An Interesting Application of Fatou's Lemma

GORDON SIMONS*

Let X and Y be any two random variables (defined on the same probability space) for which the expectation $E(X - Y)$ exists (i.e., $E(X - Y)^+ < \infty$ and/or $E(X - Y)^- < \infty$). In either of the following two situations, it seems intuitively obvious that $E(X - Y)$ should be equal to zero.

(a) X and Y are identically distributed random variables.
(b) X and Y are random variables symmetrically distributed about the same point.

Of course the conclusion $E(X - Y) = 0$ is obvious, in either situation, if EX and EY exist and are finite. The conclusion is true even if EX and EY fail to exist. It should be pointed out that neither condition implies that $X - Y$ is symmetric about zero as the following counter-example, a table of $\Pr(X = x, Y = y)$ values, demonstrates:

$y \backslash x$	-1	0	1
-1	0	$\frac{1}{3}$	0
0	0	0	$\frac{1}{3}$
1	$\frac{1}{3}$	0	0

The proof given below, while indirect, is quite elementary and appears to have pedagogical value. Most students' first exposure to the use of "truncation" occurs in a more complicated setting. Below, the power of Fatou's lemma is illustrated within an elementary context.

Proof. For any random variable Z and constant $c > 0$, let Z^c denote the truncated random variable given by

$$\begin{aligned} Z^c &= -c && \text{if} \quad Z < -c \\ &= Z && \text{if} \quad -c \leq Z \leq c \\ &= c && \text{if} \quad Z > c. \end{aligned}$$

Then $(X^c - Y^c)^+ \leq (X - Y)^+$ and $(X^c - Y^c)^- \leq (X - Y)^-$.

Now suppose $E(X - Y)$ exists. For definiteness, assume $E(X - Y)^+ < \infty$. Then, under (a) or (b), $E(X^c - Y^c) = 0$ and

$$E(X^c - Y^c)^- = E(X^c - Y^c)^+ \leq E(X - Y)^+ < \infty.$$

It follows from letting $c \to \infty$ and Fatou's lemma that

$$E(X - Y)^- \leq E(X - Y)^+ < \infty. \tag{1}$$

Since $E(X - Y)^- < \infty$, it follows as above that

$$E(X - Y)^+ \leq E(X - Y)^-. \tag{2}$$

The conclusion $E(X - Y) = 0$ follows from (1) and (2).

Finally, observe that the particular method of truncation used is integral to the proof. The more conventional method of truncation which would set $Z^c = 0$ when $|Z| > c$ will not work.

Acknowledgment: The author wishes to express his appreciation to the Associate Editor for suggesting some improvements which appear in this final version.

* Dept. of Statistics, Univ. of North Carolina, Chapel Hill, NC 27514. This research was sponsored by the Air Force Office of Scientific Research under Grant AFOSR 75-2296.

On Conditional and Partial Correlation

A. J. LAWRANCE*

In this article attention is focused on the difference between *conditional* and *partial* correlation. The *conditional correlation* between the random variables U and V given that a third random variable W is fixed, will be defined as

$$\rho_{U,V|W} = \operatorname{Corr}(U, V|W) \tag{1}$$

$$= \frac{E_{U,V|W}\{[U - E(U|W)][V - E(V|W)]\}}{[E_{U|W}\{[U - E(U|W)]^2\}E_{V|W}\{[V - E(V|W)]^2\}]^{1/2}}.$$

Here $E_{U|W}$, $E_{V|W}$ and $E_{U,V|W}$ denote expectations with respect to the marginal distributions and joint distribution of U and V conditional on W. We give results concerning the relation between the conditional and partial correlation of U and V given W when the regressions of U and V on W are linear. Summarized, these are as follows: (I) That the conditional correlation of U and V given W is not necessarily free of W and hence cannot in general equal the partial correlation of U and V on W; (II) that the two correlations

* Dept. of Mathematical Statistics, The University of Birmingham, P.O. Box 363, Birmingham B15 2TT, England.

are equal when the conditional variances and covariance of U and V given W are free of W; (III) that when the conditional correlation is non-zero and free of W, then the ratio of the partial to the conditional correlation is less than or equal to one. The result (I) contradicts a passing remark of Fleiss and Tanur (1971) in this periodical, that if the expectations of U and V conditional on W are linear, then the conditional and partial correlations are equal; in the examples of Fleiss and Tanur the two correlations were equal.

