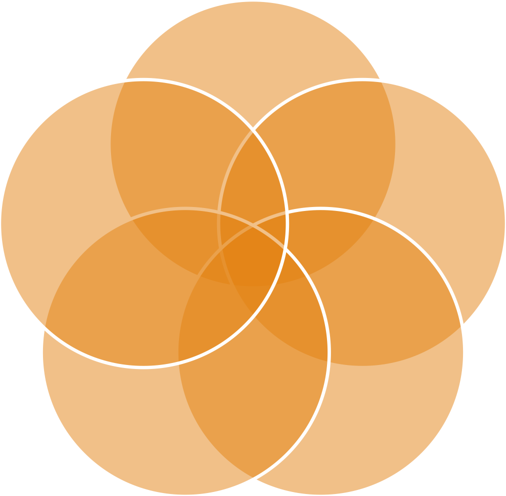
相关基础知识回顾与补充

概率论

线性系统理论

随机过程



泛函分析

矩阵分析

概率论

随机实验 → 基本事件 → 随机事件 → 样本空间 → σ-代数 → 概率测度及其基本性质 → 概率空间

随机试验、样本点与样本空间,

如果一个试验，具有下列特点:

(1) 试验可以在相同的条件下重复进行;

(2) 每次试验的结果可能不止一个, 并且能够事先明确所有的可能结果;

(3) 进行试验之前不能确定哪一个结果将会出现.

这样的试验称为**随机试验**, 简称**试验**, E. 随机试验中的每一个可能结果称为一个样本点, 记为 ω; 全体样本点所构成的集合称为**样本空间**, 记为 Ω。

Definition 3.1.1. **【随机事件】**随机试验 E 的样本空间 Ω 的子集称为随机事件, 简称**事件**. 常用大写字母 A, B, C 等表示. 事件 A 发生, 是指在试验中当且仅当集合 A 中所包含的样本点出现, 否则称 A 不发生.

特别地, 由一个样本点组成的单点集称为基本事件. 在每次试验中必然发生的事件, 称为必然事件. 由于样本空间 Ω 包含所有的样本点, 在每次试验中 Ω 的样本点总是会出现, 故 Ω 是必然事件. 在每次试验中必然不发生的事件, 称为不可能事件. 由于空集 $\varnothing$ 不含任何样本点, 故 $\varnothing$ 是不可能事件.

Definition 3.1.2. 【σ-代数】设 $\mathscr{F}$ 为空间 Ω 的子集构成的集合，满足

(1) $\Omega \in \mathscr{F}$;

(2) 若 $A \in \mathscr{F}$, 则 $\bar{A} \in \mathscr{F}$;

(3) 若 $A_n \in \mathscr{F}, n = 1, 2, \cdots$, 则 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{F}$.

那么称 $\mathscr{F}$ 为 σ-代数, 也成为 σ-域.

σ-代数具有如下性质:

(1) $\varnothing \in \mathscr{F}$;

(2) 若 $A_n \in \mathscr{F}, n = 1, 2, \cdots$, 则 $\bigcap_{n=1}^{\infty} A_n \in \mathscr{F}$;

(3) 若 $A_n \in \mathscr{F}, n = 1, 2, \cdots$, 则 $\bigcup_{n=1}^{k} A_n \in \mathscr{F}, \bigcap_{n=1}^{k} A_n \in \mathscr{F}$;

(4) 若 $A_n \in \mathscr{F}, n = 1, 2, \cdots$, 则 $\varlimsup_{n\to\infty} A_n \in \mathscr{F}, \varliminf_{n\to\infty} A_n \in \mathscr{F}$.

Definition 3.1.3. 【事件域】设 Ω 是样本空间, $\mathscr{F}$ 是由样本空间 Ω 的一些子集构成的一个 σ-代数，则称 $\mathscr{F}$ 为事件域. $\mathscr{F}$ 中的元素称为事件，Ω 称为必然事件，$\varnothing$ 称为不可能事件.

Definition 3.1.4. 【概率】设 $P(A)$ 是定义在事件域 $\mathscr{F}$ 上的实值集合函数，如果满足

(1) 非负性：对任一 $A \in \mathscr{F}$, $P(A) \geq 0$;

(2) 规范性：对必然事件 Ω, $P(\Omega) = 1$;

(3) 可列可加性：设 $A_1, A_2, \cdots$ 是两两互不相容事件序列, 即对于 $i \neq j$, $A_iA_j = \varnothing (i, j = 1, 2, \cdots)$, 则有

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

那么称 $P(A)$ 为事件域 $\mathscr{F}$ 上事件 A 的概率.

概率空间

一般称三元组 $(\Omega, \mathscr{F}, P)$ 为概率空间, 其中 Ω 是样本空间, $\mathscr{F}$ 是事件域, P 是概率. 可以容易验证概率的如下性质:

(1) $P(\varnothing)=0, P(\Omega)=1.$

(2) 对 n 个两两互不相容事件 $A_1, A_2, \cdots, A_n$, 有 $P\left(\bigcup\limits_{k=1}^{n} A_k\right)=\sum\limits_{k=1}^{n} P\left(A_k\right).$

(3) 对于任一事件 A, 有 $P(\bar{A})=1-P(A).$

(4) 对于任意两个事件 A 、B , 有 $P(A-B)=P(A)-P(AB)$. 特别地, 当 $B \subset A$ 时, 有 $P(AB)=P(A)-P(B)$, 且 $P(A) \geq P(B)$.

(5) 对于任意两个事件 A 、B , 有 $P(A \cup B)=P(A)+P(B)-P(AB).$

Definition 3.1.5. 【条件概率】设 A 、B 是两个事件, 且 $P(A)>0$, 称 $P(B|A)=\frac{P(AB)}{P(A)}$ 为在事件 A 已经发生的条件下事件 B 发生的条件概率.

