Multivariate Statistics

Lecture 02

Fudan University

Outline

Outline

Joint Distributions (Two Variables)

1. Consider two (real) random variables X and Y. Probabilities of events defined in terms of these variables can be obtained by operations involving the cumulative distribution function (cdf),

$$F(x,y) = \Pr\{X \le x, Y \le y\}.$$

defined for every pair of real numbers (x,y).

2. We are interested in cases where $F(x,y)$ is absolutely continuous; this means the following partial derivative exists almost everywhere:

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

and we have

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\mathrm{d}u\mathrm{d}v$$

3. The nonnegative function $f(x,y)$ is called the probability density function (pdf).

Joint Distributions (Two Variables)

The pair of random variables (X, Y) defines a random point in a plane. The probability that (X, Y) falls in a rectangle is

$$
\begin{aligned}
&\Pr\{x \le X \le x + \Delta x,\ y \le Y \le y + \Delta y\}\\
=&F(x + \Delta x, y + \Delta y) - F(x + \Delta x, y) - F(x, y + \Delta y) + F(x, y)\\
=&\int_{y}^{y+\Delta x}\int_{x}^{x+\Delta y} f(u, v)\mathrm{d}u\mathrm{d}v,
\end{aligned}
$$

where $\Delta x > 0$ and $\Delta y > 0$.

The probability of the random point (X, Y) falling in any set $\mathcal{E}$ for which the following integral is defined (that is, any measurable set $\mathcal{E}$) is

$$\Pr\{(X, Y) \in \mathcal{E}\} = \iint_{\mathcal{E}} f(x, y)\mathrm{d}u\mathrm{d}v.$$

Joint Distributions (Two Variables)

If $f(x, y)$ is continuous in both two variables, the probability element $f(x, y)\Delta x \Delta y$ is approximately the probability that X falls between x and $x + \Delta x$ and Y falls between y and $y + \Delta y$ for small Δx and Δy since

$$\begin{aligned} &\Pr\{x \le X \le x + \Delta x,\ y \le Y \le y + \Delta y\} \\ =&\int_y^{y+\Delta x} \int_x^{x+\Delta y} f(u, v)\mathrm{d}u\mathrm{d}v \\ =&f(x_0, y_0)\Delta x \Delta y \end{aligned}$$

for some x_0, y_0 such that $x \le x_0 \le x + \Delta x$, $y \le y_0 \le y + \Delta y$ by the mean value theorem. The continuity of f means $f(x_0, y_0)\Delta x \Delta y$ is approximately $f(x, y)\Delta x \Delta y$.

Joint Distributions (p Variables)

The cumulative distribution function of p random variables $X_1, \ldots X_p$ is

$$F(x_1, \ldots, x_p) = \Pr\{X_1 \le x_1, \ldots, X_p \le x_p\}.$$

If $F(x_1, \ldots, x_p)$ is absolutely continuous, its density function is

$$\frac{\partial^p F(x_1, \ldots, x_p)}{\partial x_1 \ldots \partial x_p} = f(x_1, \ldots, x_p)$$

(almost everywhere), and

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(u_1, \ldots, u_p) \mathrm{d}u_1 \ldots \mathrm{d}u_p.$$

Joint Distributions (p Variables)

The probability of falling in any (measurable) set $\mathcal{R}$ in the p-dimensional Euclidean space is

$$\Pr\{(X_1, \ldots, X_p) \in \mathcal{R}\} = \int \cdots \int_{\mathcal{R}} f(x_1, \ldots, x_p) \mathrm{d}x_1 \ldots \mathrm{d}x_p.$$

The probability element

$$f(x_1, \ldots, x_p) \Delta x_1 \ldots \Delta x_p$$

is approximately the probability

$$\Pr\{x_1 \leq X_1 \leq x_1 + \Delta_1, \ldots, x_p \leq X_p \leq x_p + \Delta_p\}$$

if $f(x_1, \ldots, x_p)$ is continuous.

