Chapter 2 Random Vector

Content

§ 2.1 Univariate Distribution

- § 2.1.1. random variables and probability distribution function
- § 2.1.2. types of probability distribution
- § 2.1.3. expectation and variance of random variables
- § 2.1.4. some significant univariate distributions

§ 2.1.1. random variables and probability distribution function

- Random vectors are quantity characteristics of random events, denoted by x.
- **Probability Distribution Function** (概率分布函数)of random variable x: :

$$F(a) = P(x \le a)$$

- Basic properties of distribution function:

(1) F(x) is a non-decreasing function；
(2) Bounded： $0 \le F(x) \le 1$， $F(-\infty)=0$， $F(+\infty)=1$；
(3) Right-continuous, F(x+0)=F(x).

§ 2.1.2. kinds of probability distribution

1 Discrete Distribution

- If random variables x can only take finite or countable values, then x is called **Discrete Random Variable**(离散型随机变量)**.**
- Suppose the possible values of discrete random variable x：

$$x_1,\quad x_2,\quad \ldots,\quad x_n,\quad \ldots$$

$p_i = P(x = x_i)$, $i = 1, 2, \ldots\ldots$ is called the x **Distribution Sequence(**分布列).

- Basic properties of distribution sequence：

(1) $p_i \geq 0$ （Nonnegativity）；

(2) $\sum_{i=1}^{\infty} p_i = 1$(Orthogonality)。

2 Continuous Distribution

- If the distribution function of random variable x can be expressed as,

$$F(a)=\int_{-\infty}^{a} f(x)dx$$

Density Function

when it is set up for all $a \in R$, then x is called **Continuous Random Variable**连续型随机变量.

- Properties of density function $f(x)$:

(1) $f(x) \geq 0$;

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

- Practice: 1 Are the following distribution sequences of discrete random variables are right?

(1) X: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.1 & 0.6 & 0.1 & 0.1 \end{pmatrix}$

(2) X: $\begin{pmatrix} 1 & 2 & 3 & \cdots \\ \dfrac{1}{1\times 2} & \dfrac{1}{2\times 3} & \dfrac{1}{3\times 4} & \cdots \end{pmatrix}$

2 Are the following functions are distribution functions?

(1)

$$f(x)=\begin{cases} 1/(1+x^2) & x>0 \\ 0 & x\le 0 \end{cases}$$

(2)

$$f(x)=\begin{cases} e^{-(x-a)} & x>a \\ 0 & \text{others} \end{cases}$$

§ 2.1.3. expectation and variance of random variables

- Discrete Random Variable:

(1) $$\mu = E(x) = \sum_{k=1}^{\infty} a_k p_k$$

(2) $$\sigma^2 = V(x) = E(x-\mu)^2 = \sum_{k=1}^{\infty} (a_k - \mu)^2 p_k$$

- Continuous Random Variable:

(1) $$\mu = E(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

(2) $$\sigma^2 = V(x) = E(x-\mu)^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x) dx$$

Properties of Expectation

- (1) c is a constant, $E(c) = c$.
- (2)Suppose x and y are two independent random variables, then

$$E(xy) = E(x)\,E(y)$$

- (3)Suppose k is a constant and x is a random variable, then

$$E(kx) = k\,E(x)$$

- (4)Suppose $x_1, x_2, \ldots, x_n$ are n random variables, then

$$E(x_1 + x_2 + \ldots + x_n) = E(x_1) + E(x_2) + \ldots + E(x_n)$$

Properties of Variance

- (1) Suppose c is a constant, then $V(c) = 0$.
- (2) Suppose k is a constant and x is a random variable, then

$$V(kx) = k^2 V(x)$$

- (3) Suppose $x_1, x_2, \ldots, x_n$ is n mutually independent random variables, then

$$V(x_1 + x_2 + \ldots + x_n) = V(x_1) + V(x_2) + \ldots + V(x_n)$$

Question: If $x_1, x_2, \ldots, x_n$ are not mutually independent, then

$$V(x_1 + x_2 + \ldots + x_n) = ?$$

§ 2.1.4. some significant univariate distributions

- (1) Binomial Distribution $x \sim b(n, p)$
- (2) Hypergeometric Distribution $x \sim H(M, N, n)$
- (3) Poisson Distribution $x \sim P(\lambda)$
- (4) Normal Distribution $x \sim N(\mu, \sigma^2)$
- (5) Chi-square Distribution $x \sim \chi^2(n)$
- (6) *t*-Distribution $t \sim t(n)$
- (7) *F*-Distribution $F \sim F(n, m)$

- Consider n time Bernoulli tests, also n independent tests with the successful probability is p while the failing probability is q=1-p, 0<p<1. The number of successful among n time Bernoulli tests, X, is **Binomial Distributed**(二项分布), denoted by $X \sim b(n, p)$
- There are N balls in a bag. M of them are black. Others are white. Now we take n balls from the bag randomly $n \le N$. Let the number of black balls among the n balls be X. Then, X is a discrete random variable and values among $\{0,1,2\cdots,\min(n,M)\}$,

Then the distribution sequence of X is:

$$P(X = k) = C_M^k C_{N-M}^{n-k} / C_N^n, \ \mathrm{k} = 0,1,2\cdots,\min(n,M).$$

This distribution law is called **Hypergeometric Distribution**(超几何分布), denoted by $X \sim H(M,N,n)$

- Suppose a discrete random variable values among $\{0,1,2,\cdots, \quad n,\cdots\}$,

$$P(X=n)=\frac{\lambda^n}{n!}e^{-\lambda}, n=0,1,2,\cdots, \qquad \lambda>0.$$

The above kind of distribution is called **Poisson Distribution**(泊松分布)**with parameter** λ .

For instance, The call frequency of telephone exchange is Poisson distributed.

- Suppose random variables $x_1, x_2, \cdots, x_n$ are mutually independent and commonly distributed with N(0,1), then the random variable $z_n=\sum_{i=1}^{n} x_i^2$ obeys **Chi-square Distribution(卡方分布)**, denoted by $z_n \sim \chi^2(n)$.

