Cowles Foundation Paper 705

Journal of Econometrics 38 (1988) 341–348 North-Holland

CONDITIONAL AND UNCONDITIONAL STATISTICAL INDEPENDENCE

Peter C.B. PHILLIPS*

Yale University, New Haven, CT 06520, USA

Received March 1987, final version received December 1987

Conditional independence almost everywhere in the space of the conditioning variates does not imply unconditional independence, although it may well imply unconditional independence of certain functions of the variables An example that is important in linear regression theory is discussed in detail This involves orthogonal projections on random linear manifolds, which are conditionally independent but not unconditionally independent under normality Necessary and sufficient conditions are obtained under which conditional independence does imply unconditional independence

1. Introduction

It is often useful in problems of multivariate analysis to work with conditional distributions as far as possible, leaving the integration that is required for unconditional results to the final stage of analysis. Many examples where this approach has been helpful in achieving simplifications have arisen recently in econometric distribution theory Phillips (1984) and Hillier (1985) give some specific instances from simultaneous equations theory where the gains in economy that can be achieved with this approach are rather apparent. The approach is also useful in asymptotic theory, particularly when central limit theorems are inapplicable or difficult to employ. For example, Phillips (1987), Park and Phillips (1986) and Phillips and Park (1986) use the method extensively in asymptotic studies of regression with non-ergodic processes. In these applications the limiting distribution is first expressed as a functional of vector Brownian motion. The conditional distribution of the functional is then derived for a given realization of a subvector of the stochastic process and the unconditional distribution follows by integration with respect to the marginal probability measure of the conditioning variates.

The approach is particularly helpful when selective conditioning ensures statistical independence of key remaining variates. In a series of recent articles

*I am grateful to two referees and to Arnold Zellner for comments on an earlier version of this paper My thanks go also to Glena Ames for her skill and effort in keyboarding the manuscript and to the NSF for research support under Grant no SES 8519595

in the statistical literature, Dawid (1979, 1980, 1985) has made extensive use of the concept of conditional independence, developing a rigorous framework for its use in terms of statistical operations and illustrating its application in diverse areas of statistical inference.

In view of these developments it seems likely that the use of this approach will become more widespread in econometrics. The purpose of this note is to bring attention to the following rather simple (but easily neglected) point. conditional independence almost everywhere in the space of the conditioning variates does not imply unconditional independence. We give examples to illustrate this point. The first is a simple and familiar example to multivariate analysts but is less well known to econometricians. The second example is new. It involves orthogonal projections on random linear manifolds and has interesting consequences in the study of regressions with stochastic regressors (such as models of simultaneous equations). Necessary and sufficient conditions are given under which conditional independence does imply unconditional independence.

2. Mixtures of normals

Let x be a random n-vector and w a random scalar for which $w \geq 0$ and $P(w=0)=0$. Suppose that, given w, $x \equiv N(0, wI)$. The symbol '$\equiv$' here signifies equality in distribution. Let G denote the distribution function of w Then the probability density function (pdf) of x is

$$\mathrm{pdf}(x) = \int_0^\infty N(0, wI)\,\mathrm{d}G(w), \tag{1}$$

a scale mixture of normals. Conditional on w, the elements of x are statistically independent and this holds for all $w > 0$. However, unless G assigns a probability mass of unity to some single point w_0 [in which case $\mathrm{pdf}(x) \equiv N(0, w_0 I)$] the elements of x are not unconditionally independent. Indeed, the only member of the family of distributions (1) for which the elements of x are unconditionally independent is the normal. The same result applies in the somewhat wider family of spherically symmetric distributions as shown, for example, in Muirhead (1982, theorem 1.8.3).