Joint Moments

The joint moments of the joint distribution of random variables $X_1, \ldots, X_p$ are defined as integers

$$\mathbb{E}\left[X_1^{h_1} \cdots X_p^{h_p}\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{h_1} \cdots x_p^{h_p} f(x_1, \ldots, x_p) \mathrm{d}x_1 \ldots \mathrm{d}x_p.$$

where $k_i \geq 0$ for all $i = 1, \ldots, p$.

Outline

Marginal Distributions (two variables)

Given the cdf of two random variables X, Y as being $F(x, y)$, the marginal cdf of X is

$$F(x) = \Pr\{X \leq x\} = \Pr\{X \leq x, Y \leq \infty\} = F(x, \infty).$$

Clearly, we have

$$F(x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f(u, v) \mathrm{d}v \right) \mathrm{d}u.$$

We call

$$f(u) = \int_{-\infty}^{\infty} f(u, v) \mathrm{d}v,$$

say, the marginal density of X. Then

$$F(x) = \int_{-\infty}^{x} f(u) \mathrm{d}u.$$

Marginal Distributions (two variables)

In a similar fashion we define $G(y)$ as the marginal cdf of Y and $g(y)$ as marginal density of Y, that is

$$G(y) = \int_{-\infty}^{y} \left(\int_{-\infty}^{\infty} f(u, v) \mathrm{d}u \right) \mathrm{d}v.$$

and

$$g(v) = \int_{-\infty}^{\infty} f(u, v) \mathrm{d}u.$$

Marginal Distributions (p variables)

Given $F(x_1, \ldots, x_p)$ as the cdf of $X_1, \ldots, X_p$, the marginal cdf of some of $X_1, \ldots, X_p$ say, of $X_1, \ldots, X_r$ $(r < p)$, is

$$\begin{aligned} F(X_1, \ldots, X_r) &= \Pr\{X_1 \le x_1, \ldots, X_r \le x_r\} \\ &= \Pr\{X_1 \le x_1, \ldots, X_r \le x_r, X_{r+1} \le \infty, \ldots, X_p \le \infty\} \\ &= F(x_1, \ldots, x_r, \infty, \ldots, \infty). \end{aligned}$$

The marginal density of $X_1, \ldots, X_r$ is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_r, u_{r+1} \ldots, u_p) \mathrm{d}u_{r+1} \ldots \mathrm{d}u_p.$$

The marginal distribution and density of any other subset of $X_1, \ldots, X_p$ are obtained in the obviously similar fashion.

Joint Moments

The joint moments of a subset of variables can be computed from the marginal distribution; for example,

$$
\begin{aligned}
&\mathbb{E}\left[X_1^{h_1}\cdots X_r^{h_r}\right] \\
=&\mathbb{E}\left[X_1^{h_1}\cdots X_r^{h_r}X_{r+1}^0\ldots X_p^0\right] \\
=&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{h_1}\cdots x_r^{h_r} f(x_1,\ldots,x_p)\mathrm{d}x_1\ldots\mathrm{d}x_p \\
=&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{h_1}\cdots x_r^{h_r}\left[\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1\ldots,x_p)\mathrm{d}x_{r+1}\ldots\mathrm{d}x_p\right]\mathrm{d}x_1\ldots\mathrm{d}x_r \\
=&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{h_1}\cdots x_r^{h_r} f(x_1,\ldots,x_r)\mathrm{d}x_1\ldots\mathrm{d}x_r.
\end{aligned}
$$

Statistical Independence

Two random variables X, Y with cdf $F(x,y)$ are said to be independent if

$$F(x,y) = F(x)G(y),$$

where $F(x)$ is the marginal cdf of X and $G(y)$ is the marginal cdf of Y.

This implies the density of X, Y can be written as

$$f(x,y) = f(x)g(y),$$

where $f(x)$ and $g(y)$ are the marginal densities of X and Y respectively.

Conversely, if $f(x,y) = f(x)g(y)$, then $F(x,y) = F(x)G(y)$.

