

A Modification of Linfoot's Informational Correlation Coefficient

Georgy Shevlyakov
Peter the Great St.Petersburg Polytechnic

Nikita Vasilevskiy
Itivity Group AB

Abstract

Performance of the Linfoot's informational correlation coefficient is experimentally studied at the bivariate normal distribution. It is satisfactory in the case of a strong correlation and on large samples. To reduce the bias of estimation, a symmetric version of this correlation measure is proposed. On small and large samples, this modified informational correlation coefficient outperforms Linfoot's correlation measure at the bivariate normal distribution in the wide range of the correlation coefficient.

Keywords: informational measures, correlation, bivariate normal distribution.

1. Introduction

Pearson's correlation coefficient is a well-defined measure of the linear dependence between continuous random variables X and Y , as well as the closely related to it rank measures: namely, the quadrant, Spearman and Kendall correlation coefficients. However, if one is interested either in processing discrete data or in revealing the possible nonlinear relationship between random variables, then difficulties may arise both in the implementation of those classical measures as well as in their interpretation.

In the literature, several proposals have been made to solve these problems, for instance, Gebelein's and Sarmanov's correlation coefficients (Gebelein 1941), (Sarmanov 1958). Recently, the distance correlation coefficient was proposed (see (Székely, Rizzo, and Bakirov 2007)).

In what follows, we focus on the informational measures of association between random variables (Shannon 1948). The dependence measure by Joe (1989) exploits the concept of the relative entropy that measures the similarity of two random variables with the distributions $p(x)$ and $q(x)$ in the discrete case

$$D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

Silvey (1964) uses the measure of dependence between two random variables defined by the ratio of their joint density and the product of their marginal densities

$$\varphi(x, y) = p(x, y) / [p(x)p(y)].$$

The introduced measure is defined as $\Delta = E[d(x)]$, where

$$d(x) = \int_{y: \varphi(x,y) > 1} [p(y|x) - p(y)] dy.$$

Thus, it can be rewritten as

$$\Delta = \int \int_{(x,y): \varphi(x,y) > 1} [p(x,y) - p(x)p(y)] dx dy.$$

Granger, Maasoumi, and Racine (2004) introduce another measure of dependence

$$S_p = \frac{1}{2} \int \int [p(x,y)^{1/2} - [p(x)p(y)]^{1/2}]^2 dx dy.$$

However, Joe's measure of dependence is not symmetric, whereas Silvey's and Granger's measures are hard to compute.

Mutual information ($I(X,Y)$) for any pair of discrete and continuous random variables X and Y is defined as follows

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \quad I(X,Y) = \int \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dx dy.$$

The informational correlation coefficient (ICC), firstly introduced by Linfoot (1957), is defined as follows

$$\rho_{ICC}(X,Y) = \sqrt{1 - e^{-2I(X,Y)}}. \quad (1)$$

Note that ICC is equal to the classical Pearson's correlation coefficient at the bivariate normal distribution: $\rho_{ICC}(X,Y) = \rho$.

2. Problem setting

Although the Linfoot's correlation coefficient ICC was introduced more than 60 years ago, its properties as a statistical measure of correlation have not yet been studied; it was not checked how well this measure estimates the correlation coefficient based on a sample of a given size.

Below, using Monte Carlo method on small and large samples, we experimentally examine the unbiasedness of ICC at the bivariate normal distribution over the wide range of its correlation coefficient.

Moreover, in order to improve the performance of ICC , namely to reduce its bias, we propose and study its modified symmetric version denoted as $MICC$.

3. Monte Carlo experiment

3.1. Description of the computational algorithm

All numerical experiments are performed using R language, especially its "entropy" library. The first problem is how to compute mutual information, which is used in (1). This is solved by applying a shrink-algorithm (Hausser and Strimmer 2010).

There exist several different algorithms of computing $I(X,Y)$; in our work, we choose the most precise one, not the fastest (for comparative analysis, see (Hausser and Strimmer 2010)). All experiments are performed at the standard bivariate normal distribution with density $f(x,y) = N(x,y; 0, 0, 1, 1, \rho)$.

