ON SOME INFINITE SERIES INTRODUCED BY TSCHUPROW

By J. B. D. Derksen

In his fundamental work on the principles of the theory of correlation Tschuprow introduces some infinite series, leaving certain questions regarding their convergence or divergence unsolved.[1]

As will be shown in the following note, these series are what may be termed randomly divergent,[2] that is series involving random variables which may take on values which will make the series divergent. This result is of importance: e.g. the well-known formula for the standard error of a correlation coefficient $\left(\frac{1-r^2}{\sqrt{N}}\right)$ is the first term of an infinite series for which the question of convergence has not been carefully considered.

Tschuprow finds himself confronted by infinite series, when dealing with the mathematical expectations of quotients as e.g. correlation coefficients or sums of quotients as e.g. the mean square contingency. Let us consider a two-dimensional discontinuous universe, where the variables are x and y. Let p_{ij} be the probability of the occurrence of the pair of values x_i, y_j. The probability that x assumes the value x_i equals $\sum_{j=1}^{l} p_{ij} = p_{i|}$ $(i = 1, \cdots k; j = 1, \cdots l)$. When taking a sample of N pairs of observations (x, y) the relative frequency of x_i will be $p'_{i|}$, and that of the pair (x_i, y_j) will be p'_{ij}. In accordance with Tschuprow we put $p'_{ij} = p_{ij} + dp_{ij}$ and $p'_{i|} = p_{i|} + dp_{i|}$, where dp_{ij} and $dp_{i|}$ are random variables.

As one of the simplest cases we consider the mathematical expectation of

$$\left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \frac{p_{ij}^2}{p_{i|}^2}\left(1 + \frac{dp_{ij}}{p_{ij}}\right)^2 \cdot \left(1 + \frac{dp_{i|}}{p_{i|}}\right)^{-2}.$$

Now Tschuprow develops the last factor in an infinite binomial series, getting

$$\left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \frac{p_{ij}^2}{p_{i|}^2}\left\{1 + 2\frac{dp_{ij}}{p_{ij}} + \left(\frac{dp_{ij}}{p_{ij}}\right)^2\right\} \cdot \left\{1 - 2\frac{dp_{i|}}{p_{i|}} + 3\frac{dp_{i|}^2}{p_{i|}^2} \cdots\right\}$$

$$= \frac{p_{ij}^2}{p_{i|}^2}\left[1 + 2\frac{dp_{ij}}{p_{ij}} - 2\frac{dp_{i|}}{p_{i|}} + \frac{dp_{ij}^2}{p_{ij}^2} - 4\frac{dp_{i|} \cdot dp_{ij}}{p_{i|} \cdot p_{ij}} + 3\frac{dp_{i|}^2}{p_{i|}^2} \cdots\right]. \tag{1}$$

He has given general formulae (*Biometrika*, vol. XII, p. 194 (1919)) from which the mathematical expectations of the terms of this infinite series may immediately be found. We get an infinite series containing ascending powers of N

[1] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*, Leipzig-Berlin, 1925, p. 85-97. An English translation was prepared by M. Kantorowitsch (*Principles of the Theory of Correlation*, W. Hodge & Co., 1939).

[2] Cf. my *Inleiding tot de correlatierekening*, Delft, 1935, p. 88–90.

in the denominator. Finally the convergence or divergence of this series has to be investigated, the problem left unsolved by Tschuprow.

The series expansion of $\left(1 + \frac{dp_{i|}}{p_{i|}}\right)^{-2}$ however, diverges for values $dp_{i|}$ such that $\left|\frac{dp_{i|}}{p_{i|}}\right| \geq 1$ and converges only if

$$\left|\frac{dp_{i|}}{p_{i|}}\right| < 1. \tag{2}$$

This result is not affected by the procedure of the determination of mathematical expectations. For if $f(p'_{i|}, p'_{ij})$ is the probability distribution of $(p'_{i|}, p'_{ij})$, then we have to multiply (1) by this function and to sum for all possible values of $p'_{i|}, p'_{ij}$. As the expressions

$$f(p'_{i|}, p'_{ij}) \frac{p^2_{ij}}{p^2_{i|}} \left(1 + \frac{dp_{ij}}{p_{ij}}\right)^{-2}$$

are always positive, the infinite series, which results from replacing the terms of (1) by their mathematical expectations, will also be divergent.

