



**FORDHAM UNIVERSITY
DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

2023/2024

**Taraldsen's Exact
Correlation Density**

H.D. Vinod

Fordham University

Discussion Paper No: 2023-10

November 2023

**Department of Economics
Fordham University**

Taraldsen's Exact Correlation Density

H. D. Vinod *

November 28, 2023

Abstract

This short paper serves as a vignette for the R package “practicalSigni” describing the usage of `pvTarald()` and `qTarald()`, respectively, for computing the p-values and quantiles of Taraldsen's new exact sampling distribution of Pearson correlation coefficient and its generalized versions. A new table allows practitioners to see one-sided critical values from Taraldsen's (2023) exact sampling distribution, which generalizes Fisher's z-transform. It computes p-values and quantiles for the arbitrary hypothesized value of the population correlation coefficient, $\rho \in [-1, 1]$.

*address: H. D. Vinod, Professor of Economics, Fordham University, Bronx, New York, USA 10458. E-mail: vinod@fordham.edu. JEL codes C30, C51. Keywords: Sampling distribution, Generalized correlation, Arbitrary null.

1 Introduction

Since sample correlation coefficient r_{ij} from a bivariate normal parent has a non-normal distribution, Sir R. A. Fisher developed his famous z-transformation in the 1920s. He proved that the following transformed statistic r_{ij}^T is approximately normal with a stable variance,

$$r_{ij}^T = (1/2) \log \frac{(1 + r_{ij})}{(1 - r_{ij})} \sim N(0, 1/n), \quad (1)$$

provided $r_{ij} \neq 1$. Recent work has developed the exact sampling distribution of r_{ij} . This short paper directly computes a confidence interval for the arbitrary hypothesized value $\rho \in [-1, 1]$ of the population correlation coefficient.

Let r be the empirical correlation of a random sample of size n from a bivariate normal parent. Fisher's famous z-transformation was extended by C. R. Rao in 1973. Unfortunately, Rao's elegant derivation is impractical to implement computationally, according to Taraldsen (2023). Taraldsen's Theorem 2.1 provides a computable exact density with $v = (n - 1) > 1$ as

$$f(\rho|r, v) = \frac{v(v-1)\Gamma(v-1)}{\sqrt{(2\pi)\Gamma(v+0.5)}} (1 - r^2)^{\frac{v-1}{2}} (1 - \rho^2)^{\frac{v-2}{2}} (1 - r\rho)^{\frac{1-2v}{2}} \quad (2)$$
$$F\left(\frac{3}{2}; -0.5; v + 0.5; \frac{1+r\rho}{2}\right),$$

where $F(.,.;.,.)$ denotes the Gaussian hypergeometric function, available in the R package **hypergeom** by R.K.S Hankin. Our code computes (2) over a grid of r values used in constructing our Table 1 and Figures 1 and 2 below.

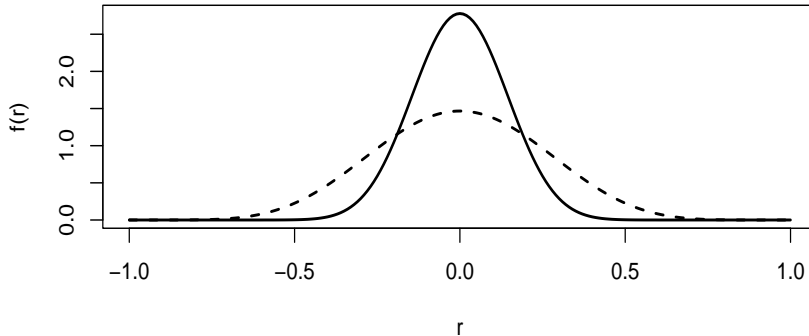
The first term, called ‘Trm1’ in our algorithm `pvTara1d(.)` inside the package’s code, involves a ratio of two gamma (factorial) functions appearing in (2). For $n > 164$, each gamma becomes infinitely large, and ‘Trm1’ becomes ‘NaN’ or not a number. Hence, our code replaces (Winsorizes) expression involving $\Gamma(n - 1)$ for $n > 164$ by $\Gamma(163)$.

