Lecture - 8

March 12, 2013

A d-dimensional normal distribution is characterized by a d-vector μ and a $d \times d$ covariance matrix Σ. We abbreviate it as N(μ, Σ), To qualify as a covariance matrix, Σ must be symmetric (i.e, Σ and Σ^T are equal) and positive semidefinite (meaning that $x^{'}\Sigma x \geq 0$ for all $x \in R^d$). This is equivalent to the requirement that all eigenvalues of Σ be nonnegative (as a symmetric matrix Σ automatically has real eigenvalues). If Σ is positive definite (meaning that strict inequality $x^{'}\Sigma x > 0$ holds for all $x \in R^d$) or equivalently that all eigenvalues of Σ are positive, then the normal distribution N(μ, Σ) has the density :

$$\phi_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^{'}\Sigma^{-1}(x-\mu)}, x \in \mathcal{R}^d$$

with $|\Sigma|$ the determinant of Σ. The standard d-dimensional normal $\mathcal{N}(0, I_d)$ with I_d the $d \times d$ identity matrix is the special case:

$$\frac{1}{(2\pi)^{d/2}} e^{-\frac{1}{2}x^{'}x}$$

If $X \sim N(\mu, \Sigma)$ (i.e, the random vector X has multivariate normal distribution) the its i-th component X_i has distribution $N(\mu_i, \sigma_i)$ with $\sigma_i^2 = \Sigma_{ii}$. The i-th and the j-th components have covariance :

$$Cov(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j)).$$

which justifies calling Σ the covariance matrix. The correlation between X_i and X_j is given by $\rho_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$.

In specifying a multivariate distribution it is sometimes convenient to use the definition is opposite direction; specify the marginal standard deviation σ_i, $i = 1, 2, \cdots, d$ and correlations ρ_{ij} from which the covariance matrix $\Sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$ is then determined. If the $d \times d$ matrix Σ is positive semidefinite but not positive definite then the rank of Σ is less than d and Σ fails to be invertible and there is no normal density with covariance matrix Σ. In this case we can define the normal distribution $\mathcal{N}(\mu, \Sigma)$ as the distribution of $X = \mu + AZ$ with $Z \sim \mathcal{N}(0, I_d)$ for any $d \times d$ matrix A satisfying $AA^T = \Sigma$. The resulting distribution is independent of which such A is chosen.

Some Properties of Multivariate normal

Linear Transformation Property: Any linear transformation of a normal vector is again normal,

$$X \sim N(\mu, \Sigma) \Rightarrow AX \sim N(A\mu, A\Sigma A^T)$$

for any d vector μ, $d \times d$ matrix Σ and any $k \times d$ matrix A for any k.

Moment Generating Function: If $X \sim N(\mu, \Sigma)$, then $E(e^{\theta^T X}) = e^{\theta^T \mu + \frac{1}{2}\theta^T \Sigma \theta}$.

Generating random number from conditional distribution:

Suppose partition vector $(X_{[1]}, X_{[2]})$ (where each $X_{[i]}$ may itself be a vector) is multivariate normal with :

$$\begin{pmatrix} X_{[1]} \\ X_{[2]} \end{pmatrix} \sim N\left(\begin{pmatrix} X_{[1]} \\ X_{[2]} \end{pmatrix}, \begin{pmatrix} \Sigma_{[11]} & \Sigma_{[12]} \\ X_{[21]} & X_{[22]} \end{pmatrix} \right)$$

Suppose, $\Sigma_{[22]}$ has full rank. Then $(X_{[1]}|X_{[2]} = x) \sim N(\mu_{[1]} + \Sigma_{[12]}\Sigma_{[22]}^{-1}(x - \mu_{[2]}), \Sigma_{[11]} - \Sigma_{[12]}\Sigma_{[22]}^{-1}\Sigma_{[21]})$

This equation gives the distribution of $X_{[1]}$ conditioned on $X_{[2]} = x$.

In bivariate set-up, from the above result we can deduce,

$(X_1|X_2 = x) \sim N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x - \mu_2), \sigma_1^2(1 - \rho^2))$ we also know $X_2 \sim N(\mu_2, \sigma_2^2)$

Therefore the algorithm will be as follows :

1. Generate two independent $Z_1, Z_2 \sim N(0, 1)$.
2. First generate X_2, i.e. set $X_2 = \mu + \sigma_2 Z_1$
3. Then generate X_1 conditional on the $X_2 = x$ generated on the previous step i.e. set $X_1 = \mu^* + \sigma^* Z_2$ where $\mu^* = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x - \mu_2)$ and $\sigma^* = \sigma_1\sqrt{(1 - \rho^2)}$

Generation by Cholesky's decomposition:

Among all such A a lower triangular one is particularly convenient be cause it reduces the calculation $\mu + AZ$ to the following :

$$\begin{aligned}
X_1 &= \mu_1 + A_{11}Z_1 \\
X2 &= \mu_2 + A_{21}Z_1 + A_{22}Z_2 \\
\cdots &= \cdots \\
X_d &= \mu_d + A_{d1}Z_1 + A_{d2}Z_2 + \cdots + A_{dd}Z_d.
\end{aligned}$$

For the case of a $d \times d$ covariance matrix Σ we need to solve:

$$\begin{pmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \vdots & \vdots & & \\ A_{d1} & A_{d2} & \cdots & A_{dd} \end{pmatrix} \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{d1} \\ & A_{22} & \cdots & A_{d2} \\ & & & \cdots \\ & & & A_{dd} \end{pmatrix} = \Sigma$$

In 2×2 case covariance matrix Σ is represented as

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}$$

Thus we can sample from a bivariate normal distribution by setting:

$$X_1 = \mu_1 + \sigma_1 Z_1$$

$$X_2 = \mu_2 + \rho\sigma_2 Z_1 + \sqrt{1-\rho^2}\sigma_2 Z_2$$

For the case of $d \times d$ covariance matrix, we get :

$$A_{ij} = \frac{(\Sigma_{ij} - \sum_{k=1}^{j-1} A_{ik}A_{jk})}{A_{jj}} \quad j < i,$$

$$A_{ii} = \sqrt{\Sigma_{ii} - \sum_{k=1}^{i-1} A_{ik}^2}$$