Properties:

$$E(z_n)=n, \qquad Var(z_n)=2n$$

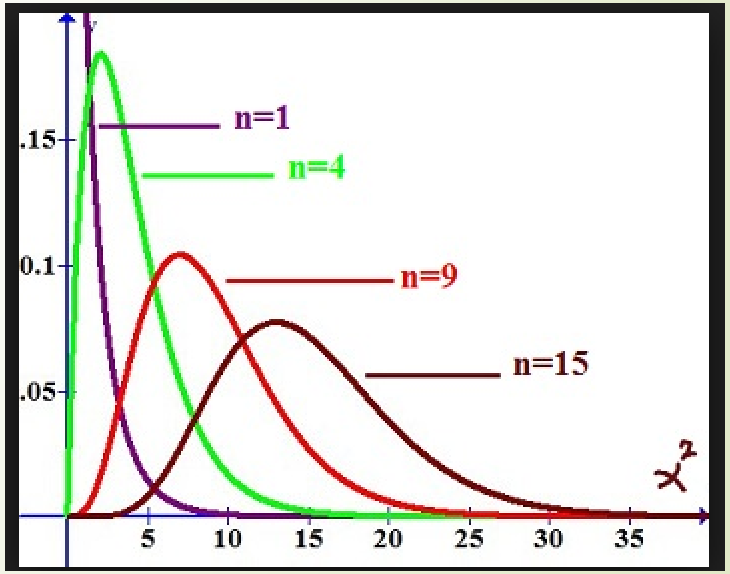
- Suppose random variables $z_n = \sum_{i=1}^{n} x_i^2$ are mutually independent and commonly distributed with N(0,1), then the random variable

obeys **Chi-square Distribution(卡方分布)**, denoted by $z_n \sim \chi^2(n)$.

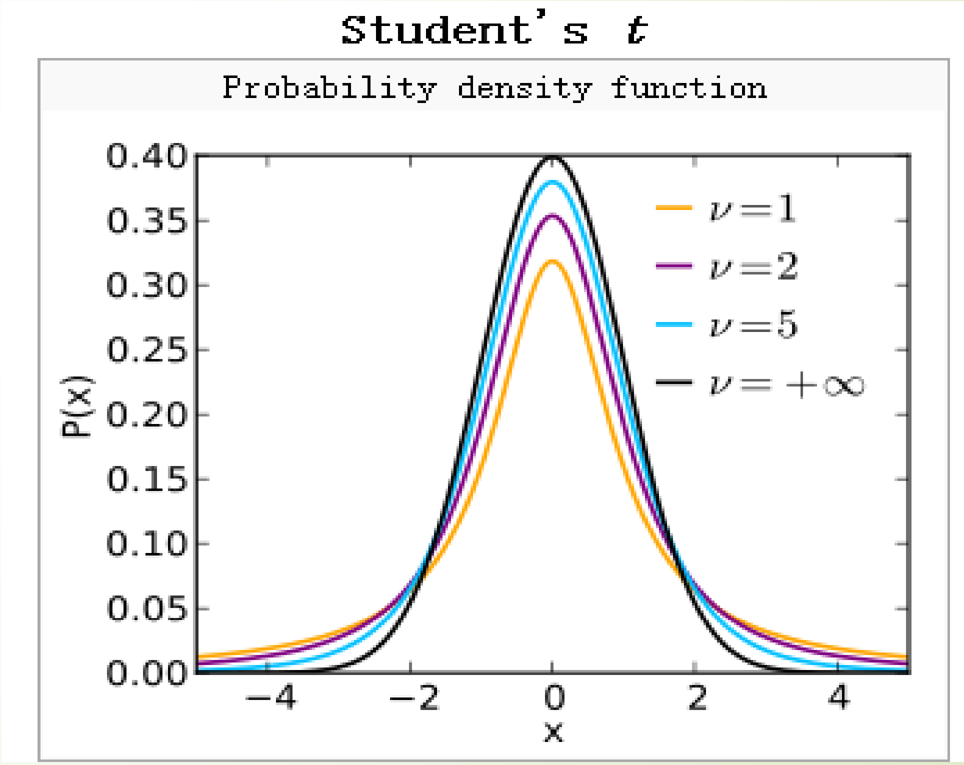
Properties: $x_1, x_2, \cdots, x_n$

$$E(z_n) = n, \qquad Var(z_n) = 2n$$

Chisquare square: probability density function



- Suppose X and Y are mutually independent, X is a standard normal distribution, Y is a chi-square variable with free degree n. Let $Z = X / \sqrt{\frac{Y}{n}}$, then t-variable Z obeys **t-Distribution(t分布)** with free degree n, denoted by $Z \sim t(n)$.



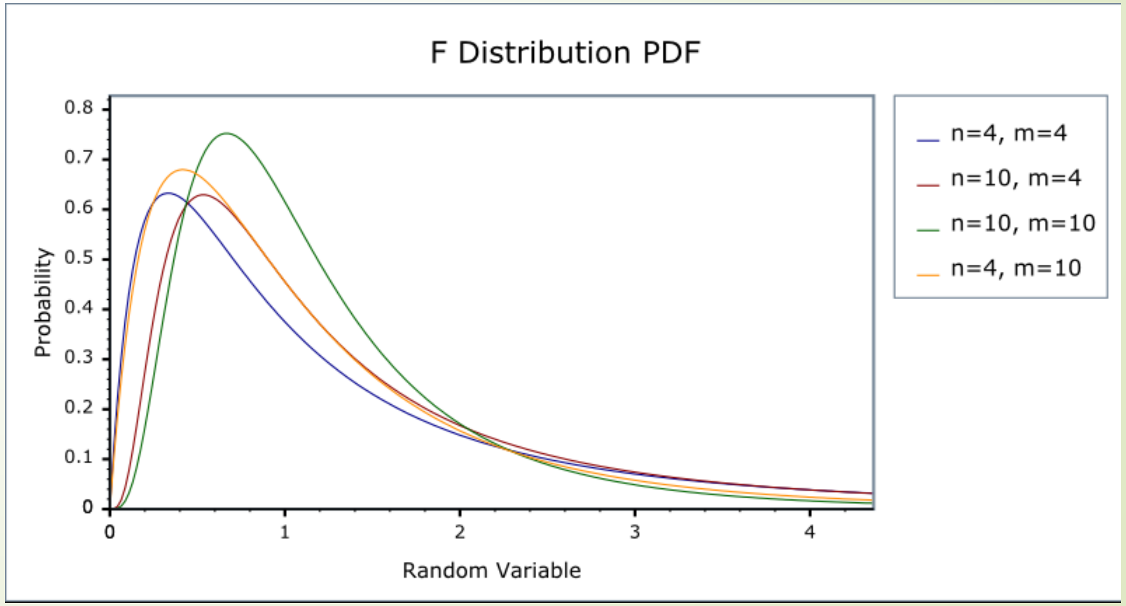
- Suppose X and Y are mutually independent, X is a χ^2 -variable with free degree n, Y is χ^2 - variable with free degree m.Let

$$Z = \frac{X}{n} / \frac{Y}{m},$$

Then F-variable Z obeys **F-Distribution(F分布)**, with free degree n and m, denoted by $Z \sim F(n, m)$.

