Partitioned Covariance Matrices and Partial Correlations

Proposition 1 *Let the $(p+q)\times(p+q)$ covariance matrix $\boldsymbol{C}>0$ be partitioned as*

$$\boldsymbol{C}=\begin{pmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{pmatrix}.$$

Then the symmetric matrix $\boldsymbol{C}^{-1}>0$ has the following partitioned form:

$$\boldsymbol{C}^{-1}=\begin{pmatrix}\boldsymbol{C}_{1|2}^{-1} & -\boldsymbol{C}_{1|2}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\\ -\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{1|2}^{-1} & \boldsymbol{C}_{22}^{-1}+\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{1|2}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\end{pmatrix}=\begin{pmatrix}\boldsymbol{C}_{1|2}^{-1} & -\boldsymbol{C}_{1|2}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\\ -\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{1|2}^{-1} & \boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\end{pmatrix},$$

where

$$\boldsymbol{C}_{1|2}=\boldsymbol{C}_{11}-\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}.$$

Proof. Introduce the auxiliary matrix

$$\boldsymbol{D}=\begin{pmatrix}\boldsymbol{I}_p & -\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\\ \boldsymbol{O} & \boldsymbol{I}_g\end{pmatrix}.$$

Note that $|\boldsymbol{D}|=1$, so $\boldsymbol{D}$ is regular. With it,

$$\boldsymbol{D}\boldsymbol{C}\boldsymbol{D}^T=\begin{pmatrix}\boldsymbol{C}_{1|2} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{C}_{22}\end{pmatrix},$$

from where

$$|\boldsymbol{C}|=|\boldsymbol{C}_{1|2}|\cdot|\boldsymbol{C}_{22}|$$

and

$$\boldsymbol{C}^{-1}=\boldsymbol{D}^T\cdot\begin{pmatrix}\boldsymbol{C}_{1|2}^{-1} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{C}_{22}^{-1}\end{pmatrix}\cdot\boldsymbol{D}$$

that implies the required result. □

Note that by the above proof, $\boldsymbol{C}>0$ implies that both $\boldsymbol{C}_{22}$ and $\boldsymbol{C}_{1|2}$ are regular (invertible) matrices, they are positive definite. Of course, $\boldsymbol{C}_{11}$ is regular (positive definite) too, as it is a principal minor of $\boldsymbol{C}$.

Theorem 1 *Let $(\mathbf{X}_1^T,\mathbf{X}_2^T)^T\sim\mathcal{N}_{p+q}(\boldsymbol{\mu},\boldsymbol{C})$ be a random vector, where the expectation $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{C}$ are partitioned (with block sizes p and q) in the following way:*

$$\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\quad \boldsymbol{C}=\begin{pmatrix}\boldsymbol{C}_{11} & \boldsymbol{C}_{12}\\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22}\end{pmatrix}.$$

Here $\boldsymbol{C}_{11}$, $\boldsymbol{C}_{22}$ are covariance matrices of $\mathbf{X}_1$ and $\mathbf{X}_2$, whereas $\boldsymbol{C}_{12}=\boldsymbol{C}_{21}^T$ is the cross-covariance matrix. Then the conditional distribution of the random vector $\mathbf{X}_1$ conditioned on $\mathbf{X}_2=\mathbf{x}_2$ is $\mathcal{N}_p(\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)+\boldsymbol{\mu}_1,\boldsymbol{C}_{1|2})$ distribution.

Proof. Observe that it suffices to prove for the $(\mathbf{X}_1^T,\mathbf{X}_2^T)^T\sim\mathcal{N}_{p+q}(\mathbf{0},\boldsymbol{C})$ case. The conditional distribution of $\mathbf{X}_1$ conditioned on $\mathbf{X}_2=\mathbf{x}_2$ is determined by the conditional pdf as follows:

$$f(\mathbf{x}_1|\mathbf{x}_2)=\frac{(2\pi)^{-\frac{p+q}{2}}|\boldsymbol{C}|^{-\frac{1}{2}}\exp[-\frac{1}{2}(\mathbf{x}_1^T,\mathbf{x}_2^T)\boldsymbol{C}^{-1}(\mathbf{x}_1^T,\mathbf{x}_2^T)^T]}{(2\pi)^{-\frac{q}{2}}|\boldsymbol{C}_{22}|^{-\frac{1}{2}}\exp[-\frac{1}{2}\mathbf{x}_2^T\boldsymbol{C}_{22}^{-1}\mathbf{x}_2]}.$$