Generalizing a construction of Fleiss and Tanur, we first consider the conditional correlation (1) when U, V and W are related by the equations $U = A + BW$ and $V = C + DW$ in which A, B, C and D are random variables distributed independently of W. Then we have the linear regressions $E(U|W) = E(A) + E(B)W$ and $E(V|W) = E(C) + E(D)W$. In a straightforward manner we can obtain the conditional variances and covariances of U and V, which can be at most quadratic in W; (1) becomes

$$\rho_{U,V|W} = \frac{\operatorname{Cov}(A, C) + \{\operatorname{Cov}(A, D) + \operatorname{Cov}(B, C)\}W + \operatorname{Cov}(B, D)W^2}{[\{\operatorname{Var}(A) + 2\operatorname{Cov}(A, B)W + \operatorname{Var}(B)W^2\} \cdot\{\operatorname{Var}(C) + 2\operatorname{Cov}(C, D)W + \operatorname{Var}(D)W^2\}]^{1/2}}. \tag{2}$$

When B and D or A and C are constants, we obtain examples of the types used by Fleiss and Tanur in which the conditional correlations are free of W, having the values $\operatorname{Corr}(A, C)$ and $\operatorname{Corr}(B, D)$ respectively. Our generalization of the Fleiss and Tanur construction will be used as a counter example to the equality of conditional and partial correlation.

Result I. The conditional correlation of random variables U and V given W, whose regressions on W are linear, is not necessarily free of W, and so cannot in general equal the partial correlation.

Proof. Suppose in the generalized construction that A, B, C and D have equal variances and are equally correlated with coefficient ρ; then (2) becomes

$$\rho_{U,V|W} = \rho(1 + W)^2/(1 + 2\rho W + W^2). \tag{3}$$

This is a function of W unless $\rho = 1$. It follows that this correlation cannot be equal to the partial correlation of U, V on W, which is by definition free of W.

Example. The random variables U, V and W have the trivariate gamma distribution with joint moment generating function

$$E\{e^{t_1U+t_2V+t_3W}\} = [(1 - t_1 - t_2 - t_3)(1 - t_1)(1 - t_2)(1 - t_3)]^{-1}.$$

They thus have gamma marginal distributions; general multivariate gamma distributions of this type are discussed by Johnson and Kotz (1972, p. 217); from their results we have

$$E(U|W) = 1 + \frac{1}{2}W, \quad \operatorname{Var}(U|W) = 1 + \frac{1}{12}W^2.$$

The distribution thus has linear regressions and quadratic variances and we now calculate its conditional correlation. The moment $E(UV|W)$ can be evaluated in four steps from the joint moment generating function: (1) Differentiate with respect to t_1 and t_2; (2) set $t_1 = t_2 = 0$; (3) invert with respect to t_3; and (4) divide by the marginal density function we^{-w} of W. It is then found that

$$E(UV|W) = \frac{1}{3}W^2 + W + 1,$$

giving the conditional correlation as

$$\rho_{U,V|W} = W^2/(W^2 + 12).$$

This certainly is never free of W, even though the regressions are linear.

