from T.W. Anderson, An Introduction to Multivariate Statistical Analysis (1958

2.5. CONDITIONAL DISTRIBUTIONS AND MULTIPLE CORRELATION COEFFICIENT

2.5.1. Conditional Distributions

In this section we find that conditional distributions derived from joint normal distributions are normal. The conditional distributions are of a particularly simple nature because the means depend only linearly on the variates held fixed and the variances and covariances do not depend at all on the values of the fixed variates. The theory of partial and multiple correlation discussed in this section was originally developed by Karl Pearson (1896) for three variables and extended by Yule (1897a, 1897b).

Let X be distributed according to $N(\mu, \Sigma)$ (with Σ nonsingular). Let us partition

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \tag{1}$$

as before into q- and $(p-q)$-component subvectors respectively. We shall use the algebra developed in Section 2.4 here. The joint density of $Y^{(1)} = X^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}X^{(2)}$ and $Y^{(2)} = X^{(2)}$ is

$$n(y^{(1)} | \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})n(y^{(2)}|\mu^{(2)}, \Sigma_{22}).$$

The density of $X^{(1)}$ and $X^{(2)}$ then can be obtained from this expression by substituting $x^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}x^{(2)}$ for $y^{(1)}$ and $x^{(2)}$ for $y^{(2)}$ (the Jacobian of this transformation being 1); the resulting density of $X^{(1)}$ and $X^{(2)}$ is

$$\begin{aligned} f(x^{(1)}, x^{(2)}) = \frac{1}{(2\pi)^{\frac{1}{2}q}\sqrt{|\Sigma_{11\cdot 2}|}} \exp\{-\tfrac{1}{2}[(x^{(1)} - \mu^{(1)}) \\ - \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})]'\Sigma_{11\cdot 2}^{-1}[(x^{(1)} - \mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})]\} \\ \cdot \frac{1}{(2\pi)^{\frac{1}{2}(p-q)}\sqrt{|\Sigma_{22}|}} \exp[-\tfrac{1}{2}(x^{(2)} - \mu^{(2)})'\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})], \end{aligned} \tag{2}$$

where

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{3}$$

This density must be $n(x|\mu, \Sigma)$. The conditional density of $X^{(1)}$ given that $X^{(2)} = x^{(2)}$ is the quotient of (2) and the marginal density of $X^{(2)}$ at the point $x^{(2)}$, which is $n(x^{(2)}|\mu^{(2)}, \Sigma_{22})$, the second factor of (2). The quotient is

$$\begin{aligned} f(x^{(1)}|x^{(2)}) = \frac{1}{(2\pi)^{\frac{1}{2}q}\sqrt{|\Sigma_{11\cdot 2}|}} \exp\{-\tfrac{1}{2}[(x^{(1)} - \mu^{(1)}) \\ - \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})]'\Sigma_{11\cdot 2}^{-1}[(x^{(1)} - \mu^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)})]\}. \end{aligned} \tag{4}$$

It is understood that $x^{(2)}$ consists of $p - q$ numbers. The density $f(x^{(1)}|x^{(2)})$ is clearly a q-variate normal density with mean

$$\mathscr{E}(X^{(1)}|x^{(2)}) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x^{(2)} - \mu^{(2)}) = \nu(x^{(2)}), \tag{5}$$

say, and covariance matrix

$$\mathscr{E}\{[X^{(1)} - \nu(x^{(2)})][X^{(1)} - \nu(x^{(2)})]'|x^{(2)}\} = \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{6}$$

It should be noted that the mean of $X^{(1)}$ given $x^{(2)}$ is simply a linear function of $x^{(2)}$, and the covariance matrix of $X^{(1)}$ given $x^{(2)}$ does not depend on $x^{(2)}$ at all.

DEFINITION 2.5.1. *The matrix* $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ *is the matrix of regression coefficients of* $X^{(1)}$ *on* $x^{(2)}$.

The i, jth element of $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ is often denoted by

$$\beta_{ij \cdot q+1, \cdots, j-1, j+1, \cdots, p}.$$

The vector $\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(x^{(2)} - \boldsymbol{\mu}^{(2)})$ is often called the *regression function.*

Let $\sigma_{ij \cdot q+1, \cdots, p}$ be the i, jth element of $\boldsymbol{\Sigma}_{11 \cdot 2}$. We call these *partial covariances.*

DEFINITION 2.5.2.

$$\rho_{ij \cdot q+1, \cdots, p} = \frac{\sigma_{ij \cdot q+1, \cdots, p}}{\sqrt{\sigma_{ii \cdot q+1, \cdots, p}}\sqrt{\sigma_{jj \cdot q+1, \cdots, p}}} \tag{7}$$

is the partial correlation between X_i *and* X_j *holding* $X_{q+1}, \cdots, X_p$ *fixed.*

The numbering of the components of X is arbitrary and q is arbitrary. Hence, the above serves to define the conditional distribution of any q components of X given any other $p - q$ components. Indeed, we can consider the marginal distribution of any r components of X and define the conditional distribution of any q components given the other $r - q$ components.

THEOREM 2.5.1. *Let the components of* X *be divided into two groups composing the subvectors* $X^{(1)}$ *and* $X^{(2)}$. *Suppose the mean* $\boldsymbol{\mu}$ *is similarly divided into* $\boldsymbol{\mu}^{(1)}$ *and* $\boldsymbol{\mu}^{(2)}$, *and suppose the covariance matrix* $\boldsymbol{\Sigma}$ *of* X *is divided into* $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{22}$, *the covariance matrices of* $X^{(1)}$, *of* $X^{(1)}$ *and* $X^{(2)}$, *and of* $X^{(2)}$ *respectively. Then if the distribution of* X *is normal, the conditional distribution of* $X^{(1)}$ *given* $X^{(2)} = x^{(2)}$ *is normal with mean* $\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(x^{(2)} - \boldsymbol{\mu}^{(2)})$ *and covariance matrix* $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

As an example of the above considerations let us consider the bivariate normal distribution and find the conditional distribution of X_1 given $X_2 = x_2$. In this case $\boldsymbol{\mu}^{(1)} = \mu_1$, $\boldsymbol{\mu}^{(2)} = \mu_2$, $\boldsymbol{\Sigma}_{11} = \sigma_1^2$, $\boldsymbol{\Sigma}_{12} = \sigma_1\sigma_2\rho$, and $\boldsymbol{\Sigma}_{22} = \sigma_2^2$. Thus the (1×1) matrix of regression coefficients is $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} = \sigma_1\rho/\sigma_2$, and the (1×1) matrix of partial covariances is

$$\boldsymbol{\Sigma}_{11 \cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \sigma_1^2 - \sigma_1^2\sigma_2^2\rho^2/\sigma_2^2 = \sigma_1^2(1 - \rho^2).$$

Thus the density of X_1 given x_2 is $n\left[x_1 \middle| \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right]$. It will be noticed that the mean of this conditional distribution increases with x_2 when ρ is positive and decreases with x_2 when ρ is negative.

A geometrical interpretation of the theory is enlightening. The density $f(x_1, x_2)$ can be thought of as a surface $z = f(x_1, x_2)$ over the x_1, x_2-plane.