Assuming that the data come from a bivariate normal parent, the sampling distribution of any correlation coefficient is (2). Hence, the sampling distribution of unequal off-diagonal elements of the matrix of generalized correlations R^* based on Vinod (2014) also follows (2). Vinod (2021) and its eight vignettes describe an R package for computing elements of the generalized correlation matrix R^* . One uses the function `gmcmtx0(mtx)` if the data are in the form of a matrix denoted as ‘mtx.’ If, on the other hand, there are two data columns (x, y) , the R function is `rstar(x, y)`, which further provides p-values using Fisher’s z-transform.

When we test the null hypothesis $H_0 : \rho = 0$, the relevant sampling distribution is obtained by plugging $\rho = 0$ in (2), depicted in Figure 1, for two selected sample sizes. Both distributions are centered at the null value $\rho = 0$.

We obtain a two-tail (95%, say) confidence interval by using the sampling distribution’s 2.5% and 97.5% quantiles. If the observed correlation coefficient r is inside the confidence interval, we say that the observed r is statistically insignificant, because it could have arisen from a population where the null value $\rho = 0$ holds.

Figure 1: Taraldsen's exact sampling density of a correlation coefficient under the null of $\rho = 0$, solid line $n=50$, dashed line $n=15$

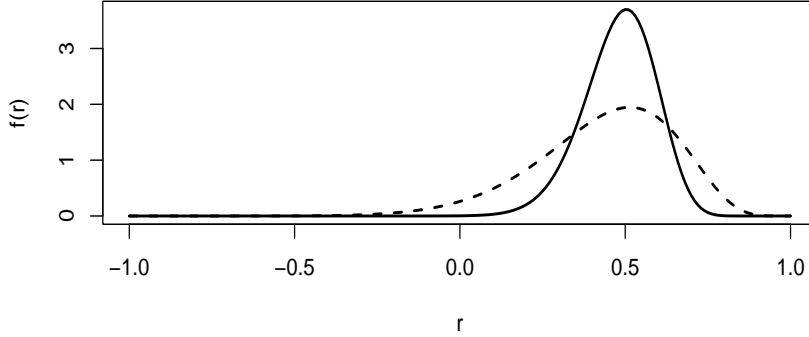


Similarly, one can test the nonzero null hypothesis $H_0 : \rho = 0.5$ using the equation obtained by plugging $\rho = 0.5$ in (2) depicted in Figure 2.

Figures 1 and 2 show that the formula (2) and our numerical implementation are ready for practical use. These exact densities depend on the sample size and the population correlation coefficient, $-1 \leq \rho \leq 1$. Given any hypothesized ρ and sample size, our computer algorithm computes the exact density, similar to Figures 1 and 2. Suppose we wish to help typical practitioners who want the tail areas helpful in testing the null hypothesis $\rho = 0$. Then, we need to create a table of a set of typical quantiles evaluated at specific cumulative probabilities and a corresponding selected set of standard sample sizes with a fixed $\rho = 0$.

Because of the complicated form of the density (2), it is not surprising that its (cumulative) distribution function $\int_{-1}^r f(\rho|r, v)$ by analytical methods is

Figure 2: Taraldsen's exact sampling density of correlation coefficient under the null of $\rho = 0.5$, solid line $n=50$, dashed line $n=15$



not available in the literature. Hence, let us compute cumulative probabilities by numerical integration defined as the rescaled area under the curve $f(r, v)$ for specified ρ . See Figure 1 for two choices of $v(= n - 1)$ for sample sizes ($n=50, 15$) and $\rho = 0$. The cumulative probability becomes a sum of rescaled areas of small-width rectangles whose heights are determined by the curve tracing $f(r, v)$. The accuracy of numerical approximation to the area is better the larger the number of rectangles (i.e., finer the evaluation grid's width).

We use $r \in [-1, 1]$ sequence created by the R command `r=seq(-1, 1, by =0.001)`, yielding 2001 rectangles. Denote the height of $f(r, v)$ by $H_f = H_{f(r,v)}$. Now, the area between any two $r \in [-1, 1]$ limits, say r_{Lo} and r_{Up} , is a summation of areas (height times width=0.001) of all rectangles. Now,

the cumulative probabilities in the range are

$$\Sigma_{r_{Lo}}^{r_{Up}} H_f / \Sigma_{-1}^1 H_f, \quad (3)$$

where the common width cancels, and where the denominator $\Sigma_{-1}^1 H_f$ converts the rectangle areas into probabilities. More generally, we can use numerical densities $f(\rho, r, v)$ for any hypothesized null value of $\rho \in [-1, 1]$.

