requires a different formulation of the assertion of Lemma 3 since (18) may be evidently senseless. Let us assume that $t_1(\gamma_1) = \infty$. If $t_1(\varepsilon_1) < \infty$, the situation is not substantially different from that one considered in Lemma 3, since we have to select only some $\gamma' \in (\gamma_1, \varepsilon_1]$ such that $t_1(\gamma') < \infty$. So we should assume that also $t_1(\varepsilon_1) = \infty$. Since in what follows (see the proof of Lemma 4 below) we will need to apply Lemma 3 in the case where $|\varepsilon_1 - \varepsilon_2| \to 0$, it is also senseless to assume that $t_1(\varepsilon_2) < \infty$. For the case where $t_1(\gamma_1) = t_1(\varepsilon_1) = t_1(\varepsilon_2) = \infty$ we easily find that in the fraction on the right-hand side of (6) the term $f_0(t_1(\varepsilon))$ will be missing. On the other hand, a simple analysis of the proof of Assertion 5 shows that $|r'_z(t)|$ is bounded. (An opposite possibility implies that for $\varepsilon \to 0$ (and, consequently, $t_1(\varepsilon) \to \infty$) the right-hand side of (6) increases above any bound while the left-hand side converges to 1.) So we can prove the assertion of Lemma 3 in a modified form (without $t_1(\varepsilon_1)$ and $t_1(\varepsilon_2)$; observe that the second term of the right-hand side of (21) disappears in this case). In order not to obscure further considerations we shall restrict ourselves to the cases where $\max\{|t_0(\gamma_1)|, |t_1(\gamma_1)|\} < \infty$.

LEMMA 4. *Let Condition* A *be fulfilled. Moreover, let the Fisher information* $J(F_0)$ *be in* $(0, \infty)$. *Then the gross-error sensitivity is*

$$\gamma^*(\psi_\varepsilon, F_\varepsilon) = (1-\varepsilon)^{-1} \cdot \gamma^*(\psi_\varepsilon, F_0)$$

and it is strictly decreasing on $(0, \varepsilon_0]$ *and strictly increasing on* $[\varepsilon_0, 1)$, *where* ε_0 *is given by the equality*

$$\left[\frac{dk(\varepsilon)}{d\varepsilon}\right]_{\varepsilon=\varepsilon_0} k^{-1}(\varepsilon_0) - \left[\frac{dJ_{F_0}(\varepsilon)}{d\varepsilon}\right]_{\varepsilon=\varepsilon_0} J_{F_0}^{-1}(\varepsilon_0) + \frac{1}{1-\varepsilon_0} = 0.$$

Proof. If Condition A holds, $-f_0'(t)/f_0(t)$ is strictly increasing, and hence we may define an inverse function $\xi\colon \boldsymbol{R} \to \boldsymbol{R}$ so that for any $\varepsilon \in (0, 1)$

$$t_i(\varepsilon) = \xi((-1)^i k(\varepsilon))$$

(see (7)). This directly implies that

$$\frac{dt_i(\varepsilon)}{d\varepsilon} = (-1)^i \frac{f_0^2(t_i(\varepsilon))}{f_0''(t_i(\varepsilon)) f_0(t_i(\varepsilon)) - [f_0'(t_i(\varepsilon))]^2} \cdot \frac{dk(\varepsilon)}{d\varepsilon}. \tag{22}$$

Now, using a well-known formula for the influence function of the location M-estimator (for general ψ and general F)

$$\mathrm{IF}(t, \psi, F) = -\psi(t) \left\{ \int \psi(v) \frac{f'(v)}{f(v)} dF(v) \right\}^{-1}$$

(see, e.g., [4], 2.3.8), substituting for ψ_ε and F_ε the respective expressions from (5) and (8), and making use of (11) we obtain

$$\gamma^*(\psi_\varepsilon, F_\varepsilon) = \frac{k(\varepsilon)}{(1-\varepsilon) J_F(\varepsilon)} = \frac{\gamma^*(\psi_\varepsilon, F_0)}{1-\varepsilon}. \tag{23}$$

