CH5350: Applied Time-Series Analysis

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Probability & Statistics: Bivariate / Multivariate case

Multivariate analysis

In the analysis of several statistical events and random signals we will be required to analyze two or more variables simultaneously. Of particular interest would be to examine the presence of linear dependencies, develop linear models and predict one RV using another set of random variables.

We shall primarily study bivariate analysis, i.e., analysis of a pair of random variables.

Bivariate analysis and joint p.d.f.

Where bivariate analysis is concerned, we start to think of **joint probability density functions**, i.e., the probability of two RVs taking on values within a rectangular cell in the 2-D real space.

Examples:

- Height and weight of an individual
- Temperature and pressure of a gas

Joint density

Consider two continuous-valued RVs X and Y. The probability that these variables take on values in a rectangular cell is given by the **joint density**

$$Pr(x_1 \leq x \leq x_2, y_1 \leq y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)\,dx\,dy$$

Joint Gaussian density

The joint Gaussian density function of two RVs is given by

$$f(x,y) = \frac{1}{2\pi|\Sigma_{\mathbf{Z}}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{Z}-\boldsymbol{\mu}_Z)^T \Sigma_Z^{-1}(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})\right) \tag{1}$$

where

$$\mathbf{Z} = \begin{bmatrix} X \\ Y \end{bmatrix} \qquad \Sigma_{\mathbf{Z}} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \tag{2}$$

The quantity σ_{XY} is known as the **covariance** and $\Sigma_{\mathbf{Z}}$ the **variance-covariance** matrix.

Marginal and conditional p.d.f.s

Associated with this joint probability (density function), we can ask two questions:

1. What is the probability $\Pr(x_1 \leq X \leq x_2)$ regardless of the outcome of Y and vice versa? (**marginal density**)
2. What is the probability $\Pr(x_1 \leq X \leq x_2)$ given Y has occurred and taken on a value $Y = y$? (**conditional density**)
   - Strictly speaking, one cannot talk of Y taking on an exact value, but only of values within an infinitesimal neighbourhood of y.

Marginal density

The marginal density is arrived at by walking across the outcome space of the "free" variable and adding up the probabilities of the free variable within infinitesimal intervals.

The **marginal density** of a RV X with respect to another RV Y is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \tag{3}$$

Likewise,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \tag{4}$$

Conditional density and Expectation

The conditional density is used in evaluating the probability of outcomes of an event given the outcome of another event

Example: What is the probability that Rahul will carry an umbrella given that it is raining?

Conditional density

The conditional density of Y given $X = x$ (strictly, between x and $x + dx$) is

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f(x)} \tag{5}$$

Conditional Expectation

In several situations we are interested in "predicting" the outcome of one phenomenon given the outcome of another phenomenon.

Conditional expectation

The conditional expectation of Y given $X = x$ is

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)\, dy \tag{6}$$

Conditional expectation and prediction

The conditional expectation of Y given $X = x$ is a function of x, *i.e.*,

$$\boxed{E(Y|X = x) = \phi(x)}$$

The conditional expectation is the best predictor of Y given X among all the predictors that minimize the mean square prediction error

---

Note: We shall prove this result later in the lecture on prediction theory.

Iterative expectation

A useful result involving conditional expectations is that of **iterative expectation**.

Iterative Expectation

$$E_X(E_Y(Y|X)) = E(Y) \quad ;E_Y(E_X(X|Y)) \quad = E(X)$$

In evaluating the inner expectation, the quantity that is fixed is treated as deterministic. The outer expectation essentially averages the inner one over the outcome space of the fixed inner variable.

Independence

Two random variables are said to be **independent** if and only if

$$\boxed{f(x, y) = f_X(x)f_Y(y)} \tag{7}$$

Alternatively

Two random variables are said to be **independent** if and only if

$$f_Y(y|x) = f_Y(y) \qquad \text{or} \qquad f_X(x|y) = f_X(x) \tag{8}$$

Covariance

One of the most interesting questions in bivariate analysis and prediction theory is if the outcomes of two random variables influence each other, i.e., whether they co-vary.