不难验证, 条件概率 $P(\cdot|A)$ 符合上面概率定义的三个条件, 即

(1) 非负性: 对于每一事件 B , 有 $P(B|A)\geqslant 0$;

(2) 规范性: 对于必然事件 Ω , 有 $P(\Omega|A)=1$;

(3) 可列可加性: 设 $B_1,B_2,\cdots$ 是两两互不相容的事件, 则有 $P\left(\bigcup_{i=1}^{+\infty} B_i|A\right)=\sum_{i=1}^{\infty} P\left(B_i|A\right)$.

由此可知, 概率的有关性质对于条件概率也都适用. 例如

$$
\begin{aligned}
&P(\bar{B}|A)=1-P(B|A)\\
&P\left(B_1\cup B_2|A\right)=P\left(B_1|A\right)+P\left(B_2|A\right)-P\left(B_1B_2|A\right)
\end{aligned}
\tag{3.1}
$$

Theorem 3.1.1. 设 A、B 是两个事件, 且 $P(A) > 0$, 有

$$\boxed{P(AB) = P(B|A)P(A)}$$

若 $P(B) > 0$, 则有 $\boxed{P(AB) = P(A|B)P(B)}$.

Theorem 3.1.2. 设 $A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一个完备事件组, 且 $P(A_i) > 0 (i = 1, 2, \cdots, n)$. 对于任意事件 B, 有

$$\boxed{P(B) = \sum_{i=1}^{n} P(B|A_i) P(A_i)} \tag{3.3}$$

Theorem 3.1.3. 设 $A_1, A_2, \cdots, A_n$ 是样本空间 Ω 的一个完备事件组, 且 $P(A_i) > 0 (i = 1, 2, \cdots, n)$. 对于任意事件 B, $P(B) > 0$, 则有

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_{i=1}^{n} P(B|A_i) P(A_i)} \tag{3.4}$$

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

随机变量

Definition 3.1.6. 【随机变量】设 $(\Omega, \mathscr{F}, P)$ 是一概率空间，$X(\omega)$ 是定义在 Ω 上的$\boxed{\text{单值实函数}}$, 若对任意实数 $x \in R$, 有

$$\{\omega | X(\omega) \le x\} \in \mathscr{F} \tag{3.6}$$

则称 $X = X(\omega)$ 为随机变量.

Definition 3.1.7. 【分布函数】设 $X = X(\omega)$ 是定义在概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变量, 对任意 $x \in R$, 称

$$\boxed{F(x) = P\{\omega | X(\omega) \leq x\}} \tag{3.7}$$

为随机变量 ω 的分布函数. 简记为 $F(x) = P\{X(\omega) \leq x\}$, $F(x)$ 有时也记为 $F_X(x)$.

对于任意实数 $x_1, x_2\,(x_1 < x_2)$, 有

$$\boxed{P\{x_1 < X \leqslant x_2\} = P\{X \leqslant x_2\} - P\{X \leqslant x_1\} = F(x_2) - F(x_1)}$$

Theorem 3.1.4. 分布函数 $F(x)$ 具有如下性质:

(1) 不减性. 若 $\forall x_1 < x_2 \in R$, 则 $F(x_1) \leqslant F(x_2)$;

(2) 规范性. $0 \leqslant F(x) \leqslant 1$ 且 $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0, F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$;

(3) 右连续性. 对 $\forall x_0 \in R$, 有 $\lim_{x \to x_0^+} F(x) = F(x_0)$.

概率密度函数

Definition 3.1.8. 【概率密度函数】如果对随机变量 X 的分布函数 $F(x)$, 存在非负可积函数 $f(x)$, 使得对任意的实数 x, 有

$$\boxed{F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(t)\mathrm{d}t} \tag{3.8}$$

其中 $f(x)$ 称为随机变量 X 的概率密度函数, 简称概率密度或密度函数.

Theorem 3.1.5. 概率密度函数具有如下性质:

(1) 非负性. $f(x) \geq 0, \forall x \in R$.

(2) 规范性. $\int_{-\infty}^{+\infty} f(x) \mathrm{d} x=1$.

(3) 对任意实数 $x_1, x_2\left(x_1 \leqslant x_2\right)$,

$$P\left\{x_1<X(\omega) \leqslant x_2\right\}=F\left(x_2\right)-F\left(x_1\right)=\int_{x_1}^{x_2} f(x) \mathrm{d} x=P\left(X(\omega) \leqslant x_2\right)-P\left(X(\omega) \leqslant x_1\right)$$

(4) 若 $F(x)$ 在 x 处是连续的，则 $F^{\prime}(x)=f(x)$.

(5) 若 X 是连续型随机变量, 则 $\forall a \in R, P\{\omega \mid X(\omega)=a\}=0$.

设 X 为一个随机变量,$g(x)$ 为定义在实数集合 I 上的实值函数，而 X 的可能取值 $x \in I$. 由高等数学中函数的定义知，函数 $g(x)$ 的值由自变量 x 确定，所以当随机变量 X 的取值具有随机性时, 函数 $Y = g(X)$ 的取值也具有随机性, 即 $Y = g(X)$ 也是随机变量. 例如 $Y = X, Y = \ln X, Y = \sin X$ 等都是随机变量.

Theorem 3.1.6. 设连续随机变量 X 的取值范围为 (a, b), 其密度函数为 $f_x(x)$. 若函数 $y = g(x)$ 在 (a, b) 内严格单调, 且其反函数 $x = g^{-1}(y)$ 有连续导数, 则 $Y = g(X)$ 的密度函数为

$$f_y(y) = \begin{cases} f_x\left(g^{-1}(y)\right)\left|\left(g^{-1}(y)\right)'\right|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases} \tag{3.10}$$

其中, $\alpha = \min\{g(a), g(b)\}, \beta = \max\{g(a), g(b)\}$.