The family of scale mixtures of normals (1) is receiving a growing amount of attention in the econometric literature. Zellner (1976), for example, showed that inferences based on the usual regression t- and F-statistics remain valid in a linear regression model whose errors follow the multivariate distribution (1). Similar robustness results have been demonstrated for other tests by Ullah and Zinde-Walsh (1984) and for predictive inferences by Chib et al (1987) Such robustness is a simple consequence of (i) the scale invariance of the statistic (which holds by construction) and (ii) the fact that, conditional on scale, the component variates of (1) are independent normal It may be said that the unconditional dependence in the component variates of (1) directly compensates for the departures from normality in these studies. Indeed, results for regression models whose errors have the same marginal distributions but are unconditionally independent are very different. This has been shown recently in Phillips and Hajivassiliou (1987). In particular, data from an n-dimensional Cauchy population [whose density is given by (1) with $1/w \equiv \chi_1^2$ and whose components are therefore *conditionally independent*] yield a t-statistic whose distribution is classical student t_{n-1} and the robustness of conventional inference applies. However, if the sample comprises *independent* draws from the same Cauchy (0, 1) marginals then the t-statistic is no longer student t_{n-1}. Its distribution is, in fact, bimodal for all n, including the asymptotic distribution. This example serves to illustrate the major differences that can result from working with conditionally independent rather than unconditionally independent variates in regression

3. Orthogonal projections in regression

Let $x \equiv N(0, I_n)$ and $y \equiv N(0, I_n)$ be independent random n-vectors. Let P_x be the projection matrix onto the range space of x and write $Q_x = I - P_x$. Clearly, $p = P_x y$ and $q = Q_x y$ are conditionally independent given x and this is true for all x. However, p and q are not unconditionally independent To see this we may examine the joint characteristic function (cf) of (p, q) and show that it does not factor. Write $P_x = hh'$ where $h = x/(x'x)^{1/2} \in V_n$, the unit sphere in R^n Then h and y are independent and by iterated expectations we find the joint cf of (p, q) as follows.

$$\begin{aligned}\mathrm{cf}(s, t) &= \mathrm{E}\left[\mathrm{E}\{\exp(is'hh'y + it'(I - hh')y)|h\}\right]\\ &= \mathrm{E}\{\exp(-s'hh's/2 - t'(I - hh')t/2)\}\\ &= \mathrm{e}^{-t't/2}\int_{V_n}\exp\{\tfrac{1}{2}(tt' - ss')hh'\}(\underline{\mathrm{d}h})\\ &= \mathrm{e}^{-t't/2}{}_0F_0^{(n)}\left(\tfrac{1}{2}(tt' - ss'), 1\right). \end{aligned} \tag{2}$$

In the above $(\underline{\mathrm{d}h})$ denotes the normalized invariant measure on the manifold V_n and ${}_0F_0^{(n)}$ is a hypergeometric function with two matrix arguments. Expression (2) has an alternative series representation in the form

$$\mathrm{e}^{-t't/2}\sum_{j=0}^{\infty}(1/j!)C_j\left(\tfrac{1}{2}(tt' - ss')\right)/C_j(I_n), \tag{3}$$

where $C_j(\cdot)$ denotes a top-order zonal polynomial of degree j, for which James (1964) gives explicit formulae. When $n > 1$, $\mathrm{cf}(s, t)$ does not admit a factorization into the marginal characteristic functions. The latter are given by

$$\mathrm{cf}(s) = {}_0F_0^{(n)}\left(-\tfrac{1}{2}ss', 1\right) = {}_1F_1\left(\tfrac{1}{2}, n/2; -\tfrac{1}{2}s's\right), \tag{4}$$

$$\mathrm{cf}(t) = \mathrm{e}^{-t't/2}{}_0F_0^{(n)}\left(\tfrac{1}{2}tt', 1\right) = \mathrm{e}^{-t't/2}{}_1F_1\left(\tfrac{1}{2}, n/2; \tfrac{1}{2}t't\right). \tag{5}$$

When $n = 1$ we have $\mathrm{cf}(s) = \mathrm{e}^{-s^2/2}$, $\mathrm{cf}(t) = 1$ and factorization is immediate since $q = 0$ almost surely. In (4) and (5) ${}_1F_1$ denotes the confluent hypergeometric function.

