ELSEVIER

Statistics & Probability Letters 57 (2002) 91–100

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Minimax variance estimation of a correlation coefficient for ε-contaminated bivariate normal distributions

Georgy L. Shevlyakov[a,*], Nikita O. Vilchevski[b]

[a]Department of Mathematics, St. Petersburg State Technical University, Nalichnaya st. 36-5-399, 199226 St. Petersburg, Russia

[b]Department of Applied Mathematics, St.Petersburg State Technical University, St.Petersburg, Russia

Received May 2001; received in revised form November 2001

Abstract

A minimax variance (in the Huber sense) estimator of a correlation coefficient for ε-contaminated bivariate normal distributions is given by the *trimmed correlation coefficient*. Consistency and asymptotic normality of this estimator are established, and the explicit expression for its asymptotic variance is obtained. The limiting cases of this estimator are the sample correlation coefficient with $\varepsilon = 0$ and the *median correlation coefficient* as $\varepsilon \to 1$. In ε-contaminated normal models, the proposed trimmed correlation coefficient is superior in efficiency than the sample correlation coefficient.

MSC: 62G35

Keywords: Minimax robustness; Correlation coefficient

1. Introduction

The aim of robust methods is to ensure high stability of statistical inferences under the deviations from the assumed distribution model. Much less attention is devoted in the literature to robust estimators of association and correlation as compared to robust estimators of location and scale (see Huber, 1981; Hampel et al., 1986). However, it is necessary to study these problems due to their widespread occurrence (estimation of the correlation and covariance matrices in regression and multivariate analysis, estimation of the correlation functions of stochastic processes, etc.), and also due to the great instability of classical methods of estimation in the presence of outliers in the data.

* Corresponding author.

E-mail address: shev@stat.hop.stu.neva.ru (G.L. Shevlyakov).

PII: S0167-7152(02)00058-5

The simplest problem of correlation analysis is to estimate the correlation coefficient ρ of a bivariate distribution density $f_{XY}(x, y)$ with the observed sample $(x_1, y_1), \ldots, (x_n, y_n)$ of a bivariate random variable (r.v.) (X, Y). Its classical estimator is given by the sample correlation coefficient

$$r = \sum (x_i - \bar{x})(y_i - \bar{y}) \Big/ \left(\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right)^{1/2}, \tag{1}$$

where $\bar{x} = n^{-1} \sum x_i$, and $\bar{y} = n^{-1} \sum y_i$ are the sample means.

The sample correlation coefficient is, on the one hand, a statistical counterpart of the correlation coefficient

$$\rho = \mathrm{Cov}(X, Y)/(\mathrm{Var}\, X \,\mathrm{Var}\, Y)^{1/2}, \tag{2}$$

where $\mathrm{Var}\, X$, $\mathrm{Var}\, Y$ and $\mathrm{Cov}(X, Y)$ are the variances and the covariance of the r.v. X and Y. On the other hand, it is an efficient maximum likelihood estimator of ρ for a bivariate normal distribution density $\mathcal{N}(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, where the parameters μ_1 and μ_2 are the means, σ_1 and σ_2 are the standard deviations of the r.v. X and Y, respectively.

In the gross error model (the traditional in robustness studies contamination scheme (Huber, 1964, 1981; Hampel et al., 1986))

$$f(x, y) = (1 - \varepsilon)\mathcal{N}(x, y; 0, 0, 1, 1, \rho) + \varepsilon\mathcal{N}(x, y; 0, 0, k, k, \rho'), \quad 0 \leqslant \varepsilon < 1, \tag{3}$$

the sample correlation coefficient is strongly biased with regard to ρ (Gnanadesikan et al., 1972; Devlin et al., 1975). For instance, estimating the correlation coefficient $\rho = 0.9$ of the main bulk of the data under the contamination with $\varepsilon = 0.1$, $k = 3$ and $\rho' = -0.99$, asymptotically (as $n \to \infty$) we have $\mathrm{E}r = -0.055$, what means that even the sign of the sample correlation coefficient is wrong (Shevlyakov, 1997). This shows that the sample correlation coefficient r is very sensitive to the presence of outliers in the data, and hence it is necessary to use its robust counterparts.

