

第一章 概率论基础

张立国

zhangliguo@hrbeu.edu.cn

哈尔滨工程大学

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.1 概率空间

● 样本空间

- 一个试验 (experiment) 所有可能出现的结果的全体称为样本空间 (sample space) , 记为 Ω .

● 样本点

- 试验的一个结果称为样本点 (sample) , 记为 ω , 即 $\Omega = \{\omega\}$.

● 随机事件

- 样本空间的某个子集称为随机事件, 简称为事件 (event) .

1.1 概率空间

● 定义1.1.1

设 Ω 是样本空间， \mathcal{F} 是 Ω 的某些子集构成的集合，如果

(1) $\Omega \in \mathcal{F}$

(2) 若 $A \in \mathcal{F}$ ，则 $\bar{A} \in \mathcal{F}$ (对差运算封闭)

(3) 若 $A_n \in \mathcal{F}, n=1,2,\dots$ ，则 $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (对并运算封闭)

那么称 \mathcal{F} 为 一事件域，也称 \mathcal{F} 为 σ 域

显然，如果 \mathcal{F} 是一事件域，那么：

(1) $\emptyset \in \mathcal{F}$

(2) 若 $A, B \in \mathcal{F}$ ，则 $A-B \in \mathcal{F}$

(3) 若 $A_n \in \mathcal{F}, n=1,2,\dots$ 则 $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

1.1 概率空间

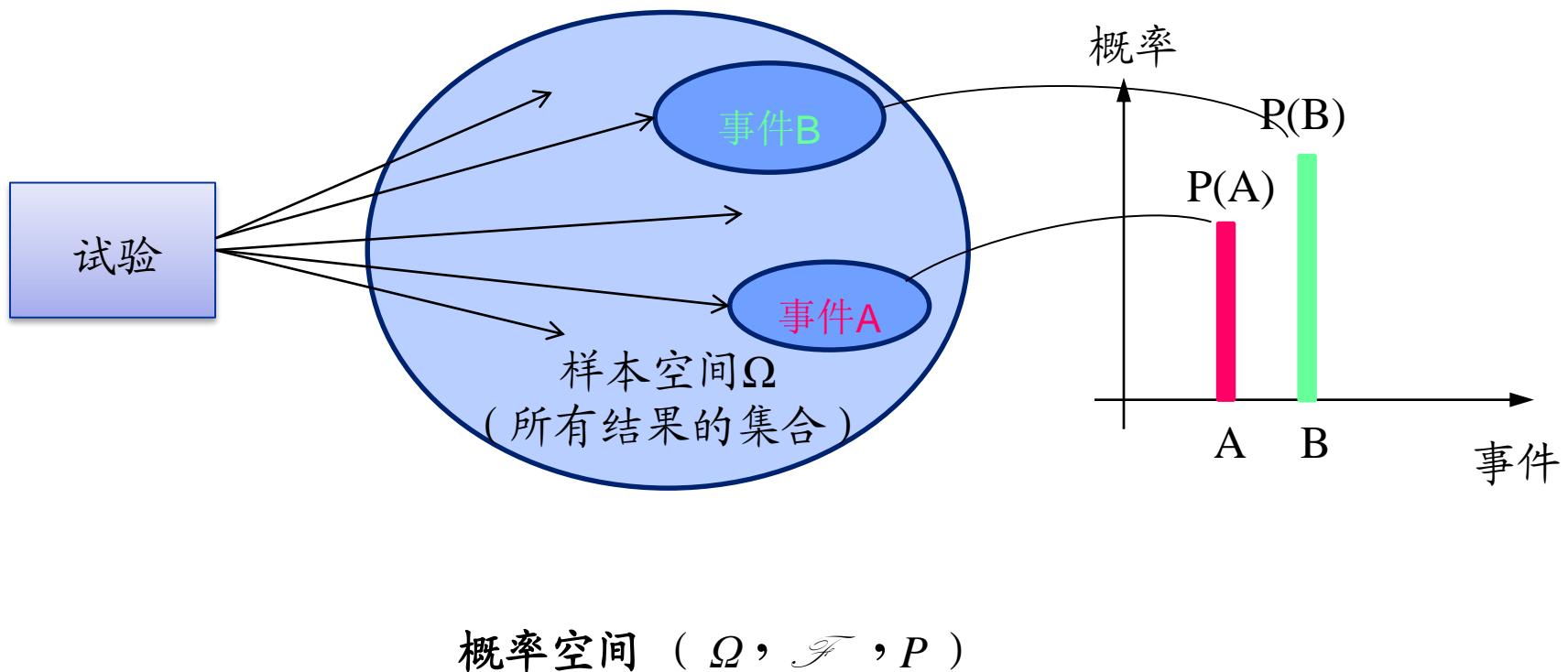
● 定义1.1.2 设 Ω 是样本空间， \mathcal{F} 是一事件域，定义在 \mathcal{F} 上的实值函数 $P(\cdot)$ ，如果满足

- (1) (Nonnegativity, 非负性) 对 \mathcal{F} 中的任意事件 A , $P(A) \geq 0$;
- (2) (Normalization, 归一性) $P(\Omega) = 1$;
- (3) (Additivity, 可列可加性) 对 \mathcal{F} 中任意事件 A_n , $n = 1, 2, \dots$, $A_i A_j = \emptyset$, $i \neq j$, $i, j = 1, 2, \dots$, 有

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

那么，称 P 是二元组 (Ω, \mathcal{F}) 上的概率 (Probability)，称 $P(A)$ 为事件 A 的概率，称三元组 (Ω, \mathcal{F}, P) 为概率空间 (Probability Space) .

1.1 概率空间



1.1 概率空间

● 概率的性质

(1) $P(\emptyset)=0$

(2) 若 $A_i \in \mathcal{F}, i=1,2,\dots,n, A_i A_j = \emptyset, i \neq j, i,j = 1,2,\dots,n,$ 则

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (\text{有限可加性})$$

(3) $A, B \in \mathcal{F}, A \subset B,$ 则 $P(B-A) = P(B)-P(A)$

(4) $A, B \in \mathcal{F}, A \subset B,$ 则 $P(A) \leq P(B)$

(5) 若 $A_n \in \mathcal{F},$ 则 $P(A) \leq 1$

(6) 若 $A_n \in \mathcal{F},$ 则 $P(\bar{A}) \leq 1 - P(A)$

(7) 若 $A_n \in \mathcal{F},$ 则 $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$

(8) 若 $A_i \in \mathcal{F}, i=1,2,\dots,n,$ 则 (包含排斥原理)

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

1.1 概率空间

- 一列事件 $A_n \in \mathcal{F}$, $n=1,2,\dots$, 称为单调递增的事件列,
如果 $A_n \subset A_{n+1}, n=1,2,\dots$ 。 一列事件 $A_n \in \mathcal{F}$, $n=1,2,\dots$,
称为单调递减的事件列, 如果 $A_n \supset A_{n+1}, n=1,2,\dots$
- **定理1.1.1** 设 $A_n \in \mathcal{F}$, $n=1,2,\dots$

(1) 若 $A_n, n=1,2,\dots$, 是单调递增的事件列, 则

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right)$$

(2) 若 $A_n, n=1,2,\dots$, 是单调递减的事件列, 则

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

1.1 概率空间

- 定理1.1.1 设 $A_n \in \mathcal{F}$, $n=1,2,\dots$

(1) 若 A_n , $n=1,2,\dots$, 是单调递增的事件列, 则

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\overline{\bigcup_{n=1}^{\infty} A_n}\right)$$

证明: (1) 令 $B_1 = A_1$, $B_n = A_n \overline{\bigcup_{i=1}^{n-1} A_i} = A_n \bar{A}_{n-1}$, $n=2,3,\dots$, 则 $B_i B_j = \emptyset$, $i \neq j$, $i, j = 1, 2, \dots$, 且 $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. 于是 $P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(B_k)$

$$= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=1}^n A_k\right) = \lim_{n \rightarrow \infty} P(A_n)$$

即:

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\overline{\bigcup_{n=1}^{\infty} A_n}\right)$$

1.1 概率空间

- 定理1.1.1 设 $A_n \in \mathcal{F}$, $n=1,2,\dots$

(2) 若 A_n , $n=1,2,\dots$, 是单调递减的事件列, 则

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

证明: (2) 若 A_n , $n=1,2,\dots$ 单调递减, 则 \bar{A}_n ($n=1,2,\dots$) 单调递增,

由(1)得 $\lim_{n \rightarrow \infty} P(\bar{A}_n) = P\left(\bigcap_{n=1}^{\infty} \bar{A}_n\right)$, 而 $P\left(\bigcup_{n=1}^{\infty} \bar{A}_n\right) = P\left(\overline{\bigcup_{n=1}^{\infty} A_n}\right) = 1 - P\left(\bigcup_{n=1}^{\infty} A_n\right)$

$$\lim_{n \rightarrow \infty} P(\bar{A}_n) = \lim_{n \rightarrow \infty} (1 - P(A_n)) = 1 - \lim_{n \rightarrow \infty} P(A_n)$$

所以

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

1.1 概率空间

- 定义1.1.3 设 (Ω, \mathcal{F}, P) 为一概率空间， $A, B \in \mathcal{F}$ 且 $P(A) > 0$ ，则称

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

为事件 A 发生的条件下事件 B 发生的条件概率.