特征函数

设 X 为随机矢量，其特征函数定义为

$$\varphi_{\mathbf{x}}(\boldsymbol{\xi}) = \int_{-\infty}^{+\infty} e^{i\mathbf{x}^T\boldsymbol{\xi}} dF_{\mathbf{x}}(\boldsymbol{\xi})$$

随机向量及其分布

Definition 3.1.9. 【随机向量】设 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 是定义在概率空间 $(\Omega, \mathscr{F}, P)$ 上的 n 个随机变量，称 $\boldsymbol{X}(\omega) = [X_1(\omega), X_2(\omega), \cdots, X_n(\omega)]^T$ 为 n 维随机向量.

n 维随机向量取值于 n 维欧几里得空间 R^n. 对 n 个实数 $x_1, x_2, \cdots, x_n$, $\{X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n\} = \bigcap_{i=1}^{n} \{X_i \leq x_i\}$ 有定义，并属于 $\mathscr{F}$.

Definition 3.1.10.【随机向量分布函数】n 元函数 $F(x_1, x_2, \cdots, x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n)$ 称为 n 维随机变量 $\boldsymbol{X}(\omega) = [X_1(\omega), X_2(\omega), \cdots, X_n(\omega)]^T$ 的 (联合) 分布函数.

Definition 3.1.11.【随机向量概率密度】若对任意的 n 个实数 $x_1, x_2, \cdots, x_n$, 存在非负实函数 $f(x_1, x_2, \cdots, x_n)$, 使随机向量 $\boldsymbol{X}$ 的分布函数

$$\begin{aligned} F(\boldsymbol{x}) &= F(x_1, x_2, \cdots, x_n) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(u_1, u_2, \cdots, u_n) \mathrm{d}u_1 \mathrm{d}u_2 \cdots \mathrm{d}u_n \\ &= \int_{-\infty}^{\boldsymbol{x}} f(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} \end{aligned} \tag{3.11}$$

则称函数 $f(\boldsymbol{x}) = f(x_1, x_2, \cdots, x_n)$ 为随机向量 $\boldsymbol{X}$ 的概率分布密度或概率密度函数.

Theorem 3.1.7. 随机向量概率密度 $f(\boldsymbol{x})$ 具有如下性质：

(1) 对 $\forall \boldsymbol{x} \in R^n$, $f(\boldsymbol{x}) \geq 0$;

(2) $\int_{-\infty}^{+\infty} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \int \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \cdots, x_n) \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n = 1$;

(3) 若 $F(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处连续，则有

$$\frac{\partial^n F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n} = f(x_1, \cdots, x_n) = f(\boldsymbol{x})$$

(4) 设 V 是 R^n 中任一区域，随机点 $\boldsymbol{x}$ 落入区域 V 内的概率为

$$P(\boldsymbol{x} \in V) = \int_V f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

注意，这里我们用 $\mathrm{d}\boldsymbol{x}$ 表示 $\mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$.

Definition 3.1.12. 【边缘分布】设 n 维随机向量 $\boldsymbol{X}$ 和 m 维随机向量 $\boldsymbol{Y}$ 的联合分布函数为 $F_{XY}(\boldsymbol{x},\boldsymbol{y})$, 其中 $\boldsymbol{x} \in R^n$, $\boldsymbol{y} \in R^m$, 那么称

$$\begin{aligned} F_X(\boldsymbol{x}) &= F_{XY}(\boldsymbol{x},+\infty) = P(\boldsymbol{X} \le \boldsymbol{x}, \boldsymbol{Y} \leqslant +\infty) \\ &= P(X_1 \le x_1, \cdots, X_n \le x_n; Y_1 \leqslant +\infty, \cdots, Y_m \leqslant +\infty) \end{aligned} \tag{3.14}$$

为随机向量 $\boldsymbol{X}$ 的边缘分布函数. 称

$$\begin{aligned} F_Y(\boldsymbol{y}) &= F_{XY}(+\infty,\boldsymbol{y}) = P(\boldsymbol{X} \leqslant +\infty, \boldsymbol{Y} \le \boldsymbol{y}) \\ &= P(X_1 \leqslant +\infty, \cdots, X_n \leqslant +\infty; Y_1 \le y_1, \cdots, Y_m \le y_m) \end{aligned} \tag{3.15}$$

为随机向量 $\boldsymbol{Y}$ 的边缘分布函数.

Definition 3.1.13. 【边缘分布密度函数】设随机向量 $\boldsymbol{X}$ 和随机向量 $\boldsymbol{Y}$ 的联合分布函数为 $F_{XY}(\boldsymbol{x}, \boldsymbol{y})$, 对应的联合概率密度函数为 $f_{XY}(\boldsymbol{x}, \boldsymbol{y})$, 即

$$F_{XY}(\boldsymbol{x}, \boldsymbol{y}) = \int_{-\infty}^{\boldsymbol{x}} \int_{-\infty}^{\boldsymbol{y}} f_{XY}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}\boldsymbol{u} \mathrm{d}\boldsymbol{v} \tag{3.16}$$

称

$$\boxed{f_X(\boldsymbol{x}) = \int_{-\infty}^{+\infty} f_{XY}(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y}} \tag{3.17}$$

为随机向量 $\boldsymbol{X}$ 的边缘分布密度函数. 称

$$\boxed{f_Y(\boldsymbol{y}) = \int_{-\infty}^{+\infty} f_{XY}(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x}} \tag{3.18}$$

为随机向量 $\boldsymbol{Y}$ 的边缘分布密度函数.