At present there exist two principal methods of design of robust estimators, i.e., the minimax method of quantitative robustness (Huber, 1981), and the method of qualitative robustness based on influence functions (Hampel et al., 1986). According to the first of these methods, we determine the least informative (favorable) distribution density minimizing Fisher information over a given class of distributions, with the subsequent construction of optimal maximum likelihood estimators for this density. This ensures that the asymptotic variance of the estimator will not exceed a certain threshold level (namely, the supremum of the asymptotic variance as a measure of quantitative robustness) which strongly depends on the characteristics of a chosen class of distributions. According to the second method, we construct an estimator with an assigned influence function whose type of behavior determines the qualitative robustness properties of the estimation procedure (such as its sensitivity to large outliers in the data, their rounding off, etc.).

Most of robust estimators of a correlation coefficient have been obtained from heuristic considerations related to the desired behavior of their influence functions (Devlin et al., 1975; Huber, 1981; Shevlyakov, 1997).

In the literature, there is only one result on applying the minimax approach to robust estimation of ρ (Huber, 1981, p. 205): the quadrant correlation coefficient is asymptotically minimax with respect to bias at the mixture $F = (1 - \varepsilon)G + \varepsilon H$ (G and H being centrosymmetric distributions in $\mathbf{R}^2$).

In this paper we extend Huber's results on robust M-estimators of location and scale in ε-contamination models (Huber, 1964, 1981) to the problem of robust estimation of a correlation coefficient for ε-contaminated bivariate normal distributions.

The class of robust estimators for ρ based on robust estimators for the variances of principal variables is described in Section 2. The corresponding class of bivariate distributions is introduced in Section 3. The minimax variance estimator for ρ in the class of ε-contaminated bivariate normal distributions is proposed in Section 4. Final remarks are made in Section 5.

2. Robust estimation of correlation via robust estimation of scale

2.1. Main groups of estimators

Most of robust estimators of a correlation coefficient are based on: (i) direct robust counterparts of the sample correlation coefficient; (ii) nonparametric measures of correlation; (iii) robust regression; (iv) robust estimation of the variances of principal variables; (v) the preliminary rejection of outliers from the data and the subsequent application of the sample correlation coefficient to the rest of the observations.

The behavior of the typical representatives of these groups in gross error model (3) was thoroughly examined in asymptotics and on finite samples ($n = 20, 30, 60$), and the estimators based on robust variances proved to be remarkably robust (Gnanadesikan and Kettenring, 1972; Devlin et al., 1975; Pasman and Shevlyakov, 1987; Shevlyakov, 1997) and therefore advantageous for the further theoretical examination.

2.2. The estimators based on robust variances

Consider the following identity for ρ

$$\rho = (\mathrm{Var}\, U - \mathrm{Var}\, V)/(\mathrm{Var}\, U + \mathrm{Var}\, V), \tag{4}$$

where

$$U = (X/\sigma_1 + Y/\sigma_2)/\sqrt{2}, \quad V = (X/\sigma_1 - Y/\sigma_2)/\sqrt{2}$$

are the principal variables such that

$$\mathrm{Cov}(U, V) = 0, \quad \mathrm{Var}\, U = 1 + \rho, \quad \mathrm{Var}\, V = 1 - \rho.$$

By introducing a robust scale functional $S(X)$: $S(aX + b) = |a|S(X)$, we can write $S^2(\cdot)$ for a robust counterpart of variance, and a robust counterpart for (4) in the form

$$\rho^*(X, Y) = [S^2(U) - S^2(V)]/[S^2(U) + S^2(V)]. \tag{5}$$

By substituting the sample robust estimates for S into (5), we obtain robust estimates for ρ

$$\hat{\rho} = [\hat{S}^2(U) - \hat{S}^2(V)]/[\hat{S}^2(U) + \hat{S}^2(V)]. \tag{6}$$

2.3. M-estimators of scale

Here we use Huber's M-estimator of scale for $S(X)$ which is defined by the following implicit relation:

$$\int \chi(x/S(X))\, \mathrm{d}F(x) = 0, \tag{7}$$

where χ is a score function, typically even $\chi(-x) = \chi(x)$ (Huber, 1981).