Notice that from (23) and Assertion 3 it follows that

$$\limsup_{\varepsilon \to 1} \gamma^*(\psi_\varepsilon, F_\varepsilon) = \infty,$$

since $\gamma^*(\psi_{\text{med}}, F) > 0$ (however, it is not possible to learn from Assertion 3 and (23) anything about the monotonicity of $\gamma^*(\psi_\varepsilon, F_\varepsilon)$). Integrating by parts in (16) we obtain

$$J_F(\varepsilon) = \int \psi'_\varepsilon(t) f(t)\, dt$$

(which leads to another well-known expression for

$$\mathrm{IF}(t, \psi, F) = \psi(t) / \int \psi'(t) f(t)\, dt;$$

see [4], 2.3.12) and using (5) and (8) we arrive at

$$J_{F_0}(\varepsilon) = \int_{t_0(\varepsilon)}^{t_1(\varepsilon)} \frac{[f_0'(t)]^2 - f_0''(t) f_0(t)}{f_0(t)}\, dt.$$

Now, making use of the continuity of the integrand and using Lemma 3, we find that (for $i = 0, 1$)

$$\lim_{\varepsilon^* \to \varepsilon} (\varepsilon^* - \varepsilon)^{-1} \int_{t_i(\varepsilon^*)}^{t_i(\varepsilon)} \left\{ \frac{[f_0'(t)]^2 - f_0''(t) f_0(t)}{f_0(t)} - \frac{[f_0'(t_i(\varepsilon))]^2 - f_0''(t_i(\varepsilon)) f_0(t_i(\varepsilon))}{f_0(t_i(\varepsilon))} \right\} dt = 0,$$

and hence

$$\frac{dJ_{F_0}(\varepsilon)}{d\varepsilon} = \frac{[f_0'(t_1(\varepsilon))]^2 - f_0''(t_1(\varepsilon)) f_0(t_1(\varepsilon))}{f_0(t_1(\varepsilon))} \frac{dt_1(\varepsilon)}{d\varepsilon} - \frac{[f_0'(t_0(\varepsilon))]^2 - f_0''(t_0(\varepsilon)) f_0(t_0(\varepsilon))}{f_0(t_0(\varepsilon))} \frac{dt_0(\varepsilon)}{d\varepsilon}$$

$$= \{f_0(t_0(\varepsilon)) + f_0(t_1(\varepsilon))\} \frac{dk(\varepsilon)}{d\varepsilon}$$

(see (22)). Let us now consider

$$\log \gamma^*(\psi_\varepsilon, F_\varepsilon) = \log k(\varepsilon) - \log J_{F_0}(\varepsilon) - \log(1 - \varepsilon).$$

Its derivative is equal to

$$\tau(\varepsilon) = \frac{dk(\varepsilon)}{d\varepsilon} k^{-1}(\varepsilon) - \frac{dJ_{F_0}(\varepsilon)}{d\varepsilon} J_{F_0}^{-1}(\varepsilon) + \frac{1}{1 - \varepsilon}.$$

Using the fact that the integrability of the function $f_0''(t) f_0(t) / f_0'(t)$ implies

$$\lim_{|t| \to \infty} f_0''(t) f_0(t) / f_0'(t) = 0,$$

we find that $\tau(\varepsilon)$ is strictly monotone on $(0, 1)$ and

$$\lim_{\varepsilon \to 0_+} \tau(\varepsilon) = -\infty \quad \text{and} \quad \lim_{\varepsilon \to 1_-} \tau(\varepsilon) = +\infty.$$

This means that $\log \gamma^*(\psi_\varepsilon, F_\varepsilon)$ is strictly convex, and the proof follows. ∎

The next exhibit gathers some results which may help to create an idea of the gross-error sensitivity of the Huber estimator under various contamination levels. The contamination levels are given in the first row, in the second one the gross-error sensitivities of ψ_ε (see (5)) with respect to the central model $F_0 = \Phi$ (normal distribution) are presented, and in the third one the values of the gross-error sensitivity of ψ_ε with respect to the least favourable distribution F_ε (see (8) and Lemma 4) are given. To explain values in the fourth row let us remind that the "tunning" constant equal to 1.5, which is sometimes recommended for the Huber estimator (see, e.g., [4]), corresponds to the contamination level 3.8%, i.e., $\varepsilon_1 = 0.038$. For the function ψ_{ε_1} the suprema

$$\gamma^*_{\text{sup}}(\psi_{\varepsilon_1}) = \sup_{G \in \mathscr{P}^*_\varepsilon} \gamma^*(\psi_{\varepsilon_1}, G), \qquad \text{where } \mathscr{P}^*_\varepsilon = \mathscr{P}_\Phi(\varepsilon),$$

have been found for ε's as given in the first row of the exhibit and put in the last row.