The statistic that measures the co-variance between two RVs is given by

$$\boxed{\sigma_{XY} = E((X - \mu_X)(Y - \mu_Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - \mu_X)(y - \mu_Y)f(x, y)\,dx\,dy} \tag{9}$$

Covariance, second-order property of the joint p.d.f., can further be shown as

$$\boxed{\sigma_{XY} = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - E(X)E(Y)} \tag{10}$$

Covariance for vector quantities

The covariance and variance of a pair of random variables X_1 and X_2 are collected in a single matrix, known as the **variance-covariance** matrix

$$\Sigma_X = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2} \\ \sigma_{X_2 X_1} & \sigma^2_{X_2} \end{bmatrix}$$

Vector case

In the general case, for a vector of random variables,

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_N \end{bmatrix}^T$$

the matrix is given by,

$$\begin{aligned} \Sigma_{\mathbf{X}} &= E((\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T) \\ &= \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_N} \\ \sigma_{X_2 X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2 X_N} \\ \vdots & \cdots & \cdots & \vdots \\ \sigma_{X_N X_1} & \sigma_{X_N X_2} & \cdots & \sigma^2_{X_N} \end{bmatrix} \end{aligned}$$

Properties of the covariance matrix

The covariance (matrix) possesses certain properties which have far-reaching implications in analysis of random processes.

- Covariance is a second-order property of the joint probability density function
- It is a **symmetric** measure: $\sigma_{XY} = \sigma_{YX}$, i.e., it is not a directional measure. **Consequently it cannot be used to sense causality** (cause-effect relation).
- The covariance matrix $\Sigma_{\mathbf{X}}$ is a **symmetric, positive semi-definite** matrix $\implies \lambda_i(\Sigma_{\mathbf{X}}) \geq 0 \;\; \forall\, i$
- **Number of zero eigenvalues of $\Sigma_{\mathbf{Z}}$ = Number of linear relationships in $\mathbf{Z}$** (cornerstone for principal component analysis and multivariate regression)

Properties of the covariance matrix

- Linear transformation of the random variables $\mathbf{Z} = \mathbf{A}\mathbf{X}$ results in

$$\Sigma_{\mathbf{Z}} = E((\mathbf{Z} - \mu_{\mathbf{Z}})(\mathbf{Z} - \mu_{\mathbf{Z}})^T) = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T \tag{11}$$

- Most importantly, **covariance is only a measure of linear relationship** between two RVs, i.e.,

$\boxed{\text{When } \sigma_{XY} = \sigma_{YX} = 0, \text{ there is no linear relationship between } X \text{ and } Y}$

Correlation

Two issues are encountered with the use of covariance in practice:

i. Covariance is sensitive to the choice of units for the random variables under investigation. Stated otherwise, **it is sensitive to scaling**.

ii. **It is not a bounded measure**, meaning it is not possible to the infer the degree of the strength of the linear relationship from the value of σ_{XY}

To overcome these issues, a normalized version of covariance known as **correlation** is introduced:

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}} \tag{12}$$

Properties of correlation

Correlation enjoys all the properties that covariance satisfies,*i.e.*, symmetricity and ability to detect linear relationships, etc.

Importantly, correlation is a **bounded** measure, i.e., $\boxed{|\rho_{XY}| \leq 1}$

Boundedness

For all bivariate distributions with finite second order moments,

$$|\rho_{XY}| \leq 1 \tag{13}$$

with equality if, with probability 1, there is a linear relationship between X and Y.

The result can be proved using **Chebyshev's inequality**

Unity correlation

Correlation measures linear dependence. Specifically,

i. $\rho_{XY} = 0 \iff X$ and Y have no linear relationship (non-linear relationship cannot be detected)

ii. $|\rho_{XY}| = 1 \iff Y = \alpha X + \beta$ (Y and X are linearly related with or without an intercept)

Unity correlation ...contd.