Thus, we have a numerical approximation to the exact (cumulative) distribution function under the bivariate normality of the parent,

$$F(\rho, r, v) = \int_{-1}^r f(\rho|r, v).$$

The transform from $f(\cdot)$ to $F(\cdot)$ is called the probability integral transform, and its inverse $F^{-1}(c|\rho, v)$ gives relevant correlation coefficients r as quantiles for specified cumulative probability c or “cum” as the argument. A computer algorithm to find such quantiles `qTarald(n, rho, cum)` is included in the package **practicalSigni**.

The computable version of the exact $F^{-1}(c|\rho, v)$ allows the construction of confidence intervals based on quantiles for each ρ and sample size. For example, a 95% two-tail confidence interval uses the 2.5% quantile $F^{-1}(c = 0.025)$ as the lower limit, and 97.5% quantile $F^{-1}(c = 0.975)$ as the upper limit. These limits depend on hypothesized ρ and sample size. Since $\rho = 0$ is a common null hypothesis for correlation coefficients, let us provide a table of $F^{-1}(c)$ quantiles for eleven sample sizes (listed in row names) and eight

cumulative probabilities listed in column titles of Table 1.

The p-values in statistical inference are defined as the probability of observing the random variable (correlation coefficient) as extreme or more extreme than the observed value of the correlation coefficient r for a given null value $\rho = 0$. Any one-tail p-values based on $f(\rho|r, v)$ of (2) for arbitrary nonzero “null” values of ρ can be similarly computed by numerical integration defined as the area under the curve.

Table 1: Critical values for powerful one-sided tests for Pearson’s $r(i,j)$ when $\rho = 0$. We report quantiles evaluated at specified cumulative probabilities (c=.) using Taraldsen’s exact sampling distribution for various sample sizes.

	c=				c=			
	c=0.01	0.025	c=0.05	c=0.1	c=0.9	c=0.95	0.975	c=0.99
n=5	-0.83	-0.75	-0.67	-0.55	0.55	0.67	0.75	0.83
n=10	-0.66	-0.58	-0.50	-0.40	0.40	0.50	0.58	0.66
n=15	-0.56	-0.48	-0.41	-0.33	0.33	0.41	0.48	0.56
n=20	-0.49	-0.42	-0.36	-0.28	0.28	0.36	0.42	0.49
n=25	-0.44	-0.38	-0.32	-0.26	0.26	0.32	0.38	0.44
n=30	-0.41	-0.35	-0.30	-0.23	0.23	0.30	0.35	0.41
n=40	-0.36	-0.30	-0.26	-0.20	0.20	0.26	0.30	0.36
n=70	-0.27	-0.23	-0.20	-0.15	0.15	0.20	0.23	0.27
n=90	-0.24	-0.20	-0.17	-0.14	0.14	0.17	0.20	0.24
n=100	-0.23	-0.20	-0.16	-0.13	0.13	0.16	0.20	0.23
n=150	-0.19	-0.16	-0.13	-0.10	0.10	0.13	0.16	0.19

For the convenience of practitioners, we explain how to use the cumulative probabilities in Table 1 in the context of testing the null hypothesis $\rho = 0$. A close look at Table 1 confirms that the distribution is symmetric around $\rho = 0$, as in Figure 1. Let us consider some examples. If $n=100$, the critical value from Table 1 for a one-tail 95% test is 0.16 (line $n=100$, column $c=0.95$). Let

the observed positive r be 0.3. Since r exceeds the critical value ($r > 0.16$), we reject $\rho = 0$. If $n=25$, the critical value for a 5% left tail in Table 1 is -0.32 . If the observed $r = -0.44$ is more extreme than the critical value -0.32 , it implies that the observed r falls (outside the confidence interval) in the left tail. Hence, we reject $\rho = 0$ to conclude that the observed $r = -0.44$ is significantly negative.