Remark 1. In this example it is interesting to construct cases where joint probabilities do not factorize into the product of the marginal probabilities. Let $n = 2$, write $h = x/(x'x)^{1/2}$ as before, and define $H = [h, k]$ to be orthogonal. Then

$$p = hh'y = hY,$$

$$q = (I - hh')y = kk'y = kZ.$$

Here Y and Z are independent $N(0,1)$ and both are independent of h (and hence k). Now consider the event $(hY \geq 0, kZ \geq 0)$, i.e., both coordinates of hY and both coordinates of kZ non-negative. Since h is orthogonal to k in the plane R^2 it is clear that this event occurs with probability zero. Thus

$$P(hY \geq 0, kZ \geq 0) = 0. \tag{6}$$

On the other hand it is easy to see that

$$P(hY \geq 0) = \tfrac{1}{4}, \qquad P(kZ \geq 0) = \tfrac{1}{4}, \tag{7}$$

and the joint probability is not equal to the product of the marginals. Note that this example does not depend on the fact that the joint probability (6) is zero. Indeed, the probability measures of h, Y and Z are continuous and the inequality between the joint probability and the product of the marginal probabilities continues as we move away from the zero probability event $(hY \geq 0, kZ \geq 0)$

Remark 2. Even though p and q are dependent, some functions of these projections may be independent. For example, $q'q|_x = y'Q_xy|_x \equiv \chi^2_{n-1}$ and is independent of x so that the unconditional distribution of $q'q$ is also χ^2_{n-1}. In this case, p and $f = q'q$ are conditionally independent as before for all x.

However, the conditional distribution of f is independent of x and this ensures that p and f are unconditionally independent. The same is true of any measurable function of q (or p) that is distributed independently of x. Let $f=f(q)$ be such a function. Then simple manipulations show that the joint cf of (f, p),

$$\begin{aligned} \mathrm{cf}(a, b) &= \mathrm{E}\left(\mathrm{E}\left(\mathrm{e}^{ia'f+ib'p}|x\right)\right) \\ &= \mathrm{E}\left(\mathrm{E}\left(\mathrm{e}^{ia'f}|x\right)\mathrm{E}\left(\mathrm{e}^{ib'p}|x\right)\right) \\ &= \mathrm{E}\left(\mathrm{e}^{ia'f}\right)\mathrm{E}\left(\mathrm{E}\left(\mathrm{e}^{ib'p}|x\right)\right) \\ &= \mathrm{cf}(a)\mathrm{cf}(b), \end{aligned}$$

factors into the marginals.

Remark 3. One interesting application of the observation in the previous remark is to the central Wishart. Let $Z(n \times T, T \geq n)$ be matrix $N_{n,T}(0, I_{nT})$. Then $ZZ' \equiv W_n(T, I_n)$. Now partition Z as

$$Z = \overset{T}{\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}} \begin{matrix} n_1 \\ n_2 \end{matrix}.$$

Then, as given by Muirhead (1982, theorem 3.2.10),

$$W_1 = Z_1Z_1' - Z_1Z_2'(Z_2Z_2')^{-1}Z_2Z_1' = Z_1Q_{Z_2'}Z_1' \equiv W_{n_1}(T - n_2, I_{n_1})$$

is independent of $W_2 = Z_2'(Z_2Z_2')^{-1}Z_2Z_1'$. To see this we simply write $W_1 = Z_1CC'Z_1'$, where C is a $T \times (T - n_2)$ matrix of orthonormal vectors that span the orthogonal complement of the range of Z_2'. Observe that $Z_1C \equiv N_{n_1, T-n_2}(0, I_{n_1(T-n_2)})$ and is independent of C (and hence Z_2). It follows that $W_1 \equiv W_{n_1}(T - n_2, I_{n_1})$ as stated. Moreover, conditional on Z_2, W_1 and W_2 are independent. But, since W_1 is independent of Z_2, the argument of the previous remark applies and we deduce that W_1 and W_2 are unconditionally independent It is easy to deduce that W_1 is also unconditionally independent of the matrices Z_2Z_1' and Z_2Z_2', as given in Muirhead's theorem 3.2.10.