By Proposition 1, after writing $\boldsymbol{C}^{-1}$ in the block-matrix form, and writing the quadratic form in the numerator as the sum of four terms, this simplifies to

$$f(\mathbf{x}_1|\mathbf{x}_2) = (2\pi)^{-\frac{p}{2}}|\boldsymbol{C}_{1|2}|^{-\frac{1}{2}} \exp[-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\mathbf{x}_2)^T \boldsymbol{C}_{1|2}^{-1}(\mathbf{x}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\mathbf{x}_2)$$

which is a p-dimensional Gaussian density with the stated parameters. □

Note that the conditional covariance matrix $\boldsymbol{C}_{1|2}$ does not depend on $\mathbf{x}_2$ of the condition. Further, for the conditional expectation, which is the expectation of the conditional distribution, we get that

$$\mathbb{E}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) + \boldsymbol{\mu}_1.$$

Therefore,

$$\mathbb{E}(\mathbf{X}_1|\mathbf{X}_2) = \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2) + \boldsymbol{\mu}_1$$

which is a linear function of the coordinates of $\mathbf{X}_2$. In the $p = q = 1$ case, it is called *regression line*, while in the $q = 1$, $p > 1$ case, *regression plane*. We will not deal with the $q > 1$, $p > 1$ case, which is the topic of the Canonical Correlation Analysis (CCA), Partial Least Squares Regression (PLS) and the Structural Equation Modeling (SEM). Summarizing, in case of the multidimensional Gaussian distribution, the regression functions are linear functions of the variables in the condition, which fact has important consequences in the multivariate statistical analysis.

Assume that $\mathbb{E}(\mathbf{X}) = \mathbf{0}$. Then with the notation $\boldsymbol{\varepsilon} = \mathbf{X}_1 - \mathbb{E}(\mathbf{X}_1|\mathbf{X}_2) = \mathbf{X}_1 - \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\mathbf{X}_2$ we can decompose $\mathbf{X}_1$ as the sum of $\mathbb{E}(\mathbf{X}_1|\mathbf{X}_2)$ and the error term $\boldsymbol{\varepsilon}$. By construction, $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\mathbf{X}_2\boldsymbol{\varepsilon}^T) = \boldsymbol{O}$. Therefore, the covariance matrix of $\mathbf{X}_1$ is decomposed as

$$\boldsymbol{C}_{11} = \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21} + \boldsymbol{C}_{1|2},$$

or else, the formula for *total variances* is applicable:

$$\mathrm{Var}(\mathbf{X}_1) = \mathrm{Var}(\mathbb{E}(\mathbf{X}_1|\mathbf{X}_2)) + \mathbb{E}(\mathrm{Var}(\mathbf{X}_1|\mathbf{X}_2)).$$

Theorem 2 *Let* $\mathbf{X} = (X_1, \dots, X_m)^T \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{C})$ *be a random vector, and let* $\Gamma := \{1, \dots, m\}$ *denote the index set of the variables,* $m \geq 3$*. Assume that* $\boldsymbol{K} = (k_{ij}) = \boldsymbol{C}^{-1} > 0$*. Then*

$$r_{X_i X_j|\mathbf{X}_{\Gamma\setminus\{i,j\}}} = \frac{-k_{ij}}{\sqrt{k_{ii}k_{jj}}} \qquad i \neq j,$$

where $r_{X_i X_j|\mathbf{X}_{\Gamma\setminus\{i,j\}}}$ *denotes the partial correlation coefficient between* X_i *and* X_j *after eliminating the effect of the remaining variables* $\mathbf{X}_{\Gamma\setminus\{i,j\}}$*. Further,*

$$k_{ii} = (\mathrm{Var}(X_i|\mathbf{X}_{\Gamma\setminus\{i\}})^{-1}, \qquad i = 1, \dots m$$

is the reciprocal of the conditional variance of X_i *conditioned on the other variables* $\mathbf{X}_{\Gamma\setminus\{i\}}$*.*

Proof. Without loss of generality we prove that

$$r_{X_1 X_2|\mathbf{X}_{\Gamma\setminus\{1,2\}}} = \frac{-k_{12}}{\sqrt{k_{11}k_{22}}}.$$

We use the result of Proposition 1 with $p = 2$, $q = m - 2$, and partition $\boldsymbol{C}$ and $\boldsymbol{K}$ as in Figure (a) and (b).