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u$$

$$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\mathrm{d}u$$

简化的记号

Theorem 3.1.8. 设随机变量 X, Y 的联合概率密度为 $f_{XY}(x,y)$, $(x,y) \in D$. 令 $U = g_1(X,Y), V = g_2(X,Y)$. 假设 $x = h_1(u,v), y = h_2(u,v)$ 对 u, v 有连续的偏导数，并且下述雅可比矩阵:

$$J(u,v) := \frac{\partial(h_1,h_2)}{\partial(u,v)} = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}$$

非奇异. 记

$$G = \{(u,v) | u = g_1(x,y), v = g_2(x,y), (x,y) \in D\}$$

那么

$$f_{UV}(u,v) = \begin{cases} f_{XY}(h_1(u,v), h_2(u,v)) |J(u,v)|^{-1}, & (u,v) \in G; \\ 0, & \text{其他}. \end{cases}$$

随机向量的数字特征（低阶统计特性）

Definition 3.1.14. **【数学期望】**设随机向量 $\boldsymbol{X}$ 的概率密度函数为 $f(\boldsymbol{x})$，如果

$$E(\boldsymbol{X})=\int_{-\infty}^{+\infty}\boldsymbol{x}f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{3.21}$$

存在，则称 $E(\boldsymbol{X})$ 为随机向量 $\boldsymbol{X}$ 的数学期望，有时又称为均值.

上述定义表明，数学期望 $E(\boldsymbol{X})$ 是与随机向量 $\boldsymbol{X}$ 同维的常向量. 经常也记为 $E\boldsymbol{X},\bar{\boldsymbol{X}},\mu(\boldsymbol{X})$ 等. 数学期望表示随机量的平均值，我们称扣除均值后的随机变量为中心随机变量，例如 $\mathring{X}=X-E(X)$.

Theorem 3.1.9. 数学期望具有如下性质：

(1) 若 $\boldsymbol{C}$ 是一个常值向量，则 $E(\boldsymbol{C}) = \boldsymbol{C}$;

(2) 若 A 是常值矩阵，$\boldsymbol{X}$ 是随机向量，则 $E(A\boldsymbol{X}) = AE(\boldsymbol{X})$;

(3) 若 $\boldsymbol{X}, \boldsymbol{Y}$ 是两个同维的随机向量，那么 $E(\boldsymbol{X} + \boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y})$.

Definition 3.1.15. **【协方差矩阵】** 设 $\boldsymbol{X}$ 是随机向量，若

$$E(\mathring{\boldsymbol{X}}\mathring{\boldsymbol{X}}^T) = \int_{-\infty}^{+\infty} \mathring{\boldsymbol{X}}\mathring{\boldsymbol{X}}^T f(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{3.22}$$

存在，则称其为随机向量 $\boldsymbol{X}$ 的协方差矩阵，记为 $\mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})$ 或 $D(\boldsymbol{X})$.

对于随机变量，即一维随机向量，上述协方差的概念退化为方差. 随机变量 X 的方差记为 $D(X)$ 或 $\sigma^2(X)$. 显然随机变量的方差是非负的，其平方根称为**均方差**、**标准差**，记为 $\sigma(X)$. 即 $\sigma(X) = \sqrt{D(X)} = \sqrt{\sigma^2(X)}$.

Theorem 3.1.10. 随机向量的协方差矩阵是对称的非负定方阵.

Definition 3.1.16. 【互协方差矩阵】设 $\boldsymbol{X},\boldsymbol{Y}$ 是两个随机向量，若

$$E(\mathring{\boldsymbol{X}}\mathring{\boldsymbol{Y}}^T)=\int_{-\infty}^{+\infty}\mathring{\boldsymbol{X}}\mathring{\boldsymbol{Y}}^T f(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}$$

存在，则称其为随机向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的互协方差矩阵，记为 $\mathrm{cov}(\boldsymbol{X},\boldsymbol{Y})$.

两个随机变量的互协方差是一个标量，简称为协方差，反映了两种之间的统计关联程度.

Definition 3.1.17. 【独立性】设 $\boldsymbol{X}, \boldsymbol{Y}$ 是两个随机向量，如果

$$\boxed{F_{XY}(\boldsymbol{x}, \boldsymbol{y}) = F_X(\boldsymbol{x}) F_Y(\boldsymbol{y})}$$

等价地 (假设相应概率密度函数存在)

$$\boxed{f_{XY}(\boldsymbol{x}, \boldsymbol{y}) = f_X(\boldsymbol{x}) f_Y(\boldsymbol{y})}$$

那么称 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 相互独立.

Definition 3.1.18. 【相关性】设 $\boldsymbol{X}, \boldsymbol{Y}$ 是两个随机向量，如果

$$E(\mathring{\boldsymbol{X}}\mathring{\boldsymbol{Y}}^T) = \mathbf{0}$$

那么称 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 不相关.

如果 X, Y 是两个不相关的随机变量，那么 $E(XY) = E(X)E(Y)$, 同时 $D(X+Y) = D(X) + D(Y)$.

Theorem 3.1.11. 如果两个随机向量相互独立，则两者不相关. 反之，则不一定.

Definition 3.1.19. 【相关系数】设 X, Y 是随机变量，两者的方差均存在，那么称

$$\boxed{r_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}}$$

为随机变量 X 与 Y 之间的相关系数.

Theorem 3.1.12. 设随机变量 X 与 Y 之间的相关系数为 r_{XY}, 那么

(1) $|r_{XY}| \leq 1$;

(2) $|r_{XY}| = 1$ 的充要条件是存在常数 a, b, 使

$$P(Y = aX + b) = 1$$

即 X 与 Y 以概率 1 存在线性关系.

(3) 如果 X 与 Y 不相关，则 $r_{XY} = 0$.