Assume $Y = \alpha X$. Then, $\mu_Y = \alpha\mu_X$; $\sigma_Y^2 = \alpha^2\sigma_X^2$

$$\begin{aligned}
\rho_{XY} &= \frac{E(XY) - E(X)E(Y)}{\sigma_X\sigma_Y} \\
&= \frac{\alpha(E(X)^2 - E(X)^2)}{|\alpha|\sigma_X^2} \\
&= \frac{\alpha}{|\alpha|} \\
&= \pm 1
\end{aligned}$$

Uncorrelated variables

Uncorrelated variables

Two random variables are said to be **uncorrelated** if $\sigma_{XY} = 0 \Longrightarrow \rho_{XY} = 0$.

Alternatively, since $\sigma_{XY} = E(XY) - E(X)E(Y)$, the condition also implies

$$\boxed{E(XY) = E(X)E(Y)} \tag{14}$$

- Uncorrelatedness $\Longleftrightarrow$ NO **linear** relationship between X and Y.
- Determining the absence of non-linear dependencies requires the test of **independence.**

Independence vs. Uncorrelated variables

Independence $\Longrightarrow$ Uncorrelated condition but NOT vice versa.

Thus independence is a stronger condition.

If the variables X and Y have a bivariate Gaussian distribution,

Independence $\Longleftrightarrow$ Uncorrelated condition.

Therefore, in all such cases, independence and lack of correlation are equivalent.

Conditional expectation and Covariance

The conditional expectation essentially gives the expectation of Y given $X = x$. Intuitively, if there is a linear relationship between Y and X, the conditional expectation should be expected to be different from the unconditional expectation. Theoretically,

$$\boxed{\textbf{Two variables are uncorrelated if } E(Y|X) = E(Y)}$$

The result can be proved using iterative expectation. See text (p. 168) for details.

Remark: The result is similar to that for independence: $f(y|x) = f(y)$ (conditional = marginal)

Correlation values less than unity - implications

The magnitude of correlation in practice is never likely to touch unity since no process is truly linear. Therefore, it is useful to quantify the effects of factors that contribute to the fall of correlation (in magnitude) below unity:

- Non-linearities and / or modelling errors
- Measurement noise and / or effects of unmeasured disturbances

A common scenario: linear relation plus noise, etc.

Assume a r.v. Y is made up of a linear effect αX and an additional element ϵ, $Y = \alpha X + \epsilon$ s.t. (i) $\mu_\epsilon = 0 \Longrightarrow \mu_Y = \alpha \mu_X$ and (ii) ϵ and X are uncorrelated, i.e., there is nothing in ϵ that can be explained by a linear function of X (reasonable assumption).

Then, it can be shown that

$$\boxed{\rho_{YX} = \pm \frac{1}{\sqrt{1 + \dfrac{\sigma_\epsilon^2}{\alpha^2 \sigma_X^2}}} \leq \pm 1}$$

Signal-to-Noise Ratio (SNR) and Non-linearities

- When ϵ represents merely the measurement error in Y, the ratio $\dfrac{\alpha^2\sigma_X^2}{\sigma_\epsilon^2}$ is known as the **signal-to-noise ratio** (SNR). Thus, even when the true relationship is linear, SNR can cause a significant dip in the correlation. In fact, as $SNR \to 0$, $\rho_{YX} \to 0$
- When ϵ represents the non-linearities and other factors that are uncorrelated with X, the ratio $\dfrac{\alpha^2\sigma_X^2}{\sigma_\epsilon^2}$ represents the variance-explained to prediction-error ratio. Once again when $\sigma_\epsilon^2 \gg \alpha^2\sigma_X^2$, $\rho_{YX} \approx 0$

In practice, ϵ contains both unexplained non-linear effects and noise. It is hard to distinguish the individual contributions, but the net effect is a drop in the correlation.

Connections b/w Correlation and Linear regression

Correlation between two RVs is naturally related to linear regression of one variable on the other.

Given two (zero-mean) RVs X and Y, consider the linear predictor of Y in terms of X

$$\hat{Y} = bX \tag{15}$$

The optimal estimate b that minimizes $E(\varepsilon^2) = E(Y - \hat{Y})^2$ is

$$\boxed{b^\star = \frac{\sigma_{XY}}{\sigma_X^2} = \rho_{XY}\frac{\sigma_Y}{\sigma_X}} \tag{16}$$

Correlation and regression ...contd.