Then regressing X_1 with $\mathbf{X}_{\Gamma\setminus\{1,2\}}$ we get that

$$\mathbb{E}(X_1|\mathbf{X}_{\Gamma\setminus\{1,2\}}) = \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbf{X}_{\Gamma\setminus\{1,2\}}$$

and the error term is

$$\varepsilon_1 = X_1 - \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbf{X}_{\Gamma\setminus\{1,2,\}}.$$

Likewise, regressing X_2 with $\mathbf{X}_{\Gamma\setminus\{1,2\}}$ we get that

$$\mathbb{E}(X_2|\mathbf{X}_{\Gamma\setminus\{1,2\}}) = \mathbf{c}_2^T \boldsymbol{C}_{22}^{-1} \mathbf{X}_{\Gamma\setminus\{1,2\}}$$

and the error term is

$$\varepsilon_2 = X_2 - \mathbf{c}_2^T \boldsymbol{C}_{22}^{-1} \mathbf{X}_{\Gamma\setminus\{1,2\}}.$$

By definition,

$$r_{X_1 X_2|\mathbf{X}_{\Gamma\setminus\{1,2\}}} = \mathrm{Corr}(\varepsilon_1, \varepsilon_2).$$

To find this correlation, we find the covariance and the variances of ε_1 and ε_2. As $\mathbb{E}(\varepsilon_1) = 0$ and $\mathbb{E}(\varepsilon_2) = 0$, with the notation of Figures (a) and (b),

$$\begin{aligned}\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = \mathbb{E}(\varepsilon_1 \varepsilon_2^T) = \mathbb{E}(X_1 X_2^T) - \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbb{E}(\mathbf{X}_{\Gamma\setminus\{1,2\}} X_2^T) - \mathbb{E}(X_1 \mathbf{X}_{\Gamma\setminus\{1,2\}}^T) \boldsymbol{C}_{22}^{-1} \mathbf{c}_2 \\ + \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbb{E}(\mathbf{X}_{\Gamma\setminus\{1,2\}} \mathbf{X}_{\Gamma\setminus\{1,2\}}^T) \boldsymbol{C}_{22}^{-1} \mathbf{c}_2 = c_{12} - \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbf{c}_2,\end{aligned}$$

where we used that $\mathbb{E}(\mathbf{X}_{\Gamma\setminus\{1,2\}} \mathbf{X}_{\Gamma\setminus\{1,2\}}^T) = \boldsymbol{C}_{22}$. Likewise,

$$\mathrm{Var}(\varepsilon_1) = \mathbb{E}(\varepsilon_1 \varepsilon_1^T) = c_{11} - \mathbf{c}_1^T \boldsymbol{C}_{22}^{-1} \mathbf{c}_1$$

and

$$\mathrm{Var}(\varepsilon_2) = \mathbb{E}(\varepsilon_2 \varepsilon_2^T) = c_{22} - \mathbf{c}_2^T \boldsymbol{C}_{22}^{-1} \mathbf{c}_2.$$

Putting these together, and using that, by Theorem 2, the upper left 2×2 block of $\boldsymbol{K}$ is: $\boldsymbol{K}_{11} = \boldsymbol{C}_{1|2}^{-1}$, with the notation of Figure (c) we get that

$$r_{X_1 X_2|\mathbf{X}_{\Gamma\setminus\{1,2\}}} = \frac{b}{\sqrt{ad}} = \frac{-\frac{-b}{\sqrt{ad-b^2}}}{\sqrt{\frac{a}{\sqrt{ad-b^2}} \frac{d}{\sqrt{ad-b^2}}}} = -\frac{k_{12}}{\sqrt{k_{11} k_{22}}}.$$

In the last equality we utilized that $\boldsymbol{K}_{11} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$ and $k_{12} = k_{21}$.