The following cases are of our particular interest:

- the choice

$$\chi(x) = -xf'(x)/f(x) - 1 \tag{8}$$

  yields the maximum likelihood estimator of σ for the scale family of densities $\sigma^{-1}f(x/\sigma)$;
- the standard deviation is an M-estimator with $\chi(x) = x^2 - 1$;
- the mean absolute deviation occurs with $\chi(x) = |x| - 1$;
- Huber (1981) proposes the choice

$$\chi(x) = \begin{cases} x_0^2 - 1 & \text{for } |x| < x_0, \\ x^2 - 1 & \text{for } x_0 \leqslant |x| \leqslant x_1, \\ x_1^2 - 1 & \text{for } |x| > x_1 \end{cases} \tag{9}$$

  for the ε-contaminated normal distribution, with some constants $x_0 = x_0(\varepsilon)$ and $x_1 = x_1(\varepsilon)$–this estimator is asymptotically equivalent to the trimmed standard deviation;
- the median absolute deviation $S = \operatorname{Med}|X|$ is a limiting case of the latter estimator as $\varepsilon \to 1$, with $\chi(x) = \operatorname{sign}(|x| - 1)$.

2.4. The median and trimmed correlation coefficients

The choice of the median absolute deviation $\hat{S} = \operatorname{MAD} x = \operatorname{med}|x_i - \operatorname{med} x|$ yields a remarkable robust estimator called the *median correlation coefficient* (Pasman and Shevlyakov, 1987)

$$r_{\text{med}} = (\operatorname{med}^2|u| - \operatorname{med}^2|v|)/(\operatorname{med}^2|u| + \operatorname{med}^2|v|), \tag{10}$$

where u and v are the robust principal variables

$$u = \frac{x - \operatorname{med} x}{\sqrt{2}\operatorname{MAD} x} + \frac{y - \operatorname{med} y}{\sqrt{2}\operatorname{MAD} y}, \quad v = \frac{x - \operatorname{med} x}{\sqrt{2}\operatorname{MAD} x} - \frac{y - \operatorname{med} y}{\sqrt{2}\operatorname{MAD} y}. \tag{11}$$

Choosing Huber's proposal (9) for $\hat{S}$, we obtain the structure of a *trimmed correlation coefficient* for $\hat{\rho}$:

$$r_{\text{tr}} = \left(\sum_{i=n_1+1}^{n-n_2} u_{(i)}^2 - \sum_{i=n_1+1}^{n-n_2} v_{(i)}^2\right) \Bigg/ \left(\sum_{i=n_1+1}^{n-n_2} u_{(i)}^2 + \sum_{i=n_1+1}^{n-n_2} v_{(i)}^2\right), \tag{12}$$

where $u_{(i)}$ and $v_{(i)}$ are the ith order statistics of the corresponding robust principal variables.

Note that the general construction (12) yields the following limiting cases: (i) the sample correlation coefficient r with $n_1 = 0$, $n_2 = 0$ and with the classical estimators (the sample means for location and the standard deviations for scale) in its inner structure; (ii) the median correlation coefficient r_{med} with $n_1 = n_2 = [0.5(n-1)]$.

3. The class of bivariate distributions

Consider the class of bivariate distribution densities corresponding to the above-introduced class of estimators (6) (the parameters of location and scale of the r.v. X and Y are assumed known: $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$)

$$f(x, y) = \frac{1}{\beta_u(\rho)} g\left(\frac{u}{\beta_u(\rho)}\right) \frac{1}{\beta_v(\rho)} g\left(\frac{v}{\beta_v(\rho)}\right), \tag{13}$$

where u and v are the principal variables

$$u = (x + y)/\sqrt{2}, \quad v = (x - y)/\sqrt{2};$$

$g(x)$ is a symmetric density $g(-x) = g(x)$ belonging to a certain class $\mathscr{G}$.