The general algorithm can be described as follows:

1. Generate a sample of a fixed size: $N = 20, 60, 100, 400, 1000, 10000$.
2. Extract x - and y -components from the sample, which are dependent random variables with the correlation coefficient ρ .
3. Construct the table of frequencies, the discrete analog of the joint distribution: we take a rectangle $[x_{min}, x_{max}] \times [y_{min}, y_{max}]$ on plane and divide it into $n_x \times n_y$ "bins" of equal size. Thus, the table of dimension n_x, n_y is built, each element of which is equal to the number of points in the corresponding bin.
4. Mutual information $I(X, Y)$ and ICC are computed using this table of frequencies.

This sequence is repeated 1000 times, allowing us to compute Monte Carlo estimates of the mean and variance of ICC : computations are performed for $\rho = 0, 0.1, 0.2, \dots, 0.9, 1$; the number of bins is taken equal to 400. Typical results are exhibited in Figure 1.

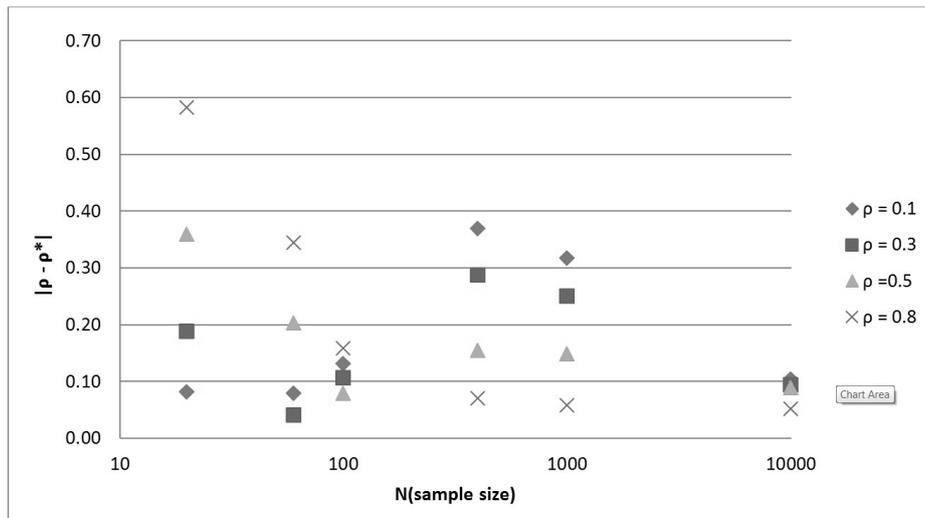


Figure 1: Monte Carlo biases of ICC

3.2. Monte Carlo results for ICC

From Figure 1 it follows that estimation biases are considerably large (on small samples, they can even be greater than 0.5). Relatively small biases are observed only on large samples $N = 1000$ and $N = 10000$.

Satisfactory performance is observed in the case of a strong correlation—the ICC biases decrease with the growth of the sample size.

We may also add that the coefficient of variance is less than 0.2 for all examined combinations of (ρ, N) .

A remark on the choice of the number of bins. The shrink-algorithm takes the table of frequencies as an input. It appeared that the algorithm performance depends on the relation N/K^2 , where K is the linear dimension of the table. We observed that results are almost independent of the changes of K , as they depend only on the coefficient $B = N/K^2$. For $\rho = 0.5$, the value $B = 7$ is optimal. Given a data sample, we can choose an appropriate value of K , which is a tuning parameter of our algorithm.

4. A symmetric version of Linfoot's correlation coefficient

4.1. Main result

Mutual entropy, also known as the Kullback-Leibler distance, has a serious disadvantage – it is not symmetric, i.e., $D_{KL}(p||q) \neq D_{KL}(q||p)$. Thus, the Kullback-Leibler divergence is

used (Kullback 1959)

$$\text{Div}(p||q) = D_{KL}(p||q) + D_{KL}(q||p). \quad (2)$$

Analogously, a symmetric version of mutual information can be introduced as it is natural to use a symmetric measure of correlation for estimation of the Pearson correlation coefficient, a symmetric measure of interdependence between random variables—in this case we expect lesser biases of estimation.

$$\begin{aligned} J(X, Y) &= I(X, Y) + I^*(X, Y) \\ &= \int \int f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy + \int \int f(x)f(y) \log \frac{f(x)f(y)}{f(x, y)} dx dy. \end{aligned}$$

Our idea is to repeat Linfoot's derivation of Equation (1), replacing the mutual information $I(X, Y)$ with its symmetric version $J(X, Y)$. In this case, the following result holds.