不难验证，条件概率 $P(\cdot | A)$ 符合定义1.1.2中的三个条件

(1) $\forall B \in \mathcal{F}, P(B|A) \geq 0$

(2) $P(\Omega | A) = 1$

(3) 设 $B_n \in \mathcal{F}, n = 1, 2, \dots, B_i B_j = \emptyset, i \neq j, i, j = 1, 2, \dots$ ，则

$$P\left(\bigcup_{n=1}^{\infty} B_n | A\right) = \sum_{n=1}^{\infty} P(B_n | A)$$

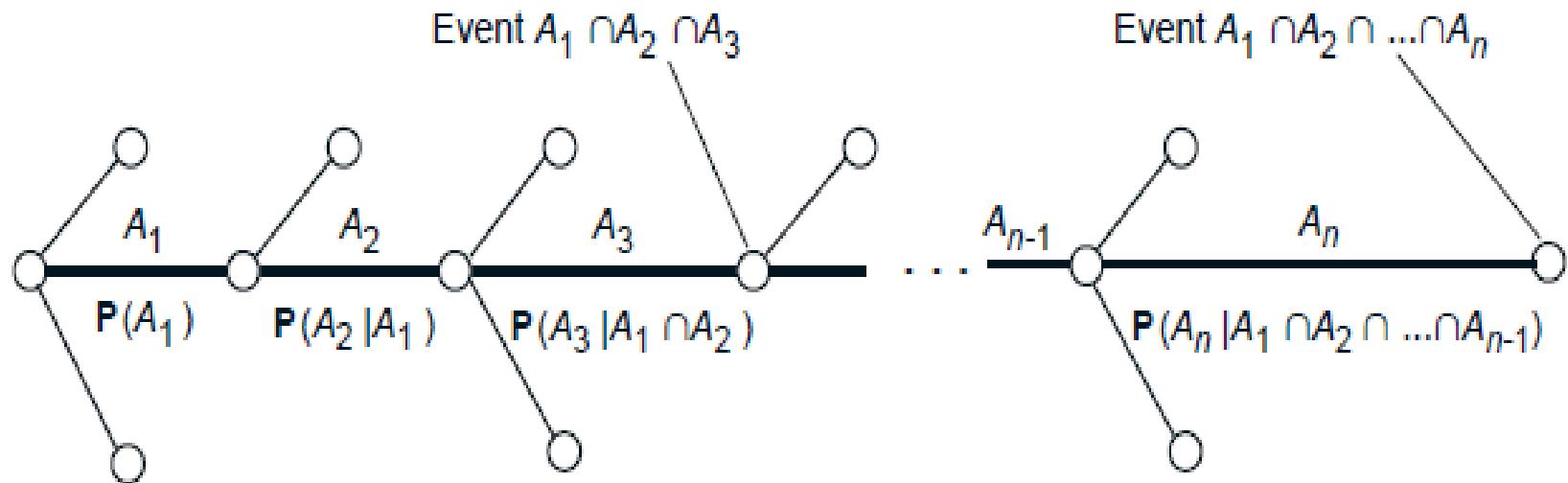
1.1 概率空间

● 定理 1.1.2 设 (Ω, \mathcal{F}, P) 是一概率空间，有：

(1) (乘法公式) 若 $A_i \in \mathcal{F}$, $i=1, 2, \dots, n$, 且 $P(A_1 A_2 \dots A_n) > 0$,

则 $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 A_2) \dots P(A_n | A_1 A_2 \dots A_{n-1})$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2)$$

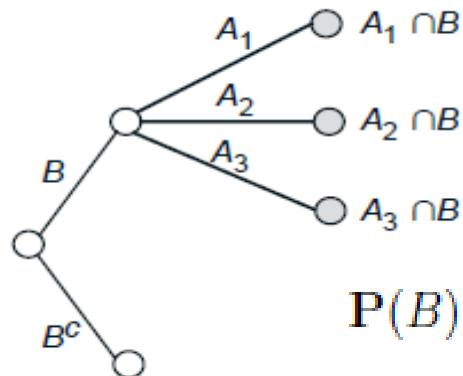
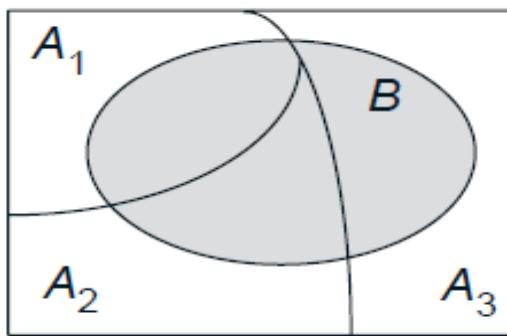


$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

1.1 概率空间

- (2) (全概率公式) 设 $B \in \mathcal{F}$, $A_i \in \mathcal{F}$, $P(A_i) > 0, i=1, 2, \dots$,
且 $A_i A_j = \emptyset, i \neq j, i, j = 1, 2, \dots$, $\bigcup_{i=1}^{\infty} A_i \supset B$, 则

$$P(B) = \sum_{i=1}^{\infty} P(A_i)P(B | A_i)$$



$$P(A_i \cap B) = P(A_i)P(B | A_i)$$

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B).$$

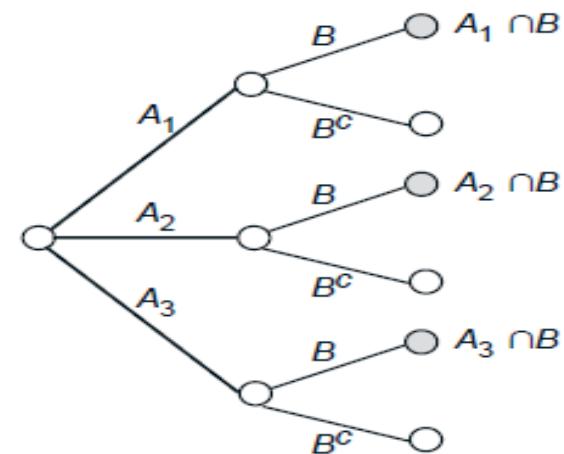
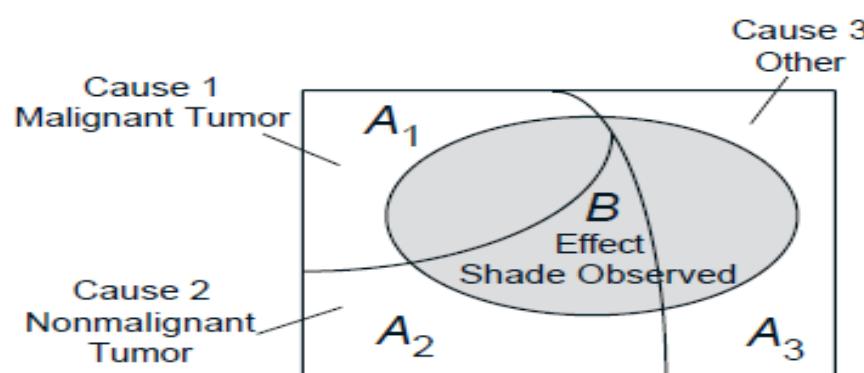
$$B = (A_1 \cap B) \cup \cdots \cup (A_n \cap B) \quad P(B) = P(A_1)P(B | A_1) + \cdots + P(A_n)P(B | A_n)$$

1. 1 概率空间

- (3) (贝叶斯(Bayes)公式) 设 $B \in \mathcal{F}$, $P(B) > 0$, $A_i \in \mathcal{F}$,

$P(A_i) > 0, i = 1, 2, \dots$, 且 $A_i A_j = \emptyset, i \neq j, i = 1, 2, \dots$, $\bigcup_{i=1}^{\infty} A_i \supset A$,

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B | A_j)}, i = 1, 2, \dots$$



$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + P(A_3)P(B | A_3)}, \quad i = 1, 2, 3.$$

1.1 概率空间

- **定义1.1.4** 设 (Ω, \mathcal{F}, P) 为一概率空间， $A_i \in \mathcal{F}$ ， $i=1, 2, \dots, n$ ，如果对于任意的 $k (1 < k \leq n)$ 及任意的 $1 \leq i_1 < i_2 < \dots < i_k \leq n$ ，有

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$$

则称事件 $A_1 A_2 \dots A_n$ 相互独立 (independent) .

- **定理1.1.3** 设 $A, B \in \mathcal{F}$ 相互独立，则 A 所生成的 σ 域
 $\mathcal{F}_A = \{A, \bar{A}, \emptyset, \Omega\}$ 中的任意一个事件和 B 所生成的 σ 域
 $\mathcal{F}_B = \{B, \bar{B}, \emptyset, \Omega\}$ 中的任意一个事件都相互独立。这时，
我们称这两个 σ 域 \mathcal{F}_A 和 \mathcal{F}_B 相互独立.