We next give a brief explanation of partial correlation. The *partial correlation* $\rho_{U,V.W}$ of U and V on W is by definition (c.f. Cramer (1946, Section 23.4) for instance) the ordinary correlation between the 'residual' variables $U_r = U - \alpha - \beta(W - \mu)$ and $V_r = V - \gamma - \delta(W - \mu)$ where $\mu = E(W)$, $\alpha = E(U)$, $\beta = \operatorname{Cov}(U, W)/\operatorname{Var}(W)$, $\gamma = E(V)$ and $\delta = \operatorname{Cov}(V, W)/\operatorname{Var}(W)$. These values of α, β, γ and δ are those which minimize $E_{U,W}\{U - \alpha - \beta(W - \mu)\}^2$ and $E_{V,W}\{V - \gamma - \delta(W - \mu)\}^2$, and so U_r and V_r are the variables U and V after their linear dependence on W has been removed. The above definition of partial correlation implies the definitions of partial variances and partial covariance of U and V as the ordinary variances and covariance of U_r and V_r; these will be denoted by $\operatorname{Pavar}(U.W)$, $\operatorname{Pavar}(V.W)$ and $\operatorname{Pacov}(U,V.W)$, respectively. Their calculations lead to the well known formula

$$\rho_{U,V.W} = \frac{\rho_{UV} - \rho_{UW}\rho_{VW}}{[(1 - \rho_{UW}{}^2)(1 - \rho_{VW}{}^2)]^{1/2}} \tag{4}$$

in terms of the ordinary pairwise correlations between U, V and W; this is often used as the definition of partial correlation. We shall now obtain some simple connections between partial and conditional correlation when the regressions are linear.

Results II. If $E(U|W) = \alpha + \beta(W - \mu)$, $E(V|W) = \gamma + \delta(W - \mu)$, where α, β, γ, and δ are constants, then

(i) $E_W\{\operatorname{Var}(U|W)\} = \operatorname{Pavar}(U, W)$, $E_W\{\operatorname{Var}(V|W)\} = \operatorname{Pavar}(V, W)$;

(ii) $E_W\{\operatorname{Cov}(U, V|W)\} = \operatorname{Pacov}(U,V.W)$, where E_W denotes expectation with respect to the marginal distribution of W;

(iii) the conditional correlation (1) equals the partial correlation (4) when the conditional variances and covariance of U and V given W are free of W.

Proof. It will be evident that $\alpha = E(U)$, $\beta = \operatorname{Cov}(U, W)/\operatorname{Var}(W)$, $\gamma = E(V)$ and $\delta = \operatorname{Cov}(V, W)/\operatorname{Var}(W)$. In (i) the conditional variance of U given W is

$$\operatorname{Var}(U|W) = E_{U|W}[\{U - E(U|W)\}^2] \tag{5}$$
$$= E_{U|W}[\{U - \alpha - \beta(W - \mu)\}^2],$$

and the last equality holds only when the regression is linear. Now the partial variance is by definition

$$\operatorname{Pavar}(U.W) = E_{U,W}[\{U - \alpha - \beta(W - \mu)\}^2] \tag{6}$$

and there is equality of the terms inside the conditional expectation of (5) and the joint expectation of (6). It is this which allows us to take expectations of (5) with respect to W and to see that

$$E_W\{\operatorname{Var}(U|W)\} = \operatorname{Pavar}(U.W) \tag{7}$$

and similarly for the second equation in (i). We proceed similarly for (ii). Since

$$\operatorname{Cov}(U, V|W) \tag{8}$$
$$= E_{U,V|W}[\{U - E(U|W)\}\{V - E(V|W)\}]$$
$$= E_{U,V|W}[\{U - \alpha - \beta(W - \mu)\} \cdot\{V - \gamma - \delta(W - \mu)\}]$$

is the conditional covariance and $E_{U,V,W}[\{U - \alpha - \beta(W - \mu)\}\{V - \gamma - \delta(W - \mu)\}]$ is the partial covariance, it follows that $E_W\{\operatorname{Cov}(U, V|W)\} = \operatorname{Pacov}(U, V.W)$. The principal conclusion (iii) follows by inspection.