If the variance of the density g exists ($\sigma_g^2 = \int x^2 g(x)\,\mathrm{d}x < \infty$) then the straightforward calculation yields

$$\operatorname{Var} X = \operatorname{Var} Y = (\beta_u^2 + \beta_v^2)\sigma_g^2/2, \quad \operatorname{Cov}(X, Y) = (\beta_u^2 - \beta_v^2)\sigma_g^2/2$$

and the correlation coefficient of distributions (13) depends on the scale parameters β_u and β_v as follows:

$$\rho = (\beta_u^2 - \beta_v^2)/(\beta_u^2 + \beta_v^2). \tag{14}$$

Now assume that the variances of the r.v. X and Y do not depend on the unknown ρ

$$\operatorname{Var} X = \operatorname{Var} Y = \sigma^2 = \text{const.}(\rho).$$

Setting for convenience $\sigma_g = 1$, we obtain for β_u and β_v that

$$\beta_u = \sigma\sqrt{1 + \rho}, \quad \beta_v = \sigma\sqrt{1 - \rho}$$

and hence for densities (13)

$$f(x, y) = \frac{1}{\sigma\sqrt{1 + \rho}} g\left(\frac{u}{\sigma\sqrt{1 + \rho}}\right) \frac{1}{\sigma\sqrt{1 - \rho}} g\left(\frac{v}{\sigma\sqrt{1 - \rho}}\right). \tag{15}$$

It is important that class (13) and its subclass (15) contain the standard bivariate normal distribution density $f(x, y) = \mathscr{N}(x, y|0, 0, 1, 1, \rho)$ if $\beta_u(\rho) = \sqrt{1 + \rho}$, $\beta_v(\rho) = \sqrt{1 - \rho}$ and $g(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$.

Using other forms of univariate distribution densities, say the Laplace or even the long-tailed Cauchy (with the apparent modification of the definition for ρ), we can construct the bivariate analogs for the corresponding univariate distributions.

Henceforth we shall use subclass (15).

Now we recall the principal idea of introducing class (13): for any random pair (X, Y) the transformation $U = X + Y$, $V = X - Y$ gives the uncorrelated random principal variables (U, V) (actually independent for densities (13)), and estimation of their scale solves the problem of estimation of correlation between X and Y.

4. Main results

4.1. The estimation procedure

Given $(x_1, y_1), \ldots, (x_n, y_n)$, we propose the following estimation procedure for the correlation coefficient:

- transform the initial data

$$u_i = (x_i + y_i)/\sqrt{2}, \quad v_i = (x_i - y_i)/\sqrt{2}, \;\; i = 1, \ldots, n;$$

- evaluate the M-estimates of scale $\hat{\beta}_u$ and $\hat{\beta}_v$ as the solutions to the equations

$$\sum \chi(u_i/\hat{\beta}_u) = 0, \quad \sum \chi(v_i/\hat{\beta}_v) = 0, \tag{16}$$

where $\chi(\cdot)$ is some score function;
- evaluate the estimate of ρ of the form

$$\hat{\rho}_n = (\hat{\beta}_u^2 - \hat{\beta}_v^2)/(\hat{\beta}_u^2 + \hat{\beta}_v^2). \tag{17}$$

The choice of a score function in (16) will be made by applying the minimax approach in Section 4.3.

4.2. Consistency and asymptotic normality

The asymptotic properties of the proposed estimator (17) are completely determined by the asymptotic properties of M-estimators of scale (16). The sufficient conditions of regularity providing the desired properties are put on the densities g and score functions χ (Hampel et al., 1986, pp. 125, 139):
(g1) g is twice continuously differentiable and satisfies $g(x) > 0$ for all x in $\mathbf{R}^1$.
(g2) Fisher information for scale $I(g)$ satisfies $0 < I(g) < \infty$.
(χ1) χ is well-defined and continuous on $\mathbf{R}^1 \backslash C(\chi)$, where $C(\chi)$ is finite. In each point of $C(\chi)$ there exist finite left- and right-limits of χ which are different. Also $\chi(-x) = \chi(x)$ if $(-x, x) \subset \mathbf{R}^1 \backslash C(\chi)$, and there exists $d > 0$ such that $\chi(x) \leqslant 0$ on $(0, d)$ and $\chi(x) \geqslant 0$ on (d, ∞).
(χ2) The set $D(\chi)$ of points in which χ is continuous but in which χ' is not defined or not continuous is finite.
(χ3) $\int \chi \, \mathrm{d}G = 0$ and $\int \chi^2 \, \mathrm{d}G < \infty$.
(χ4) $0 < \int x\chi'(x) \, \mathrm{d}G(x) < \infty$.