Property: $F_{1-\alpha}(n,m) = \dfrac{1}{F_\alpha(n,m)}$, （2.1.15）

where $F_\alpha(n,m)$ is listed on F-distribution table



§ 2.2 Multivariate Distribution

- § 2.2.1 multivariate probability distribution
- § 2.2.2 two common discrete multivariate distribution
- § 2.2.3 multivariate probability density function
- § 2.2.4 marginal distribution
- § 2.2.5 conditional distribution
- § 2.2.6 independency

§ 2.2.1 multivariate probability distribution

- A vector is called **Random Vector(随机向量)** if all components of it are random variables.
- **Distribution Function(分布函数)** of random variable *x*:

$$F(a) = P(x \le a)$$

- Distribution function of random vector $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$:

$$F(a_1, a_2, \cdots, a_p) = P(x_1 \le a_1, x_2 \le a_2, \cdots, x_p \le a_p)$$

- Properties of multivariate distribution functions:

(i) $F(x_1, x_2, \cdots, x_p)$ is non-decreasing and right-continuous function for each variable $x_i\,(i=1,\cdots,p)$;

(ii) $0 \le F(x_1, x_2, \cdots, x_p) \le 1$

(iii) $F(-\infty, x_1, x_2, \cdots, x_p) = F(x_1, -\infty, x_3, \cdots, x_p) = F(x_1, x_2, \cdots, x_{p-1}, -\infty,) = 0$

(iv) $F(\infty, \infty, \cdots, \infty) = 1.$

§ 2.2.2 two common discrete multivariate distribution

- **1、Multinomial(多项) Distribution**

Example (throwing a dice)： A dice has 6 sides with 6 counts respectively, from 1 to 6. Suppose the probabilities of occurrence of every side are denoted by (such as a irregular shaped dice) $p_1, \cdots, p_6$. Throw the dice repeatedly n times. Calculate the probability when the number of occurrence times of 1 to 6 are respectively $k_i \, (i = 1, \cdots, 6)$?

Solution: Let x_i denote the number of occurrence times of the ith side after throwing the dice n times, then the probability of the above event can be showed as $p(x_1 = k_1, x_2 = k_2, \cdots, x_6 = k_6) = ?\ C_n^{k_1} p_1^{k_1} C_{n-k_1}^{k_2} p_2^{k_2} \cdots C_{n-k_1-k_2-k_5}^{k_6} p_6^{k_6}$

where $\sum_{i=1}^{6} k_i = n, \sum_{i=1}^{n} p_i = 1.$

- If the distribution sequence of $\mathbf{x} = (x_1, x_2, \cdots, x_m)'$ is as follows:

$$P(x_1 = k_1, x_2 = k_2, \cdots, x_m = k_m) = \frac{n!}{k_1! k_2! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m}$$

$$k_i = 0, 1, \cdots, n \qquad i = 1, 2, \cdots, m \qquad k_1 + k_2 + \cdots + k_m = n$$

$$\text{where} 0 < p_i < 1, \qquad i = 1, 2, \cdots, m \qquad p_1 + p_2 + \cdots + p_m = 1$$

Then $\mathbf{x} = (x_1, x_2, \cdots, x_m)'$ obeys multinomial distribution.

2、Multivariate Hypergeometric Distribution

Example(摸麻将): A pair of mahjong include 34 different kinds of cards in addition to "flower" and the number of each color card is four, so a mahjong content a total of 136 cards. Assume each player gets 13 cards on the play.

What is the probability distribution sequence of the event?

Solution: Number the 34 kinds of cards as1,2,…,34. Let x_i denote the number of times when the ith card appears(i=1,2…,34, random variables).

Therefore, the player’s cards can be described by the value of random vector $x=(x_1=k_1,x_2=k_2,\cdots,x_{34}=k_{34})'$, i.s.

$$p(x_1=k_1,x_2=k_2,\cdots,x_{34}=k_{34})=?\quad \frac{\binom{4}{k_1}\cdots\binom{4}{k_{34}}}{\binom{136}{13}},k_i=0,1,2,3,4,i=1,\cdots,34,\sum_{i=1}^{13}k_i=13$$

The above distribution sequence obeys

Multivariate Hypergeometric Distribution(多元超几何分布)

Multivariate Hypergeometric Distribution(摸球模型)

There are N balls in some bag with the number of ball $i(i=1,\cdots,m)$ being N_i such that $N_i+\cdots N_m=N$. One hopes to draw n balls randomly from this bag. Let x_i be the number of occurrence of the i-th ball , find the distribution of $\mathrm{x}=(x_1,x_2,\cdots,x_m)'$:

$$P(x_1=k_1,x_2=k_2,\cdots,x_m=k_m)=\frac{\binom{N_1}{k_1}\cdots\binom{N_m}{k_m}}{\binom{N}{n}}$$

$$k_i=0,1,\cdots\min(n,N_i)\quad i=1,2,\cdots,m \qquad k_1+k_2+\cdots+k_m=n$$

$$N_1+N_2+\cdots+N_m=N$$

We call $\mathbf{x}=(x_1,x_2,\cdots,x_m)'$ obeys **Multivariate Hypergeometric Distribution**.

§ 2.2.3 multivariate probability density function

- Univariate:

$$F(a)=\int_{-\infty}^{a} f(x)\,\mathrm{d}x, \qquad f(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

- Multivariate:

$$F(a_1,\cdots,a_p)=\int_{-\infty}^{a_1}\cdots\int_{-\infty}^{a_p} f(x_1,\cdots,x_p)\,\mathrm{d}x_1\cdots\mathrm{d}x_p$$

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

- Properties of $f(x_1,\cdots,x_p)$:

(1) $f(x_1,\cdots,x_p)\geq 0$, 对一切实数$x_1,\cdots,x_p$;

(2) $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\cdots,x_p)\,\mathrm{d}x_1\cdots\mathrm{d}x_p=1$。

