Marginal and Conditional Distributions

Theorem. If $X \sim\sim N_p(\mu, \Sigma)$, then the marginal distribution of any subset of k components of X is k-variate normal.

Proof. Partition as follows:

$$X = \begin{pmatrix} X^{(1)}_{k\times 1} \\ X^{(2)}_{(p-k)\times 1} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)}_{k\times 1} \\ \mu^{(2)}_{(p-k)\times 1} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}.$$

Note that $X^{(1)} = (I_k|0) \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim N\left(\mu^{(1)}, \Sigma_{11}\right)$. Since marginals (without independence) do not determine the joint distribution, the converse is not true.

Example. $Z \sim N(0,1)$ independent of U which takes values 1 and -1 with equal probability. Then $Y = UZ \sim N(0,1)$ since

$$\begin{aligned} P(Y \le y) &= P(UZ \le y) \\ &= \frac{1}{2}P(Z \le y|U=1) + \frac{1}{2}P(-Z \le y|U=-1) \\ &= \frac{1}{2}\Phi(y) + \frac{1}{2}\Phi(y) = \Phi(y). \end{aligned}$$

Therefore, (Z,Y) has a joint distribution under which the marginals are normal. However, it is not bivariate normal. Consider $Z + Y =$
$Z + UZ = \begin{cases} 2Z & 1/2 \\ 0 & 1/2 \end{cases}$. Since $P(Z+Y=0) = 1/2$ (i.e., a point mass at 0, and $Z + Y = 2Z \sim N(0,1)$ with probability 1/2, it cannot be normally distributed.

Result. Let $X_{p\times 1} = \begin{pmatrix} X^{(1)}_{k\times 1} \\ X^{(2)}_{(p-k)\times 1} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu^{(1)}_{k\times 1} \\ \mu^{(2)}_{(p-k)\times 1} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right)$.
Then $X^{(1)}$ and $X^{(2)}$ are independent iff $\Sigma_{12} = 0$.

Proof. Only if: Independence implies that $Cov(X^{(1)}, X^{(2)}) = \Sigma_{12} = 0$.

If part: Suppose that $\Sigma_{12} = 0$. Then, note that

$$\begin{aligned}
& M_{(X^{(1)},X^{(2)})}(s_1, s_2) \\
&= E(\exp(s_1' X^{(1)} + s_2' X^{(2)}) = E(\exp\left(\left(\begin{array}{c} s_1 \\ s_2 \end{array}\right)' X\right)) \\
&= E(\exp\left(\left(\begin{array}{c} s_1 \\ s_2 \end{array}\right)' \left(\begin{array}{c} \mu^{(1)} \\ \mu^{(2)} \end{array}\right) + \frac{1}{2}\left(\begin{array}{c} s_1 \\ s_2 \end{array}\right)' \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{array}\right) \left(\begin{array}{c} s_1 \\ s_2 \end{array}\right)\right)) \\
&= \exp\left(s_1'\mu^{(1)} + s_2'\mu^{(2)} + \frac{1}{2}s_1'\Sigma_{11}s_1 + \frac{1}{2}s_2'\Sigma_{22}s_2 + s_1'\Sigma_{12}s_2\right) \\
&= \exp\left(s_1'\mu^{(1)} + \frac{1}{2}s_1'\Sigma_{11}s_1\right) \exp\left(s_2'\mu^{(2)} + \frac{1}{2}s_2'\Sigma_{22}s_2\right) \\
&= M_{X_1}(s_1) M_{X_2}(s_2),
\end{aligned}$$

for all s_1 and s_2 iff $\Sigma_{12} = 0$.

Result. Suppose $X \sim N_p(\mu, \Sigma)$ and let $U = AX$, $V = BX$. Then U and V are independent iff $Cov(U, V) = A\Sigma B' = 0$.

Proof. Same as above, since $\left(\begin{array}{c} U \\ V \end{array}\right) = \left(\begin{array}{c} A \\ B \end{array}\right) X \sim N(.,.)$.

Theorem. If $X \sim N_p(\mu, \Sigma)$ and Σ is p.d. then

$$f_X(x) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \; x \in \mathcal{R}^p.$$

Proof. Let $\Sigma = CC'$ where $C = \Sigma^{1/2}$ is nonsingular. Then $X = CZ + \mu$, $Z \sim N(0, I_p)$. Since Z_i are i.i.d $N(0, 1)$,

$$f_Z(z) = (2\pi)^{-p/2} \exp(-\frac{1}{2}\sum_{i=1}^{p} z_i^2) = (2\pi)^{-p/2} \exp(-\frac{1}{2}z'z).$$

Since $X = CZ + \mu$, $Z = C^{-1}(X - \mu)$. Jacobian of the transformation is $dz = |C|^{-1}\, dx = |\Sigma|^{-1}\, dx$. Therefore,

$$\begin{aligned}
f_X(x) &= (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'(C')^{-1}C^{-1}(x-\mu)\right) \\
&= (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right).
\end{aligned}$$

Note. $f_X(x)$ is constant on the ellipsoid, $\{x : (x-\mu)'\Sigma^{-1}(x-\mu) = r^2\}$.

Ex. Check for $p = 2$ to see if the above results agree with those of the bivariate normal.

Theorem. Let $X \sim N_p(\mu, \Sigma)$, $\Sigma > 0$ (i.e., p.d.), and let

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where X_1 and μ_1 are of length k. Also, let $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Then $\Sigma_{11.2} > 0$ and,
(i) $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \sim N_k(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11.2})$ and is independent of X_2;
(ii) The conditional distribution of X_1 given X_2 is
$N_k\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11.2}\right)$.

Proof. (i) Let $C = \begin{pmatrix} I_k & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-k} \end{pmatrix}$. Then

$$CX = \begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \\ X_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}, C\Sigma C'\right).$$

$$\begin{aligned} C\Sigma C' &= \begin{bmatrix} I_k & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-k} \end{bmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{bmatrix} I_k & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{p-k} \end{bmatrix} \\ &= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{bmatrix} I_k & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_{p-k} \end{bmatrix} = \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}. \end{aligned}$$

Now, independence of $X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ and X_2 follows from the fact that $Cov(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2, X_2) = 0$.
(ii) Note that $X_1 = (X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2) + \Sigma_{12}\Sigma_{22}^{-1}X_2$. Therefore, from the independence of these two parts, $X_1|(X_2 = x_2) = \Sigma_{12}\Sigma_{22}^{-1}x_2 + (X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2) \sim N(\Sigma_{12}\Sigma_{22}^{-1}x_2 + \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11.2})$.

Remark. Under multivariate normality, the best regression is linear. If we want to predict X_1 based on X_2, the best predictor is $E(X_1|X_2)$, which is equal to $\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 + \Sigma_{12}\Sigma_{22}^{-1}x_2$. The prediction error, however, is independent of X_2.