1.1 概率空间

- 定理1.1.4 设 $A, B, C \in \mathcal{F}$ 相互独立，则

- (1) A 与 BC 相互独立；
- (2) A 与 $B \cup C$ 相互独立；
- (3) A 与 $B-C$ 相互独立；
- (4) A 所生成的 σ 域中的任一事件与 B 和 C 所生成的 σ 域.

$\mathcal{F}_{B,C} = \{ B, \bar{B}, C, \bar{C}, BC, B\bar{C}, C\bar{B}, \bar{B}\bar{C}, (C\bar{B} \cup B\bar{C}), (BC \cup \bar{B}\bar{C}), B \cup C, \bar{B} \cup C, \bar{B} \cup \bar{C}, B \cup \bar{C}, \Omega, \emptyset \}$ 中的任意一个事件都相互独立.

1.1 概率空间

- **推论1.1.1** 设 $A, B, C \in \mathcal{F}$ 相互独立，将 A, B, C 任意分为两组， 则他们各自生成的 σ 域仍然相互独立.
- **定理1.1.5** 设 $A_i \in \mathcal{F}, i=1, 2, \dots, n$ 相互独立， 将 $A_i, i=1, 2, \dots, n$ ， 任意分成 $m (m \leq n)$ 组，并对各组中的事件施以积、和、逆运算后，所得到的事件 B_1, B_2, \dots, B_m 也是相互独立的. 从而这 m 组事件各自所生成的 σ 域也是相互独立的.

1.1 概率空间

- 定理1.1.5 蕴含以下的有用的具体结论:

(1) 若 A_1, A_2, A_n 相互独立, 则 $\overline{A_1}, \overline{A_2} \dots \overline{A_n}$ 也相互独立, 从而

$$P\left(\bigcup_{i=1}^n A_i\right) = 1 - P\left(\bigcup_{i=1}^n \overline{A_i}\right) = 1 - P\left(\bigcap_{i=1}^n \overline{A_i}\right) = 1 - P(\overline{A_1})P(\overline{A_2}) \cdots P(\overline{A_n})$$

(2) 一列独立事件中的任何一部分事件也相互独立.

(3) 若一列事件相互独立, 则将其中任一部分改写为对立事件, 所得的事件也相互独立.

1.1 概率空间（小结）

- 概率空间

- 样本空间
- 事件域
- 概率

- 条件概率

- 乘法公式
- 全概率公式
- Bayes 公式

- 事件的独立性

第1章 概率论基础

- 1.1 概率空间
- **1.2 随机变量及分布**
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.2 随机变量及其分布

- 随机变量是概率论的主要研究对象.
- 在概率模型中，试验结果是数值化的，或与某些数值相联系. 当我们讨论这些数值的时候，通常给这些数值确定概率. 我们通过随机变量实现这个任务.
- 随机变量是试验结果的一个实值函数.
- 描述随机变量的概率特性用分布函数.

1.2 随机变量及其分布

- 定义1.2.1 设 (Ω, \mathcal{F}, P) 为一概率空间，定义在 Ω 上的实函数 $X(\cdot)$ ，如果 $\forall x \in \mathbb{R}$

$$\{\omega | X(\omega) \leq x\} \in \mathcal{F}$$

则称 X 是 \mathcal{F} 的随机变量（Random Variable, r.v.）.

称

$$F(x) = P(X \leq x) \quad -\infty < x < +\infty$$

为随机变量 X 的分布函数（Cumulative Distribution Function, CDF）.

1.2 随机变量及其分布

- 分布函数 $F(x)$ 的性质：

(1) $F(x)$ 是单调不减函数，即若 $x_1 < x_2$ 则 $F(x_1) \leq F(x_2)$

(2) $F(x)$ 是右连续函数，即 $\forall x \in R, F(x+0) = F(x)$

(3) $F(-\infty) \stackrel{\Delta}{=} \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) \stackrel{\Delta}{=} \lim_{x \rightarrow +\infty} F(x) = 1$

- 同时可以证明，设 $F(x), x \in R$ 是单调不减、右连续的函数，并且 $F(-\infty) = 0, F(+\infty) = 1$ ，则必存在概率空间 (Ω, \mathcal{F}, P) 及其上的一个随机变量 X 使得 X 以 $F(x)$ 为其分布函数

1.2 随机变量及其分布

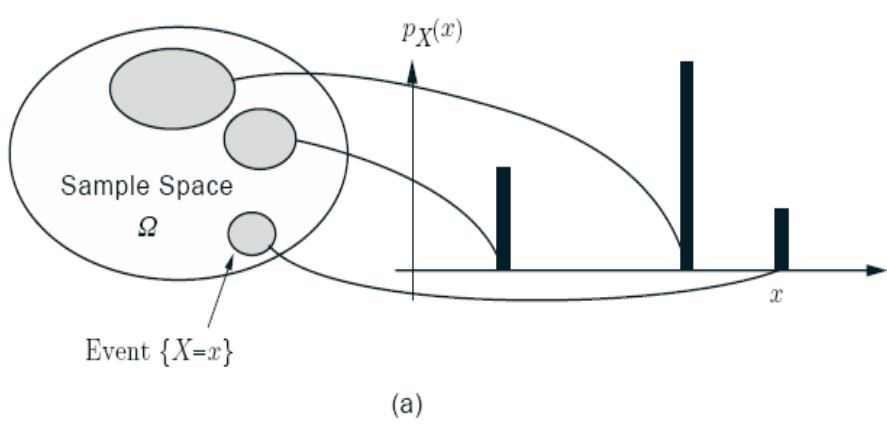
随机变量有两种类型： 离散型和连续型随机变量

- 随机变量 X 的可能取值为有限 (finite) 个或可列无限个 (countably infinite), 则称 X 为离散型随机变量 (discrete r.v.) .
- 离散型随机变量 X 的分布可用分布律 (Probability Mass Function, PMF) 来描述, 即

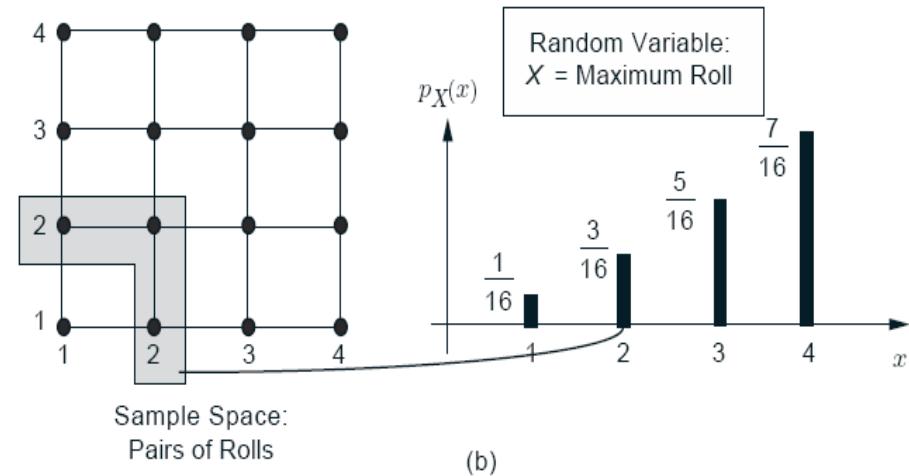
$$P(X = x_i) = p_i, \quad i=1,2,\dots$$

这时, X 的分布函数为 $F(x) = \sum_{x_i \leq x} p_i, x \in R$

1.2 随机变量及其分布



(a)



(b)

Figure 2.2: (a) Illustration of the method to calculate the PMF of a random variable X . For each possible value x , we collect all the outcomes that give rise to $X = x$ and add their probabilities to obtain $p_X(x)$. (b) Calculation of the PMF p_X of the random variable $X = \text{maximum roll}$ in two independent rolls of a fair 4-sided die. There are four possible values x , namely, 1, 2, 3, 4. To calculate $p_X(x)$ for a given x , we add the probabilities of the outcomes that give rise to x . For example, there are three outcomes that give rise to $x = 2$, namely, (1,2), (2,1), (2,2). Each of these outcomes has probability $1/16$, so $p_X(2) = 3/16$, as indicated in the figure.