Theorem *A symmetric analog of the Linfoot's informational correlation coefficient (1) called as the modified informational correlation coefficient (MICC) is given by*

$$\rho_{MICC} = \sqrt{1 - \frac{2}{W(2e^{2(J+1)})}} \quad (3)$$

with the particular case $\rho_{MICC} = \rho$ at the bivariate normal distribution.

Here W is the Lambert's function—the inverse function for xe^x ; it cannot be expressed in terms of elementary functions. Its properties are well developed, and there are special methods to compute it (see (Corless, Gonnet, Hare, Jeffrey, and Knuth 1996)).

Proof The first step is to express the mutual information via the correlation coefficient similarly to that obtained by Linfoot (1957)

$$I(X, Y) = -\frac{1}{2} \log(1 - \rho^2).$$

Compute $I^*(X, Y)$ at the bivariate normal with density $f(x, y) = N(x, y; 0, 0, \sigma_x^2, \sigma_y^2, \rho)$

$$\begin{aligned} I^*(X, Y) &= \int \int f(x)f(y) \log \frac{f(x)f(y)}{f(x, y)} dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \int \int \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \left(\frac{\rho^2}{1-\rho^2} \frac{x^2}{2\sigma_x^2} + \frac{\rho^2}{1-\rho^2} \frac{y^2}{2\sigma_y^2} - \frac{\rho xy}{\sigma_x\sigma_y}\right) dx dy. \end{aligned}$$

Consider the following three integrals:

$$\begin{aligned} I_1(X, Y) &= \frac{1}{2\pi\sigma_x\sigma_y} \int \int \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \frac{\rho^2}{1-\rho^2} \frac{x^2}{2\sigma_x^2} dx dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \frac{\rho^2}{1-\rho^2} \int \frac{x^2}{2\sigma_x^2} e^{-\frac{x^2}{2\sigma_x^2}} dx \int e^{-\frac{y^2}{2\sigma_y^2}} dy \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \frac{\rho^2}{1-\rho^2} \frac{1}{2\sigma_x^2} \frac{\sqrt{\pi}}{2} 2^{\frac{3}{2}} \sigma_x^3 \sqrt{2\pi} \sigma_y \\ &= \frac{\rho^2}{2(1-\rho^2)}. \end{aligned}$$

Analogously, we have

$$I_2(X, Y) = \frac{1}{2\pi\sigma_x\sigma_y} \int \int \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \frac{\rho^2}{1-\rho^2} \frac{y^2}{2\sigma_y^2} dx dy$$

$$= \frac{\rho^2}{2(1 - \rho^2)}.$$

Next,

$$I_3(X, Y) = -\frac{\rho}{\sigma_x \sigma_y} \int x \exp\left(-\frac{x^2}{2\sigma_x^2}\right) dx \int y \exp\left(-\frac{y^2}{2\sigma_y^2}\right) dy = 0.$$

Thus, we get

$$I^*(X, Y) = I_1 + I_2 + I_3 = \frac{\rho^2}{1 - \rho^2},$$

$$J(X, Y) = -\frac{\log(1 - \rho^2)}{2} + \frac{\rho^2}{1 - \rho^2}.$$

Now we express the correlation coefficient ρ via the symmetrized mutual information $J(X, Y)$ inverting the above Equation:

$$J(X, Y) = J = -\frac{\log(1 - \rho^2)}{2} + \frac{\rho^2}{1 - \rho^2}.$$

Set $1 - \rho^2 = t$. Then we get

$$J = -\frac{\log t}{2} + \frac{1 - t}{t} = -\frac{\log t}{2} + \frac{1}{t} - 1.$$

Further

$$-2(J + 1) = \log t - \frac{2}{t} \Rightarrow \frac{\exp(-2(J + 1))}{2} = \frac{t}{2} \exp(-2/t),$$

and setting $2/t = p$, we finalize the derivation

$$\frac{1}{2e^{2(J+1)}} = \frac{1}{p} e^{-p} \Rightarrow 2e^{2(J+1)} = pe^p \Rightarrow p = W(2e^{2(J+1)}),$$

where W is the Lambert function, namely the inverse function for pe^p . Finally, we arrive at Equation (3).