Definition 3.1.20. 对所有使 $f_Y(\boldsymbol{y}) > 0$ 的 $\boldsymbol{y}$, 给定 $\boldsymbol{Y} = \boldsymbol{y}$ 条件下的 $\boldsymbol{X}$ 的条件分布函数和条件密度函数分别为

$$F_{X|Y}(\boldsymbol{x}|\boldsymbol{y}) = \int_{-\infty}^{\boldsymbol{x}} \frac{f_{XY}(\boldsymbol{u}, \boldsymbol{y})}{f_Y(\boldsymbol{y})} \mathrm{d}\boldsymbol{u} \tag{3.29}$$

$$\boxed{f_{X|Y}(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{XY}(\boldsymbol{x}, \boldsymbol{y})}{f_Y(\boldsymbol{y})}} \tag{3.30}$$

同样地，对所有使 $f_X(\boldsymbol{x}) > 0$ 的 $\boldsymbol{x}$, 给定 $\boldsymbol{X} = \boldsymbol{x}$ 条件下的 $\boldsymbol{Y}$ 的条件分布函数和条件密度函数分别为

$$F_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{y}} \frac{f_{XY}(\boldsymbol{x}, \boldsymbol{v})}{f_X(\boldsymbol{x})} \mathrm{d}\boldsymbol{v} \tag{3.31}$$

$$\boxed{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = \frac{f_{XY}(\boldsymbol{x}, \boldsymbol{y})}{f_X(\boldsymbol{x})}} \tag{3.32}$$

Theorem 3.1.13. 【贝叶斯法则】设 $f_Y(\boldsymbol{y}) > 0$，有

$$\boxed{f_{X|Y}(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{Y|X}(\boldsymbol{y}|\boldsymbol{x})f_X(\boldsymbol{x})}{f_Y(\boldsymbol{y})}} \tag{3.33}$$

这是我们后面要经常用到的一个结论，以后会称 $f_{X|Y}(\boldsymbol{x}|\boldsymbol{y})$ 为验后概率密度，称 $f_X(\boldsymbol{x})$ 为验前概率密度.

常用的随机分布

Definition 3.1.21. 【均匀分布】设随机变量 X 具有概率密度

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

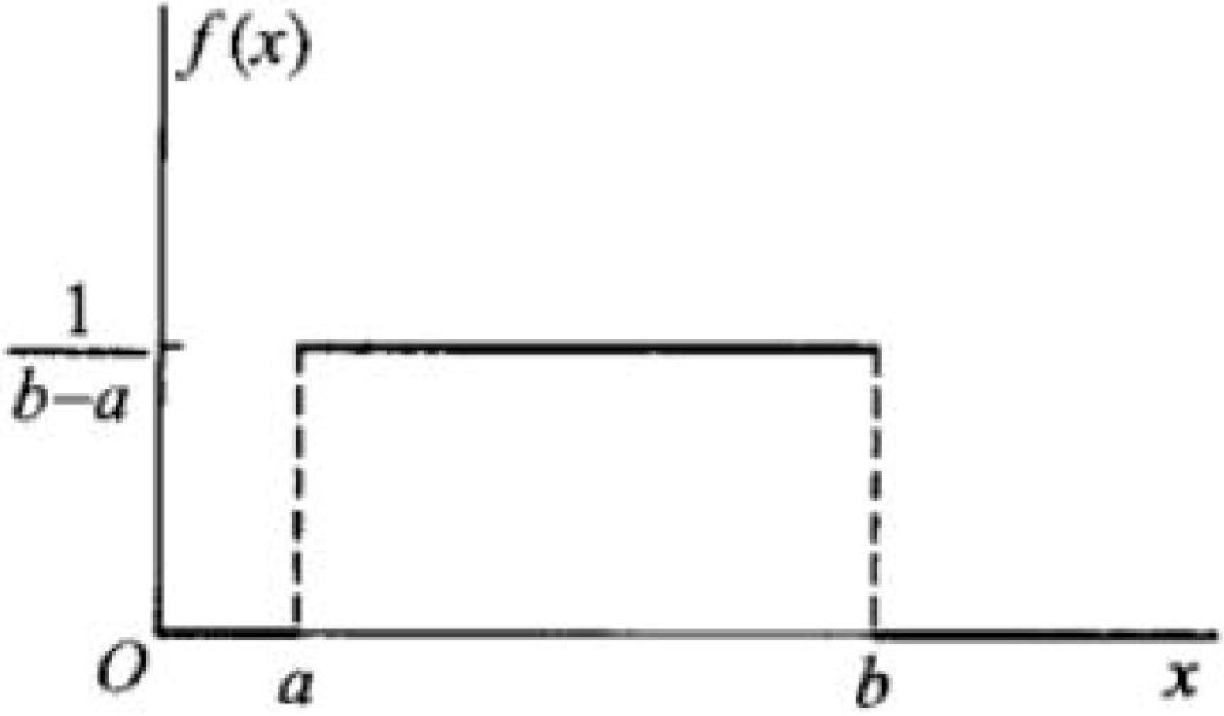
则称 X 在区间 (a,b) 上服从均匀分布, 记为 $X \sim U(a,b)$.

Theorem 3.1.14. 设 $X \sim U(a,b)$，那么

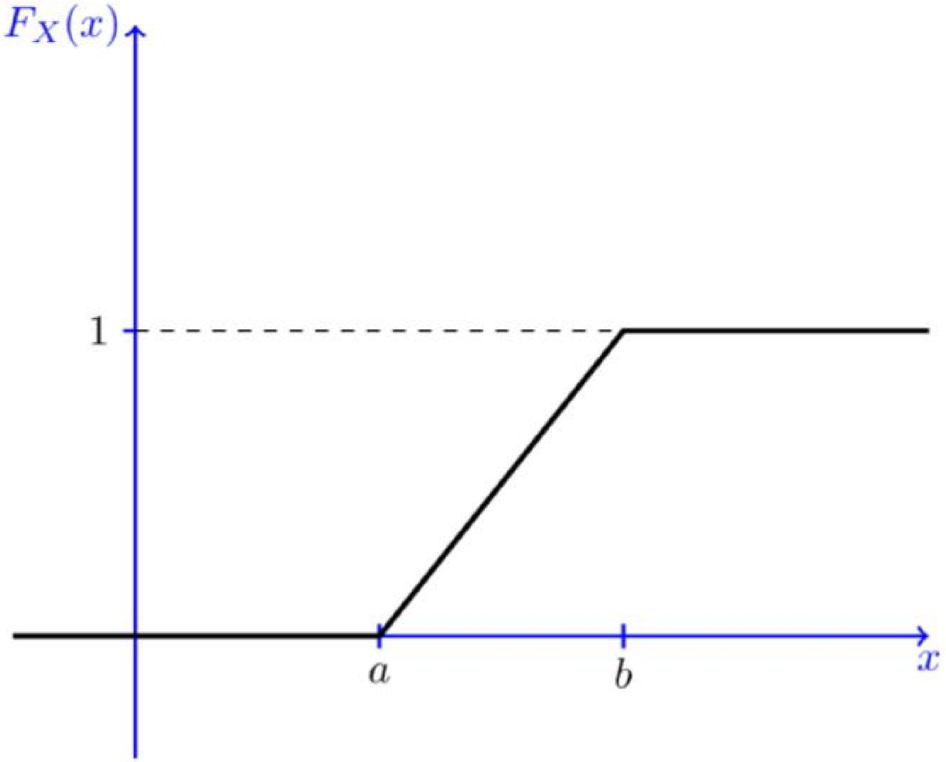
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leqslant x < b \\ 1, & x \geqslant b \end{cases}$$