1.2 随机变量及其分布

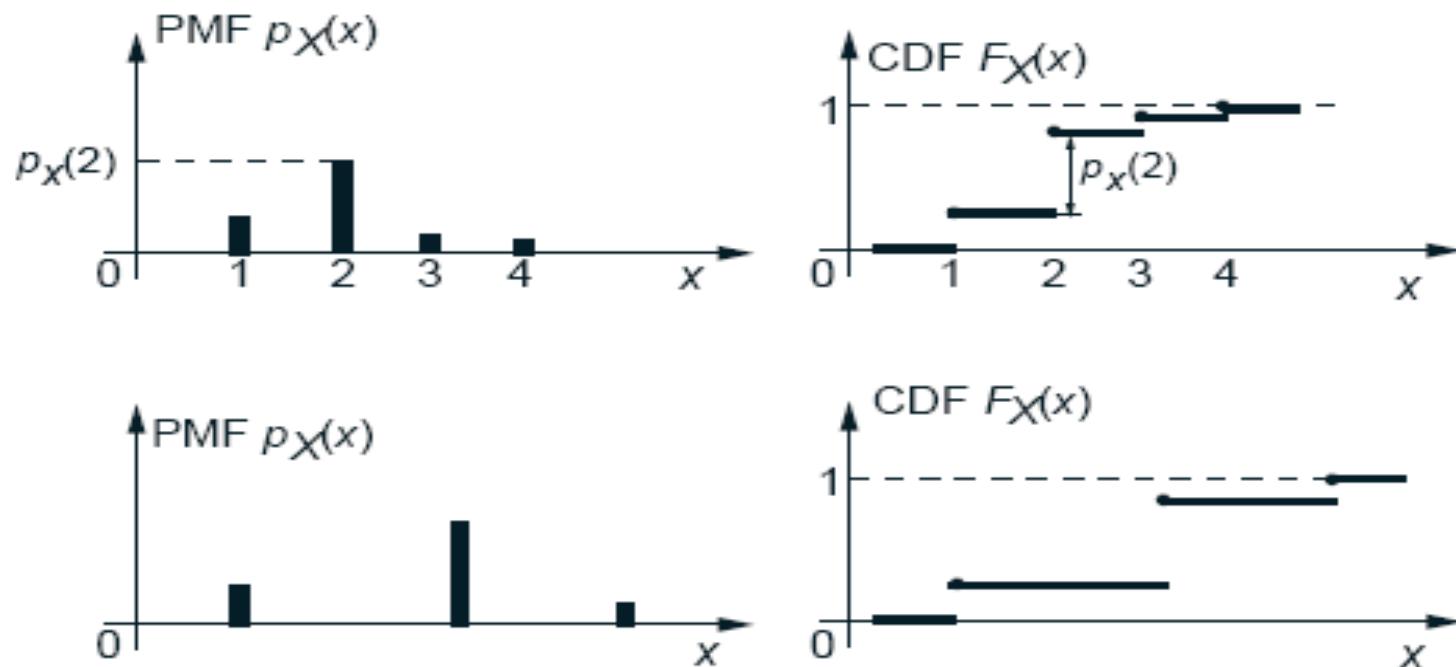


Figure 3.6: CDFs of some discrete random variables. The CDF is related to the PMF through the formula

$$F_X(x) = \mathbf{P}(X \leq x) = \sum_{k \leq x} p_X(k),$$

and has a staircase form, with jumps occurring at the values of positive probability mass. Note that at the points where a jump occurs, the value of F_X is the larger of the two corresponding values (i.e., F_X is continuous from the right).

1.2 随机变量及其分布

- 设随机变量 X 的分布函数为 $F(x)$, 如果存在非负可积函数 $f(x)$, 使得

$$F(x) = \int_{-\infty}^x f(t)dt, x \in R$$

则称 X 为连续型随机变量 (Continuous r.v.) , $f(x)$ 为连续型随机变量 X 的概率密度函数 (Probability Density Function, PDF) .

1.2 随机变量及其分布

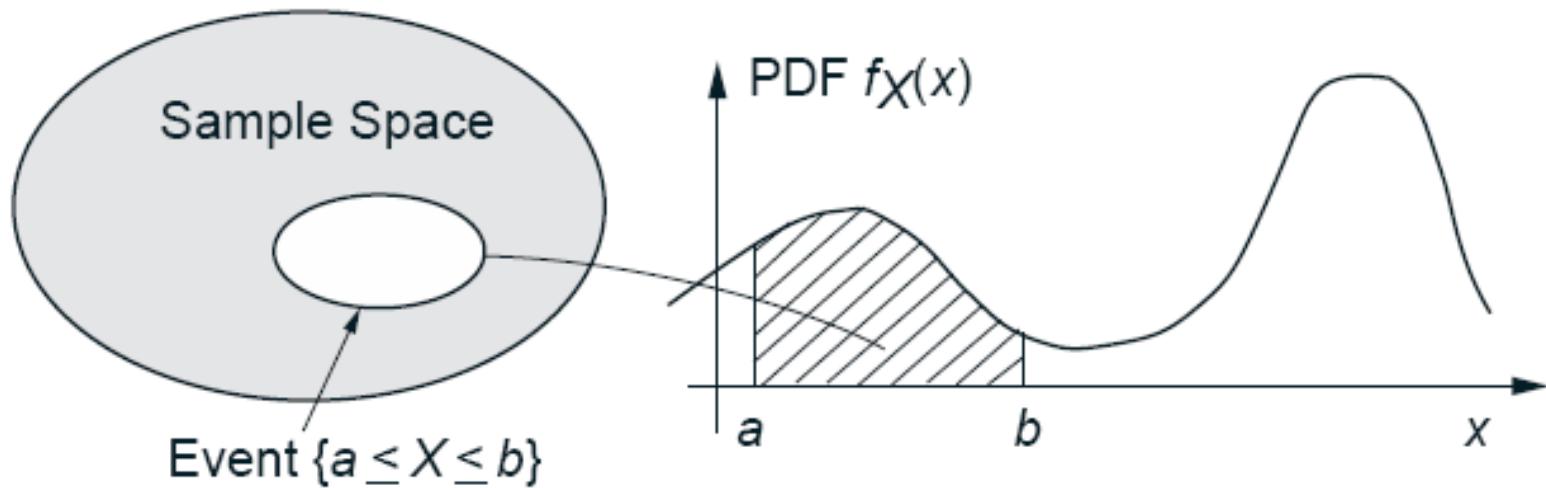


Figure 3.1: Illustration of a PDF. The probability that X takes value in an interval $[a, b]$ is $\int_a^b f_X(x) dx$, which is the shaded area in the figure.

1.2 随机变量及其分布

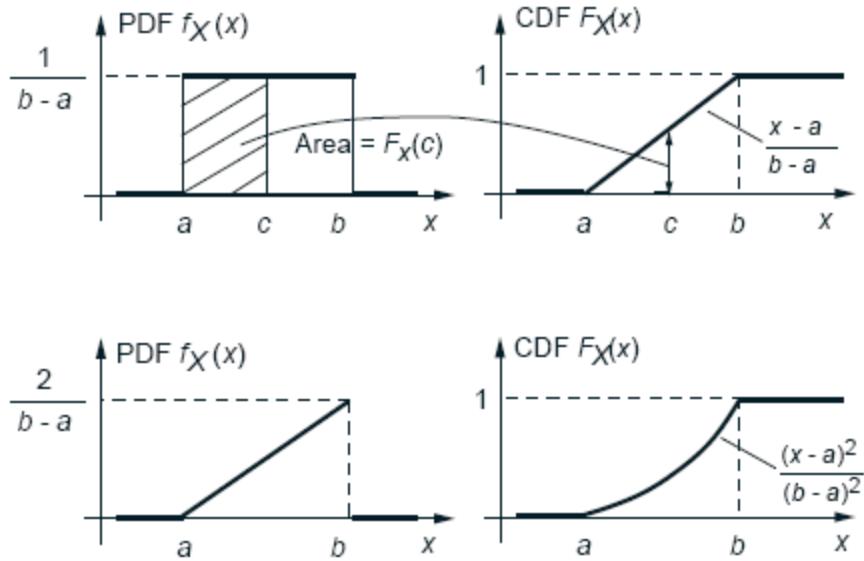


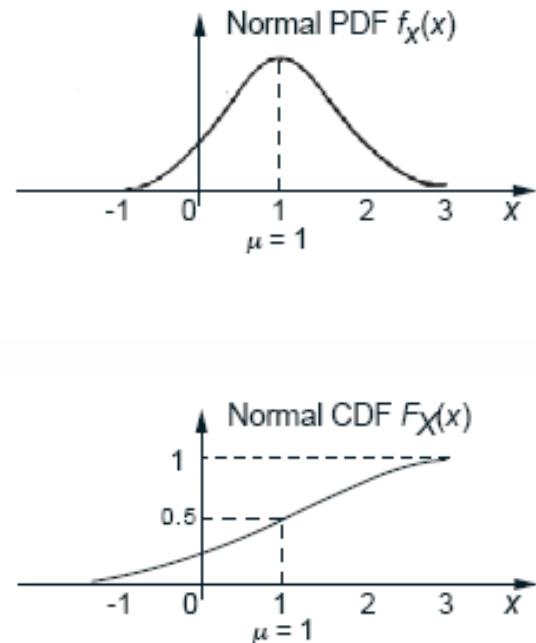
Figure 3.7: CDFs of some continuous random variables. The CDF is related to the PDF through the formula

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

Thus, the PDF f_X can be obtained from the CDF by differentiation:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

For a continuous random variable, the CDF has no jumps, i.e., it is continuous.



A normal PDF and CDF, with $\mu = 1$ and $\sigma^2 = 1$.

1.2 随机变量及其分布

- 定义1.2.2 设 (Ω, \mathcal{F}, P) 为一概率空间，定义在 Ω 上的 n 元实函数 $X(\cdot) = (X_1(\cdot), X_2(\cdot), \dots, X_n(\cdot))$, 如果 $\forall x = (x_1, x_2, \dots, x_n) \in R$,
 $\{\omega | X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\} \in \mathcal{F}$
则称 $X = (X_1, X_2, \dots, X_n)$ 为 n 维随机变量 或 n 维随机向量.
称

$$F(x) = F(x_1, x_2, \dots, x_n) + P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$
$$x = (x_1, x_2, \dots, x_n) \in R^n$$

为 X 的 联合分布函数 (Joint CDFs) .