Statistical Independence

The statistical independence of X and Y implies

$$\begin{aligned}
&\Pr\{x_1 \le X \le x_2, y_1 \le Y \le y_2\} \\
=& \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(u, v) \mathrm{d}u \mathrm{d}v \\
=& \int_{y_1}^{y_2} f(u) \mathrm{d}u \int_{x_1}^{x_2} g(v) \mathrm{d}v \\
=& \Pr\{x_1 \le X \le x_2\} \Pr\{y_1 \le Y \le y_2\}.
\end{aligned}$$

Note that we say X and Y are uncorrelated if

$$\begin{aligned}
& \mathrm{Cov}(X, Y) \triangleq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = 0 \\
\iff & \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

Independent $\neq$ Uncorrelated

Note that

X are Y are independent implies X are Y uncorrelated.

However,

X are Y are uncorrelated do NOT implies X are Y are independent.

Mutually Independence

If the cdf of $X_1, \dots, X_p$ is $F(x_1, \dots, x_p)$, the set of random variables is said to be mutually independent if

$$F(x_1, \dots, x_p) = F_1(x_1) \dots F(x_p),$$

where $F_i(x_i)$ is the marginal cdf of X_i, $i = 1, \dots, p$.

The set $X_1, \dots, X_r$ is said to be independent of the set $X_{r+1}, \dots, X_p$ if

$$F(x_1, \dots, X_p) = F(x_1, \dots, x_r, \infty, \dots, \infty) F(\infty, \dots, \infty, x_{r+1}, \dots, x_p).$$

Conditional Distributions

If A and B are two events such that the probability of A and B occurring simultaneously is $P(AB)$ and the probability of B occurring is $P(B) > 0$, then the conditional probability of A occurring given that B has occurred is

$$\frac{P(AB)}{P(B)}.$$

Conditional Distributions

Suppose the event A is X falling in the $[x_1, x_2]$ and the event B is Y falling in $[y_1, y_2]$. Then the conditional probability that X falls in $[x_1, x_2]$, given that Y falls in $[y_1, y_2]$, is

$$\begin{aligned} &\Pr\{x_1 \leq X \leq x_2 \mid y_1 \leq Y \leq y_2\} \\ =&\frac{\Pr\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\}}{\Pr\{y_1 \leq Y \leq y_2\}} \\ =&\frac{\int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u, v)\mathrm{d}v\mathrm{d}u}{\int_{y_1}^{y_2} g(v)\mathrm{d}v}. \end{aligned}$$

Conditional Distributions

For y such that $g(y) > 0$, we define $\Pr\{x_1 \le X \le x_2 \mid Y = y\}$ as the probability that X lies between x_1 and x_2 given that Y is y. Then

$$\Pr\{x_1 \le X \le x_2 \mid Y = y\} = \int_{x_1}^{x_2} f(u \mid y)\mathrm{d}u,$$

where $f(u \mid y) = \dfrac{f(u, y)}{g(y)}$.

For given y, $f(\cdot \mid y)$ is a density function and is called the conditional density of X given y.

If X and Y are independent, we have $f(x \mid y) = f(x)$.

Conditional Distributions

In the general case of $X_1, \ldots, X_p$ with cdf $F(X_1, \ldots, X_p)$, the conditional density of $X_1, \ldots, X_r$, given $X_{r+1} = x_{r+1}, \ldots, X_p = x_p$ is

$$\frac{f(x_1, \ldots, x_p)}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_r, x_{r+1}, \ldots, x_p)} \mathrm{d}u_1 \cdots \mathrm{d}u_r.$$

Outline

Transformation of Variables

Let the density of p dimensional random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top$ be $f(\mathbf{x})$.

Consider the random vector p dimensional random vector $\mathbf{y} = [y_1, \ldots, y_p]^\top$ such that $y_i = u_i(\mathbf{x})$ for $i = 1, \ldots, p$. Let the density function of $\mathbf{y}$ be $g(\mathbf{y})$.

