A Radical Stable M-Estimator of a Correlation Coefficient for a Bivariate Normal Distribution

G. L. Shevlyakov[1], V. I. Shin[1], and K. Kim[1]

[1] Gwangju Institute of Science and Technology, 1 Oryong-dong, Buk-gu, Gwangju, 500-712, South Korea

Keywords: Robustness, Stable Estimation, Correlation Coefficient

1 Introduction

The simplest problem of correlation analysis is to estimate the correlation coefficient ρ of a bivariate distribution with density $f_{XY}(x,y)$ from the sample $(x_1,y_1),\dots,(x_n,y_n)$ of a bivariate r.v. (X,Y). Its classical estimator is given by the sample correlation coefficient $\mathbf{r}$ being the 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 μ_1 and μ_2 are the means, σ_1 and σ_2 are the standard deviations of the r.v.'s X and Y, respectively.

The sample correlation coefficient $\mathbf{r}$ is known to be very sensitive to the presence of outliers in the data Devlin et al. (1975), and hence it is necessary to use its robust counterparts.

At present there exist two principal approaches to the design of robust estimators, namely Huber's and Hampel's. The both approaches have been used to obtain robust estimators of a correlation coefficient of the bivariate normal distribution (see, e.g., Devlin et al. (1975), Shevlyakov et al. (2002)).

A few years ago, the so called *stable estimation* approach to designing of robust estimators was proposed by Shurygin (1994). This approach can be described as follows: (i) in the class of M-estimators, a new asymptotical characteristic of estimation called *stability* is introduced by the appropriately scaled Lagrange functional derivative of the asymptotic variance with respect to a distribution density; (ii) this characteristic measures the local sensitivity of an M-estimator to the variations of a model distribution and is complementary to the classical characteristic of efficiency similarly lying in the $[0,1]$ range; (iii) to design stable M-estimators of location and scale with desired levels of efficiency and stability, several criteria of optimization by the suitable choice of a score function (e.g., efficiency under restricted stability, the linear and nonlinear combinations of efficiency and stability, etc.) have been proposed.

2 Stable Estimation of a Distribution Parameter

For the data $(x_1,y_1),\dots,(x_n,y_n)$ from a bivariate distribution with density $f(x,y;\theta)$ defined on the support $\mathbf{R^2}$, consider the M-estimators $\widehat{\theta}$ of an unknown scalar parameter θ in the conventional form $\sum\psi(x_i,y_i;\widehat{\theta})=0$, where $\psi(x,y;\theta)$ is a score function belonging to a certain class Ψ. In particular, the maximum likelihood estimator of θ is given by the M-estimator with the score function $\psi_{\mathrm{ML}}(x,y;\theta)=\partial\log f(x,y;\theta)/\partial\theta$. Under general conditions of regularity put upon densities f and score functions ψ, M-estimators are consistent and asymptotically normal with the well-known expression for the asymptotic variance $\operatorname{Var}\theta=n^{-1}V(\psi,f)=\mathrm{E}_F[\psi^2]/(n\,\mathrm{E}_F^2[\partial\psi/\partial\theta])$. The minimum of the asymptotic variance is attained at the maximum likelihood score function $V^*=V(\psi_{\mathrm{ML}},f)$, and the efficiency of an M-estimator is commonly defined by $\operatorname{Eff}\widehat{\theta}=V^*/V(\psi,f)$.

A new measure of stability of estimation, the Lagrange functional derivative of the asymptotic variance $W(\psi,f)=\partial V(\psi,f)/\partial f$ is introduced and it takes the following form

$$W(\psi,f)=\mathrm{E}_F^{-2}[\partial\psi/\partial\theta]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi^2(x,y;\theta)\,dx\,dy \tag{1}$$

An M-estimator and the corresponding score function are called *stable* if there exists the integral $\int_{\mathcal{X}} \psi^2(x, y; \theta)\, dx\, dy < \infty$ and *unstable* otherwise. Note that for M-estimators of location, the latter requirement leads to highly robust *redescending M-estimators*.