Theorem 1. *Under the conditions of regularity* (g1)–(χ4), *estimator* (17) *is consistent and asymptotically normal with the following variance*:

$$\operatorname{Var} \hat{\rho}_n = \frac{2(1-\rho^2)^2}{n} V(\chi, g), \quad V(\chi, g) = \frac{\int \chi^2(x) g(x) \, \mathrm{d}x}{(\int x\chi'(x) g(x) \, \mathrm{d}x)^2}. \tag{18}$$

Proof. Consistency of estimator (17) immediately follows from consistency of M-estimators of scale: as $\hat{\beta}_u$ and $\hat{\beta}_v$ tend to $\beta_u = \sigma\sqrt{1+\rho}$ and $\beta_v = \sigma\sqrt{1-\rho}$ in probability, hence $\hat{\rho}_n$ tends to ρ in probability.

We have asymptotic normality by the reasoning of Huber (1964, p. 78, Lemma 5): the numerator of the fraction in (17) is asymptotically normal, the denominator tends in probability to the positive constant $c = \beta_u^2 + \beta_v^2$, hence $n^{1/2}\rho_n$ is asymptotically normal (Cramér, 1946, p. 20.6).

The precise structure of (18) is obtained by a direct routine calculation using the following formula for the variance of a fraction of r.v. ξ and η (Kendall and Stuart, 1962)

$$\operatorname{Var}\frac{\xi}{\eta} = \left(\frac{\mathrm{E}\xi}{\mathrm{E}\eta}\right)^2 \left(\frac{\operatorname{Var}\xi}{\mathrm{E}^2\xi} + \frac{\operatorname{Var}\eta}{\mathrm{E}^2\eta} - \frac{2\operatorname{Cov}(\xi,\eta)}{\mathrm{E}\xi\,\mathrm{E}\eta}\right) + \mathrm{o}(1/n), \qquad (19)$$

where $\xi = \hat{\beta}_u^2 - \hat{\beta}_v^2$ and $\eta = \hat{\beta}_u^2 + \hat{\beta}_v^2$.

By the independence of $\hat{\beta}_u$ and $\hat{\beta}_v$ we have the following components of (19):

$$\mathrm{E}\xi = \beta_u^2 - \beta_v^2 + \sigma_u^2 - \sigma_v^2, \quad \mathrm{E}\eta = \beta_u^2 + \beta_v^2 + \sigma_u^2 + \sigma_v^2,$$

$$\operatorname{Var}\xi = \operatorname{Var}\eta = 4(\beta_u^2\sigma_u^2 + \beta_v^2\sigma_v^2) + \mathrm{o}(1/n),$$

$$\operatorname{Cov}(\xi,\eta) = 4(\beta_u^2\sigma_u^2 - \beta_v^2\sigma_v^2) + \mathrm{o}(1/n),$$

where

$$\beta_u^2 = \sigma^2(1+\rho), \quad \beta_v^2 = \sigma^2(1-\rho), \quad \sigma_u^2 = \beta_u^2 V(\chi,g)/n, \quad \sigma_v^2 = \beta_v^2 V(\chi,g)/n.$$

By substituting these components into (19) we obtain (18). □

Example 1. From (18) we get the expression for asymptotic variance of the sample correlation coefficient for a bivariate normal distribution with $\chi(x) = x^2 - 1$ and $g(x) = \phi(x)$: $\operatorname{Var} r = (1-\rho^2)^2/n$.

Example 2. The choice $\chi(x) = \operatorname{sign}(|x| - 1)$ and $g(x) = \phi(x)$ yields asymptotic variance for the median correlation coefficient

$$\operatorname{Var} r_{\mathrm{med}} = \frac{(1-\rho^2)^2}{8n\phi^2(\zeta_{3/4})\zeta_{3/4}^2}, \quad \zeta_{3/4} = \Phi^{-1}(3/4),$$

where $\Phi(x)$ is a standard normal cumulative, $\phi(x) = \Phi'(x)$.