§ 2.2.4 marginal distribution

- Suppose $\boldsymbol{x}$ is a p-dimensional random vector. The distribution of subvector $\boldsymbol{x}_{(1)}$ that consists of $\boldsymbol{x}$'s $q(<p)$ components is called $\boldsymbol{x}$ to $\boldsymbol{x}_{(1)}$ **Marginal Distribution(边缘分布), where** $x = \begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix}$
- For continuous distribution, The density function of $\boldsymbol{x}_{(1)} = \left(x_1, \cdots, x_q\right)'$

$$f_{(1)}(x_1, \cdots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_p) \mathrm{d} x_{q+1} \cdots \mathrm{d} x_p$$

$$F_{(1)}(a_1, \cdots, a_q) = P(x_1 \leq a_1, \cdots, x_q \leq a_q)$$

$$= P(x_1 \leq a_1, \cdots, x_q \leq a_q, x_{q+1} \leq +\infty, \cdots, x_p \leq +\infty)$$

$$= \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_q} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \cdots, x_p) dx_1 \cdots dx_p$$

$$= \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_q} f_{(1)}(x_1, \cdots, x_q) dx_1 \cdots dx_q$$

§ 2.2.5 conditional distribution

- Suppose $\boldsymbol{x} = \left(x_1, \cdots, x_p\right)'$ is a continuous *p*- random vector. Given the condition $\boldsymbol{x}_{(2)} = \left(x_{q+1}, \cdots, x_p\right)' \left(f_{(2)}\left(\boldsymbol{x}_{(2)}\right) > 0 \right)$, the **Conditional Density** of (条件密度) $\boldsymbol{x}_{(1)} = \left(x_1, \cdots, x_q\right)'$ is defined as

$$f\left(x_1, \cdots, x_q \mid x_{q+1}, \cdots, x_p\right) = \frac{f\left(x_1, \cdots, x_p\right)}{f_{(2)}\left(x_{q+1}, \cdots, x_p\right)}$$

or denoted by

$$f\left(\boldsymbol{x}_{(1)} \mid \boldsymbol{x}_{(2)}\right) = \frac{f\left(\boldsymbol{x}\right)}{f_{(2)}\left(\boldsymbol{x}_{(2)}\right)}$$

Example 2.2.3 Suppose the p.d.f. of $x=(x_1,x_2)'$ is denoted by

$$f(x_1,x_2)=\begin{cases}\frac{6}{5}x_1^2(4x_1x_2+1), & 0<x_1<1, 0<x_2<1\\ 0, & \text{others}\end{cases}$$

Now try to calculate the conditional density function $f(x_1|x_2)$ and $f(x_2|x_1)$.

Solution:

$$f_1(x_1)=\int_{-\infty}^{+\infty} f(x_1,x_2)dx_2=\int_0^1\frac{6}{5}x_1^2(4x_1x_2+1)dx_2$$

$$=\frac{24}{5}x_1^3\int_0^1 x_2dx_2+\frac{6}{5}x_1^2=\frac{12}{5}x_1^3+\frac{6}{5}x_1^2, 0<x_1<1$$

$$f_2(x_2)=\int_{-\infty}^{+\infty} f(x_1,x_2)dx_1=\int_0^1\frac{6}{5}x_1^2(4x_1x_2+1)dx_1$$

$$=\frac{24}{5}x_2\int_0^1 x_1^3dx_1+\frac{6}{5}\int_0^1 x_1^2dx_1$$

$$=\frac{6}{5}x_2+\frac{2}{5}, 0<x_2<1$$

Therefore for $0 < x_2 < 1,$

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{3x_1^2(4x_1x_2 + 1)}{3x_2 + 1}, 0 < x_1 < 1$$

for $0 < x_1 < 1,$

$$f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{4x_1x_2 + 1}{2x_1 + 1}, 0 < x_2 < 1$$

§ 2.2.6 independency

- Two continuous random vectors are **Independent**(独立), if and only if

$$f(\boldsymbol{x}, \boldsymbol{y}) = f_x(\boldsymbol{x}) f_y(\boldsymbol{y})$$

- *n* continuous random vectors are mutually independent, if and only if

$$f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = f_1(\boldsymbol{x}_1) \cdots f_n(\boldsymbol{x}_n)$$

- In real application, if the values of random vectors do not affect each other, then they are thought as mutually independent.

Example: Suppose the p.d.f of $x=(x_1,x_2,x_3)$ is as follows

$$f(x_1,x_2,x_3)=\begin{cases} e^{-(x_1+x_2+x_3)}, & x_1,x_2,x_3>0 \\ 0, & other \end{cases}$$

Try to prove x_1,x_2,x_3 are mutually independent.

Proof:

$$f_1(x_1)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2,x_3)dx_2dx_3=e^{-x_1}$$

Similarly, $f_2(x_2)=e^{-x_2}$ $\quad f_3(x_3)=e^{-x_3}$

then $f(x_1,x_2,x_3)=f_1(x_1)f_2(x_2)f_3(x_3)$

So, x_1,x_2,x_3 are mutually independent.

§ 2.3 Numerical characteristics of random vectors

- 1. Expectation （mean）
- 2. Covariance matrix
- 3. Correlation matrix
- *4. Generalized Variance

1. Expectation (mean)

- The **expectation** of random vector $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$ is defined as

$$E(\boldsymbol{x}) = \left[E(x_1), E(x_2), \cdots, E(x_p)\right]'$$

noted as $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_p)'$。

- The **expectation** of random matrix $\boldsymbol{X}=(x_{ij})$ is defined as

$$E(\boldsymbol{X}) = \left(E(x_{ij})\right) = \begin{pmatrix} E(x_{11}) & E(x_{12}) & \cdots & E(x_{1q}) \\ E(x_{21}) & E(x_{22}) & \cdots & E(x_{2q}) \\ \vdots & \vdots & & \vdots \\ E(x_{p1}) & E(x_{p2}) & \cdots & E(x_{pq}) \end{pmatrix}$$

Properties of Expectation of Random Matrix $\boldsymbol{X}$

- (1) Let a is a constant, then

$$E(a\boldsymbol{X})=aE(\boldsymbol{X})$$

- (2) Let $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ is a constant matrix, then

$$E(\boldsymbol{AXB}+\boldsymbol{C})=\boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}+\boldsymbol{C}$$

- Especially, for a random vector $\boldsymbol{x}$**,**

$$E(\boldsymbol{Ax})=\boldsymbol{A}E(\boldsymbol{x})$$

- (3) Let $\boldsymbol{X}_{\mathbf{1}},\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n$ is n random matrices with the same order, then

$$E(\boldsymbol{X}_1+\boldsymbol{X}_2+\cdots+\boldsymbol{X}_n)=E(\boldsymbol{X}_1)+E(\boldsymbol{X}_2)+\cdots+E(\boldsymbol{X}_n)$$

