1 Normal Theory and the Precision Matrix

Letting x be a p-vector with a multivariate Gaussian/normal, $x \sim N(m, \Sigma)$, with pdf

$$p(x) = \left((2\pi)^p |\Sigma|\right)^{-1/2} \exp\left(-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)\right).$$

Σ is referred to as the covariance matrix, and $\Phi = \Sigma^{-1}$ is known as the precision matrix (when Σ^{-1} exists). Writing the pdf in terms of the precision matrix, we get

$$p(x) \propto |\Phi|^{1/2} \exp\left(-\frac{1}{2}(x-m)^T \Phi (x-m)\right).$$

1.1 Partitioned Normal Distribution

The p-dimensional normal random quantity $x \sim N(m, \Sigma)$ can be partitioned as

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The mean and covariance matrix are partitioned conformably as follows

$$m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} V_1 & R \\ R^T & V_2 \end{pmatrix}.$$

Of course, the partitions satisfy $x_i \sim N(m_i, V_i)$.
Standard results from linear algebra give

$$\Phi = \Sigma^{-1} = \begin{pmatrix} K_1 & H \\ H^T & K_2 \end{pmatrix},$$

where

- $K_1^{-1} = V_1 - R V_2^{-1} R^T$,
- $H = -K_1 R V_2^{-1}$,
- $K_2^{-1} = V_2 - R^T K_1^{-1} R$.

1.2 Conditional Normal Distributions

Consider the zero-mean case: $x \sim N(0, \Phi^{-1})$ with $p(x) \propto \exp(-Q(x)/2)$, where

$$Q(x) = x^T \Phi x = x_1^T K_1 x_1 + 2x_1^T H x_2 + x_2^T K_2 x_2.$$

By inspection we have

$$p(x_1|x_2) \propto \exp(-Q(x_1|x_2)/2),$$

where

$$Q(x_1|x_2) = x_1^T K_1 x_1 + 2x_1^T H x_2.$$

This yields the conditional distribution

$$x_1|x_2 \sim N(E[x_1|x_2], V(x_1|x_2)),$$

where $V(x_1|x_2) = K_1^{-1}$,
and $E[x_1|x_2] = A_1 x_2$, with $A_1 = RV_2^{-1} = -K_1^{-1}H$. In the general case we have

$$(x_1|x_2) \sim N(m_1 + A_1(x_2 - m_2), K_1^{-1}).$$

The expression $A_1 = RV_2^{-1}$ expresses how the regression of x_1 on x_2 is based on the covariance elements R being rotated and scaled by the variance V_2 of the conditioning variables. The second expression for A_1 relates to the elements H of the precision matrix Φ and has critical implications.

1.3 Precision Matrix and Conditional Representation

Consider the partition where x_1 is a scalar and $x_2 = x_{2:p} = x_{-1}$. Denote the elements of the precision matrix Φ as $\Phi_{i,j}$. From the previous section, $p(x_1|x_{-1})$ is normally distributed with moments

$$V(x_1|x_{-1}) = 1/\Phi_{1,1},$$

and

$$E[x_1|x_{-1}] = m_1 - \frac{1}{\Phi_{1,1}} H(x - m_2).$$

Now, $H = -K_{1,1}RV_2^{-1} = (\Phi_{1,1}, \ldots, \Phi_{1,p})$. From this, we get

$$p(x_1|x_{-1}) = m_1 - \sum_{j=2}^{p} \frac{\Phi_{1,j}}{\Phi_{1,1}}(x_j - m_{2,j}).$$

Of course, all of this generalizes to any scalar partition with x_1 being the i^{th} element and $x_2 = x_{-i}$. This shows explicitly how the elements of the precision matrix Φ, for the full joint distribution, determines the conditional structure.

- Zeros in the precision matrix Φ define, and are defined by, the conditional independencies in $p(x)$. That is, the precision $\Phi_{i,j} = 0$ iff the complete conditional distribution of x_i does not depend on x_j. More explicitly, $\Phi_{i,j} = 0$ iff $x_i \perp x_j$ conditional on $x_{-(i,j)}$.
- This is the basic underlining idea in the field of Gaussian Graphical Models.