Exhibit 1

ε	0.050	0.075	0.100	0.125	0.150	0.175
$\gamma^*(\psi_\varepsilon, \Phi)$	1.67	1.58	1.53	1.49	1.46	1.43
$\gamma^*(\psi_\varepsilon, F_\varepsilon)$	1.76	1.71	1.70	1.70	1.71	1.73
$\gamma^*_{\text{sup}}(\psi_{\varepsilon_1})$	1.82	1.87	1.92	1.98	2.03	2.10

ε	0.200	0.225	0.250	0.275	0.300	0.325
$\gamma^*(\psi_\varepsilon, \Phi)$	1.41	1.39	1.38	1.36	1.35	1.34
$\gamma^*(\psi_\varepsilon, F_\varepsilon)$	1.76	1.80	1.84	1.88	1.93	1.99
$\gamma^*_{\text{sup}}(\psi_{\varepsilon_1})$	2.16	2.23	2.31	2.38	2.47	2.56

ε	0.350	0.375	0.400	0.425	0.450	0.475	0.500
$\gamma^*(\psi_\varepsilon, \Phi)$	1.33	1.32	1.32	1.31	1.30	1.30	1.29
$\gamma^*(\psi_\varepsilon, F_\varepsilon)$	2.05	2.12	2.19	2.28	2.37	2.47	2.59
$\gamma^*_{\text{sup}}(\psi_{\varepsilon_1})$	2.66	2.77	2.88	3.01	3.14	3.29	3.46

Remark 9. It is clear from Exhibit 1 that even in the case where we (considerably) underestimate the contamination level the increase of the gross-error sensitivity of *B*-robust estimators will not be dramatic. The values in the third row of Exhibit 1 also "confirm" the assertion of the previous lemma. The precise value of ε_0 (see Lemma 4) for this setup is 0.1084. On the other hand, the values in the last row show that the gross-error sensitivity of the usually used Huber's estimator, i.e., the estimator generated by ψ_{ε_1} (with "tunning" constant equal to 1.5), is increasing with increasing level of contamination. So its behaviour is quite different from that one which we can adversely expect on the base of Assertion 3.

Conclusions. We have shown that a natural feeling that some number of atypical observations may represent a contamination of data may be mathematically reflected as the contamination level. This notion may then serve as a basis for the formulation of the well-known problems in the robust statistics.

We have also shown that when the contamination level overcomes 10.84% (to which the "tunning" constant $k(\varepsilon) = 1.1$ for the Huber function corresponds) the gross-error sensitivity $\gamma^*(\psi_\varepsilon, F_\varepsilon)$ becomes increasing. This means that a further decrease of the "tunning" constant will imply the increase of the gross-error sensitivity.

On the other hand, the changes of the gross-error sensitivity (as demonstrated in Exhibit 1) as well as of the asymptotic efficiency (see Exhibit 2 below) are so small that the proper selection of the "tunning" constant for the given data is not evidently the problem of the (dramatic) loss of the efficiency or of a (serious) increase of the gross-error sensitivity, but much more the problem of estimation of such a model which is acceptable for practical purposes. One easily finds that with changing "tunning" constant we obtain for the given data various estimates which may be considerably different from each other (see [10] or [11]). So we have to solve the problem of selection of one estimate from the whole set of estimates which were calculated for the same data, for the same model (i.e., for the same set of regressors) and all of them should be "near to the true model" because of the asymptotic consistency. One of the possibilities is – instead of choosing one of these models – to combine them into a new estimate. The possibility was studied in [8]. Another possibility how to cope with the problem may be to compare the (kernel) estimates of density of the residuals for (two) distinct subsamples of data, and select the one for which the density estimates are similar each to other "as much as possible" (e.g., in the sense of the Hellinger distance, see [10]).