1.2 随机变量及其分布

- 设 X 是 n 维随机变量，则 X 的联合分布函数具有下列性质

(1) $F(x_1, x_2, \dots, x_n)$ 对任一 $x_i (i=1, 2, \dots, n)$ 是单调不减函数；

(2) $F(x_1, x_2, \dots, x_n)$ 对任一 $x_i (i=1, 2, \dots, n)$ 是右连续函数；

(3) $F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0, i = 1, 2, \dots, n,$

$$F(+\infty, +\infty, \dots, +\infty) = 1$$

(4) 设 $x_i \leq y_i, i=1, 2, \dots, n$, 则

$$F(y_1, y_2, \dots, y_n) - \sum_{i=1}^n F(y_1, y_2, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_n)$$

$$+ \sum_{1 \leq i < j \leq n} F(y_1, y_2, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_{j-1}, x_j, y_{j+1}, \dots, y_n)$$

$$- \dots + (-1)^n F(x_1, x_2, \dots, x_n) \geq 0$$

1.2 随机变量及其分布

- 若 n 维随机变量 X 的可能取值为有限对或可列无限对，则称 n 维随机变量为离散型 n 维随机变量.

离散型 n 维随机变量 $X=(X_1, X_2, \dots, X_n)$ 的分布可用联合分布律 (Joint PMFs) 来描述，即

$$P(X_1=x_1, X_2=x_2, \dots, X_n=x_n)$$

其中 $x_i \in I_i$, I_i 是离散集, $i=1, 2, \dots, n$ 这时 X 的联合分布函数为

$$F(u_1, u_2, \dots, u_n) = \sum_{x_1 \leq u_1} \sum_{x_2 \leq u_2} \cdots \sum_{x_n \leq u_n} P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

1.2 随机变量及其分布

- 设 n 维随机变量 X 的联合分布函数为 $F(x_1, x_2, \dots, x_n)$, 如果存在非负可积函数 $f(x) = f(x_1, x_2, \dots, x_n)$, $x \in R^n$ 使得

$$F(u_1, u_2, \dots, u_n) = \int_{-\infty}^{u_1} \int_{-\infty}^{u_2} \cdots \int_{-\infty}^{u_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots d_n$$

则称 X 为连续型 n 维随机变量, $f(x_1, x_2, \dots, x_n)$, 称为连续型 n 维随机变量 X 的联合概率密度函数 (Joint PDFs) .

1.2 随机变量及其分布

保留 $k (1 \leq k < n)$ 个 x_i ，比如 x_1, x_2, \dots, x_k ，而令其他 x_i 都趋于 $+\infty$ 得到 k 维 边缘分布函数 (marginal CDF)

$$F(x_1, x_2, \dots, x_k, +\infty, \dots, +\infty)$$

若 X 是连续型 n 维随机变量，则有

$$\begin{aligned} & F(x_1, x_2, \dots, x_k, +\infty, +\infty, \dots, +\infty) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_k} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y_1, y_2, \dots, y_n) dy_1 dy_2 \cdots dy_n \end{aligned}$$

可见 $F(x_1, x_2, \dots, x_k) = F(x_1, x_2, \dots, x_k, +\infty, \dots, +\infty)$ 也是连续型 k 维随机变量的联合分布函数，其联合概率密度函数为：

$$g(x_1, x_2, \dots, x_k) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_{k+1} \cdots dx_n$$

1.2 随机变量及其分布

特别的，当 $k=1$ 时， n 维随机变量 $X=(X_1, X_2, \dots, X_n)$ 的 n 个边缘分布函数和 n 个边缘概率密度函数分别为

$$F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$$

和

$$f_{X_1}(x_1), f_{X_2}(x_2), \dots, f_{X_n}(x_n)$$

1.2 随机变量及其分布

- 定义1.2.3 设 $X=(X_1, X_2, \dots, X_n)$ 是一 n 维随机变量，其联合分布函数和边缘概率密度函数分别为 $F(x_1, x_2, \dots, x_n)$ 和 $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$ 如果对于任意的 $x_1, x_2, \dots, x_k \in R$ 有 $F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$ 则称随机变量 X_1, X_2, \dots, X_n 相互独立 (independent) .

1.2 随机变量及其分布

- 若 $X = (X_1, X_2, \dots, X_n)$ 是 离散型 n 维随机变量，则 X_1, X_2, \dots, X_n 相互独立的充要条件是 联合分布律是各边缘分布律的乘积，即

$$P(X_1=x_1, X_2=x_2, \dots, X_n=x_n) = P(X_1=x_1)P(X_2=x_2) \cdots P(X_n=x_n)$$

- 若 $X = (X_1, X_2, \dots, X_n)$ 是 连续型 n 维随机变量，则 X_1, X_2, \dots, X_n 相互独立的充要条件是 联合概率密度函数是各边缘概率密度函数的乘积，即

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n), \quad x_1, x_2, \dots, x_n \in R$$

1.2 随机变量及其分布

- 如果 X 是随机变量（一维或多维）， $g(x)$ 是已知的连续函数，则 $g(X)$ 也是随机变量。关于 $g(X)$ 的分布，有以下定理：
- 定理1.2.1** 设连续型 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的联合概率密度函数为 $f_X(x_1, x_2, \dots, x_n)$ ， n 元函数 $y_i = y_i(x_1, x_2, \dots, x_n)$ ，满足：
(1) 存在唯一的反函数 $x_i = x_i(y_1, y_2, \dots, y_n)$ ，即方程组

$$\begin{cases} y_1(x_1, x_2, \dots, x_n) = y_1 \\ y_2(x_1, x_2, \dots, x_n) = y_2 \\ \vdots \\ y_n(x_1, x_2, \dots, x_n) = y_n \end{cases}$$

存在唯一的实数解 $x_i = x_i(y_1, y_2, \dots, y_n)$ ， $i = 1, 2, \dots, n$ ；

1.2 随机变量及其分布

(2) $y_i = y_i(x_1, x_2, \dots, x_n)$, 及 $x_i = x_i(y_1, y_2, \dots, y_n)$, $i=1, 2, \dots, n$ 都是连续的.

(3) $\frac{\partial x_i}{\partial y_j}, \frac{\partial y_i}{\partial x_j}, i, j = 1, 2, \dots, n$ 存在且连续, 令

$$J = \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(y_1, y_2, \dots, y_n)}$$

则, n 维随机变量 $Y = (Y_1, Y_2, \dots, Y_n)$, $Y_i = y_i(X_1, X_2, \dots, X_n)$, $i=1, 2, \dots, n$ 的联合概率密度函数为

$$f_Y(y_1, y_2, \dots, y_n) = f_X(x_1(y_1, y_2, \dots, y_n), x_2(y_1, y_2, \dots, y_n), \dots, x_n(y_1, y_2, \dots, y_n)) |J|$$

1.2 随机变量及其分布

如果(1)的方程有多个解

$$\begin{cases} x_1^{(l)} = x_1^{(l)}(y_1, y_2, \dots, y_n) \\ x_2^{(l)} = x_2^{(l)}(y_1, y_2, \dots, y_n) \\ \vdots \\ x_n^{(l)} = x_n^{(l)}(y_1, y_2, \dots, y_n) \end{cases}$$

$l=1, 2, \dots, n$ ，则 n 维随机变量 $\mathbf{Y}=(Y_1, Y_2, \dots, Y_n)$ 的联合概率密度函数为

$$f_Y(y_1, y_2, \dots, y_n) = \sum_l f_X(x_1^{(l)}(y_1, y_2, \dots, y_n), x_2^{(l)}(y_1, y_2, \dots, y_n), \dots, x_n^{(l)}(y_1, y_2, \dots, y_n)) |J|$$

1.2 随机变量及其分布

特别的，若 X 为 连续型一维随机变量，其概率密度函数为 $f_X(x)$ ，则对于 $Y=g(X)$ 的概率密度函数，有下列结果：

(1) 若 $g(x)$ 是严格单调可微函数，则 $Y=g(X)$ 的概率密度函数为

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & y \in I \\ 0, & y \notin I \end{cases}$$

其中 $h(y)$ 是 $y=g(x)$ 的反函数， I 是使 $h(y)$ 有定义、 $h'(y)$ 有定义及 $f_X(h(y))>0$ 的 y 的取值的公共部分。

1.2 随机变量及其分布

(2) 若 $g(x)$ 不是严格单调的可微函数，则将 $g(x)$ 在其定义域分成若干个单调分支，在每个单调分支上应用(1)的结果，得 $Y=g(X)$ 的概率密度函数为

$$f_Y(y) = \begin{cases} f_X(h_1(y))|h_1'(y)| + f_X(h_2(y))|h_2'(y)| + \dots, & y \in I \\ 0, & y \notin I \end{cases}$$

其中 I 是在每个单调分支上按照(1)确定的 y 的取值的公共部分.