Formula (18) for asymptotic variance has two factors: the first depends only on ρ, the second $n^{-1}V(\chi,g)$ is the asymptotic variance of M-estimators of scale (Huber, 1981, p. 123). Thus, we can directly apply the minimax variance estimators of scale in the gross error model (Huber, 1964, 1981) for the minimax variance estimation of a correlation coefficient.

4.3. Minimax variance estimators

Huber (1981) showed that under rather general conditions of regularity $(g1)$–$(\chi 4)$ M-estimators $\hat{\beta}_n$ (16) are consistent, asymptotically normal and possess the minimax property with regard to $V(\chi,g) = n\operatorname{Var}\hat{\beta}_n$

$$V(\chi^*,g) \leqslant V(\chi^*,g^*). \qquad (20)$$

Here g^* is the least informative (favorable) density minimizing Fisher information $I(g)$ for scale over a certain class $\mathscr{G}$

$$g^* = \arg\min_{g\in\mathscr{G}} I(g), \quad I(g) = \int (-xg'(x)/g(x) - 1)^2 g(x)\,\mathrm{d}x \tag{21}$$

and the score function $\chi^*(x)$ is given by (8).

Inequality (20) shows that the estimator $\hat{\beta}_n$ determined by the score function χ^* provides the guaranteed level of the accuracy of estimation for all g in $\mathscr{G}$

$$V(\chi^*, g) \leqslant V(\chi^*, g^*) = 1/(nI(g^*)).$$

For the class of ε-contaminated univariate normal distributions

$$\mathscr{G} = \{g\colon g(x) \geqslant (1-\varepsilon)\phi(x), 0 \leqslant \varepsilon < 1\}, \tag{22}$$

the minimax variance M-estimator of scale is defined by (9), where the parameters $x_0(\varepsilon)$ and $x_1(\varepsilon)$ are tabulated in (Huber, 1981, p. 121). This M-estimator is asymptotically equivalent to the L-estimator of the trimmed standard deviation type (see Huber, 1981, p. 122) with the levels of trimming $n_1(\varepsilon)$ and $n_2(\varepsilon)$ satisfying the following equations:

$$n_1(\varepsilon) = [(0.5 - G^*(x_0))n], \quad n_2(\varepsilon) = [G^*(-x_1)n], \tag{23}$$

where G^* is the least informative cumulative distribution.

The following result is obtained by the direct application of the above solution.

Theorem 2. *In class* (15) *of ε-contaminated bivariate normal distributions*

$$f(x,y) \geqslant (1-\varepsilon)\mathscr{N}(x,y|0,0,1,1,\rho), \quad 0 \leqslant \varepsilon < 1, \tag{24}$$

the minimax variance estimator of ρ is the trimmed correlation coefficient (12), *where the numbers $n_1 = n_1(\varepsilon)$ and $n_2 = n_2(\varepsilon)$ of the trimmed smallest and greatest order statistics $u_{(i)}$ and $v_{(i)}$ depend on the value of the contamination parameter ε through the auxiliary parameter $\gamma = 1 - \sqrt{1-\varepsilon}$. The precise character of the dependencies $n_1 = n_1(\gamma)$ and $n_2 = n_2(\gamma)$ can be found in* (*Huber*, 1981, *p*. 5.6).

Proof. It suffices to check that densities (24) belong to class (15).

From (22) we have the following inequalities:

$$\frac{1}{\sigma\sqrt{1+\rho}}\, g\left(\frac{u}{\sigma\sqrt{1+\rho}}\right) \geqslant (1-\gamma)\frac{1}{\sigma\sqrt{1+\rho}}\,\phi\left(\frac{u}{\sigma\sqrt{1+\rho}}\right),$$

$$\frac{1}{\sigma\sqrt{1-\rho}}\, g\left(\frac{v}{\sigma\sqrt{1-\rho}}\right) \geqslant (1-\gamma)\frac{1}{\sigma\sqrt{1-\rho}}\,\phi\left(\frac{v}{\sigma\sqrt{1-\rho}}\right).$$

By multiplying them we obtain the restriction of the class of ε-contaminated bivariate normal distributions (24), where $\varepsilon = 2\gamma - \gamma^2$. □

In the limiting case as $\varepsilon \to 1$, n_1 and n_2 tend to $[n/2]$, the estimates of scale $\hat{\beta}_u$ and $\hat{\beta}_v$ tend to the median absolute deviations MAD u and MAD v, respectively, and hence $\hat{\rho}$ tends to the median correlation coefficient r_{med}. If $\varepsilon = 0$ then the trimmed correlation coefficient is asymptotically equivalent to the sample correlation coefficient (see Section 2.4).