Further, it is shown that an optimal score function minimizing $W(\psi, f)$ is given by

$$\psi_*(x, y; \theta) = \arg\min_{\psi \in \Psi} W(\psi, f) = \partial f(x, y; \theta)/\partial\theta + \beta\, f(x, y; \theta), \tag{2}$$

where the constant β is obtained from the condition providing the consistency of M-estimators.

An estimator with the score function (2) is called the estimator of maximal stability when $W_* = W(\psi_*, f)$, and a new characteristic called the *stability* of an M-estimator is introduced as the following ratio

$$\operatorname{Stb} \widehat{\theta} = W_*/W(\psi, f),$$

naturally lying in the $[0, 1]$ range.

Setting different weights for efficiency and stability, various criteria of optimization of estimation can be proposed (see Shurygin (1994)). In particular, a reasonable choice is associated with the equal weights of the efficiency and stability functionals, i.e., when $\operatorname{Eff} \widehat{\theta} = \operatorname{Stb} \widehat{\theta}$: this estimator is called *radical*, and in this case, the score function of the radical M-estimator is given by

$$\psi_{\text{rad}}(x, y; \theta) = [\partial \log f(x, y; \theta)/\partial\theta + \beta]\ \sqrt{f(x, y; \theta)}, \tag{3}$$

where the constant β is obtained from the condition of consistency.

Note that the factor $\sqrt{f(x, y; \theta)}$ in Eq. (3) yields relatively lesser weights to the relatively greater observations in the equation $\sum \psi_{\text{rad}}(x_i, y_i; \widehat{\theta}) = 0$ for M-estimators and therefore provides robustness of the estimation procedure to gross errors.

3 Stable Estimation of a Correlation Coefficient

Now we apply the aforementioned results to stable estimation of the correlation coefficient of the bivariate normal distribution density $\mathcal{N}(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Assume that the means μ_1 and μ_2, the standard deviations σ_1 and σ_2 are given, and without loss of generality set them to zero and unity, respectively. Thus, the model distribution density is $f(x, y; \rho) = \mathcal{N}(x, y; 0, 0, 1, 1, \rho) = (2\pi)^{-1}(1 - \rho^2)^{-1/2} \exp\left\{-(x^2 - 2\rho xy + y^2)/[2(1 - \rho^2)]\right\}$. From Eq. (3) it follows that the radical M-estimator of ρ is the solution to the equation $\sum (\partial \log f(x_i, y_i; \rho)/\partial\rho + \beta)\ \sqrt{f(x_i, y_i; \rho)} = 0$, where $\beta = -\rho/[3(1-\rho^2)]$. Its asymptotic variance is $\operatorname{Var} \mathbf{r}_{\text{rad}} = 81(9+10\rho^2)(1-\rho^2)^2/[512(1+\rho^2)^2\, n]$ with the values of efficiency and stability $\operatorname{Eff} \mathbf{r}_{\text{rad}} = \operatorname{Stb} \mathbf{r}_{\text{rad}} = g(\rho^2) = 512(1 + \rho^2)/[81(9 + 10\rho^2)]$ varying in a narrow range from $g_{\min} = g(1) = 0.6654$ to $g_{\max} = g(0) = 0.7023$. Thus the radical estimator $\mathbf{r}_{\text{rad}}$ possesses the reasonable levels of efficiency and stability.

The performance of the radical estimator is compared with many other estimators of a correlation coefficient proposed in Devlin et al. (1975) and in Shevlyakov et al. (2002 (e.g., the sample correlation coefficient, the quadrant and Spearman rank correlation coefficients, the trimmed and median correlation coefficients, etc.) in gross error models both on finite samples and in asymptotics. Summarizing these studies, we state that, with respect to variance, the radical estimator is the best among the considered competitors.

References

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

G.L. Shevlyakov and N.O. Vilchevsky (2002). Minimax variance estimation of a correlation coefficient for ε-contaminated bivariate normal distributions. *Statist. Probab. Lett.* 57, 91–100.

A.M. Shurygin (1994). New approach to optimization of stable estimation. In *Proceedings of the 1st US/Japan Conference on Frontiers of Statistical Modeling*, pp. 315–340, Kluwer, Netherlands.