2. Covariance Matrix

- **Covariance**(协方差) is defined as

$$\mathrm{Cov}(x,y)=E\left[x-E(x)\right]\left[y-E(y)\right]$$

- If Cov(x,y)=0, x and y is **Uncorrelated** (不相关)。
- Any two of independent random variables are uncorrelated, but two uncorrelated random variables may not independent.
- When x=y, covariance is **Variance**(方差), i.s.

$$\mathrm{Cov}(x,x)=V(x)$$

- **Covariance Matrix** (协方差矩阵) of $\boldsymbol{x}=\left(x_1,x_2,\cdots,x_p\right)'$ and $y=\left(y_1,y_2,\cdots,y_p\right)'$ is defined as

$$\operatorname{Cov}(\boldsymbol{x},\boldsymbol{y})=\begin{pmatrix}\operatorname{Cov}(x_1,y_1) & \operatorname{Cov}(x_1,y_2) & \cdots & \operatorname{Cov}(x_1,y_q)\\ \operatorname{Cov}(x_2,y_1) & \operatorname{Cov}(x_2,y_2) & \cdots & \operatorname{Cov}(x_2,y_q)\\ \vdots & \vdots & & \vdots\\ \operatorname{Cov}(x_p,y_1) & \operatorname{Cov}(x_p,y_2) & \cdots & \operatorname{Cov}(x_p,y_q)\end{pmatrix}$$

$$=\begin{pmatrix}E\left[x_1-E(x_1)\right]\left[y_1-E(y_1)\right] & \cdots & E\left[x_1-E(x_1)\right]\left[y_q-E(y_q)\right]\\ \vdots & & \vdots\\ E\left[x_p-E(x_p)\right]\left[y_1-E(y_1)\right] & \cdots & E\left[x_p-E(x_p)\right]\left[y_q-E(y_q)\right]\end{pmatrix}$$

$$=E\begin{pmatrix}x_1-E(x_1)\\ \vdots\\ x_p-E(x_p)\end{pmatrix}\left(y_1-E(y_1),\cdots,y_q-E(y_q)\right)$$

$$=E\left[\boldsymbol{x}-E(\boldsymbol{x})\right]\left[\boldsymbol{y}-E(\boldsymbol{y})\right]'$$

- Covariance matrix of $\boldsymbol{y}$ and $\boldsymbol{x}$ is the transposition of the covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$, then

$$\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y})=\left[\mathrm{Cov}(\boldsymbol{y},\boldsymbol{x})\right]'$$

- When $\boldsymbol{x}=\boldsymbol{y}$, the covariance matrix Cov($\boldsymbol{x}$,$\boldsymbol{x}$) is called the covariance matrix of $\boldsymbol{x}$, denoted by *Cov*($\boldsymbol{x}$), i.s.

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{x}) &= E\left[\boldsymbol{x}-E(\boldsymbol{x})\right]\left[\boldsymbol{x}-E(\boldsymbol{x})\right]' \\ &= \begin{pmatrix} V(x_1) & \mathrm{Cov}(x_1,x_2) & \cdots & \mathrm{Cov}(x_1,x_p) \\ \mathrm{Cov}(x_2,x_1) & V(x_2) & \cdots & \mathrm{Cov}(x_2,x_p) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(x_p,x_1) & \mathrm{Cov}(x_p,x_2) & \cdots & V(x_p) \end{pmatrix}\end{aligned} \qquad (2\text{-}32)$$

Cov($\boldsymbol{x}$) can also be written as $\boldsymbol{\Sigma}=(\sigma_{ij})$, where $\sigma_{ij}=\mathrm{Cov}(x_i,x_j)$。

Example 2.13(Computing the covariance matrix) Find the covariance matrix for the two random variable X_1 and X_2 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

x_1 \ x_2	0	1	$p_1(x_1)$
-1	.24	.06	.3
0	.16	.14	.3
1	.40	.00	.4
$p_2(x_2)$	.8	.2	1

- Solution: $E(X_1) = \sum_{\text{all } x_1} x_1 p_1(x_1) = (-1)(.3) + (0)(.3) + (1)(.4) = .1$ and $E(X_2) = \sum_{\text{all } x_2} x_2 p_2(x_2) = (0)(.8) + (1)(.2) = .2.$ Thus,

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}$$

In addition,

$$\sigma_{11} = E(X_1 - \mu_1)^2 = \sum_{\text{all } x_1} (x_1 - .1)^2 p_1(x_1)$$

$$= (-1 - .1)^2(.3) + (0 - .1)^2(.3) + (1 - .1)^2(.4) = .69$$

x_1 \ x_2	0	1	$p_1(x_1)$
−1	.24	.06	.3
0	.16	.14	.3
1	.40	.00	.4
$p_2(x_2)$	.8	.2	1

In addition,

$$\sigma_{22} = E(X_2 - \mu_2)^2 = \sum_{\text{all } x_2} (x_2 - .2)^2 p_2(x_2)$$
$$= (0 - .2)^2(.8) + (1 - .2)^2(.2)$$
$$= .16$$

$$\sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2) = \sum_{\text{all pairs } (x_1, x_2)} (x_1 - .1)(x_2 - .2)p_{12}(x_1, x_2)$$
$$= (-1 - .1)(0 - .2)(.24) + (-1 - .1)(1 - .2)(.06)$$
$$+ \cdots + (1 - .1)(1 - .2)(.00) = -.08$$
$$\sigma_{21} = E(X_2 - \mu_2)(X_1 - \mu_1) = E(X_1 - \mu_1)(X_2 - \mu_2) = \sigma_{12} = -.08$$

Then, we have

$$\mathbf{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$$
$$= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix}$$
$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix}$$
$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .69 & -.08 \\ -.08 & .16 \end{bmatrix}$$

- Example When a random vector is divided into two subvectors, its covariance matrix is separated into four blocks correspondingly .

$$Cov\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} Cov(\boldsymbol{x}) & \mathrm{Cov}(\boldsymbol{x},\boldsymbol{y}) \\ \mathrm{Cov}(\boldsymbol{y},\boldsymbol{x}) & Cov(\boldsymbol{y}) \end{pmatrix}$$

The two diagonal blocks are the covariance matrices of the two subvectors respectively and any one of the two non-diagonal blocks which have the same form is the covariance matrix between the two subvectors. It is beneficial to be familiar with the meaning of the four submatrices.