The same argument is true, when we consider for instance the mathematical expectation of the Pearson-Bravais correlation coefficient. Denoting by μ_{11}, μ_{20}, μ_{02} the population values of the product moment and the second order moment of x and y, and by μ'_{11}, μ'_{20}, μ'_{02} the values observed in a sample, the mathematical expectation of the correlation coefficient may be found from

$$E(r) = E\left[\frac{\mu'_{11}}{(\mu'_{20}\,\mu'_{02})^{\frac{1}{2}}}\right] = E\left[\frac{\mu_{11} + d\mu_{11}}{(\mu_{20} + d\mu_{20})^{\frac{1}{2}}(\mu_{02} + d\mu_{02})^{\frac{1}{2}}}\right]$$

$$= \frac{\mu_{11}}{(\mu_{20}\mu_{02})^{\frac{1}{2}}} E\left[\frac{1 + \frac{d\mu_{11}}{\mu_{11}}}{\left(1 + \frac{d\mu_{20}}{\mu_{20}}\right)^{\frac{1}{2}}\left(1 + \frac{d\mu_{02}}{\mu_{02}}\right)^{\frac{1}{2}}}\right]$$

where $d\mu_{11}$, $d\mu_{20}$, and $d\mu_{02}$ are random variables. Tschuprow expands the denominator in binomial series. However if $d\mu_{20}$ and $d\mu_{02}$ take on values such that $\left|\frac{d\mu_{20}}{\mu_{20}}\right| \geq 1$ or $\left|\frac{d\mu_{02}}{\mu_{02}}\right| \geq 1$, these series will again be divergent. Analogous difficulties arise in all other cases, where Tschuprow makes use of binomial expansions.

It should also be remarked that the well-known formulae, given by the Biometric School for the standard errors of regression and correlation coefficients are equal to the mathematical expectation of the first terms of infinite series, which, as explained above, are divergent for certain values of the random variables. Therefore the question arises as to what effect the divergence for some of the values of the random variables has on those formulae.

This question can be cleared up by the introduction of Slutsky's *conditionally aleatory* variables.[3] These are defined as follows. Suppose that an aleatory variable z can assume the values $z_1, z_2, \cdots z_n$, with probabilities $p_1, p_2, \cdots p_n$. Now we put some of these probabilities equal to 0, dividing the remaining ones by $1 - Q$, if Q represents the total of all the reduced probabilities. The variable z then becomes the so-called conditionally aleatory variable z'. Moreover we assume that z converges stochastically to some limit. Then Slutsky has shown, that if Q converges to 0, z' will converge to the same stochastical limit as z. Moreover the ratio of corresponding moments of the distributions of z and z' will tend to unity.

Now let us consider for example

$$z = \left(\frac{p'_{ij}}{p'_{i|}}\right)^2 = \left(\frac{p_{ij} + dp_{ij}}{p_{i|} + dp_{i|}}\right)^2.$$

Omitting the values for which $| dp_{i|} | \geq p_{i|}$, we get a conditionally aleatory variable z' instead of z. However, according to the theorem mentioned before z' and z will converge to the same stochastical limit, since the probability that $| dp_{i|} | \geq p_{i|}$ converges stochastically to zero as the number of observations increases indefinitely.