Theorem 3.1.15. 设 $X \sim U(a,b)$，那么

$$E(X) = \frac{1}{2}(a+b)$$
$$\sigma^2(X) = \frac{1}{12}(b-a)^2$$



(a) 概率密度函数



(b) 概率分布函数

正态分布

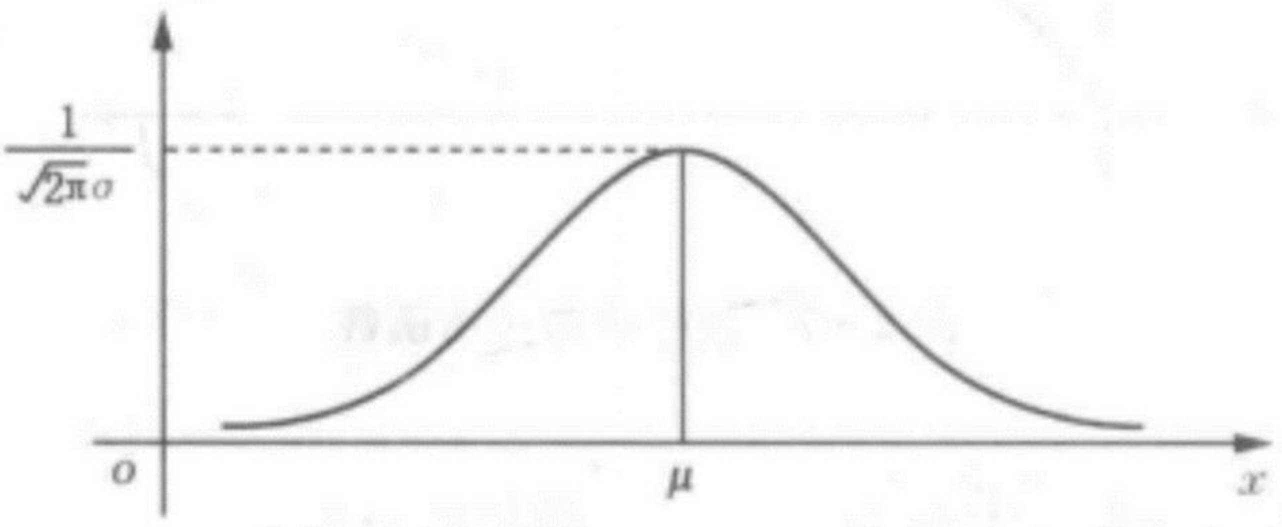
Definition 3.1.22. 【正态分布】设随机变量 X 具有概率密度

$$\boxed{f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty} \tag{3.38}$$

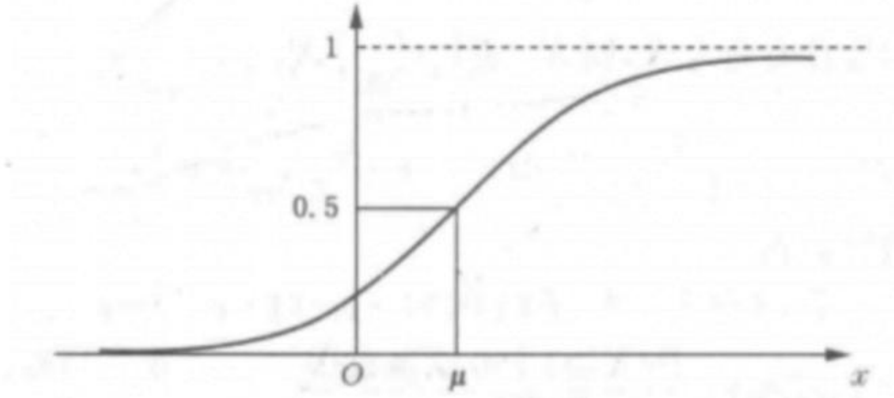
其中 $\mu,\sigma(\sigma>0)$ 为常数, 则称 X 服从参数为 μ,σ 的正态分布或高斯 $(Gauss)$ 分布, 记为 $X\sim N\left(\mu,\sigma^2\right)$.

Theorem 3.1.16. 设 $X\sim N\left(\mu,\sigma^2\right)$, 那么 $E(X)=\mu, D(X)=\sigma^2$.