Assume the transformation $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), \ldots, u_p(\mathbf{x})]^\top : \mathbb{R}^p \to \mathbb{R}^p$ from the space of $\mathbf{x}$ to the space of $\mathbf{y}$ is smooth and one-to-one.

Then we have $f(\mathbf{x}) = g(\mathbf{u}(\mathbf{x}))|\det(\mathbf{J}(\mathbf{x}))|$ where

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial u_1(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_1(\mathbf{x})}{x_2} & \cdots & \dfrac{\partial u_1(\mathbf{x})}{\partial x_p} \\ \dfrac{\partial u_2(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_2(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial u_2(\mathbf{x})}{\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial u_p(\mathbf{x})}{\partial x_1} & \dfrac{\partial u_p(\mathbf{x})}{\partial x_2} & \cdots & \dfrac{\partial u_p(\mathbf{x})}{\partial x_p} \end{bmatrix}.$$

Transformation of Variables

Similarly, we also have $g(\mathbf{y}) = f(\mathbf{u}^{-1}(\mathbf{y}))|\det(\mathbf{J}^{-1}(\mathbf{y}))|$ where

$$\mathbf{J}^{-1}(\mathbf{y}) = \begin{bmatrix} \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_1^{-1}(\mathbf{y})}{\partial y_p} \\ \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_2^{-1}(\mathbf{y})}{\partial y_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_1} & \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial u_p^{-1}(\mathbf{y})}{\partial y_p} \end{bmatrix}.$$

Outline

Random Matrix

A random matrix

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \ddots & \cdots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

is a matrix of random variables $z_{11}, \ldots, z_{mn}$.

Random Matrix

We define

$$\mathbb{E}[\mathbf{Z}] = \begin{bmatrix} \mathbb{E}[z_{11}] & \mathbb{E}[z_{12}] & \dots & \mathbb{E}[z_{1n}] \\ \mathbb{E}[z_{21}] & \mathbb{E}[z_{22}] & \dots & \mathbb{E}[z_{2n}] \\ \vdots & \ddots & \dots & \vdots \\ \mathbb{E}[z_{m1}] & \mathbb{E}[z_{m2}] & \dots & \mathbb{E}[z_{mn}]. \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Random Vector and Mean Vector

For random vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \in \mathbb{R}^p,$$

the expected value

$$\mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \mathbb{E}[x_2] \\ \vdots \\ \mathbb{E}[x_p] \end{bmatrix} \in \mathbb{R}^p,$$

is the mean or mean vector of $\mathbf{x}$.

We shall usually denote the mean vector $\mathbb{E}[\mathbf{x}]$ by $\boldsymbol{\mu}$.

Random Vector and Covariance Matrix

For random vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$ and its mean vector $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$, the expected value of the random matrix $(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top$ is

$$\mathrm{Cov}(\mathbf{x}) = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\right],$$

the covariance or covariance matrix of $\mathbf{x}$.

1. The i-th diagonal element of this matrix, $\mathbb{E}\left[(x_i - \mu_i)^2\right]$, is the variance of x_i.
2. The i, j-th off-diagonal element $(i \neq j)$, $\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$ is the covariance of x_i and x_j.

Random Vector and Covariance Matrix

Note that

$$
\begin{aligned}
\operatorname{Cov}(\mathbf{x}) =& \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\right] \\
=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top - \boldsymbol{\mu}\mathbf{x}^\top - \mathbf{x}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right] \\
=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \mathbb{E}\left[\boldsymbol{\mu}\mathbf{x}^\top\right] - \mathbb{E}\left[\mathbf{x}\boldsymbol{\mu}^\top\right] + \mathbb{E}[\boldsymbol{\mu}\boldsymbol{\mu}^\top] \\
=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \boldsymbol{\mu}\mathbb{E}\left[\mathbf{x}^\top\right] - \mathbb{E}\left[\mathbf{x}\right]\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
=& \mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\top,
\end{aligned}
$$

where we have used the following lemma:

Lemma

If $\mathbf{Z}$ is an $m \times n$ random matrix, $\mathbf{D}$ is a fixed $l \times m$ real matrix, $\mathbf{E}$ is a fixed $n \times q$ real matrix, and $\mathbf{F}$ is a fixed $l \times q$ real matrix, then

$$\mathbb{E}[\mathbf{DZE}+\mathbf{F}] = \mathbf{D}\mathbb{E}[\mathbf{Z}]\mathbf{E}+\mathbf{F}.$$

Outline

Univariate Normal Distribution

A random variable X is normally distributed with mean μ and standard deviation $\sigma > 0$ can be written in the following notation

$$X \sim \mathcal{N}(\mu, \sigma).$$

The probability density function of univariate normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The standard normal distribution is a normal distribution with a mean of 0 and standard deviation of 1.

The Central Limit Theorem

The sum of many random variables will have an approximately normal distribution.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with the same arbitrary distribution, zero mean, and variance σ^2.

Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the random variable

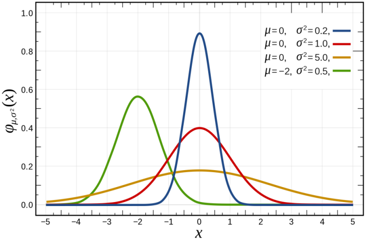
$$Z = \lim_{n\to\infty} \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

is a standard normal distribution.

What about multivariate case?

Normal Distribution

正态分布



~~正太分布~~

词语起源

播报 编辑

"正太"一词最初出现于日本《ファンロード(Fanroad)》杂志中的"Q&A"栏目。在该栏目中，当被问及「喜欢男孩的女性应该被称作什么」时，该杂志的编辑"あるイニシャル・K"回答「喜欢"正太郎"的正太控(ショタコン)」。[2]

该回答所提及的"正太郎"，源于漫画家横山光辉的作品《铁人28号》主角"金田正太郎"的名字。[2]

此后，"正太控"一词开始流行。在传播过程中，"正太控"中的"正太"二字逐渐被分离出来，成为了形容"年龄小的男生"的词汇。[2]

金田正太郎

Multivariate Normal Distribution

The multivariate normal distribution of a p-dimensional random vector $\mathbf{x} = [x_1, \ldots, x_p]^\top$ can be written in the following notation:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

or to make it explicitly known that $\mathbf{x}$ is p-dimensional.

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with p-dimensional mean vector

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}] = \begin{bmatrix} \mathbb{E}[x_1] \\ \vdots \\ \mathbb{E}[x_p] \end{bmatrix} \in \mathbb{R}^p$$

and covariance matrix

$$\boldsymbol{\Sigma} = \mathbb{E}\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top\right] \in \mathbb{R}^{p \times p}.$$

Multivariate Normal Distribution

The density function of univariate normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

where μ is the mean and σ^2 is the variance with $\sigma > 0$.

The density function of non-singular p-dimensional multivariate normal distribution is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean and $\boldsymbol{\Sigma} \succ \mathbf{0}$ is the $p \times p$ covariance matrix.

Multivariate Normal Distribution

The density function of non-singular p-dimensional multivariate normal distribution is

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean and $\boldsymbol{\Sigma} \succ \mathbf{0}$ is the $p \times p$ covariance matrix.

When the covariance matrix $\boldsymbol{\Sigma}$ is singular, we call the distribution is singular (degenerate) normal distribution and we cannot write its density function.

We first focus on the case of $\boldsymbol{\Sigma} \succ \mathbf{0}$.

How to obtain the pdf of multivariate normal distribution?

We generalize the form of pdf for univariate normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

to the multivariate case

$$f(\mathbf{x}) = K \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})^\top \mathbf{A}(\mathbf{x}-\mathbf{b})\right),$$

where $\mathbf{A}$ is symmetric positive definite.