Similarly, the optimal coefficient of the *reverse* predictor, $\hat{X} = \tilde{b}Y$, is

$$\tilde{b}^\star = \frac{\sigma_{XY}}{\sigma_Y^2} = \rho_{XY}\frac{\sigma_X}{\sigma_Y} \tag{17}$$

We make an interesting observation,

$$\boxed{\rho_{XY}^2 = b^\star \tilde{b}^\star} \tag{18}$$

Correlation captures linear effects in **both directions.**

Correlation and regression ... contd.

Further, (for the $X \to Y$ model)

$$\text{cov}(\varepsilon, X) = \text{cov}(Y - b^\star X, X) = 0 \qquad (\textbf{residual} \perp \textbf{regressor}) \tag{19}$$

$$\sigma_\varepsilon^2 = \sigma_Y^2 - b^{\star 2}\sigma_X^2 = \sigma_Y^2(1 - \rho_{XY}^2) \tag{20}$$

so that the standard theoretical measures of fit for both (directional) models are

$$\boxed{R^2_{X \to Y} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_Y^2} = \rho_{XY}^2 = R^2_{Y \to X}} \tag{21}$$

Zero correlation implies no linear fit in either direction.

Remarks, limitations, ...

- **Correlation is only a mathematical / statistical measure.** It does not take into account any physics of the process that relates X to Y.
- High values of correlation only means that a linear model can be fit between X and Y. It does not mean that in reality there exists a linear process that relates X to Y
- Correlation is symmetric, i.e., $\rho_{XY} = \rho_{YX}$. Therefore, it is not a cause-effect measure, meaning it cannot detect direction of relationships
- Correlation is primarily used to determine if a linear model can explain the relationship:

Remarks, limitations, ... contd.

- High values of **estimated** correlation may imply linear relationship over the experimental conditions, but a non-linear relationship over a wider range.
- **Absence of correlation only implies that no linear model can be fit.** Even if the true relationship is linear, noise can be high, thus limiting the ability of correlation to detect linearity
- **Correlation measures only linear dependencies**. Zero correlation means $E(XY) = E(X)E(Y)$. In contrast, independence implies $f(x, y) = f(x)f(y)$

Despite its limitations, correlation and its variants remain one of the most widely used measures in data analysis