1.2 随机变量及其分布

- **例1.2.1** 设 $X \sim U\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $Y = \tan X$, 试求 Y 的概率密度函数 $f_Y(y)$ 。

解 由于 $y = \tan x$, 故其反函数

$$h(y) = \arctan y, \quad -\infty < y < +\infty,$$

并且 $h'(y) = \frac{1}{1+y^2}, -\infty < y < +\infty$

因此, Y 的概率密度函数为

$$f_Y(y) = \frac{1}{\pi} \frac{1}{1+y^2}, -\infty < y < +\infty$$

1.2 随机变量及其分布

- 例1.2.2 设 $X \sim N(0,1)$, 求 $Y=X^2$ 的概率密度函数 $f_Y(y)$

解 由于 $y=x^2$ 有两个单调分支, 其反函数分别为

$$h_1(y) = -\sqrt{y}, y \geq 0, h_2(y) = \sqrt{y}, y \geq 0 \text{ 并且}$$

$$h_1'(y) = -\frac{1}{2\sqrt{y}}, y > 0, h_2'(y) = \frac{1}{2\sqrt{y}}, y > 0$$

因而 $Y=X^2$ 的概率密度函数为

$$f_Y(y) = \begin{cases} f_X(h_1(y))|h_1'(y)| + f_X(h_2(y))|h_2'(y)| = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

1.2 随机变量及其分布

- **例1.2.3** 设 (X, Y) 为二维随机变量，其中 X, Y 相互独立并且都服从正态分布 $N(0, \sigma^2)$ ，记 Z 为 (X, Y) 的模， Θ 为 (X, Y) 的辅角，求 (Z, Θ) 的联合概率密度函数及边缘概率密度函数。
- 从此例可得到工程上的一个重要结论：若二维随机变量的两个分量是相互独立且同服从正态分布 $N(0, \sigma^2)$ 的随机变量，则该二维随机变量的模和辐角也是相互独立的随机变量，并且模服从参数为 σ 的Rayleigh分布，辐角服从区间 $(-\pi, \pi)$ 上的均匀分布。

1.2 随机变量及其分布

解 由于

$$X \sim N(0, \sigma^2), f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, -\infty < x < +\infty$$

$$Y \sim N(0, \sigma^2), f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}, -\infty < y < +\infty$$

X, Y 相互独立，因此

$$f(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x^2+y^2)}, -\infty < x < +\infty, -\infty < y < +\infty$$

又因为方程组

$$\begin{cases} z = \sqrt{x^2 + y^2} \\ \theta = \arctan \frac{y}{x} \end{cases}$$

有唯一解(反函数)

1.2 随机变量及其分布

$$\begin{cases} x = z \cos \theta \\ y = z \sin \theta \end{cases}, z > 0, -\pi < \theta < \pi$$

$$J = \frac{\partial(x, y)}{\partial(z, \theta)} = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -z \sin \theta \\ \sin \theta & z \cos \theta \end{vmatrix} = z$$

所以 (Z, Θ) 的联合概率密度函数为

$$g(z, \theta) = f(z \cos \theta, z \sin \theta) |J| = \frac{z}{2\pi\sigma^2} e^{-\frac{z^2}{2\sigma^2}}, z > 0, -\pi < \theta < \pi$$

1.2 随机变量及其分布

故

$$g_Z(z) = \int_{-\infty}^{+\infty} g(z, \theta) d\theta = \begin{cases} \int_{-\pi}^{\pi} \frac{z}{2\pi\sigma^2} e^{-\frac{z^2}{2\sigma^2}} d\theta = \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}}, & z > 0 \\ 0, & z \leq 0 \end{cases}$$

$$g_\Theta(\theta) = \int_{-\infty}^{+\infty} g(z, \theta) dz = \begin{cases} \int_0^{+\infty} \frac{z}{2\pi\sigma^2} e^{-\frac{z^2}{2\sigma^2}} dz = \frac{1}{2\pi}, & -\pi < \theta < \pi \\ 0, & \text{其它} \end{cases}$$

从 Z, Θ 的概率密度函数可以看出， Z 服从参数为 σ 的 **Rayleigh** 分布， Θ 服从区间 $(-\pi, \pi)$ 上的均匀分布，并且 $g(z, \theta) = g_Z(z)g_\Theta(\theta)$ 所以 Z 和 Θ 是相互独立的。

1.2 随机变量及其分布（小结）

- 随机变量（向量）及分布
 - 随机变量（向量），分布函数
 - 离散型随机变量（向量），概率分布律
 - 连续性随机变量（向量），概率密度函数
- 联合分布与边缘分布
 - 联合分布函数，边缘分布函数
 - 联合分布律，边缘分布律
 - 联合概率密度函数，边缘概率密度函数
- 随机变量的独立性
- 随机变量函数的分布

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.3 随机变量的数字特征

- 随机变量的分布函数是随机变量概率分布的完整描述，但是要找到随机变量的分布函数是一件不容易的事。
- 在实际问题中描述随机变量的概率特性，不一定都要求出它的分布函数，往往需要求出描述随机变量概率特征的几个表征值就够了。
- 这就需要引入随机变量的数字特征。

1.3 随机变量的数字特征

- 定义1.3.4 设 X 是一随机变量, $F(x)$ 是其分布函数, 若

$$\int_{-\infty}^{+\infty} |x| dF(x) < +\infty$$

则称

$$EX \stackrel{def}{=} \int_{-\infty}^{+\infty} x dF(x)$$

为随机变量 X 的数学期望 (Expectation) 或均值 (Mean) .

- 若 X 是离散型随机变量, 其分布律为 $P(X=x_i)=p_i, i=1, 2, \dots$

则

$$EX = \sum_i x_i p_i$$

1.3 随机变量的数字特征

- 若 X 是连续型随机变量，其概率密度函数为 $f(x)$ ，则

$$\underline{EX = \int_{-\infty}^{+\infty} xf(x)dx}$$

- 定理1.3.1** 设 X 是一随机变量，其分布函数为 $F(x)$, $y=g(x)$ 是连续函数，如果 $\int_{-\infty}^{+\infty} g(x)dF(x)$ 存在，则

$$\underline{EY = E[g(X)] = \int_{-\infty}^{+\infty} g(x)dF(x)}$$

上述定理可推广到 n 维随机变量的场合.

1.3 随机变量的数字特征

- 定理1.3.2 设 $X=(X_1, X_2, \dots, X_n)$ 是 n 维随机变量，其联合分布函数为 $F(x_1, x_2, \dots, x_n)$, $g(x_1, x_2, \dots, x_n)$ 是连续函数，如果

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \dots, x_n) dF(x_1, x_2, \dots, x_n)$$

存在，则

$$E[g(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, x_2, \dots, x_n) dF(x_1, x_2, \dots, x_n)$$

1.3 随机变量的数字特征

- 定义1.3.5 设 X 是随机变量，若 $E/X^2 < +\infty$ ，则称

$$DX \stackrel{def}{=} E(X - EX)^2 = EX^2 - (EX)^2$$

为随机变量 X 的方差（ Variance ）。

- 定义1.3.6 设 X, Y 是随机变量，若 $E/X^2 < +\infty, E/Y^2 < +\infty$ ，

则称

$$\text{cov}(X, Y) \stackrel{def}{=} E[(X - EX)(Y - EY)] = E(XY) - EXEY$$

为随机变量 X, Y 的协方差（ Covariance ）。

1.3 随机变量的数字特征

- 若 $DX > 0, DY > 0$, 则称

$$\rho_{XY} \stackrel{def}{=} \frac{\text{cov}(X, Y)}{\sqrt{DX} \sqrt{DY}}$$

为随机变量 X, Y 的相关系数 (Correlation Coefficient) .

若 $\rho_{XY}=0$, 则称 X, Y 不相关.

1.3 随机变量的数字特征

- 根据定理1.3.1，若 X 的分布函数为 $F(x)$ 则

$$DX = e(X - EX)^2 = \int_{-\infty}^{+\infty} (x - EX)^2 dF(x)$$

- 当 X 是离散型随机变量时，其分布律为

$$P(X=x_i) = p_i, \quad i=1, 2, \dots$$

则

$$DX = \sum_i (x_i - EX)^2 p_i$$

- 当 X 是连续型随机变量时，其概率密度为 $f(x)$ ，则

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

1.3 随机变量的数字特征

- 根据定理1.3.2，若 (X, Y) 的联合分布函数为 $F(x, y)$ ，则

$$\text{cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)(y - EY) dF(x, y)$$

- 当 (X, Y) 是离散型随机变量时，其联合分布律为

$$P(X=x_i, Y=y_j) = p_{ij}, i, j = 1, 2, \dots$$

则

$$\text{cov}(X, Y) = \sum_i \sum_j (x_i - EX)(y_j - EY) p_{ij}$$

- 当 (X, Y) 是连续型随机变量时，其联合概率密度为 $f(x, y)$ ，

则

$$\text{cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)(y - EY) f(x, y) dx dy$$