In the same way we consider

$$r = \frac{\mu'_{11}}{(\mu'_{20}\mu'_{02})^{\frac{1}{2}}} = \frac{\mu_{11} + d\mu_{11}}{(\mu_{20} + d\mu_{20})^{\frac{1}{2}}(\mu_{02} + d\mu_{02})^{\frac{1}{2}}}$$

and omit those values for which $| d\mu_{20} | \geq \mu_{20}$ and $| d\mu_{02} | \geq \mu_{02}$.

If now we consider the binomial expansions for the conditionally aleatory variables and determine the mathematical expectations of the terms, these new series will converge. All terms in these convergent series will be smaller than the corresponding terms in Tschuprow's series, because we have omitted the larger values of the dp's and the $d\mu$'s. However if the number of observations increases indefinitely the ratios between corresponding terms tend to unity, because the probabilities, that e.g. $| d\mu_{20} | \geq \mu_{20}$ or $| d\mu_{02} | \geq \mu_{02}$ converge to zero.

Let us now turn again to the infinite series given by Tschuprow (loc. cit. p. 90) for the square of the standard error of a correlation coefficient.

$$\sigma_r^2 = E(r - E(r))^2 = \frac{t_1}{N} + \frac{t_2}{N^2} + \frac{t_3}{N^3} + \cdots \tag{3}$$

Here $t_1, t_2, t_3 \cdots$ represent rather lengthy expressions, for which we may refer to Tschuprow's book (loc. cit. p. 88–90). As we have seen before, this series is randomly divergent. However, by introducing a conditionally aleatory variable in the way described above, expanding it into an infinite series and

---

[3] E. Slutsky, "Uber stochastische Asymptoten und Grenzwerte," *Metron* 1925, Vol. V. No. 3, p. 79. Also my *Inleiding tot de correlatierekening*, Ch. V. and VI.

determining the mathematical expectations of its terms, we get a convergent series, say:

$$\sigma_{r'}'^2 = \frac{t_1'}{N} + \frac{t_2'}{N^2} + \frac{t_3'}{N^3} + \cdots \tag{4}$$

From Slutsky's theorem, mentioned before, it follows that if N increases the ratio of σ_r^2 and $\sigma_{r'}'^2$ will tend to unity. Moreover, if we take N sufficiently large, it will always be possible to fulfill the following inequalities:

$$\left|\frac{t_k'}{t_k}\right| > 1 - \epsilon_k \qquad (k = 1, 2, \cdots n)$$

where ϵ_k $(k = 1, 2, \cdots n)$ and n are arbitrary. Therefore, when n and N are sufficiently large the ratio between the first n terms of the infinite series (3) and the true value of σ_r^2 will differ from 1 by an arbitrary small number. Though the series (3) is divergent for any N, however large, the first n terms of this series will give an approximation of σ_r^2 by taking N sufficiently large.

In this paper we have shown that the procedures which have been followed by the Biometric School and Tschuprow to establish formulas for the standard errors of correlation and regression coefficients and in analogous problems can be made rigorous by the use of conditionally aleatory variables. It was found that their infinite expansions are divergent for some of the values of the random variables involved, however large the number of observations (N) may be. Yet it could be demonstrated, that the first n terms of these series will give an approximation, as close as is wanted, if N is sufficiently large. For practical purposes the case $n = 1$ is the most important.

NETHERLANDS CENTRAL BUREAU OF STATISTICS,
THE HAGUE

A NOTE ON FIDUCIAL INFERENCE

BY R. A. FISHER

In a recent paper [1] Bartlett has written a further justification of his criticism of the test of significance for the difference between means of two samples from normal populations not supposedly of equal or related variance. This test was originally put forward by W. V. Behrens [2], and later [3] found to be very simply derivable by the method of fiducial probability.

It is unfortunate that Bartlett did not restate his own views on this topic without making misleading allusions to mine. Thus, on p. 135 in [1]:

"It is sufficient to note that the distribution certainly provides us with an exact inference of fiducial type, as Fisher himself confirmed [9], p. 375."

I do now know, and Bartlett does not specify, what unguarded statement of mine could be used to justify this assertion. From the time I first introduced