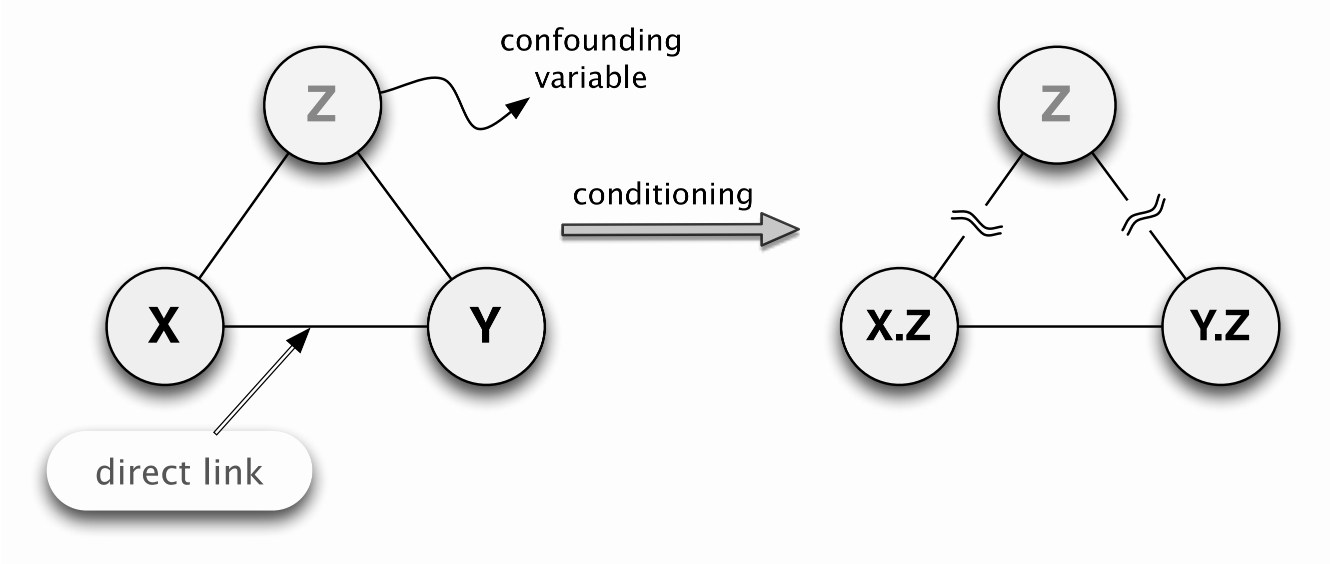
Confounding

When two variables X and Y are correlated, a question that begs attention is:

Q: Are X and Y connected to each other **directly** or **indirectly**?

Conditional measures provide the answer.

Conditional correlation



- If the **conditional correlation** vanishes, then the connection is purely indirect, else there exists a direct relation

Correlation measures total (linear) connectivity, whereas conditional or partial version measures "direct" association.

Partial covariance

The conditional or partial covariance is defined as

$$\sigma_{XY.Z} = \text{cov}(\epsilon_{X.Z}, \epsilon_{Y.Z}) \qquad \text{where} \;\; \epsilon_{X.Z} = X - \hat{X}^{\star}(Z), \;\; \epsilon_{Y.Z} = Y - \hat{Y}^{\star}(Z)$$

where $\hat{X}^{\star}(Z)$ and $\hat{Y}^{\star}(Z)$ are the **optimal predictions** of X and Y using Z.

Partial correlation

Partial correlation (PC)

$$\boxed{\begin{aligned}\rho_{XY.Z} &= \frac{\sigma_{XY.Z}}{\sigma_{\epsilon_{X.Z}}\sigma_{\epsilon_{Y.Z}}} \\ &= \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{(1-\rho_{XZ}^2)}\sqrt{(1-\rho_{ZY}^2)}}\end{aligned}} \tag{22}$$

- Can be extended to more-than-three variables case as well.
- **Partial correlation** $\equiv$ Analysis in the **inverse domain.**

Partial correlation: Example

Consider two random variables $X = 2Z + 3W$ and $Y = Z + V$ where V, W and Z are zero-mean RVs. Further, it is known that W and V are uncorrelated with Z as well as among themselves, *i.e.*, $\sigma_{VW} = 0$.

Evaluating the covariance between X and Y yields

$$\sigma_{YX} = E((2Z + 3W)(Z + V)) = 2E(Z^2) = 2\sigma_Z^2 \neq 0$$

although X and Y are not "directly" correlated.

Partial correlation: Example

Now,

$$\rho_{YX} = \frac{\sigma_Z^2}{(\sigma_Z^2 + \sigma_Y^2)(4\sigma_Z^2 + \sigma_W^2)}; \qquad \rho_{YZ} = \frac{1}{\sqrt{1 + \frac{\sigma_V^2}{\sigma_Z^2}}}; \qquad \rho_{XZ} = \frac{1}{\sqrt{1 + \frac{\sigma_V^2}{4\sigma_Z^2}}};$$

Next, applying (22), it is easy to see that

$$\rho_{YX.Z} = 0$$

Partial correlation by inversion (from covariance)

Partial correlation coefficients can also be computed from the inverse of **covariance** matrix as follows.

Assume we have M RVs $X_1, X_2, \cdots, X_M$. Let $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix}^T$.