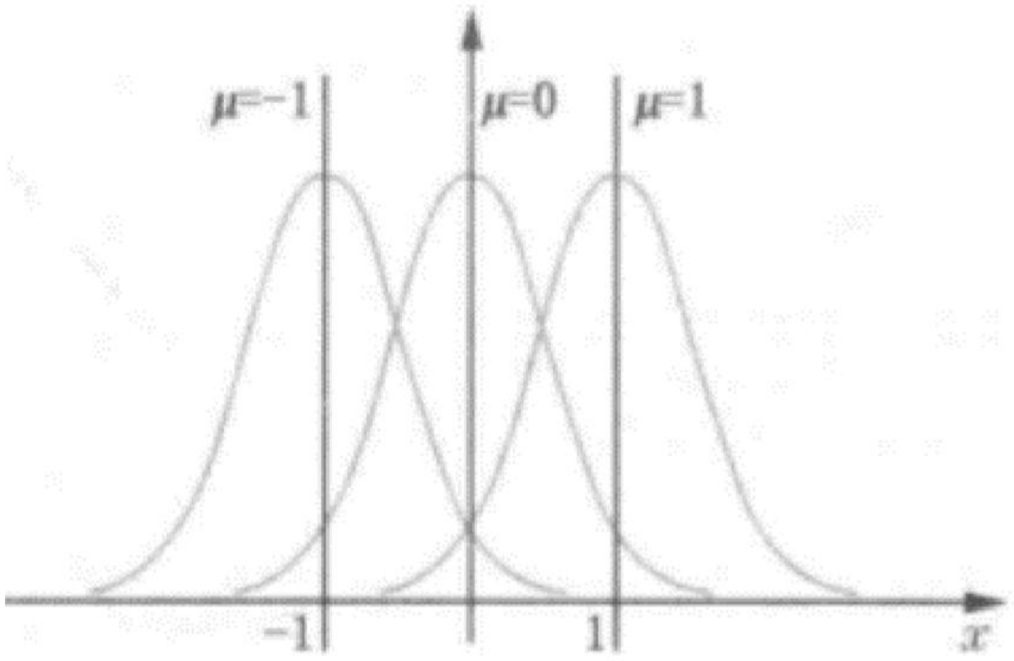
Theorem 3.1.17. 设 $X\sim N\left(\mu,\sigma^2\right)$, 则 $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$.