Table 1 can also be used as follows for constructing two-tail 95% confidence intervals. If the sample size is 30, we see along the row $n=30$, that column $c=0.025$ gives -0.35 as the lower confidence limit, and column $c=0.975$ gives 0.35 as the upper confidence limit. In other words, for $n=30$, any correlation coefficient smaller than 0.35 in absolute value is statistically insignificant.

1.1 Confidence Intervals Without Normality

Now, we outline the basic idea behind using bootstrap replicated data to construct sampling distribution elements without assuming bivariate normality. The method is well known and applies to Pearson's correlation coefficient and generalizations.

A bootstrap creates a large number $J = 999$, say, versions of data $(X_{i\ell}, X_{j\ell})$ variables. They are identified by a marker $(\ell = 1, \dots, J)$. We use the maximum entropy bootstrap (R package **meboot**) designed for dependent data, Vinod and López-de-Lacalle (2009). Each version of data from a maximum entropy density yields a new Pearson correlation $r(ij; \ell)$ and

generalized correlation $r^*(i|j; \ell), r^*(j|i; \ell)$ values. The large set of J replicates of these correlations gives a numerical approximation to the sampling distribution of these correlations. Note that such a bootstrap sampling distribution is data-driven. It does not assume bivariate normality needed for the construction of Table 1 based on (2).

The sampling distribution is empirically realized by sorting the replicated correlations from the smallest to the largest. Sorting yields their “order statistics” by inserting parentheses around ℓ as $r(ij; (\ell)), r^*(i|j; (\ell)),$ and $r^*(j|i; (\ell))$. The smallest value is identified by ($\ell = 1$) and the largest by ($\ell = J$). By definition, a left-tail 95% confidence interval leaves a 5% probability mass in the left tail. The resulting approximate 95% left tail interval when $J = 999$ starts at the 50-th order statistic and ends at unity, $[rij; (50)), 1]$. If the hypothesized $\rho = 0$ is inside the one-tail interval, one fails to reject (accept) the null hypothesis $H_0 : \rho = 0$. Two-sided 95% intervals start at the 25th order statistic and end at the 975-th order statistic described above when $J = 999$.

The tabulation of Taraldsen’s exact sampling distribution of correlation coefficients in Table 1 is new and deserves greater attention. It is an improvement over standard significance tests of correlation coefficients based on Fisher’s z-transform, especially for a nonzero null hypothesis on ρ .

2 Examples & Tests

This section considers some examples and tests for illustration of the software. Our first example deals with fuel economy in automobile design. R software comes with ‘mtcars’ data on ten aspects of automobile design and performance for 32 automobiles. We consider two design features for illustration: miles per gallon (*mpg*), and horsepower (*hp*). Vinod (2014) reports the Pearson correlation coefficient $r(\text{mpg}, \text{hp}) = -0.78$ in his Figure 2. The negative sign correctly shows that one gets reduced *mpg* when a car has larger horsepower *hp*. Table 2 in Vinod (2014) reports two generalized correlation coefficients obtained by using kernel regressions as $r^*(\text{mpg}|\text{hp}) = -0.938$ and $r^*(\text{hp}|\text{mpg}) = -0.853$.

Three R commands for the computation of generalized correlation coefficients are `library(generalCorr)`; `attach(mtcars)`; `rstar(mpg, hp)`. Note that the implicit assumption of linearity makes Pearson’s correlation coefficient ($=-0.78$) numerically “smaller” in magnitude than both generalized correlations r^* . There is no advantage in assuming linearity.

Now, let’s consult Table 1 for inference regarding the “mtcars” data. See the row marked “n=30,” and the column entitled “c=0.05”. A left-tail critical value is -0.30 . The observed correlation $r^*(\text{mpg}|\text{hp}) = -0.938$ is more negative than the critical value, or falls in the rejection region. Thus, the negative dependence of fuel economy (*mpg*) on the car’s horsepower is statistically significant. We re-confirm the significance by computing the one-tail p-value

(= 1e-16) using an R command `pvTarald(n=32,rho=0, obsr=-0.938)`. An extremely small p-value suggests strong statistical significance.

3 Taraldsen and practicalSigni Package

The **practicalSigni** package generally suggests giving less importance to formal statistical significance (p-values) and reports the practical significance as indicated by thirteen magnitudes m1 to m13. See the first vignette of the package for details.

