More on conditional independence

Do the 2 examples from Whittaker
`http://www.statlab.uni-heidelberg.de/people/eichler/graphs/graph3.pdf`

Independence graphs (Whittaker)

Def: Let $X = (X_1, \ldots., X_k)$ be a k-tuple of random vectors. A graph G with k nodes is called the independence graph of X if non-adjacent pairs of variables are independent conditional on the remaining variables.
This condition is called the pairwise Markov property.
Compare to a Markov chain, where the future is independent of the past conditional on the present.

It is equivalent to either of the following two statements:

Local Markov property: Any variable is independent of all remaining variables conditional only on its adjacent variables.

Global Markov property: Any two subsets of variables separated by a third set of variables, are independent conditional only on the variables in the third set.

Partial variance, covariance, correlation

Now, suppose $X_1, \ldots, X_k$ are random vectors. Let's look at the independence graph of the X_i vectors. When an entry of the Σ is 0, this only means that I and j are independent.
Let all X_i be normalized (i.e., the mean was subtracted). Then

$$var(X_i) = X^T * X$$

and

$$cov(X_i, X_j) = X_i^T * X_j$$

We study a prediction problem: Estimate Y from X: $Y \approx b^T X$.
Define partial variance from geometry (board).

Partial covariance

We want to define the partial covariance of Y,Z given X. Let $\hat{Y}(X)$ be the estimate of Y from X and $\hat{Z}(X)$ the estimate of Z from X. The partial covariance $cov(Y, Z|X)$ is defined as

$$cov(Y - \hat{Y}(X), Z - \hat{Z}(X))$$

.

The partial correlation coefficient is its scaled version:

$$\rho_{Y,Z|X} = \frac{cov(Y, Z|X)}{\sqrt{var(Y - \hat{Y}(X))var(Z - \hat{Z}(X))}}$$

Relation to inverse of the covariance matrix

The inverse of the covariance matrix is called precision matrix or concentration matrix.

The (i, j) entry of the inverse of Σ is 0 exactly if the partial covariance of X_i and X_j with respect to the rest is 0.

In fact, this holds true for any partition into three subsets - block diagonal structure of the inverse covariance matrix.

Gaussian Graphical Models

When the RVs in the nodes of the graph are the marginals of a Gaussian with covariance matrix Σ, then the entries of Σ^{-1}, which are 0, correspond exactly to the edges not contained in the independence graph. We call this a Gaussian Graphical Model.
In fact, the missing edges denote exactly the conditional independencies.
So it is "easy" to compute the independence graph, because the pairwise Markov property is fulfilled.
We'll say more about "easy" later, because this matrix inversion is not always so easy.

Partial correlation coefficient

The partial correlation coefficient between X_1 and X_2 is the correlation coefficient of the residual vectors arising from regression of X_1 vs the $X_3, \ldots, X_K$, and the residual vectors arising from regressing X_2 vs $X_3, \ldots, X_k$.

Partial correlation coefficients can be computed from the entries of the inverse of the variance-covariance matrix: Theorem: Let $X_1, \ldots ., X_k$ be random vectors and $P = \Sigma^{-1}$.

$$\rho_{i,j|rest} = -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}}$$

Covariance matrix is the product of (normalized) datamatrix times its transpose.
Depending on the shape (rank) of the datamatrix, the product will be singular. How to invert a singular matrix: (remember the normal equation from least squares: $Y^T X = b^T X^T X$
Singular value decomposition of matrix A , of shape $n \times m$:

$$A = UWV^T$$

U is $m \times m$ and contains the eigenvectors of AA^T.
W is a diagonal matrix and contains the singular values. They are the square roots of the eigenvalues of AA^T (which are also the eigenvalues of $A^T A$).
V is $n \times n$.
Pseudoinverse: $A^+ = V^T W^{-1} U$ The trick is that the singular values on Ws diagonal are sorted and the last ones are 0 or near-zero. The inversion is done by only taking the inverse for the the singular values that are not 0.
See SVD function in R.

Shrinkage methods: Strimmers method

See R package corpcor. Estimate a b̈etterc̈ovariance matrix. Mix the covariance matrix with another matrix (e.g., for a multivariate Gaussian, with little weight off-diagonal). This should make it invertible and not change reality too much. See Schäfer and Strimmer, 2005.