1. Construct the covariance matrix $\Sigma_{\mathbf{X}}$.
2. Determine the inverse of covariance matrix, $\mathbf{S_X} = \Sigma_{\mathbf{X}}^{-1}$.
3. The partial correlation between X_i and X_j conditioned on all $\mathbf{X}\backslash\{X_i, X_j\}$ is then:

$$\rho_{X_i X_j . \mathbf{X}\backslash\{X_i, X_j\}} = -\frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}} \tag{23}$$

Partial correlation by inversion (from correlation)

Partial correlation coefficients can also be computed from the inverse of **correlation** matrix as follows.

Assume we have M RVs $X_1, X_2, \cdots, X_M$. Let $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix}^T$.

1. Construct the correlation matrix $\Omega_{\mathbf{X}}$.
2. Determine the inverse of correlation matrix, $\mathbf{P}_{\mathbf{X}} = \Omega_{\mathbf{X}}^{-1}$.
3. The partial correlation between X_i and X_j conditioned on all $\mathbf{X}\backslash\{X_i, X_j\}$ is then:

$$\rho_{X_i X_j . \mathbf{X}\backslash\{X_i, X_j\}} = -\frac{p_{ij}}{\sqrt{p_{ii}}\sqrt{p_{jj}}} \tag{24}$$

Partial correlation and Linear regression

Partial correlation enjoys a relation with the regression models as does correlation.

Consider three random variables X_1, X_2, X_3. To understand the PC between X_1 and X_2 conditioned on X_3, consider the "forward" and "reverse" linear predictors:

$$\hat{X}_1 = b_{12}X_2 + b_{13}X_3, \qquad \hat{X}_2 = b_{21}X_1 + b_{23}X_3 \tag{25}$$

The optimal estimates (that minimize the MSEs respectively) can be derived as

$$b_{12}^{\star} = \frac{\sigma_1}{\sigma_2}\left(\frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{23}^2}\right), \qquad b_{21}^{\star} = \frac{\sigma_2}{\sigma_1}\left(\frac{\rho_{12} - \rho_{13}\rho_{23}}{1 - \rho_{13}^2}\right) \tag{26}$$

PC and linear regression

We have an interesting result:

The squared-PC between X_1 and X_2 is the product of optimal coefficients

$$\boxed{\rho^2_{12.3} = b^\star_{12} b^\star_{21}} \tag{27}$$

- Partial correlation captures "direct" dependence in both directions.
- As with correlation, it is also, therefore, an symmetric measure.

Uses of partial correlation

- In time-series analysis, partial correlations are used in devising what are known as partial auto-correlation functions, which are used in determining the order of auto-regressive models
- The partial cross-correlation function and its frequency-domain counterpart, known as partial coherency function, find good use in time-delay estimation
- In model-based control applications, partial correlations are used in quantifying the impact of model-plant mismatch in model-predictive control applications.

Commands in R

Commands	Utility
`mean,var,sd`	sample mean, variance and standard deviation
`colMeans,rowMeans`	means of columns and rows
`median,mad`	sample median and median absolute deviation
`cov,corr,cov2cor`	covariance, correlation and covariance-to-correlation
`lm, summary`	linear regression and summary of fit

Computing partial correlation in R

Use the `ppcor` package (due to Seongho Kim)

- `pcor`: Computes partial correlation for each pair of variables given others.

  Syntax: `pcormat <- pcor(X)`

  where $\mathbf{X}$ is a matrix with variables in columns. The default is Pearson correlation, but it can also compute Kendall and Spearman (partial) correlations. Result is given in `pcormat$estimate`

Computing semi-partial correlation in R

Use the `ppcor` package (due to Seongho Kim)

- `spcor`: Computes **semi-partial correlation**

  Syntax: `spcormat <- spcor(X)`

  The semi-partial correlation between X and Y given Z is computed as the correlation between X and $Y.Z$, i.e., only Y is conditioned on Z. The matrix of semi-partial correlations is *asymmetric*

Sample usage

```
w <- rnorm(200); v <- rnorm(200); z <- rnorm(200);      # Generate w,v and z
x <- 2*z + 3*w; y = z+ v;       # Generate x and y
cor(cbind(x,y))         # Compute correlation matrix between x and y
pcor(cbind(x,y,z))$estimate     # Compute partial correlation matrix
```