Some Facts about Joint Normal Distribution

Zhentao Shi

February 2, 2012

It is arguable that normal distribution is the most frequently encountered distribution in statistical inference, as it is the asymptotic distribution of many popular estimators. Moreover, it boasts some unique features that facilitates the calculation of objects of interest. This note summaries a few of them.

An $n \times 1$ random vector X follows a joint normal distribution $N(\mu, \Sigma)$, where μ is a $n \times 1$ vector and Σ is an $n \times n$ symmetric positive definite matrix. The probability density function is

$$f_X(x) = (2\pi)^{-n/2} (\det(\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right)$$

and the moment generating function is

$$M_X(t) = \exp\left(t'\mu + \frac{1}{2} t'\Sigma t\right).$$

Theorem 1 states the fact that a linear transformation of X still follows a joint normal distribution.

Theorem 1. *If $X \sim N(\mu, \Sigma)$, then $Y := AX + b \sim N(A\mu + b, A\Sigma A')$.*

We will discuss the relationship between two components of a random vector. To fix notation,

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

where X_1 is a $m \times 1$ vector, and X_2 is a $(n-m) \times 1$ vector. μ_1 and μ_2 are the corresponding mean vectors, and Σ_{ij}, $j = 1, 2$ are the corresponding variance and covariance matrices. From now on, we always maintain the assumption that $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is jointly normal.

Theorem 1 immediately implies a convenient feature of the normal distribution. Generally speaking, if we are given a joint pdf of two random variables and intend to find the marginal distribution of one random variables, we need to integrate out the other variable from the joint pdf. However, if the variables are jointly normal, the information of the other random variable is irrelevant to the marginal distribution of the random variable of interest. We only need to know the partial information of the part of interest, say the mean μ_1 and the variance Σ_{11} to decide the marginal distribution of X_1.

Theorem 2. *The marginal distribution* $X_1 \sim N\left(\mu_1, \Sigma_{11}\right)$.

This result is very convenient if we are interested in some component if an estimator, but not the entire vector of the estimator. For example, the OLS estimator of the linear regression model $y = Z\beta + \epsilon$, under the commonly used assumptions, is

$$\widehat{\beta}_n = \left(Z'Z\right)^{-1} Z'y,$$

and the asymptotic distribution of $\widehat{\beta}_n$ is

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow{\mathrm{d}} N\left(0, \left(Z'Z\right)^{-1} \sigma^2\right)$$

where β_0 is the true parameter. If we are interested in the inference of only the j-th component of $\beta_0^{(j)}$, then from Theorem 2,

$$\sqrt{n}\left(\widehat{\beta}_n^{(j)} - \beta_0^{(j)}\right) \xrightarrow{\mathrm{d}} N\left(0, \left(Z'Z\right)_{jj}^{-1} \sigma^2\right)$$

where $\left(Z'Z\right)_{jj}^{-1}$ is the j-th diagonal element of $\left(Z'Z\right)^{-1}$. The marginal distribution is independent of the other components. This saves us from integrating out the other components, which could be troublesome if the dimension of the vector is high.

Again, generally zero covariance of two random variables only indicates that they are uncorrelated, whereas full independence is a much stronger requirement. However, if X_1 and X_2 are jointly normal, then zero covariance is equivalent to full independence.

Theorem 3. *If* $\Sigma_{12} = 0$*, then* X_1 *and* X_2 *are independent.*

The last result, which is useful in linear regression, is that if X_1 and X_2 are jointly normal, the conditional distribution of X_1 on X_2 is still jointly normal, with the mean and variance specified as in the following theorem. This is the general formula for the calculation of Question 2 of Problem Set 1.

Theorem 4. $X_1|X_2 \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(X_2 - \mu_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$.