If desired, the Taraldsen method can be applied to three measures (of thirteen) involving correlation coefficients. One can then estimate more exact p-values of the following correlation measures using the R function `pvTarald(.)` discussed here.

m4 is the Pearson correlation measure. For example, between fuel economy *mpg* and one of three regressors (weight, cylinders, and horsepower).

m5 is the larger of the two generalized correlation coefficients. For example $\max(r^*(mpg|x), r^*(x|mpg))$ for $x = (cyl, hp, wt)$

m6 is the generalized partial correlation coefficient (GPCC). It measures the generalized nonlinear nonparametric correlation between two variables after removing the effect of other variables. For example, $\max(r^*(mpg|x), r^*(x|mpg))$ for one of three (cyl, hp, wt) while explicitly removing the effect of the other two regressors.

The practical significance (magnitude importance) ranking of three re-

gressors (weight, cylinders, and horsepower) in determining the fuel economy (mpg) as reported in the first vignette of this package from the ‘mtcars’ data is ($wt = 0.868 > cyl = 0.852 > hp = 0.776$) for **m4**, the Pearson correlations with mpg.

It is customary to regard a regressor with “small” p-values as more statistically significant (hence, more important). The following results from R show that significance ranking ($wt > cyl > hp$) retains the magnitude ranking, even when one uses p-values.

```
> pvTarald(n=32,rho=0,obsr=0.776)
[hp] 3.374095e-08
> pvTarald(n=32,rho=0,obsr=0.852)
[cyl] 7.837853e-11
> pvTarald(n=32,rho=0,obsr=0.868)
[wt] 1.40794e-11
```

Actually, unchanging ranking is intuitively obvious from (2) and related figures when the null is $\rho = 0$. In other words, Taraldsen’s sampling distribution is a nicety, not particularly insightful for reporting practical significance when the hypothesized null value is $\rho = 0$.

Comparable **m5** values $\max(r^*(mpg|x), r^*(x|mpg))$ for $x = (cyl, hp, wt)$ are (0.943, 0.938, 0.917), respectively. The generalized partial correlation coefficient GPCC estimates denoted **m6** are (0.0019, 0.3886, 0.4812) with *mpg* for the same regressors (*cyl*, *hp*, *wt*), respectively. Using the new function `pvTarald(n=32, rho=0, obsr)` for **m5**, **m6** estimates as ‘obsr’ or

observed correlations keep the practical significance ranking based on magnitudes of these observed correlations unchanged.

If there are theoretical reasons for hypothesizing another null value, $\rho \in [-1, 1]$, then Taraldsen's becomes a new tool in assessing practical significance of a regressor in a model.

4 Final Remarks

Table 1 provides new critical values for powerful one-sided tests for Pearson's $r(i, j)$ and generalized $r^*(i, j)$ when $\rho = 0$ under bivariate normality. The reported cumulative probabilities are based on a recent generalization of Fisher's famous z-transformation by Taraldsen (2023).

We claim that one-tail p-values of the Taraldsen's density can overcome conventional Pearson correlation's inference inaccuracy based on older Fisher's z-transform. However, it does not matter for reporting practical significance, unless there are theoretical reasons for hypothesizing a nonzero null value in the range $\rho \in [-1, 1]$.

References

- Taraldsen, G. (2023), "The Confidence Density for Correlation," *Sankhya A: The Indian Journal of Statistics*, 85, 600–616.
- Vinod, H. D. (2014), "Matrix Algebra Topics in Statistics and Economics

Using R,” in “Handbook of Statistics: Computational Statistics with R,”
, eds. Rao, M. B. and Rao, C. R., New York: North-Holland, Elsevier
Science, vol. 34, chap. 4, pp. 143–176.

— (2021), *generalCorr: Generalized Correlations and Initial Causal Path*,
Fordham University, New York, R package version 1.2.6, 2023, has 8 vi-
gnettes, URL <https://CRAN.R-project.org/package=generalCorr>.

Vinod, H. D. and López-de-Lacalle, J. (2009), “Maximum Entropy Bootstrap
for Time Series: The meboot R Package,” *Journal of Statistical Software*,
29, 1–19, URL <http://www.jstatsoft.org/v29/i05/>.