In the next exhibit we use the following notation:

$\text{var}(\psi_\varepsilon, \Phi)$ – the asymptotic variance, with respect to the central model $F_0 = \Phi$ (normal distribution), of the location estimator generated by the function ψ_ε;

$\text{var}(\psi_\varepsilon, F_\varepsilon)$ – the asymptotic variance, with respect to the least favourable model F_ε, of the location estimator generated by the function ψ_ε;

$\text{var}_{\text{sup}}(\psi_{\varepsilon_1})$ – the supremum of the asymptotic variances $\text{var}(\psi_{\varepsilon_1}, G)$ over $\{G \in \mathscr{P}_\Phi(\varepsilon)\}$;

efficiency $(\psi_\varepsilon, \psi_{\varepsilon_1})$ – the efficiency of the location estimator generated by the function ψ_{ε_1} with respect to the optimal location estimator, i.e.

$$\frac{\text{var}(\psi_\varepsilon, F_\varepsilon)}{\sup_{G \in \mathscr{P}_\Phi(\varepsilon)} \text{var}(\psi_{\varepsilon_1}, G)}.$$

Exhibit 2

ε	0.050	0.075	0.100	0.125	0.150	0.175
$\text{var}(\psi_\varepsilon, \Phi)$	1.05	1.06	1.08	1.10	1.11	1.13
$\text{var}(\psi_\varepsilon, F_\varepsilon)$	1.26	1.37	1.49	1.61	1.74	1.89
$\text{var}_{\text{sup}}(\psi_{\varepsilon_1})$	1.26	1.38	1.52	1.67	1.84	2.03
efficiency $(\psi_\varepsilon, \psi_{\varepsilon_1})$	1.00	0.99	0.98	0.96	0.95	0.93

ε	0.200	0.225	0.250	0.275	0.300	0.325
$\operatorname{var}(\psi_\varepsilon, \Phi)$	1.14	1.15	1.17	1.18	1.19	1.20
$\operatorname{var}(\psi_\varepsilon, F_\varepsilon)$	2.05	2.21	2.40	2.60	2.82	3.07
$\operatorname{var}_{\sup}(\psi_{\varepsilon_1})$	2.23	2.46	2.71	3.00	3.31	3.67
efficiency $(\psi_\varepsilon, \psi_{\varepsilon_1})$	0.92	0.90	0.88	0.87	0.85	0.84

ε	0.350	0.375	0.400	0.425	0.450	0.475	0.500
$\operatorname{var}(\psi_\varepsilon, \Phi)$	1.22	1.23	1.24	1.25	1.27	1.28	1.29
$\operatorname{var}(\psi_\varepsilon, F_\varepsilon)$	3.34	3.65	4.00	4.39	4.83	5.34	5.93
$\operatorname{var}_{\sup}(\psi_{\varepsilon_1})$	4.07	4.53	5.05	5.65	6.33	7.13	8.05
efficiency $(\psi_\varepsilon, \psi_{\varepsilon_1})$	0.82	0.81	0.79	0.78	0.76	0.75	0.74

REFERENCES

[1] C. Daniel, *Applications of Statistics to Industrial Experimentation*, Willey, New York 1976.

[2] — and F. S. Wood, *Fitting Equations to Data*, Willey, New York 1980.

[3] F. R. Hampel, *Contributions to the theory of robust estimation*, Ph. D. Thesis, University of California, Berkeley 1968.

[4] — E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics — The Approach Based on Influence Functions*, Willey, New York 1986.

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

[6] P. J. Rousseeuw, *A new infinitesimal methods in robust statistics*, Ph. D. Thesis, Vrije Universiteit, Brussels 1981.

[7] A. M. Rubio, F. Quintana and J. Á. Víšek, *Combining estimators*, submitted to Statist. Decisions (1993).

[8] — *Combining the M-estimators*, submitted to Statist. Decisions.

[9] A. M. Rubio and J. Á. Víšek, *Efficiency rate and local deficiency of Huber's location estimators and of the α-estimators*, Trabajos Estadist. 6 (1991), pp. 67–85.

[10] — *Diagnostics of regression model*: *Test of goodness-of-fit*, in: M. Hushová and P. Mandl (Eds.), *Transactions of the Fifth Prague Symposium on Asymptotic Statistics*, Springer Verlag, 1994, pp. 423–432.

[11] J. Á. Víšek, *Stability of regression models estimates with respect to subsamples*, Comput. Statist. 7 (1992), pp. 183–203.

Institute of Information Theory and Automation
Academy of Sciences of Czech Republic
Pod vodárenskou věží 4
182 08 Praha 8, Czech Republic

Received on 14.12.1993