To prove the statement for k_{ii}, it suffices to prove it for k_{11}. By Theorem 1, the conditional distribution of X_1 conditioned on $\mathbf{X}_{\Gamma\setminus\{1\}}$ is univariate $\mathcal{N}_1(\tilde{\mathbf{c}}_{12} \tilde{\boldsymbol{C}}_{22}^{-1} \mathbf{X}_{\Gamma\setminus\{1\}}, \tilde{\boldsymbol{C}}_{1|2})$ distribution, where

$$\mathrm{Var}(X_1|\mathbf{X}_{\Gamma\setminus\{1\}}) = \tilde{\boldsymbol{C}}_{1|2} = c_{11} - \tilde{\mathbf{c}}_{12}^T \tilde{\boldsymbol{C}}_{22}^{-1} \tilde{\mathbf{c}}_{12} = k_{11}^{-1}$$

if we apply the result of Proposition 1 with $p = 1$, $q = m - 1$, see Figure (d). From Proposition 1 is also clear that $\boldsymbol{C} > 0$ implies that $k_{11} \neq 0$. This completes the proof. $\square$

Definition 1 *Assume* $\mathbf{X} \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{C})$ *with* $\boldsymbol{C} > 0$. *Consider the regression line*

$$\mathbb{E}(X_1|\mathbf{X}_{\Gamma\setminus\{1\}} = \mathbf{x}_{\Gamma\setminus\{1\}}) = \sum_{j\in\Gamma\setminus\{1\}} \beta_{1j|\Gamma\setminus\{1\}} x_j,$$

where x_j*'s are the coordinates of* $\mathbf{x}_{\Gamma\setminus\{1\}}$. *Then we call the coefficient* $\beta_{1j|\Gamma\setminus\{1\}}$ j*th* ***partial regression coefficient*** *for* $j \in \Gamma \setminus \{1\}$.

The definition naturally extends to $\beta_{ij|\Gamma\setminus\{i\}}$ for $j \in \Gamma \setminus \{i\}$.

Theorem 3 $\beta_{1j|\Gamma\setminus\{1\}} = -\frac{k_{1j}}{k_{11}}$ *for* $j \in \Gamma \setminus \{1\}$.

To prove the theorem, first we need a lemma.

Lemma 1 $\boldsymbol{C}_{12} = \boldsymbol{O}$ *is equivalent to* $\boldsymbol{K}_{12} = \boldsymbol{O}$ *and*

$$\boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12} = -\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}.$$

Proof of the Lemma. By Proposition 1, an easy calculation shows that

$$\boldsymbol{K}_{11}^{-1}\boldsymbol{K}_{12} = \boldsymbol{C}_{1|2}\boldsymbol{K}_{12} = \boldsymbol{C}_{1|2}(-\boldsymbol{C}_{1|2}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}) = -\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}$$

that finishes the proof. □

Proof of Theorem 3. With the help of Figure (d), the equation of the regression line is

$$\mathbb{E}(X_1|\mathbf{X}_{\Gamma\setminus\{1\}} = \mathbf{x}_{\Gamma\setminus\{1\}}) = \tilde{\mathbf{c}}_{12}\tilde{\boldsymbol{C}}_{22}^{-1}\mathbf{x}_{\Gamma\setminus\{1\}} = \sum_{j\in\Gamma\setminus\{1\}} (\mathbf{c}_{12}^T\tilde{\boldsymbol{C}}_{22}^{-1})_j x_j$$
$$= \sum_{j\in\Gamma\setminus\{1\}} (-k_{11}^{-1}\boldsymbol{K}_{12})_j x_j = \sum_{j\in\Gamma\setminus\{1\}} -\frac{k_{1j}}{k_{11}} x_j,$$

where in the last line we used the Lemma. This finishes the proof. □

Corollary 1 *An important consequence of Theorem 3 is that*

$$\beta_{ij|V\setminus\{i\}} = -\frac{k_{ij}}{k_{ii}} = r_{X_iX_j|\mathbf{X}_{V\setminus\{i,j\}}}\sqrt{\frac{k_{jj}}{k_{ii}}}, \quad \textit{for} \quad j \neq i.$$

So only the variables X_j's whose partial correlation with X_i (after elimination the effect of the remaining variables) is 0, enter into the regression of X_i with the other variables. We sometimes say that the partial regresion coeficient $\beta_{ij|\Gamma\setminus\{i\}}$ shows the change of X_i under a unit change of X_j keeping the other variables fixed, in latin wirds, 'ceterum paribus'.

If we form a graph $\mathcal{G}$ on the vertex-set Γ and

$$i \sim j \Leftrightarrow k_{ij} \neq 0, \quad i \neq j,$$

then only the variables corresponding to vertices connected to vertex i enter into the regression of X_i with the other variables. This is called *Gaussian graphical model* and it is the base of the Path Analysis.