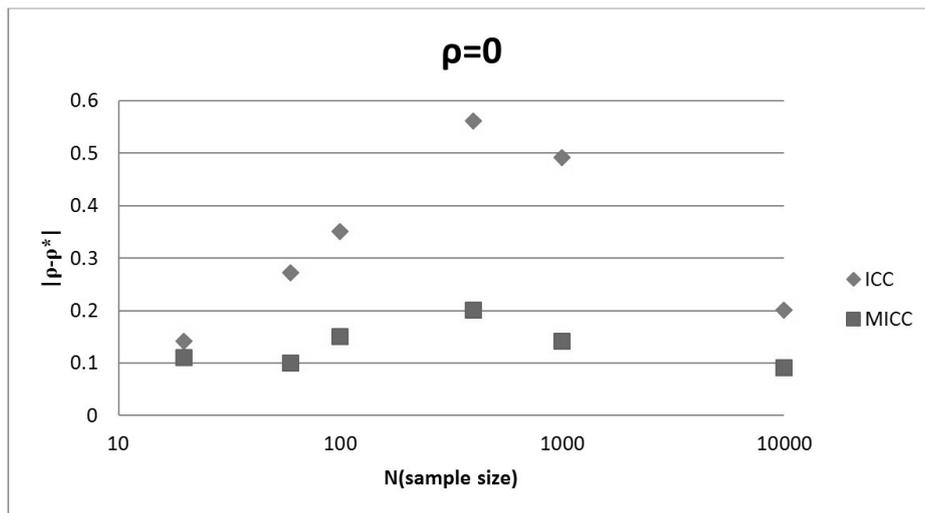
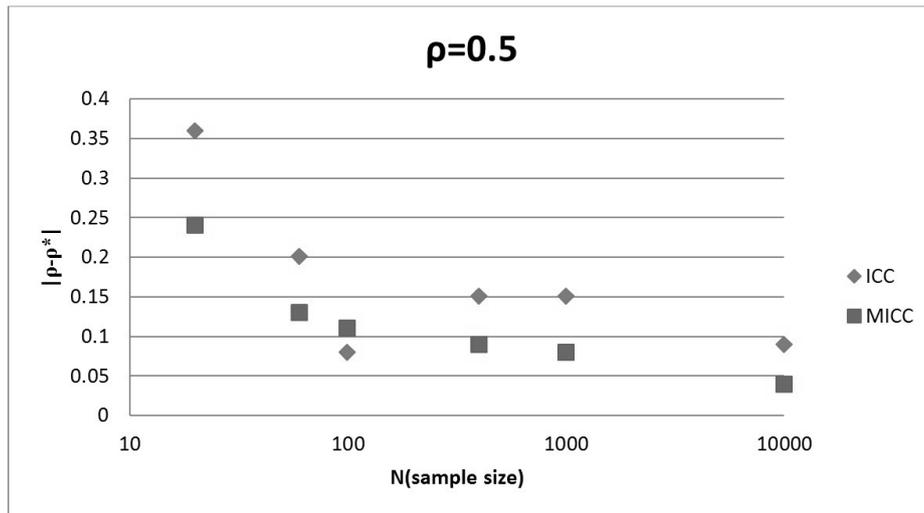
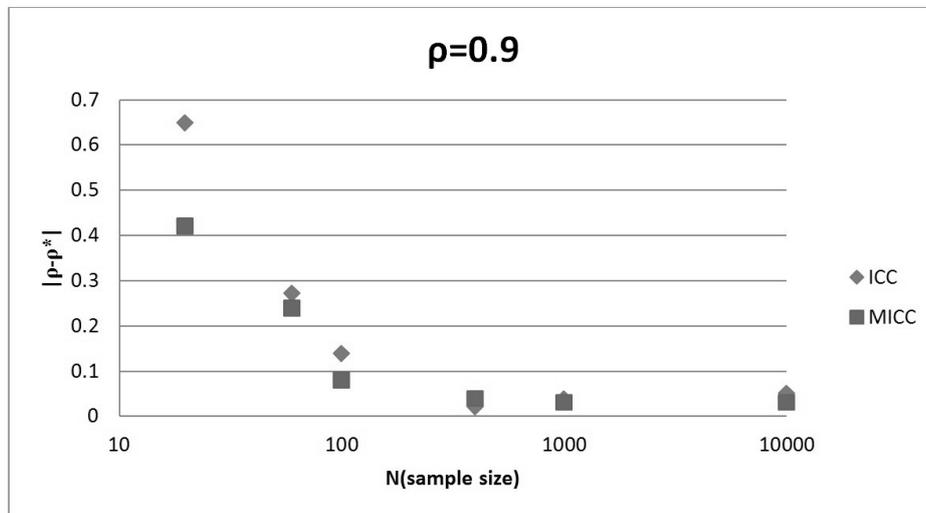