Examples. The most well known distribution satisfying the assumptions of Result II(iii) is the trivariate normal distribution; then in the standardized case, the conditional distribution of U and V is of the bivariate normal form with means $\rho_{UW}W$, $\rho_{VW}W$ and variances $1 - \rho_{UW}^2$, $1 - \rho_{VW}^2$ free of W, and with conditional correlation (1) equal to the partial correlation (4). The Fleiss and Tanur construction when B and D are constants gives other trivariate distributions, such as a trivariate Poisson distribution with linear regressions and constant variances for which the two correlations are equal. Finally we obtain some more general connections between partial and conditional correlation.

Results III. If $E(U|W) = \alpha + \beta(W - \mu)$, $E(V|W) = \gamma + \delta(W - \mu)$ where α, β, γ and δ are constants, then

(i) $$\rho_{U,V.W} = \frac{E_W\{\rho_{U,V|W}[\operatorname{Var}(U|W)\ \operatorname{Var}(V|W)]^{1/2}\}}{[E_W\{\operatorname{Var}(U|W)\}E_W\{\operatorname{Var}(V|W)\}]^{1/2}};$$

(ii) $$\frac{\rho_{U,V.W}}{\rho_{U,V|W}} = \frac{E_W\{[\operatorname{Var}(U|W)\operatorname{Var}(V|W)]^{1/2}\}}{[E_W\{\operatorname{Var}(U|W)\}E_W\{\operatorname{Var}(V|W)\}]^{1/2}}$$
under the assumption that the conditional correlation $\rho_{U,V|W}$ is non-zero and free of W;

(iii) $\rho_{U,V.W}/\rho_{U,V|W} \leq 1$ under the assumptions in (ii);

(iv) $\rho_{U,V.W} = \rho_{U,V|W}$ under the assumptions in (ii) when

$$E_W\{[\operatorname{Var}(U|W)\ \operatorname{Var}(V|W)]^{1/2}\} = [E_W\{\operatorname{Var}(U|W)\}E_W\{\operatorname{Var}(V|W)\}]^{1/2}.$$

Proof. To see that (i) is true we have only to note that $\operatorname{Cov}(U, V|W) = \rho_{U,V|W}[\operatorname{Var}(U|W)\ \operatorname{Var}(V|W)]^{1/2}$. Taking expectations with respect to W and using Results II(ii) we have $\operatorname{Pacov}(U,V.W) = E_W\{\rho_{U,V|W}[\operatorname{Var}(U|W)\ \operatorname{Var}(V|W)]^{1/2}\}$. Division of both sides by the square roots of the partial variances and Result II(i) gives the required result. To prove (ii) we merely take $\rho_{U,V|W}$ out of the expectation. The inequality in (iii) follows from a simple application of Schwarz's inequality and is intuitively plausible, while the equality in (iv) is immediate from (ii). The result (iv) will incidentally be true, for instance, whenever the conditional variances are equal or proportional, no matter the actual type of dependence on W.

Examples. An example of Result III(ii) when the conditional correlation (free of W) and partial correlation are not equal can be obtained from the generalized Fleiss and Tanur construction when A and D are constant and W is a positive random variable. In this case,

$$\operatorname{Cov}(U, V|W) = \operatorname{Cov}(B, C)W$$
$$\operatorname{Var}(U|W) = \operatorname{Var}(B)W^2$$
$$\operatorname{Var}(V|W) = \operatorname{Var}(C)$$

and so

$$\rho_{U,V|W} = \operatorname{Corr}(B, C)$$

which is free of W; further from Result III(ii),

$$\rho_{U,V.W}/\rho_{U,V|W} = E(W)/[E(W^2)]^{1/2},$$

which is less than one. A particular case is when A and D are zero, B and C have any joint log normal distribution, and W has a log normal distribution independent of B and C. Then U, V and W have a trivariate log normal distribution with the required property; the regression on W of the variable V is however free of W. We next give three examples of Result III(iv) when the conditional variances and covariance are not free of W.