The following example illustrates the application of Theorem 2 to the choice of the levels of trimming n_1 and n_2 in the minimax variance estimator.

Example 3. Consider the case of heavy contamination with $\varepsilon = 0.2$. Then we have for the auxiliary parameter γ: $\gamma = 1 - \sqrt{0.8} \approx 0.1$. Using formula (23) and the table for the parameters of the least informative distribution G^* (see Huber, 1981, p. 121), we get for the parameters of trimming n_1 and n_2

$$n_1 = 0, \quad n_2 = [0.098n].$$

Note that the level of trimming appears to be rather moderate as compared with the level of contamination.

5. Discussion

- The minimax variance estimator was obtained in the setting, where the parameters of location and scale were assumed known. In real-life applications, this assumption, as a rule, is not true. Hence one should use robust estimates, namely the sample median and the sample median absolute deviation, for those parameters, or, robust principal variables (11) for (u_i, v_i), $i = 1, \dots, n$ in Eqs. (16).
- On finite samples, the comparison of estimators' behavior can be made using Monte Carlo modeling. Below we exhibit some experimental results on this issue.
  The mixture of normal populations (3) ($\varepsilon = 0.2$, $\rho = 0.5$, $\rho' = -0.99$, $k = 3$) was simulated on samples $n = 30$. The experiment was repeated 1000 times for the sample correlation coefficient, the trimmed correlation coefficient with $n_1 = 0$ and $n_2 = 3$, and the median correlation coefficient as the estimators of the correlation coefficient $\rho = 0.5$. We have the following values for their means and variances, respectively:

  $$r: \; -0.70, 0.06; \quad r_{tr}: \; 0.45, 0.02; \quad r_{\text{med}}: \; 0.48, 0.05.$$

  Here we again confirm the extremely poor behavior of the classical sample correlation coefficient. Its robust counterparts show a good quality of estimation, the best in variance for the optimal trimmed correlation coefficient. The median correlation coefficient is the best with regard to bias. It can also be seen from the above results that bias seems to be a more informative characteristic than variance. Thus the problem of designing a minimax bias estimator of a correlation coefficient is important, but this issue deserves a separate consideration.
- The confidence intervals for estimator (17) can be constructed using the Fisher transformation $z = \ln[(1+\hat{\rho})/(1-\hat{\rho})]/2$. In this case, the variance of the transformed variable z does not depend on ρ: $\operatorname{Var} z = 2V(\chi, g)/(n-3)$.
- The median correlation coefficient has the highest qualitative robustness properties (Shevlyakov, 1997): its breakdown point is $\frac{1}{2}$. Thus, the median correlation coefficient r_{med} may be regarded as a correlation analog of the sample median and the median absolute deviation—these well-known robust estimators of location and scale also possess both quantitative minimax and highest qualitative robustness properties.

References

Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ.

Devlin, S.J., Gnanadesikan, R., Kettenring, J.R., 1975. Robust estimation and outlier detection with correlation coefficient. Biometrika 62, 531–545.

Gnanadesikan, R., Kettenring, J.R., 1972. Robust estimates, residuals and outlier detection with multiresponse data. Biometrics 28, 81–124.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics. The Approach Based on Influence Functions. Wiley, New York.

Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 36, 1–72.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Kendall, M.G., Stuart, A., 1962. The advanced theory of statistics, Vol. 1, Distribution Theory. Charles Griffin, London.

Pasman, V.R., Shevlyakov, G.L., 1987. Robust methods of estimation of a correlation coefficient. Automat. i Telemekh. 27, 70–80 (in Russian).

Shevlyakov, G.L., 1997. On robust estimation of a correlation coefficient. J. Math. Sci. 83, 90–94.