1.3 随机变量的数字特征

- 数学期望和方差的性质：

(1) 设 a,b 是任意的常数，则 $E(aX+bY)=aEX+bEY;$

(2) 设 X,Y 相互独立，则 $EXY=EXEY;$

(3) 设 a,b 是任意的常数， X,Y 相互独立，则

$$\underline{D(aX+bY)=a^2DX+b^2DY}$$

(4) 设 $E/X^2<+\infty, E/Y^2<+\infty$ ，则

$$(EXY)^2 \leq EX^2 + EY^2; \quad (\text{Schwarz不等式})$$

(5) 设 $X_n \geq 0, n=1,2,\dots$ ，则

$$E(\underline{\lim_{n \rightarrow \infty}} X_n) \leq \underline{\lim_{n \rightarrow \infty}} EX_n \leq \overline{\lim_{n \rightarrow \infty}} EX_n \leq E(\overline{\lim_{n \rightarrow \infty}} X_n)$$

1.3 随机变量的数字特征

- 例1.3.1 设 X 是随机变量，若 $E|X|^r < +\infty$ ， $r > 0$ 则称 EX^r 为随机变量的 r 阶矩，设随机变量 X 的 r 阶矩存在，则

$$\forall \varepsilon > 0 \quad P(|X| \geq \varepsilon) \leq \frac{E|X|^r}{\varepsilon^r}$$

证明 设 X 的分布函数为 $F(X)$, 则

$$P(|X| \geq \varepsilon) = \int_{|x| \geq \varepsilon} dF(x) \leq \int_{|x| \geq \varepsilon} \frac{|x|^r}{\varepsilon^r} dF(x) \leq \frac{1}{\varepsilon^r} \int_{-\infty}^{+\infty} |x|^r dF(x) = \frac{E|x|^r}{\varepsilon^r}$$

即
$$P(|X| \geq \varepsilon) \leq \frac{E|X|^r}{\varepsilon^r}$$

称不等式 $P(|X| \geq \varepsilon) \leq \frac{E|X|^r}{\varepsilon^r}$ 为马尔可夫不等式

1.3 随机变量的数字特征

特别地，在马尔可夫不等式中令 $r=2$ ，将 X 换成 $X-EX$ 可得重要的 Chebyshv 不等式.

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}$$

- 定理1.3.3 设 X 是随机变量，则 $DX=0$ 的充要条件是

$$P(X=C)=1 \quad (C \text{ 是常数}) .$$

1.3 随机变量的数字特征

- 对于多个随机变量，方差和协方差之间具有下列重要的性质：设 X_1, X_2, \dots, X_n 是 n 个随机变量，则

$$D\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n DX_i + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j)$$

- 例1.3.2 (Montmort配对问题)** n 个人将自己的帽子放在一起，充分混合后每人随机地取出一顶帽子，试求出选中自己帽子的人数的均值和方差.

1.3 随机变量的数字特征

解 设 X 表示选中自己帽子的人数，令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 个人选中自己的帽子} \\ 0, & \text{否则} \end{cases}$$

$i=1,2,\dots,n$, 则

$$X = \sum_{i=1}^n X_i$$

又

$$P(X_i = 1) = \frac{1}{n}, P(X_i = 0) = \frac{n-1}{n}$$

从而

$$\underline{EX_i = \sum_j x_j p_j = 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n}}, i = 1, 2, \dots, n$$

1.3 随机变量的数字特征

所以

$$EX = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n EX_i = 1$$

由 $EX_i^2 = \sum_j x_j^2 p_j = \frac{1}{n}$, 得

$$DX_i = EX^2 - (EX_i)^2 = \frac{1}{n} - \frac{1}{n^2} = \frac{n-1}{n^2}, i = 1, 2, \dots, n$$

而当 $i \neq j$ 时

$$E(X_i, X_j) = P(X_i = 1, X_j = 1) = P(X_i = 1)P(X_j = 1 | X_i = 1) = \frac{1}{n(n-1)}$$

$$\text{cov}(X_i, X_j) = E(X_i X_j) - EX_i EX_j = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}$$

1.3 随机变量的数字特征

所以

$$\begin{aligned}DX &= D\left(\sum_{i=1}^n X_i\right) \\&= \sum_{i=1}^n DX_i + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \\&= n \cdot \frac{n-1}{n^2} + 2C_n^2 \frac{1}{n^2(n-1)} = 1\end{aligned}$$

1.3 随机变量的数字特征

- 定义1.3.7 设 $X=(X_1, X_2, \dots, X_n)$ 是 n 维随机变量，则称

$$EX = (EX_1, EX_2, \dots, EX_n) \stackrel{def}{=}$$

为 n 维随机变量 $X=(X_1, X_2, \dots, X_n)$ 的均值向量。称

$$B = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{cov}(X_n, X_n) \end{bmatrix} \stackrel{def}{=}$$

n 维随机变量 $X=(X_1, X_2, \dots, X_n)$ 的协方差矩阵.

1.3 随机变量的数字特征

● 常见的离散随机变量的期望和方差

分布	分布律	期望	方差
0-1分布	$P(X=1)=p, P(X=0)=q,$ $0 < p < 1, p+q=1$	p	pq
二项分布	$P(X=k) = C_n^k p^k q^{n-k},$ $0 < p < 1, p+q=1, k=0,1,\dots,n$	np	npq
泊松分布	$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0, k=0,1,\dots$	λ	λ
几何分布	$P(X=k) = pq^{k-1},$ $0 < p < 1, p+q=1, k=0,1,\dots$	$1/p$	q/p^2
负二项分布	$P(X=j) = C_{j-1}^{k-1} p^k q^{n-k},$ $0 < p < 1, p+q=1, j \geq k$	k/p	kq/p^2
离散均匀分布	$P\left(X=a+i\frac{b-a}{n}\right) = \frac{1}{n+1}, \quad i=0,1,\dots,n$	$(a+b)/2$	$\frac{(n+2)(b-a)^2}{12n}$

1.3 随机变量的数字特征

● 常见的连续型随机变量的期望和方差

分布	概率密度	期望	方差
均匀分布	$f(x) = \begin{cases} 1/(b-a), & a < x < b \\ 0, & \text{其它} \end{cases}$	$(a+b)/2$	$(b-a)^2/12$
正态分布	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	μ	σ^2
指数分布	$f(x) = \lambda e^{-\lambda x}, \quad \lambda > 0$	$1/\lambda$	$1/\lambda^2$
瑞利分布	$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad \sigma > 0$	$\sqrt{\pi/2}\sigma$	$(2 - \pi/2)\sigma^2$
Γ 分布	$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\beta}\right), \quad \alpha, \beta > 0$	$\beta\alpha$	$\beta^2\alpha$
χ^2 分布	$f(x) = \frac{x^{(N/2)-1}}{2^{N/2} \Gamma(N/2)} \exp\left(-\frac{x}{2}\right), \quad N > 0$	N	$2N$
β 分布	$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{其它} \end{cases}, \quad \alpha, \beta > 0$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

1.3 随机变量的数字特征（小结）

- 数学期望
- 随机变量函数的数学期望
- 方差，协方差，相关系数
- 常用随机变量的期望和方差
- 均值向量，协方差矩阵

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.4 随机变量的特征函数

- 随机变量的分布函数是其概率分布的完整描述.
- 分布函数一般来说不具有连续性、可微性等良好的分析性质, 这给利用分布函数研究随机变量带来困难.
- 引入特征函数, 它与分布函数一一对应, 既能完整地描述随机变量的概率分布, 又有良好的分析特性.

1.4 随机变量的特征函数

- 定义1.4.1 设 (Ω, \mathcal{F}, P) 是一概率空间， X, Y 都是 \mathcal{F} 的实值随机变量，则称

$$Z = X + jY, \quad j = \sqrt{-1} \quad \stackrel{\text{def}}{=}$$

为复随机变量.

- 复随机变量 Z 的数学期望定义为

$$EZ = EX + jEY \quad \stackrel{\text{def}}{=}$$

- 若X是实值随机变量，则 e^{jtX} 应是复随机变量.

1.4 随机变量的特征函数

- 定义1.4.2 设 X 是(实)随机变量, 其分布函数为 $F(x)$ 则称

$$\varphi(t) \stackrel{\text{def}}{=} E[e^{jtX}] = \int_{-\infty}^{+\infty} e^{jtX} dF(x), -\infty < t < +\infty$$

为随机变量 X 的特征函数.

- 由于 $e^{jtX} = \cos tX + j \sin tX$, X 的特征函数也可以表示为

$$\varphi(t) = E[\cos tX] + jE[\sin tX], -\infty < t < +\infty$$

- 由于

$$E[|e^{jtX}|] = \int_{-\infty}^{+\infty} |e^{jtx}| dF(x) = \int_{-\infty}^{+\infty} dF(x) = 1 < +\infty$$

随机变量 X 的特征函数 $\varphi(t)$ 总存在.