If the Gaussian graphical model is decomposable (see Graphical models in the Information Theory and Statistics course), then the maximal cliques, together with their separators (with multiplicities), form a so-called *junction tree* structure. Denote $\mathcal{C}$ the set of the maximal cliques and $\mathcal{S}$ the set of their separators in $\mathcal{G}$. Then the ML estimator of $\boldsymbol{K}$ can be calculated based on the $\boldsymbol{S}$ matrices (see the ML estimators of the multivariate normal parameters).

Let n be the sample size for the underlying m-variate normal distribution, $\Gamma = \{1, \dots, m\}$, and assume that $m < n$. For the maximal clique $C \in \mathcal{C}$, let $[\boldsymbol{S}_C]^\Gamma$ denote n times the empirical covariance matrix corresponding to the variables $\{X_i : i \in C\}$ complemented with zero entries to have an $m \times m$ (symmetric, positive semidefinite) matrix. Likewise, for the separator $S \in \mathcal{S}$, let $[\boldsymbol{S}_S]^\Gamma$ denote n times the empirical covariance matrix corresponding to the variables $\{X_i : i \in S\}$ complemented with zero entries to have an $m \times m$ (symmetric, positive semidefinite) matrix. Then the ML estimator of the mean vector is the sample average (as usual), while the ML estimator of the concentration matrix is

$$\hat{\boldsymbol{K}} = n \left\{ \sum_{C \in \mathcal{C}} [\boldsymbol{S}_C^{-1}]^\Gamma - \sum_{S \in \mathcal{S}} [\boldsymbol{S}_S^{-1}]^\Gamma \right\};$$

further,

$$|\hat{\boldsymbol{K}}| = n^m \cdot \frac{\prod_{S \in \mathcal{S}} |\boldsymbol{S}_S|}{\prod_{C \in \mathcal{C}} |\boldsymbol{S}_C|}.$$

Testing hypothesis about partial correlations. For $i \neq j$ we want to test

$$H_0 \, : \, r_{X_i X_j | \mathbf{X}_{\Gamma \setminus \{i,j\}}} = 0,$$

i.e., that X_1 and X_2 are contitionally independent conditioned on the remaining variables. Equivalently, H_0 means that there is no edge in $\mathcal{G}$ between vertices i and j, or equivalently, $\beta_{ij|\Gamma\setminus\{i\}} = 0$, $\beta_{ji|\Gamma\setminus\{j\}} = 0$, or simply, $k_{ij} = k_{ji} = 0$ ($\boldsymbol{C} > 0$ should be assumed).

To test H_0 in some form, several exact tests are known that are usually based on likelihood ratio tests or on Basu's theorem (see below). The following test uses the empirical partial correlation coefficient, denoted by $\hat{r}_{X_i X_j | \mathbf{X}_{\Gamma \setminus \{i,j\}}}$, and the following statistic based on it:

$$B = 1 - (\hat{r}_{X_i X_j | \mathbf{X}_{\Gamma \setminus \{i,j\}}})^2 = \frac{|\boldsymbol{S}_{\Gamma \setminus \{i,j\}}| \cdot |\boldsymbol{S}_\Gamma|}{|\boldsymbol{S}_{\Gamma \setminus \{i\}}| \cdot |\boldsymbol{S}_{\Gamma \setminus \{j\}}|}.$$

It can be proven that under H_0, the test statistic

$$t = \sqrt{n - m} \cdot \sqrt{\frac{1}{B} - 1} = \sqrt{n - m} \cdot \frac{\hat{r}_{X_i X_j | \mathbf{X}_{\Gamma \setminus \{i,j\}}}}{\sqrt{1 - (\hat{r}_{X_i X_j | \mathbf{X}_{\Gamma \setminus \{i,j\}}})^2}}$$

is distributed as Student's t with $n - m$ degrees of freedom. Therefore, we reject H_0 for large values of $|t|$.

Theorem 4 (Basu) *If $T = T(\mathbf{X})$ is a sufficient and complete statistic with respect to the family $\mathcal{P}$ of distributions, and the distribution of $Y = s(\mathbf{X})$ does not depend on $\mathbb{P} \in \mathcal{P}$, then T and Y are independent.*