Remark 4. The result of the previous remark no longer applies when W is non-central Wishart. To see this suppose $Z \equiv N_{n,T}(M, I_{nT})$ and partition M into submatrices M_1 and M_2 conformably with Z. Now

$Z_1C \equiv N_{n_1\, T-n_2}(M_1C, I_{n_1(T-n_2)})$ and conditional on Z_2, W_1 is non-central $W_{n_1}(T-n_2, I, \Omega)$ with non-centrality matrix $\Omega = M_1CC'M_1'$. This distribution depends on Z_2 (unless $M_1 = 0$) and, hence, W_1 and W_2 are in general unconditionally dependent although they are conditionally independent.

4. Necessary and sufficient conditions

In Remark 2 above we saw that, if the conditional distribution of f is independent of x, then this is sufficient to ensure that conditional independence of f and p (given x) implies the unconditional independence of f and p. This condition is sufficient but not necessary. To find necessary and sufficient conditions it is helpful to use a more abstract framework.

Let x and y be random elements defined on the probability space $(\Omega, \mathscr{F}, P)$ Let $p = p(x, y)$, $q = q(x, y)$, $f = f(q)$, $g = g(p)$ be measurable functions of (x, y), q and p, respectively. Let $\mathscr{F}_\alpha$ denote the sub-σ-field of $\mathscr{F}$ that is generated by α $(= x, y, p, q, f, g)$ and let $\mathscr{F}_\alpha \vee \mathscr{F}_\beta$ be the smallest sub-σ-field of $\mathscr{F}$ that contains $\mathscr{F}_\alpha$ and $\mathscr{F}_\beta$. Finally, let us suppose that p and q are conditionally independent relative to $\mathscr{F}_x$. This means that for any sets $\Lambda_p \in \mathscr{F}_p$, $\Lambda_q \in \mathscr{F}_q$ we have

$$P(\Lambda_p \cap \Lambda_q | \mathscr{F}_x) = P(\Lambda_p | \mathscr{F}_x) P(\Lambda_q | \mathscr{F}_x).$$

In what follows, we use Λ_α to denote any set in the field $\mathscr{F}_\alpha$ and 1_α to denote the indicator of the set Λ_α Then we have.

Theorem. Conditional independence of f and g relative to $\mathscr{F}_x$ implies unconditional independence of f and g iff

$$\mathrm{E}\big(\mathrm{E}(1_f | \mathscr{F}_x) 1_g\big) = \mathrm{E}(1_f)\mathrm{E}(1_g), \tag{8}$$

and

$$\mathrm{E}\big(\mathrm{E}(1_g | \mathscr{F}_x) 1_f\big) = \mathrm{E}(1_g)\mathrm{E}(1_f). \tag{9}$$

Proof. We first observe that conditional independence of f and g relative to $\mathscr{F}_x$ implies that

$$P(\Lambda_f | \mathscr{F}_g \vee \mathscr{F}_x) = P(\Lambda_f | \mathscr{F}_x) \quad \text{a.s.}$$

and

$$P(\Lambda_g | \mathscr{F}_f \vee \mathscr{F}_x) = P(\Lambda_g | \mathscr{F}_x) \quad \text{a.s.}$$

[see Chung (1974, theorem 9.2.1, p. 307)]. Write these relations in terms of conditional expectations as

$$\mathrm{E}(1_f|\mathscr{F}_g \vee \mathscr{F}_x) = \mathrm{E}(1_f|\mathscr{F}_x) \quad \text{a.s.} \tag{10}$$

$$\mathrm{E}(1_g|\mathscr{F}_f \vee \mathscr{F}_x) = \mathrm{E}(1_g|\mathscr{F}_x) \quad \text{a.s.} \tag{11}$$

Multiply (10) by 1_g, (11) by 1_f and take expectations giving

$$P(\Lambda_g \cap \Lambda_f) = \mathrm{E}(\mathrm{E}(1_f|\mathscr{F}_x)1_g), \tag{12}$$

and

$$P(\Lambda_f \cap \Lambda_g) = \mathrm{E}(\mathrm{E}(1_g|\mathscr{F}_x)1_f). \tag{13}$$

(12) and (13) necessarily hold when f and g are conditionally independent relative to $\mathscr{F}_x$. If f and g are unconditionally independent then (8) and (9) necessarily hold as well, since $P(\Lambda_f \cap \Lambda_g) = \mathrm{E}(1_f 1_g) = \mathrm{E}(1_f)\mathrm{E}(1_g)$ under independence. Conversely, if (8) and (9) hold then we deduce directly from (12) or (13) (and hence from the conditional independence of f and g) that f and g are unconditionally independent. □