- Let $x = (x_1, x_2, x_3, x_4)'$ be a random vector, then its covariance matrix is

$$Cov(x) = \left(\begin{array}{cc|cc} V(x_1) & Cov(x_1, x_2) & Cov(x_1, x_3) & Cov(x_1, x_4) \\ Cov(x_2, x_1) & V(x_2) & Cov(x_2, x_3) & Cov(x_2, x_4) \\ \hline Cov(x_3, x_1) & Cov(x_3, x_2) & V(x_3) & Cov(x_3, x_4) \\ Cov(x_4, x_1) & Cov(x_4, x_2) & Cov(x_4, x_3) & V(x_4) \end{array}\right)$$

Properties of Covariance Matrix

- （1）Covariance Matrix is Nonnegtive definite matrix(非负定阵)，即 $\boldsymbol{\Sigma}\geq 0$。
- Deduction　If $|\boldsymbol{\Sigma}|\neq 0$, then $\boldsymbol{\Sigma}>0$。(see 1.7 (4), P21)
- Example $|\boldsymbol{\Sigma}| = 0 \Leftrightarrow \boldsymbol{x}'$s components have a linear relationship (with probability 1).
- In practical problems, when there is a linear relationship between indicators, or one is the summary values of other indicators, then $|\boldsymbol{\Sigma}|=0$. We can delete "redundant" indicators to ensure $|\boldsymbol{\Sigma}|\neq 0$ and then ensures the existence of $\boldsymbol{\Sigma}^{-1}$, thus math problems can be simplified.

- (2) Let $\boldsymbol{A}$ be a constant matrix, $\boldsymbol{b}$ is a constant vector, then

$$Cov(\mathbf{Ax}+\mathbf{b}) = \mathbf{A}Cov(\mathbf{x})\mathbf{A}'$$

- When p=1, the above equation is known as the following equation:

$$V(ax+b) = a^2V(x)$$

- Example 1 Assume the expectation and covariance matrix of random vector $x=(x_1,x_2,x_3)'$ are respectively

$$\boldsymbol{\mu} = \begin{pmatrix} 5 \\ -2 \\ 7 \end{pmatrix} \quad and \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{pmatrix}$$

Let $y_1=2x_1-x_2+4x_3$, $y_2=x_2-x_3$, $y_3=x_1+3x_2-2x_3$, try to calculate the expectation and covariance matrix of $y=(y_1,y_2,y_3)'$.

- (3) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be constant matrices, then

$$\mathrm{Cov}(\boldsymbol{Ax},\boldsymbol{By}) = \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{B}'$$

- (4) Let $\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$和$\boldsymbol{B}_1,\boldsymbol{B}_2,\cdots,\boldsymbol{B}_m$ be constant matrices, then,

$$\mathrm{Cov}\left(\sum_{i=1}^{n}\boldsymbol{A}_i\boldsymbol{x}_i,\sum_{j=1}^{m}\boldsymbol{B}_j\boldsymbol{y}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\boldsymbol{A}_i\,\mathrm{Cov}(\boldsymbol{x}_i,\boldsymbol{y}_j)\boldsymbol{B}_j'$$

- Deduction

$$\mathrm{Cov}\left(\sum_{i=1}^{n}\boldsymbol{x}_i,\sum_{j=1}^{m}\boldsymbol{y}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{Cov}(\boldsymbol{x}_i,\boldsymbol{y}_j)$$

- 5） Let $k_1, k_2, \cdots, k_n$ be n constants, $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ be n mutual independent p dimensional random vectors, then

$$Cov\left(\sum_{i=1}^{n} k_i \boldsymbol{x}_i\right) = \sum_{i=1}^{n} k_i^2 Cov(\boldsymbol{x}_i)$$

- 5） Let $k_1,k_2,\cdots,k_n$ be n constants, $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ be n mutual independent p dimensional random vectors, then

$$Cov\left(\sum_{i=1}^{n} k_i \boldsymbol{x}_i\right) = \sum_{i=1}^{n} k_i^2 Cov(\boldsymbol{x}_i)$$

3. Correlation Matrix

- **Correlation coefficients** (相关系数) of random variables x and y is defined as

$$\rho = \rho(x, y) = \frac{\mathrm{Cov}(x, y)}{\sqrt{V(x)V(y)}}$$

- **Correlation Matrix** (相关阵) of $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$ and $\boldsymbol{y} = (y_1, y_2, \cdots, y_q)'$ is defined as

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \begin{pmatrix} \rho(x_1, y_1) & \rho(x_1, y_2) & \cdots & \rho(x_1, y_q) \\ \rho(x_2, y_1) & \rho(x_2, y_1) & \cdots & \rho(x_2, y_q) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(x_p, y_1) & \rho(x_p, y_1) & \cdots & \rho(x_p, y_q) \end{pmatrix}$$

- If $\rho(\boldsymbol{x},\boldsymbol{y})=0$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are uncorrelated.
- When $\boldsymbol{x}=\boldsymbol{y}$, the covariance matrix $\rho(\boldsymbol{x},\boldsymbol{x})$ is called the correlation matrix of $\boldsymbol{x}$, denoted by $\boldsymbol{\rho}=(\rho_{ij})$ with $\rho_{ij}=\rho(x_i,x_j)$, $\rho_{ii}=1$. Then,

$$\boldsymbol{\rho}=\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}, \qquad (2\text{-}34)$$

where $\rho_{ij}=\dfrac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$ with $\sigma_{ij}=Cov(x_i,x_j)$.

The relationship between $\boldsymbol{\rho}=(\rho_{ij})$ and $\boldsymbol{\Sigma}=(\sigma_{ij})$ is

$$\rho=\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix} \tag{2-34}$$

$$=\begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$$

Rewrite this relationship in term of matrix

$$\boldsymbol{\rho}=\boldsymbol{V}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{V}^{-1/2} \tag{2-37}$$

with $V^{1/2}=\mathrm{diag}(\sqrt{\sigma_{11}},\sqrt{\sigma_{22}},\cdots,\sqrt{\sigma_{pp}})$

4. Multivariate summary statistics

4.1 Sample mean, covariance, and correlation

Let $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_N}$ be a random sample from a multivariate distribution that has a mean vector, $\boldsymbol{\mu_{p\times 1}}$, a covariance matrix, $\boldsymbol{\Sigma_{p\times p}}$, and a correlation matrix, $\boldsymbol{\rho}$. We define sample mean and sample covariance as follows:

$$\overline{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i \quad \text{and } \boldsymbol{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'$$

where $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$ is a p-dimensional vector and $\overline{\boldsymbol{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$

Usually, we use $\overline{\boldsymbol{x}}$ and $\boldsymbol{S}$ as the unbiased estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, denoted as $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}$, respectively.

