2 Properties of the multivariate normal distribution

The multivariate normal distribution is the basis for many of the classical techniques in multivariate analysis.

Recall the central role that the univariate normal distribution played in the development of univariate statistical methods (and why):

1. For many continuous random variables the assumption of normality is appropriate;

2. Numerous statistics, particularly those that can be expressed as sums or averages, have normal sampling distributions regardless of the underlying population distribution;

3. Normal distribution is easy to work with mathematically and many useful results are available.

Items (1) and (2) above are due to the Central Limit Theorem (CLT). Let us recall CLT for a univariate case.

Central Limit Theorem (CLT) If $Y_1, \dots, Y_n$ is a sequence of independent and identically distributed (i.i.d.) random variables each with mean μ and variance σ^2, then

$$Z_n = \frac{\sum_{i=1}^n Y_i - n\mu}{\sigma\sqrt{n}} \sim N(0,1).$$

That is,

$$\lim_{n\to\infty} \frac{F_{Z_n}(z)}{\Phi(z)} \to 1,$$

where $F_{Z_n}(z)$ is the cumulative density function (cdf) of Z_n and $\Phi(z)$ is the cdf of a standard normal random variable.

How large n should be?

The multivariate normal distribution plays an even more central role in multivariate analysis. The justification/motivation is the same, plus (4).

4. There are few alternatives in the multivariate setting, i.e. few tractable multivariate distributions as alternatives.

Recall the **univariate normal distribution**, we write $X \sim N(\mu, \sigma^2)$ to signify that the univariate r.v. X has the normal distribution with mean μ and variance σ^2. The cumulative distribution function (cdf) of X is

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

While F_X is not available in closed-form but the standard normal version (with μ of 0 and σ^2 of 1) is tabled.

The probability density function (pdf) of X is

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}exp\Big[-\frac{(x-\mu)^2}{2\sigma^2}\Big],$$

which implies that for two values $x_1 < x_2$ the area under the graph of the pdf between $f_X(x_1)$ and $f_X(x_2)$ gives $P(x_1 < X < x_2)$.

The moment generating function of X is

$$\psi_X(t) = E\big[e^{tX}\big] = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

and and the characteristic function is

$$\phi_X(t) = E\big[e^{itX}\big] = e^{it\mu - \frac{1}{2}t^2\sigma^2}.$$

In a case of the **multivariate normal distribution**, for the vector $\mathbf{X}$, we write $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ to signify that $\mathbf{x}$ is a p-dimensional random vector with mean $\boldsymbol{\mu}$ and variance Σ. Then cdf of $\mathbf{X}$ is

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= P\big(\mathbf{X} \leq \mathbf{x}\big) = P(X_1 \leq x_1, \ldots, X_p \leq x_p) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(\mathbf{t})d\mathbf{t} \end{aligned}$$

and pdf of $\mathbf{X}$ is

$$f(\mathbf{X}) = \frac{\partial^p}{\partial x_1 \ldots \partial x_p}F_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}exp\Big[-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\Big]$$

It is now the **volume** under the pdf which represents the joint probability that $X_1, \ldots, X_p$ fall within given intervals.

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by $\mathbf{X}$ such that

$$(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) = c^2.$$

Notice that these ellipsoids are centered at $\boldsymbol{\mu}$ and have axes at $\pm c\sqrt{\lambda}e_i$ where λ_i and $\mathbf{e}_i$ are eigenvalues and eigenvectors of Σ, respectively.

We will show that for $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$

$$(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p).$$

This result implies that the solid ellipsoid of $\mathbf{X}$ values satisfying

$$(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}) \leq \chi^2_\alpha(p)$$

has probability $1-\alpha$.

Formal Definition 1. An $p \times 1$-random vector X has a normal distribution iff for **every** $n \times 1$-vector a the one-dimensional random vector $\mathbf{a}^T\mathbf{X}$ has a normal distribution.

Note that $E\mathbf{a}^T\mathbf{X} = \mathbf{a}^T\boldsymbol{\mu}$ and $\mathrm{Var}\big(\mathbf{a}^{\mathrm{T}}\mathbf{X}\big) = \mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}$.