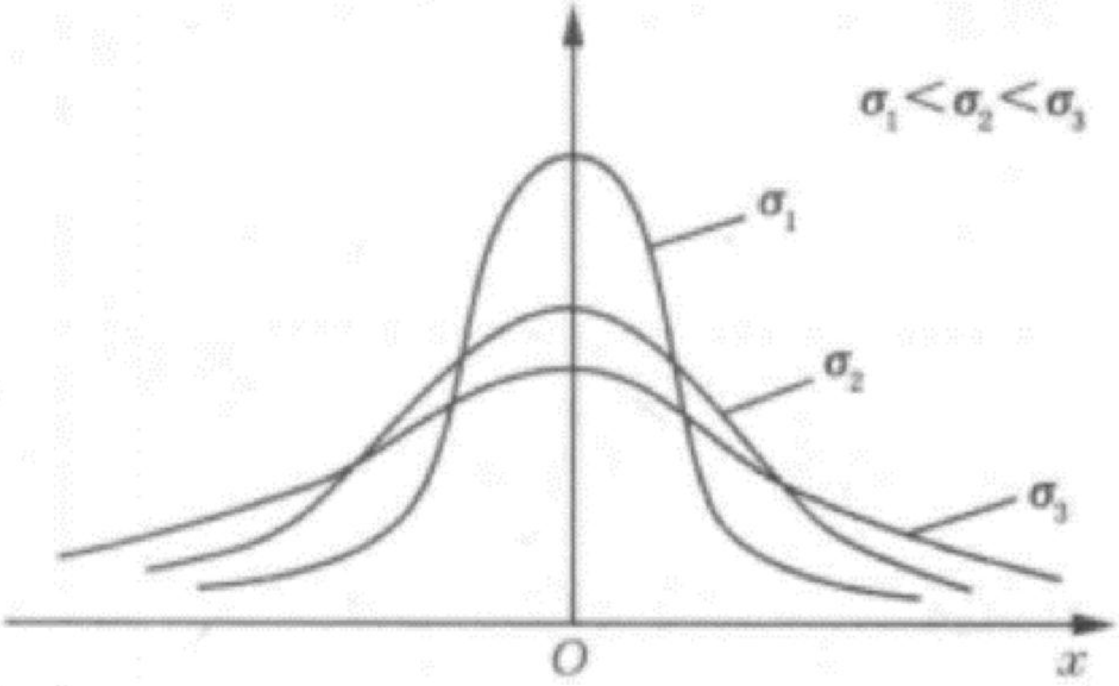
(a) 概率密度函数



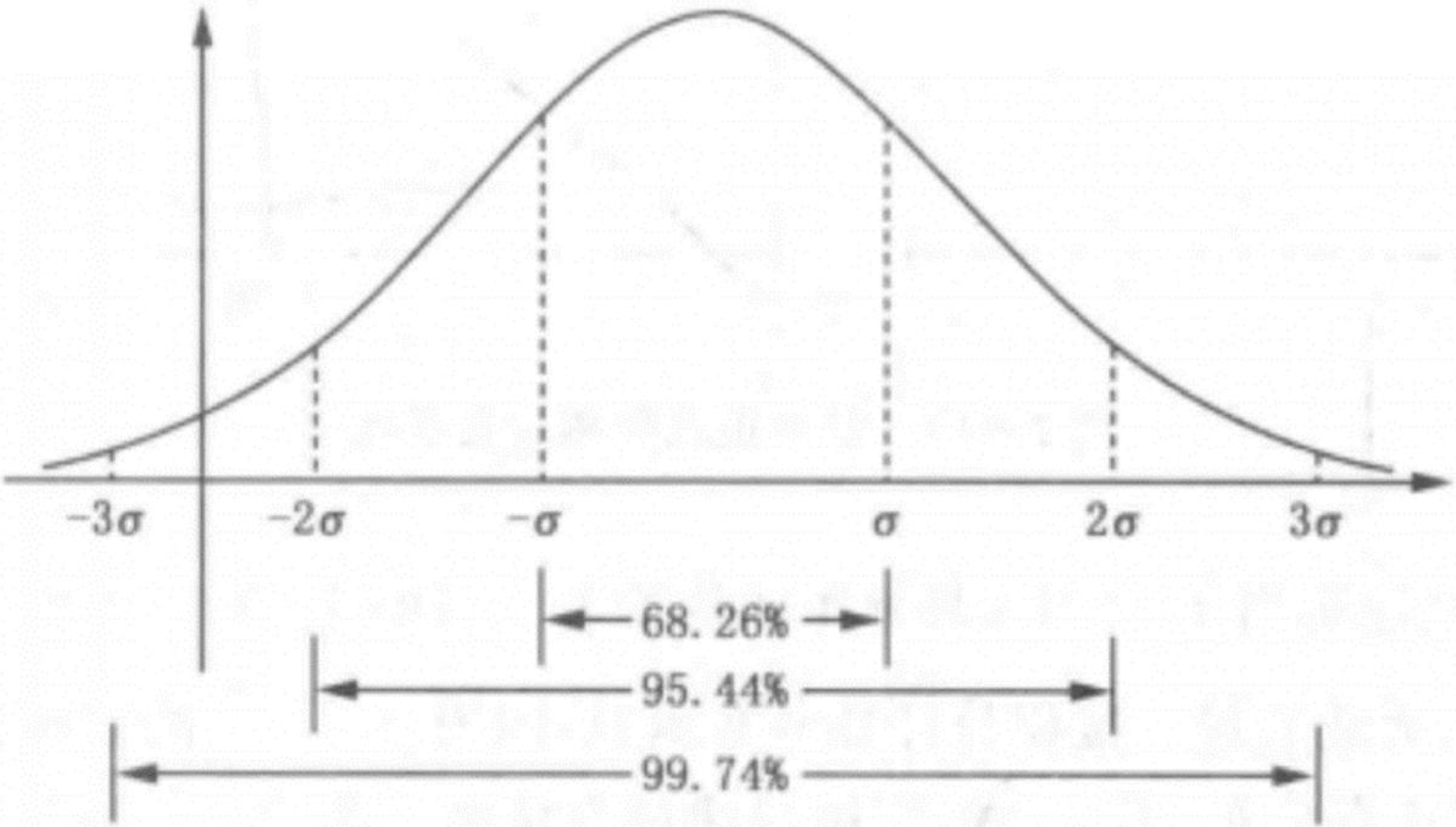
(b) 概率分布函数



$\mu=-1$
$\mu=0$
$\mu=1$
-1
1
x



$\sigma_1<\sigma_2<\sigma_3$
σ_1
σ_2
σ_3
O
x



−3σ
−2σ
−σ
σ
2σ
3σ
68. 26%
95. 44%
99. 74%

Definition 3.1.23. 设随机向量 $\boldsymbol{X}$ 的均值向量为 $\boldsymbol{\mu}$, 记

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$$

如果 $\boldsymbol{X}$ 的概率密度为

$$\boxed{f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n |P|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T P^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\}}$$

则 $\boldsymbol{X}$ 服从正态分布，记为 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, P)$. 其中 $P = \mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})$, 而且 P 对称正定.

Theorem 3.1.18. 设 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 分别服从正态分布，两者相互独立的充要条件是两者不相关，即

$$\boxed{f_{XY}(\boldsymbol{x},\boldsymbol{y}) = f_X(\boldsymbol{x})f_Y(\boldsymbol{y}) \iff \operatorname{cov}(\boldsymbol{X},\boldsymbol{Y}) = 0.}$$

对一维正态分布随机变量来说，即为 $\boxed{f_{XY}(x,y) = f_X(x)f_Y(y) \iff E(XY) = E(X)E(Y).}$

Theorem 3.1.19. 设 n 维随机向量 $\boldsymbol{X} \sim N(\boldsymbol{\mu}_X, P_X)$, $M \in R^{m\times n}$ 是常值矩阵, $\boldsymbol{b} \in R^m$ 是常值向量, 则 m 维随机向量 $\boldsymbol{Y} = M\boldsymbol{X} + \boldsymbol{b}$ 也服从正态分布, 且 $\boldsymbol{\mu}_Y = M\boldsymbol{\mu}_X + \boldsymbol{b}, P_Y = MP_XM^T$, 即 $\boldsymbol{Y} \sim N(M\boldsymbol{\mu}_X + \boldsymbol{b}, MP_XM^T)$.

Theorem 3.1.20 (中心极限定理). 设 $\boldsymbol{X}^i(i=1,2,\cdots,r)$ 是一组相互独立、同分布的 n 维随机向量，具有有限均值 $E(\boldsymbol{X}^i)$ 和协方差矩阵 P^i，令

$$\boldsymbol{Y}^r = \sum_{i=1}^{r} \boldsymbol{X}^i \tag{3.40}$$

$$\boldsymbol{Z}^r = (P^r)^{-\frac{1}{2}}(\boldsymbol{Y}^r - \bar{\boldsymbol{Y}}^r) \tag{3.41}$$

其中

$$\bar{\boldsymbol{Y}}^r = \sum_{i=1}^{r} \boldsymbol{X}^r, \quad P^r = \sum_{i=1}^{r} P^i$$

那么

$$\boxed{\lim_{r\to+\infty} f(\boldsymbol{z}^r) = \frac{1}{\sqrt{(2\pi)^n}} \exp\{-\frac{\boldsymbol{z}^T\boldsymbol{z}}{2}\}} \tag{3.42}$$

上述定理说明，当 $r\to+\infty$ 时，$\boldsymbol{Z}^r$ 趋于标准正态分布的随机向量。因此，大量的微观上独立随机因素之和在宏观上可以用正态分布来描述.

符号简记

到目前为止，我们分别用 $X, \boldsymbol{X}$ 表示随机变量和随机向量. 在不会产生歧义情况下，后面我们将分别用 $x, \boldsymbol{x}$ 表示随机变量和随机向量. 例如，$x \sim N(\mu, \sigma^2)$ 表示服从均值为 μ、方差为 σ^2 的正态分布随机变量 $x, \boldsymbol{x} \sim N(\bar{\boldsymbol{x}}, P_x)$ 表示服从均值为 $\bar{\boldsymbol{x}}$、协方差矩阵为 P_x 的正态分布随机向量 $\boldsymbol{x}$.