4. Multivariate summary statistics

4.1 Sample mean, covariance, and correlation

1. Let the random sample (size n, with each observation being a $p \times 1$ vector) be

$$\boldsymbol{X}_{n\times p} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

Then the sample mean vector is

$$\bar{\boldsymbol{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{X}_p \end{bmatrix} = \frac{1}{n}\boldsymbol{X}'\mathbf{1}$$

2. The sample covariance matrix is (**exercise 1**)

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}'(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\boldsymbol{X}$$

4. Descriptive multivariate techniques

4.1 Sample mean, covariance, and correlation

3. The sample correlation matrix is

$$\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{S}\boldsymbol{D}^{-1/2}$$

where

$$\boldsymbol{D}^{1/2} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix}$$

and

$$\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

4. Descriptive multivariate techniques

4.2 Some results of sample mean and sample covariance

1. Let $\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_n$ be a random sample from a joint distribution that has mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

   (a) For the sample mean

      i. $E(\bar{\boldsymbol{X}}) = \boldsymbol{\mu}$

      ii. $Cov(\bar{\boldsymbol{X}}) = \frac{1}{n}\boldsymbol{\Sigma}$

   (b) For the sample covariance (**Exercise 2**)

$$E(\boldsymbol{S}) = \boldsymbol{\Sigma}$$

Standardized Transformation(标准化变换)

- In data processing, since the unit of each variable is not often exactly the same, we need take Standardization Transformation for each variable. The most commonly used standardized transformation is

$$x_i^* = \frac{x_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \cdots, p$$

- Let $\boldsymbol{x}^* = (x_1^*, x_2^*, \cdots, x_p^*)'$, then

$$E\left(\boldsymbol{x}^*\right) = \mathbf{0}, \quad V\left(\boldsymbol{x}^*\right) = \boldsymbol{R}$$

i.s. the covariance matrix after standardization transformation is exactly the correlation matrix of the original vector. Obviously, correlation matrix ***R*** is also a nonnegative definite matrix.

Exercises 3:

Let the covariance matrix of a random vector $\boldsymbol{x} = (x_1, x_2, x_3)'$ be

$$\boldsymbol{\Sigma} = \begin{bmatrix} 16 & -4 & 3 \\ -4 & 4 & -2 \\ 3 & -2 & 9 \end{bmatrix}$$

Please find the correlation matrix of $\boldsymbol{x}$.

Assignment:

See the assignment file.

§ 2.4 Euclidean Distance and Mahalanobis Distance

- § 2.4.1 Euclidean Distance
- § 2.4.2 Mahalanobis Distance

§ 2.4.1 Euclidean Distance

- The **Euclidean Distance**(欧氏距离)between $\mathbf{x} = (x_1, x_2, \cdots, x_p)'$ and $\mathbf{y} = (y_1, y_2, \cdots, y_p)'$ is the straight-line distance between them, denoted by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2}$$

- Square Euclidean Distance:

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2$$
$$= (\boldsymbol{x} - \boldsymbol{y})'(\boldsymbol{x} - \boldsymbol{y})$$

- The square Euclidean distance between $\boldsymbol{x} = \left(x_1, x_2, \cdots, x_p\right)'$ and the population π is defined as

$$d^2(\boldsymbol{x}, \pi) = (\boldsymbol{x} - \boldsymbol{\mu})'(\boldsymbol{x} - \boldsymbol{\mu})$$

$$= \left(x_1 - \mu_1\right)^2 + \left(x_2 - \mu_2\right)^2 + \cdots + \left(x_p - \mu_p\right)^2$$

where $\mu = (\mu_1, \cdots, \mu_p)'$ is the center of π.

mean size

$$E\left(x_1 - \mu_1\right)^2 \quad E\left(x_2 - \mu_2\right)^2 \quad \cdots \quad E\left(x_p - \mu_p\right)^2$$

$$V\left(x_1\right) \quad V\left(x_2\right) \quad \cdots \quad V\left(x_p\right)$$

- The units of all the vectorial components are not the same, then the above Euclidean distance generally meaningless.
- Even though the units are all the same, if the variations of all the components are hugely different, the components with big variations play a decisive role in Euclidean distance, however the components with small variations are almost useless.
- In real application, in order to remove the effect of units and handle each component equably, we usually take standardized transformation of every component first and then calculate Euclidean distance.

Euclidean Distance after Standardization

- Suppose $x_i^* = \dfrac{x_i - \mu_i}{\sqrt{\sigma_{ii}}}, i = 1, \cdots, p, \quad \mathbf{x}^* = \left(x_1^*, \cdots, x_p^*\right)'$, then

$$d^2\left(\mathbf{x}^*, \pi\right) = \mathbf{x}^{*\prime}\mathbf{x}^* = x_1^{*2} + \cdots + x_p^{*2}$$

- Since $E\left(x_i^{*2}\right) = V\left(x_i^*\right) = 1, i = 1, 2, \cdots, p$, the average value of each term among the sum of squares $X_1^{*2} + \cdots + X_p^{*2}$ is 1. Therefore, the terms make an average effect.
- After standardization, the Euclidean distance is removed the influence of the diversity of variables' variations, however, the effect of dependency cannot be eliminated.

The correlation among variables cannot be eliminated after standardizing.

$$\mathrm{cov}(x_1^*, x_2^*) = \mathrm{cov}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}, \frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)$$

$$= E\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}} \frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) - E\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right) E\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)$$

$$= \frac{E(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

$$= \frac{\mathrm{cov}(x_1, x_2)}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

- Limitations of Euclidean Distance

§ 2.4.2 Mahalanobis Distance

- The **Square Mahalanobis Distance**（平方马氏距离）between $\mathbf{x} = (x_1, x_2, \cdots, x_p)'$ and $\mathrm{y} = (y_1, y_2, \cdots, y_p)'$ from the same population (μ, Σ) is defined as

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})$$

- The Square Mahalanobis Distance between $\mathrm{x} = (x_1, x_2, \cdots, x_p)'$ and the population π is defined as

$$d^2(\boldsymbol{x}, \pi) = (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

- **Property 1:** Mahalanobis distance is not influenced by the units of variables. It has no unit. (P62 Practice 2.11)

