FACTORISING THE THE NORMAL DISTRIBUTION

The joint distribution of x and y can be factored as the product of the marginal distribution of x and the conditional distribution of y given x:

$$N(y,x) = N(y|x)N(x). \tag{1}$$

The following notation may be adopted:

$$z = \begin{bmatrix} y - E(y) \\ x - E(x) \end{bmatrix}, \qquad w = \begin{bmatrix} y - E(y|x) \\ x - E(x) \end{bmatrix} = \begin{bmatrix} \varepsilon \\ x - E(x) \end{bmatrix}. \tag{2}$$

We may assume that w is a linear function of z. Then, the mapping from z to $w = Qz$ may be represented by

$$\begin{bmatrix} \varepsilon \\ x - E(x) \end{bmatrix} = \begin{bmatrix} I & -B' \\ 0 & I \end{bmatrix} \begin{bmatrix} y - E(y) \\ x - E(x) \end{bmatrix}, \tag{3}$$

wherein

$$\varepsilon = y - E(y|x) = y - E(y) - B'\{x - E(x)\}. \tag{4}$$

The following dispersion matrices are defined:

$$D(z) = \Sigma_{zz} = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}, \qquad D(w) = \Sigma_{zz} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix}. \tag{5}$$

The off-diaogonal blocks of $D(w)$, which are $C\{y - E(y|x), x\} = 0$ and its transpose, bear witness to the fact that the prediction error $\varepsilon = y - E(y|x)$ is uncorrelated with x, which is the instrument of the prediction. This is the consequence of the factorisation, whereby the conditional distribution is independent of the mariginal distribution.

The quadratic exponent of the joint distribution of x and y may be expressed in terms either of z or w. Thus, $z'\Sigma_{zz}^{-1}z = w'\Sigma_{ww}^{-1}w = z'Q'\Sigma_{ww}^{-1}Qz$, which indicates that $\Sigma_{zz} = Q^{-1}\Sigma_{ww}Q'^{-1}$. This is written more explicitly as

$$\begin{aligned} \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} &= \begin{bmatrix} I & B' \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} + B'\Sigma_{xx}B & B'\Sigma_{xx} \\ \Sigma_{xx}B & \Sigma_{xx} \end{bmatrix}. \end{aligned} \tag{6}$$

The equation is solved for

$$B = \Sigma_{xx}^{-1}\Sigma_{xy} \qquad \text{and} \qquad \Sigma_{\varepsilon\varepsilon} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}. \tag{7}$$

Therefore, the joint density function of x and $y|x$ can be written as

$$N(x;\mu_x, \Sigma_{xx})N(y|x;\mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}\mu_x, \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}). \tag{8}$$

Integrating the conditional distribution $N(y|x)$ with respect to x gives the marginal distribution $N(y;\mu_y, \Sigma_{yy})$.

The linear function

$$\begin{aligned} E(y|x) &= E(y) + C(y,x)D^{-1}(x)\{x - E(x)\} \\ &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x), \end{aligned} \tag{9}$$

which defines the expected value of x for given values of y, is described as the regression of y on x. The matrix $B = \Sigma_{xx}^{-1}\Sigma_{xy}$ is the matrix of the regression coefficients.