Normal Conditioning

Chase Abram*

September 1, 2022

This goal is to provide some intuition (and a quick derivation) for why normal conditioning works so cleanly. Let X be an n-dimensional random vector

$$X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Suppose $n_1 + n_2 = n$, and we want to consider regressing the last n_2 componenent of X, call it X_2, on the first n_1 components of X, call it X_1. Let W be length n and consider the same decomposition into W_1 and W_2, where

$$W \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Assume $\boldsymbol{\Sigma}$ is positive definite. Then we can uniquely Cholesky-decompose $\boldsymbol{\Sigma} = CC'$, where C is lower triangular. Now note that we may consider X as being constructed in the following way

$$\begin{aligned} X &= \boldsymbol{\mu} + CW \\ \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \end{aligned}$$

Where C_{ij} are blocks in the C matrix. Note that C_{ij} has dimensions $n_i \times n_j$ (in particular C_{12} may be not square). Also note this implies

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} C_{11}C_{11} & C_{11}C_{12} \\ C_{21}C_{11} & C_{21}C_{12} + C_{22}C_{22} \end{bmatrix}$$

Then

$$\begin{aligned} X_1 &= \mu_1 + C_{11}W_1 \\ \Rightarrow W_1 &= C_{11}^{-1}(X_1 - \mu_1) \\ X_2 &= \mu_2 + C_{21}W_1 + C_{22}W_2 \\ &= \mu_2 + C_{21}C_{11}^{-1}(X_1 - \mu_1) + C_{22}W_2 \end{aligned}$$

In higher dimensions, I'm not aware of cleaner form than this one for the conditional expectation. In the one variable case ($n_1 = n_2 = 1$) this becomes

*Contact abram@uchicago.edu with errors or concerns.

$$X_2 = \mathbb{E}[X_2] + \frac{\text{Cov}(X_1, X_2)}{\mathbb{V}(X_1)}(X_1 - \mathbb{E}[X_1]) + \epsilon$$

Now we can find the distribution of X_2 conditioned on X_1, since we know that if two vectors are jointly normal, then the conditional distribution of either, given the other, is also normal. Then we only require the mean and covariance matrix.

$$\begin{aligned}
\mathbb{E}[X_2 \mid X_1] &= \mu_2 + C_{21}C_{11}^{-1}(X_1 - \mu_1) \\
\mathbb{V}(X_2 \mid X_1) &= C_{22}C_{22} \\
&= \Sigma_{22} - C_{21}C_{12} \\
&= \Sigma_{22} - C_{21}C_{11}(C_{11}^{-1}C_{11}^{-1})C_{11}C_{12} \\
&= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}
\end{aligned}$$