We can verify that if $\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$ and $\mathrm{Cov}[\mathbf{x}] = \boldsymbol{\Sigma}$, then

$$K = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}}, \quad \mathbf{b} = \boldsymbol{\mu}, \quad \mathbf{A} = \boldsymbol{\Sigma}^{-1}.$$

How to obtain the pdf of multivariate normal distribution?

We first show

$$K = \frac{\sqrt{\det(\mathbf{A})}}{\sqrt{(2\pi)^p}}$$

by considering the random vector

$$\mathbf{y} = \mathbf{C}^{-1}(\mathbf{x} - \mathbf{b}) \in \mathbb{R}^p,$$

where $\mathbf{C} \in \mathbb{R}^{p \times p}$ satisfies $\mathbf{C}^\top \mathbf{A} \mathbf{C} = \mathbf{I}$.

How to obtain the pdf of multivariate normal distribution?

Then, we show $\mathbf{b} = \boldsymbol{\mu}$ and $\mathbf{A} = \boldsymbol{\Sigma}^{-1}$ by using the following lemma.

Lemma

1. If $\mathbf{Z}$ is an $m \times n$ random matrix, $\mathbf{D}$ is an $l \times m$ real matrix, $\mathbf{E}$ is an $n \times q$ real matrix, and $\mathbf{F}$ is an $l \times q$ real matrix, then

$$\mathbb{E}[\mathbf{DZE} + \mathbf{F}] = \mathbf{D}\mathbb{E}[\mathbf{Z}]\mathbf{E} + \mathbf{F}.$$

2. If $\mathbf{y} = \mathbf{Dx} + \mathbf{f} \in \mathbb{R}^l$, where $\mathbf{D}$ is an $l \times m$ real matrix, $\mathbf{x} \in \mathbb{R}^m$ is a random vector, then

$$\mathbb{E}[\mathbf{y}] = \mathbf{D}\mathbb{E}[\mathbf{x}] + \mathbf{f}$$

and

$$\mathrm{Cov}[\mathbf{y}] = \mathbf{D}\mathrm{Cov}[\mathbf{x}]\mathbf{D}^\top.$$

Multivariate Normal Distribution

If the density of a p-dimensional random vector $\mathbf{x}$ is

$$K \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})^\top \mathbf{A}(\mathbf{x}-\mathbf{b})\right),$$

where $\mathbf{A} \in \mathbb{R}^{p\times p}$ is symmetric positive definite. Then the expectation of $\mathbf{x}$ is $\mathbf{b}$ and its covariance matrix is $\mathbf{A}^{-1}$.

Conversely, given a vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and a positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, there is a multivariate normal density

$$n(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right).$$

Correlation Coefficient

We consider the bivariate normal distribution $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

The covariance matrix can be written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

where σ_1^2 is the variance of x_1, σ_2^2 is the variance of x_2 and ρ is the correlation between x_1 and x_2.

We can verify that $-1 < \rho < 1$ if $\boldsymbol{\Sigma} \succ \mathbf{0}$ and

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{bmatrix}$$

Correlation Coefficient

The density of such normal distribution is constant on ellipsoids

$$(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c$$

for every positive value of c.

We transform coordinates by $y_i = (x_i - \mu_i)/\sigma_i$ for $i = 1, 2$, then

$$\frac{1}{1-\rho^2}\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right) = c.$$

Correlation Coefficient

We transform coordinates by $y_i = (x_i - \mu_i)/\sigma_i$ for $i = 1, 2$, then

$$\frac{1}{1-\rho^2}\left(y_1^2 - 2\rho y_1 y_2 + y_2^2\right) = c.$$

The intercepts on the y_1-axis and y_2-axis are equal.

1. If $\rho > 0$, the major axis is along the 45° line with a length of $2\sqrt{c(1+\rho)}$, and the minor axis has a length of $2\sqrt{c(1-\rho)}$.
2. If $\rho < 0$, the major axis is along the 135° line with a length of $2\sqrt{c(1-\rho)}$, and the minor axis has a length of $2\sqrt{c(1+\rho)}$.