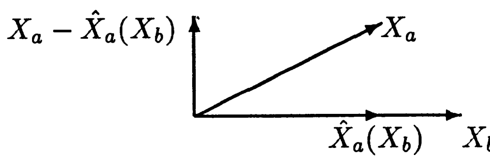
This device of embedding the dashed graph into a CI graph with "latent variables" certainly solves some problems. It also indicates why latent variables in highly structured graphs allow marginal empirical dependences to determine the statistical analysis. A prime example of this is the graphical analysis of the state space model underlying the Kalman filter.

ROLE OF THE PARTIAL VARIANCE (SCHUR COMPLEMENT)

The technical conditions for conditional independence in multivariate normal distributions, for instance, that $X_1 \perp\!\!\!\perp X_2|X_3$ is characterised by a zero in the inverse variance matrix of (X_1, X_2, X_3), appear somewhat bizarre at a first acquaintance. A good understanding requires an interpretation of the elements of this inverse variance matrix, and I found it useful in writing Chapter 5 of my book (Whittaker, 1990) to use the concept of the partial variance as the vehicle for this explanation. For instance, slightly extending the notation of the CW paper, when a vector X with variance Σ is partitioned into (X_a, X_b) the block in the inverse variance Σ^{-1} corresponding to X_a is $\Sigma^{aa} = (\Sigma^{-1})_{aa}$ (and *not* $(\Sigma_{aa})^{-1}$), the essential content of the inverse variance lemma is that

$$(1) \qquad \sum^{aa} = \mathrm{var}(X_a|X_b)^{-1}.$$

Here $\mathrm{var}(X_a|X_b)$ is the partial or residual variance of X_a having regressed out X_b, and defined by $\mathrm{var}(X_a - \hat{X}_a(X_b))$ where $\hat{X}_a(X_b)$ is the fitted (multivariate) regression of X_a on X_b. These entities can be represented in the Pythagorean vector diagram



The notion of a partial variance permits the diagonal elements of the inverse variance matrix to be interpreted as functions of the multiple correlation coefficient: if $a = \{i\}$ is 1-dimensional, so that b denotes the $p - 1$ remaining variables, then (1) becomes

$$\sum^{ii} = \mathrm{var}(X_i|X_{rest})^{-1} = \mathrm{var}(X_i)^{-1} / (1-R^2(i))$$

where $R(i)$ is the multiple correlation coefficient of X_i with the remaining variables. In consequence, the larger Σ^{ii} in relation to $\mathrm{var}(X_i)$ the more predictable is X_i from the other variables. By choosing $a = \{i, j\}$ to be 2-dimensional, formula (1) enables an explicit expression for the off-diagonal elements of the inverse variance in terms of the partial correlation of X_i and X_j given the remaining variables. In point of fact $\Sigma^{ij} / \sqrt{\Sigma^{ii}\Sigma^{jj}} = -\mathrm{corr}\,(X_i, X_j|X_{rest})$.

The inverse variance lemma, which is by no means new, is really just statistical interpretation of inverting a partitioned matrix. In fact $\mathrm{var}(X_a|X_b)$ can be computed from $\mathrm{var}(X_a) - \mathrm{cov}(X_a, X_b)\mathrm{var}(X_b)^{-1}\mathrm{cov}(X_b, X_a)$ which in the mathematical literature is well known as the Schur complement of the matrix

$$\begin{bmatrix} \mathrm{var}(X_a) & \mathrm{cov}(X_a, X_b) \\ \mathrm{cov}(X_b, X_a) & \mathrm{var}(X_b) \end{bmatrix}.$$

The determinant represents the squared length (volume) of the residual vector in the Pythagorean vector diagram above. This quantity is denoted by $\Sigma_{aa|b}$ in CW as in many books on the multivariate normal distribution, but such a notation obscures various elementary properties such as $\mathrm{var}(AX_a|X_b) = A\mathrm{var}(X_a|X_b)A'$ where A is a fixed linear transform, and if B is invertible, $\mathrm{var}(X_a|BX_b) = \mathrm{var}(X_a|X_b)$ expressing the invariance of the partial variance to a change of units in the regressor variables.

Various forms of the lemma exist and a frequent application is to Bayesian analysis for instance, in the analysis of linear models by Lindley and Smith (1972), in standard treatments of factor analysis, and in Kalman filtering.

Rejoinder

D. R. Cox and Nanny Wermuth

We are grateful to all the contributors for their thoughtful and constructive contributions. There is rather little with which we disagree so that our reply is brief.

While to some extent the use of the word *causal* is a matter of convention, we much prefer to restrict the word to situations in which we have knowledge of some underlying process. We reassure Dempster that we are deeply concerned with the elucidation of processes that might have generated the data, but are cautious about what conclusions can be drawn from single investigations or even repeated investigations, especially but