The random variables U, V and W have the trinomial distribution with joint probability generating function

$$E(\theta_1{}^U\theta_2{}^V\theta_3{}^W) = [p_1\theta_1 + p_2\theta_2 + (1 - p_1 - p_2)\theta_3]^N.$$

The conditional distribution of U and V given W is then a binomial distribution with probability generating function

$$E(\theta_1{}^U\theta_2{}^V|W) = \left[\frac{p_1}{p_1 + p_2}\theta_1 + \frac{p_2}{p_1 + p_2}\theta_2\right]^{N-W}.$$

Hence we have

$$E(U|W) = (N - W)\left(\frac{p_1}{p_1 + p_2}\right),$$

$$\operatorname{Var}(U|W) = (N - W)\left(\frac{p_1}{p_1 + p_2}\right)\left(\frac{p_2}{p_1 + p_2}\right),$$

$$\operatorname{Cov}(U, V|W) = -(N - W)\left(\frac{p_1}{p_1 + p_2}\right)\left(\frac{p_2}{p_1 + p_2}\right)$$

which are all linear in W; the conditional correlation is seen to be -1 and this is obvious anyway since conditional on W, $U + V = N - W$. Use of (4) and the pairwise correlations shows that the partial correlation is also -1.

Another example is the negative trinomial (trivariate negative binomial) with joint probability generating function

$$E(\theta_1{}^U\theta_2{}^V\theta_3{}^W) = [1 + p_1 + p_2 + p_3 - p_1\theta_1 - p_2\theta_2 - p_3\theta_3]^{-N}.$$

The conditional distribution of U, V given W is then the bivariate negative binomial distribution with probability generating function

$$E(\theta_1{}^U\theta_2{}^V|W) = \left[1 + \frac{p_1}{1+p_3} + \frac{p_2}{1+p_3} - \left(\frac{p_1}{1+p_3}\right)\theta_1 - \left(\frac{p_2}{1+p_3}\right)\theta_2\right]^{-(N+W)}.$$

Thus it may be found

$$E(U|W) = (N+W)\left(\frac{p_1}{1+p_3}\right),$$

$$\mathrm{Var}(U|W) = (N+W)\left(\frac{p_1}{1+p_3}\right)\left(1 + \frac{p_1}{1+p_3}\right),$$

$$\mathrm{Corr}(U, V|W) = \left[\frac{p_1p_2}{(1+p_1+p_3)(1+p_2+p_3)}\right]^{1/2}.$$

This is also the value of the corresponding partial correlation as found using (4) and pairwise correlations such as $\mathrm{Corr}(U, V) = [p_1p_2/(1+p_1)(1+p_2)]^{1/2}$.

A final example is the trivariate Pareto distribution with probability density function

$$f_{U,V,W}(u, v, w) = a(a+1)(a+2)(\lambda_1\lambda_2\lambda_3)^{-1} \cdot (\lambda_1{}^{-1}u + \lambda_2{}^{-1}v + \lambda_3{}^{-1}w - 2)^{-(a+3)}$$

over $u > \lambda_1 > 0$, $v > \lambda_2 > 0$, $w > \lambda_3 > 0$. The marginal distributions are of the Pareto form. General multivariate Pareto distributions are described by Johnson and Kotz (1972, p. 285); they include the results

$$E(U|W) = \lambda_1[1 + (\theta_3 a)^{-1}W],$$

$$\mathrm{Var}(U|W) = (\lambda_1/\lambda_3)^2(a+1)a^{-2}(a-1)^{-1}W^2.$$

The distribution thus has linear regressions but quadratic variances. Johnson and Kotz also indicate that the conditional distribution of U and V given W will be of the bivariate Pareto form, and from this we may deduce that the conditional correlation is $(a+1)^{-1}$. Agreement with the corresponding partial correlation is found again.

Acknowledgment. I thank the Associate Editor and referees for their constructive comments.

REFERENCES

Cramer, H.: *Mathematical Methods of Statistics,* Princeton University Press (1946).

Fleiss, J. L. and Tanur, J. M.: "A note on the partial correlation coefficient", *The American Statistician,* 25 (February 1971), 43–45.

Johnson, N. L. and Kotz, S.: *Distributions in Statistics: Continuous Multivariate Distributions,* Wiley, New York (1972).