Corollary *Sufficient conditions for conditional independence of f and g relative to $\mathscr{F}_x$ to imply unconditional independence are either*

$$\mathrm{E}(1_f|\mathscr{F}_x) = \mathrm{E}(1_f) \quad a.s. \tag{14}$$

or

$$\mathrm{E}(1_g|\mathscr{F}_x) = \mathrm{E}(1_g) \quad a.s. \tag{15}$$

Proof. We need only verify (8) and (9) If (14) holds then (8) follows immediately. To verify (9) note that

$$\begin{aligned}\mathrm{E}(\mathrm{E}(1_g|\mathscr{F}_x)1_f) &= \mathrm{E}(\mathrm{E}(1_g|\mathscr{F}_x)\mathrm{E}(1_f|\mathscr{F}_x)) \\ &= \mathrm{E}(1_f)\mathrm{E}(1_g),\end{aligned}$$

as required In a similar way, (8) and (9) also hold under (15) □

Remark 5 Since $1_f = 1(\Lambda_f)$ and (14) holds for all sets $\Lambda_f \in \mathscr{F}_f$, (14) is equivalent to the statement that the conditional distribution of f is independent of x. Thus, the Corollary gives the condition for unconditional independence discussed earlier in Remark 2 in the context of regression projections [where $g = p$ is conditionally independent of $f = f(q)$ given x and also unconditionally independent of f since f is independent of x].

Remark 6. Theorem and Corollary hold for quite general random elements. Thus, we can take $x = \{x_t\}_{-\infty}^{\infty}$, $y = \{y_t\}_{-\infty}^{\infty}$ as time series and p, q, f and g as measurable functions of them. The outcomes p and q (and hence f and g) may then be conditionally independent given a third factor x but are not necessarily independent when we no longer condition on x. In this context the results obtained seem useful in evaluating whether independence between observed series is due to (possibly implicit) conditioning on other variables. Such considerations are particularly important in tests of causality and feedback between economic time series.

References

Chib, S , R C Tiwari and S Rao Jammalamadaka, 1987, Bayes prediction in regressions with elliptical errors, Journal of Econometrics, forthcoming

Dawid, A P , 1979, Conditional independence in statistical theory, Journal of the Royal Statistical Society B 41, 1–31

Dawid, A P , 1980, Conditional independence for statistical operations, Annals of Statistics 8, 598–617

Dawid, A P , 1985, Invariance and independence in multivariate distribution theory, Journal of Multivariate Analysis 17, 304–315

Hillier, G H , 1985, On the joint and marginal densities of instrumental variable estimators in a general structural equation, Econometric Theory 1, 53–72

James, A T , 1954, Normal multivariate analysis and the orthogonal group, Annals of Mathematical Statistics 25, 40–75

James, A T , 1964, Distribution of matrix variates and latent roots derived from normal samples, Annals of Mathematical Statistics 35, 475–501

Muirhead, R , 1982, Aspects of multivariate statistical analysis (Wiley, New York)

Park, J Y and P C B Phillips, 1986, Statistical inference in regressions with integrated regressors, Part 1, Cowles Foundation discussion paper no 811

Phillips, P C B., 1984, The exact distribution of LIML I, International Economic Review 25, 249–261

Phillips, P C B , 1987, Partially indentified econometric models, Cowles Foundation discussion paper no 845

Phillips, P C B and J Y Park, 1986, Asymptotic equivalence of OLS and GLS in regressions with integrated processes, Cowles Foundation discussion paper no 802 and forthcoming in the Journal of the American Statistical Association

Phillips, P C B and V Hajivassiliou, 1987, Bimodal *t*-ratios, Cowles Foundation discussion paper no 842

Ullah, A and V Zinde-Walsh, 1984, On the robustness of LM, LR and W tests in regression models, Econometrica 52, 1055–1066

Zellner, A , Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms, Journal of the American Statistical Association 71, 400–405