Figure 2: Monte Carlo biases of *ICC* and *MICC* at $\rho = 0$

4.2. Monte Carlo results for *MICC*

The results of comparison of these two correlation measures are exhibited in Figures 2–4: from them it follows that the modified informational correlation coefficient *MICC* outperforms its classical Linfoot’s analog *ICC* in almost all of the combinations of sample sizes and correlation

Figure 3: Monte Carlo biases of *ICC* and *MICC* at $\rho = 0.5$ Figure 4: Monte Carlo biases of *ICC* and *MICC* at $\rho = 0.9$

coefficients. The observed improvement is more considerable on small samples and low values of the correlation coefficient—just in the most difficult cases for *ICC*.

5. Concluding remarks

- The statistical performance of the Linfoot's informational correlation coefficient is studied: considerable biases of estimation are observed, especially on small samples.
- To reduce the biases of *ICC*, a modified symmetric version of it, namely *MICC*, is proposed, which proved to provide much lesser estimation biases as compared to its prototype.
- The proposed modified informational correlation coefficient *MICC* is recommended for processing Big Data, as the obtained results show that its best performance is achieved on large samples.
- Work in process: since the problem of estimation of nonlinear dependencies between random variables still remains important, it seems advantageous to use informational measures of correlation in this case, and the comparative study of the performance of those and other measures of nonlinear correlation, such as the coefficient of determination, Sarmanov's correlation coefficient (Sarmanov 1958) and distance correlation

coefficient of (Székely *et al.* 2007), is in process.

References

- Corless R, Gonnet G, Hare D, Jeffrey D, Knuth D (1996). “On the Lambert W Function.” *Advances in Computational Mathematics*, **11**, 329–359.
- Gebelein H (1941). “Das Statistische Problem der Korrelation als Variations und Eigenwertproblem und Sein Zusammenhang mit der Ausgleichsrechnung.” *Z. Angewandte Math. Mech.*, **21**, 364–379.
- Granger C, Maasoumi E, Racine J (2004). “A Dependence Metric for Possibly Nonlinear Processes.” *Journal of Time Series Analysis*, **25**, 649–669.
- Hausser J, Strimmer K (2010). “Entropy Inference and the James-Stein Estimator with Application to Nonlinear Gene Association Networks.” *Journal of Machine Learning Research*, **10**, 1469–1484.
- Joe H (1989). “Relative Entropy Measures of Multivariate Dependence.” *Journal of the American Statistical Association*, **84**, 157–164.
- Kullback S (1959). *Information Theory and Statistics*. Wiley.
- Linfoot E (1957). “An Informational Measure of Correlation.” *Information and Control*, **1**, 85–89.
- Sarmanov O (1958). “Maximum Correlation Coefficient (Symmetric Case).” *Doklady Akad. Nauk SSSR*, **120**, 715–718.
- Shannon C (1948). “A Mathematical Theory of Communication.” *Bell System Technical Journal*, **27**, 379–423.
- Silvey S (1964). “A Measure of Association.” *The Annals of Mathematical Statistics*, **35**, 1157–1166.
- Székely G, Rizzo M, Bakirov N (2007). “Measuring and Testing Dependence by Correlation of Distances.” *Ann. Statist.*, **35**, 2769–2794.

Affiliation:

Georgy Shevlyakov

Peter the Great St. Petersburg Polytechnic University

Polytechnicheskaya, 29, 195251 St. Petersburg, Russia

Institute for Problems in Mechanical Engineering, Russian Academy of Sciences

V.O., Bolshoy pr., 61, 199178 St. Petersburg, Russia

E-mail: Georgy.Shevlyakov@phmf.spbstu.ru