Properties of a Multivariate Distribution

1. Let $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$. If $Y = \mathbf{a}^T\mathbf{X}$, then $Y \sim N\big(\mathbf{a}^T\boldsymbol{\mu}, \mathbf{a}^T\Sigma\mathbf{a}^T\big)$. Therefore,

$$\psi_{\mathbf{X}}(\mathbf{a}) = Ee^{\mathbf{a}^T X} = \psi_Y(1) = e^{\mathbf{a}^T\boldsymbol{\mu}+\frac{1}{2}\mathbf{a}^T\Sigma\mathbf{a}}.$$

   Hence, the moment generating function of $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ is

$$\psi_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}}.$$

2. In the same way we can find the characteristic function of $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$

$$\phi_{i\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}^T\boldsymbol{\mu}-\frac{1}{2}\mathbf{t}^T\Sigma\mathbf{t}}.$$

3. Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ and Σ be a diagonal covariance matrix with σ_i^2 on the main diagonal

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \sigma_p^2 \end{bmatrix}$$

and $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then $X_1, X_2, \ldots, X_p$ are independent normal variables. *Proof is an exercise.*

4. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then every component X_k is one-dimensional normal. To prove consider the k-th ort.

5. $X_1 + X_2 + \ldots X_p$ is one-dimensional normal. Note that the terms in the sum need not be independent.

6. Every marginal distribution of k variables $(1 \leq k < p)$ is normal. To prove this we consider any k variables $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ and then take a such that $a_j = 0$ for $j \neq i_1, \ldots i_k$ and then apply Definition 1.

7. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{Y} = B_{q\times p}\mathbf{X} + \mathbf{b}_{q\times 1}$. Then

$$\mathbf{Y} \sim N_q(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B^T).$$

   Proof is an exercise.

8. Let $\mathbf{X}_1$ be a random vector in R^m, $\mathbf{X}_2$ be a random vector in R^n, and $(\mathbf{X}_1, \mathbf{X}_2)$ have an $(m+n)$-dimensional normal distribution. Then,

   - $\mathbf{X}_1$ has an m-dimensional normal distribution
   - $\mathbf{X}_2$ as an n-dimensional normal distribution.
   - Random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $cov(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$ (i.e. the $m \times n$-zero matrix).

9. Let $\mathbf{X}_1$ be a random vector in R^m, $\mathbf{X}_1 \sim N_m(\boldsymbol{\mu}_1, \Sigma_{11})$, and $\mathbf{X}_2$ be a random vector in R^n, $\mathbf{X}_2 \sim N_n(\boldsymbol{\mu}_2, \Sigma_{22})$, and let $\mathbf{X}_1$ and $\mathbf{X}_t$ be independent, then

$$\begin{pmatrix} \mathbf{X}_1 \\ \cdots \\ \mathbf{X}_2 \end{pmatrix} \sim N_{m+n} \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \cdots \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \vdots & \mathbf{0} \\ \cdots & \cdots & \cdots \\ \mathbf{0} & \vdots & \Sigma_{22} \end{pmatrix} \right]$$

   If we combine m-dimensional multivariate normally distributed components and n-dimensional multivariate normally distributed components into a $(m+n)$-single vector, the result is not necessarily multivariate normal. However, if the components are multivariate normal and independent, then the result will be multivariate normal.

10. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, where

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \cdots \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \cdots \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \cdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix}$$

with $|\Sigma_{22}| > 0$ then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is multivariate normal with mean

$$\mu_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$$

and covariance

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Notice that the conditional covariance does not depend on the value of the conditioning variable.

This property is very important and widely used. For example, it is used for the theoretical justification of a Kalman Filter in time series analysis and the forward filtering backward sampling (FFBS) algorithm in Markov Chain Monte Carlo methods.

11. **Diagonalization** Let $\mathbf{X} \sim N_p(\mathbf{0}, \Sigma)$. We know that Σ is symmetric and positive definite. Hence, there exists an orthogonal matrix P such that $P^T\Sigma P = \Lambda$ where Λ is a diagonal matrix. (Note that P is formed by orthogonal eigenvectors of Σ and Λ has eigenvalues of Σ on its main diagonal.) Then if $\mathbf{Y} = P^T\mathbf{X}$, we have $\mathbf{Y} \sim N_p(\mathbf{0}, \Lambda)$. I.e., $\mathbf{Y}$ is a Gaussian vector and has independent components. This method of producing independent Gaussians has several important applications. One of these is the **principal component analysis**.

12. If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$

$$(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$$

To prove consider use the **diagonalization** concept.