- Proof of Property 1 $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ become $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ after transformation, then

$$\boldsymbol{y}_1 = \boldsymbol{C}\boldsymbol{x}_1 + \boldsymbol{b}, \quad \boldsymbol{y}_2 = \boldsymbol{C}\boldsymbol{x}_2 + \boldsymbol{b}$$

$$\boldsymbol{\Sigma}_y = \boldsymbol{C}\boldsymbol{\Sigma}_x\boldsymbol{C}'$$

$$(y_1 - y_2)'\Sigma_y^{-1}(y_1 - y_2)$$

$$=[(Cx_1+b)-(Cx_2+b)]'(C\Sigma_x C')^{-1}[(Cx_1+b)-(Cx_2+b)]$$

$$=(x_1-x_2)'C'C'^{-1}\Sigma_x^{-1}C^{-1}C(x_1-x_2)$$

$$=(x_1-x_2)'\Sigma_x^{-1}(x_1-x_2)$$

Unit Transformation with constant terms

- Example　The reduction formula between centigrade degree and Fahrenheit degree:

  $F = (C \times 9/5) + 32$,　$C = (F - 32) \times 5/9$

  where F——Fahrenheit degree, C——centigrade degree

- The metric unit transformation of vectors can be formulated as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} c_1 x_1 + b_1 \\ c_2 x_2 + b_2 \\ \vdots \\ c_p x_p + b_p \end{pmatrix} = \begin{pmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_p \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}$$

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{b}$$

- Property 2 Mahalanobis distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is their standardized Euclidean distance, i.s.

$$d^2(\boldsymbol{x},\boldsymbol{y})=\left(\boldsymbol{x}^*-\boldsymbol{y}^*\right)'\left(\boldsymbol{x}^*-\boldsymbol{y}^*\right)$$

where $\boldsymbol{x}^*=\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu}),\boldsymbol{y}^*=\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{y}-\boldsymbol{\mu})$.

Their mean are both **0** and their covariance $Cov(\boldsymbol{x}^*,\boldsymbol{y}^*)=I$.

- Property 3 Suppose $\boldsymbol{\Sigma}=\mathrm{diag}\left(\sigma_{11},\sigma_{22},\cdots,\sigma_{pp}\right)$, then

$$d^2(\boldsymbol{x},\boldsymbol{y})=\frac{\left(x_1-y_1\right)^2}{\sigma_{11}}+\frac{\left(x_2-y_2\right)^2}{\sigma_{22}}+\cdots+\frac{\left(x_p-y_p\right)^2}{\sigma_{pp}}$$

i.s. when the components are mutually uncorrelated, their Mahalanobis distance is the Euclidean distance of the components after standardization.

§ 2.5 Transformation of Random Vectors

- **Transformation of random variable**

Suppose the density function of continuous random variable *x* is $f_x(x)$,

$y=\varphi(x)$ is strictly monotonous, and its inverse function $x=\psi(y)$ has continuous derivative, then the p.d.f. of *y* is as follows

$$f_y(y) = f_x(\psi(y))\,|\,\psi'(y)\,|$$

where the value range of *y* is corresponding to that of *x*.

- Example 2.5.1 Suppose random variable ***x*** obeys the uniform distribution U(0,1), then its p.d.f. is

$$f_x(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{others}\end{cases}$$

Try to give the distribution of $y=-\frac{1}{\lambda}\ln x.(\lambda>0)$

Solution: the value range of $x=e^{-\lambda y}$ is $[0,\infty)$, so

$$f_y(y)=f_x(e^{-\lambda y})\,|(e^{-\lambda y})'|=1\times|-\lambda e^{-\lambda y}|$$

$$=\lambda e^{-\lambda y},\qquad y\ge 0$$

The distribution of y is called **Exponential Distribution(**指数分布) with parameter λ, denoted by $y\sim \mathrm{E}(\lambda)$.

Transformation of Random Vectors

Suppose the density function of $x=(x_1,\cdots,x_p)'$ *is* $f(x_1,\cdots,x_p)$ and function group is denoted by $y_i=\varphi_i(x_1,\cdots,x_p),\ i=1,2,\cdots,p$. The inverse transformation exists as $x_j=\phi_j(y_1,\cdots,y_p)$, then the density function of $y=(y_1,\cdots,y_p)'$ is denoted by

$$g(y_1,\cdots,y_p)=f(\varphi_1(y_1,\cdots,y_p),\cdots,\varphi_p(y_1,\cdots,y_p))\bullet|J|$$

where J is Jacobi Determinant,

Note :

$$J(x\to y)=\frac{\partial(x_1,x_2,\cdots,x_p)}{\partial(y_1,y_2,\cdots,y_p)}=\begin{vmatrix}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_p}\\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_p}\\ \vdots & \vdots & & \vdots\\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \cdots & \frac{\partial x_p}{\partial y_p}\end{vmatrix}$$

$$J(y\to x)=1/J(x\to y)$$

Example 2.5.2

- Suppose X and Y are independent and exponentially distributed with parameter λ .Let $Z = X + Y, W = \frac{X}{X+Y}$, then the value range of Z is the same as X or Y, $(0, \infty)$, but the value range of W is (0,1). Please calculate the density of (Z,W) and the marginal density of W.

Solution： the inverse transformation of z=x+y, $W = \frac{X}{X+Y}$ is x=zw, y=z(1-w), therefore the Jacobi determinant is

$$J(z,w) = \begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = -z$$

Since the density of (X,Y) is $\lambda^2 \mathrm{e}^{-\lambda(x+y)}$, $x > 0, y > 0.$

the density of (Z,W) is

$$g(z,w) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0, 0 < w < 1 \\ 0, & z \le 0 \end{cases}$$

Therefore, then density of Z=X+Y:

$$g(z,w) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0 \\ 0, & z \le 0 \end{cases}$$

Practice 2.1’ : Suppose random variables X and Y are independent identically distributed and their joint distribution density is

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0 & , x \leq 0. \end{cases}$$

Prove X+Y and X/Y are mutually independent.(Hint: calculate the joint distribution of (X+Y,X/Y), then prove the joint density of (X+Y,X/Y) is the product of the marginal density of X+Y and X/Y.

Chapter 2: assignments

2 (For your choices). Suppose X and Y are mutually independent standard normal variables. (X,Y) can be regarded as a coordinate of a random point in rectangular coordinate system on 2-d plane. Its polar coordinate is

$$x = r\cos\theta, \qquad y = r\sin\theta$$

where the value range of r is $(0, \infty)$ and that of θ is $[0,2\pi)$. Please give the joint distribution of (r,θ)