1.4 随机变量的特征函数

- 当 X 是离散型随机变量时，其分布律为

$$P(X=x_i) = p_i, i=1, 2, \dots$$

则

$$\varphi(t) = E[e^{jtx}] = \sum_i e^{jtx_i} p_i$$

- 当 X 是连续型随机变量时，其概率密度函数为 $f(x)$ ，则

$$\varphi(t) = E[e^{jtx}] = \int_{-\infty}^{+\infty} e^{jtx} f(x) dx$$

1.4 随机变量的特征函数

- **例1.4.1** 设 X 服从单点分布，即 $P(X=c)=1$ ，其中 c 为常数，则 X 的特征函数

$$\varphi(t) = E[e^{jtX}] = e^{jtc}$$

- **例1.4.2** 设 $X \sim B(n,p)$ （二项分布，**Binomial**），即

$$P(X=k) = C_n^k p^k q^{n-k} \quad k=0,1,2,\dots,n, \quad 0 < p < 1, \quad q = 1-p,$$

则 X 的特征函数

$$\varphi(t) = E[e^{jtX}] = \sum_{k=0}^n e^{jtk} C_n^k p^k q^{n-k} = \sum_{k=0}^n C_n^k (pe^{jt})^k q^{n-k} = (pe^{jt} + q)^n$$

特别地，当 $n=1$ 时， X 服从0-1分布，其特征函数为

$$\varphi(t) = pe^{jt} + q$$

1.4 随机变量的特征函数

- 例1.4.3 设 X 服从泊松分布 (Poisson) , 即

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k=0,1,2,\dots, \quad \lambda > 0, \text{ 则 } X \text{ 的特征函数}$$

$$\varphi(t) = E[e^{jtX}] = \sum_{k=0}^{\infty} e^{jtk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{jt})^k}{k!} = e^{-\lambda} \underline{e^{\lambda e^{jt}}} = e^{\lambda(e^{jt}-1)}$$

- 例1.4.4 设 X 服从区间 $[a,b]$ 上的均匀分布 (Uniform) , 即

X 的概率密度函数为

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

则 X 的特征函数

$$\varphi(t) = E[e^{jtX}] = \int_{-\infty}^{+\infty} e^{jtx} f(x) dx = \int_a^b e^{jtx} \frac{1}{b-a} dx = \frac{1}{jt(b-a)} (e^{jtb} - e^{jta})$$

1.4 随机变量的特征函数

- 例1.4.5 设 $X \sim N(\mu, \sigma^2)$ (正态分布, Normal), 即 X 的概率密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty, -\infty < \mu < +\infty, \sigma > 0$$

则 X 的特征函数

$$\begin{aligned}\varphi(t) &= E[e^{jtx}] = \int_{-\infty}^{+\infty} e^{jtx} f(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{jtx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\stackrel{\substack{u=\frac{x-\mu}{\sigma} \\ \text{---}}}{}= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{jt(\sigma u + \mu)} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} e^{j\mu t - \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{+\infty} e^{-\frac{(u-j\sigma t)^2}{2}} du = e^{j\mu t - \frac{1}{2}\sigma^2 t^2}\end{aligned}$$

$$U = \frac{X - \mu}{\sigma}, \quad U \sim N(j\sigma t, 1), \quad f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-j\sigma t)^2}{2}}, \quad \int_{-\infty}^{+\infty} f_U(u) du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(u-j\sigma t)^2}{2}} du = 1$$

1.4 随机变量的特征函数

- 特别地，若 $X \sim N(0,1)$ （标准正态分布，Standard Normal），则其特征函数

$$\varphi(t) = e^{-\frac{t^2}{2}}$$

- 例1.4.6** 设 X 服从参数为 $\lambda (\lambda > 0)$ 的指数分布（Exponential），即 X 的概率密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则 X 的特征函数

$$\varphi(t) = E[e^{jtx}] = \int_{-\infty}^{+\infty} e^{jtx} f(x) dx = \int_0^{+\infty} e^{jtx} \lambda e^{-\lambda x} dx = \int_0^{+\infty} \lambda e^{(jt-\lambda)x} dx = \frac{\lambda}{\lambda - jt} = (1 - \frac{jt}{\lambda})^{-1}$$

1.4 随机变量的特征函数

- 随机变量的特征函数 $\varphi(t)$ 具有下列 7 条性质

(1) $|\varphi(t)| \leq \varphi(0) = 1.$

(2) $\overline{\varphi(t)} = \varphi(-t)$, 其中 $\overline{\varphi(t)}$ 表示 $\varphi(t)$ 的共轭.

(3) 设随机变量 $Y=aX+b$, 其中 a, b 是常数, 则

$$\varphi_Y(t) = e^{ibt} \varphi_X(at)$$

其中 $\varphi_X(t), \varphi_Y(t)$ 分别表示随机变量 X, Y 的特征函数.

(4) $\varphi(t)$ 在 $(-\infty, +\infty)$ 上一致连续.

1.4 随机变量的特征函数

(5) 设随机变量 X, Y 相互独立，又 $Z=X+Y$ ，则

$$\varphi_Z(t) = \varphi_X(t)\varphi_Y(t)$$

此式表明两个相互独立的随机变量之和的特征函数等于各自特征函数的乘积.

(6) $\varphi(t)$ 是非负定的，即对于任意的正整数 n ，任意复数 z_1, z_2, \dots, z_n 和任意实数 t_1, t_2, \dots, t_n ，有

$$\sum_{l=1}^n \sum_{k=1}^n \varphi(t_l - t_k) z_l \overline{z_k} \geq 0$$

1.4 随机变量的特征函数

(7) 设随机变量 X 的 n 阶原点矩存在, 则 $\varphi(t)$ 存在 $k(k \leq n)$ 阶导数, 且

$$\underline{\varphi^{(k)}(0) = j^k EX^k, k \leq n}$$

- 例1.4.7 设 $X \sim \pi(\lambda)$ (Poisson分布), 求 EX, EX^2, DX .

解 由于 $X \sim \pi(\lambda)$, 因而

$$\varphi(t) = e^{\lambda(e^{jt}-1)}, \varphi'(t) = j\lambda e^{jt} e^{\lambda(e^{jt}-1)}, \varphi''(t) = -(\lambda e^{jt} + \lambda^2 e^{2jt}) e^{\lambda(e^{jt}-1)}$$

故

$$EX = \frac{\varphi'(0)}{j} = \lambda, \quad EX^2 = \frac{\varphi''(0)}{j^2} = \lambda + \lambda^2$$

$$DX = EX^2 - (EX)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

1.4 随机变量的特征函数

- 例1.4.8 设 $X \sim N(0, \sigma^2)$, 求 EX^n

解 因为

$$\varphi(t) = e^{-\frac{1}{2}\sigma^2 t^2} = \sum_{k=0}^{\infty} \frac{(-\frac{1}{2}\sigma^2 t^2)^k}{k!} = \sum_{k=0}^{\infty} \left(-\frac{\sigma^2}{2}\right)^k \frac{t^{2k}}{k!}$$

所以

$$\varphi^{(2k)}(0) = \left(-\frac{\sigma^2}{2}\right)^k \frac{(2k)!}{k!} = (-1)^k \sigma^{2k} (2k-1)!! , k = 1, 2, \dots$$

$$\varphi^{(2k-1)}(0) = 0, k = 1, 2, \dots$$

从而

$$EX^n = \begin{cases} \sigma^{2k} (2k-1)!! , & n = 2k, k = 1, 2, \dots \\ 0, & n = 2k-1, k = 1, 2, \dots \end{cases}$$

1.4 随机变量的特征函数

- 如果已知随机变量的特征函数, 怎样确定它的分布以及它所对应的分布是否唯一?
- 对连续性随机变量, 已知概率密度函数 $f(x)$, 特征函数

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{jtx} f(x) dx$$

- 在 $\varphi(t)$ 绝对可积的条件下, 根据积分理论, 有反演公式

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-jtx} \varphi(t) dt$$

且反演是唯一的.

1.4 随机变量的特征函数

- **定理1.4.1** 设随机变量 X 的分布函数为 $F(x)$, 特征函数为 $\varphi(t)$, 则对 $F(x)$ 的连续点 x_1, x_2 , 有

$$F(x_2) - F(x_1) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \frac{e^{-jtx_1} - e^{-jtx_2}}{jt} \varphi(t) dt$$

- **定理1.4.2** 随机变量 X 的分布函数 $F(x)$ 被它的特征函数 $\varphi(t)$ 惟一地确定.

由此定理可见, 随机变量的分布函数与特征函数是一一对应的.

1.4 随机变量的特征函数

- 例1.4.9 设 X_1, X_2, \dots, X_n 相互独立，且 $X_k \sim \pi(\lambda_k)$, $k=1, 2, \dots, n$

试用特征函数证明

$$\sum_{k=1}^n X_k \sim \pi\left(\sum_{k=1}^n \lambda_k\right).$$

证明 由于 X_1, X_2, \dots, X_n 相互独立， $X_k \sim \pi(\lambda_k)$, $k=1, 2, \dots, n$

故

$$\varphi_{X_k}(t) = e^{\lambda_k(e^{jt}-1)}, k=1, 2, \dots, n$$

从而

$$\varphi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \varphi_{X_k}(t) = \exp\left[\sum_{k=1}^n \lambda_k(e^{jt}-1)\right]$$

所以

$$\sum_{k=1}^n X_k \sim \pi\left(\sum_{k=1}^n \lambda_k\right)$$

独立泊松分布的和还是泊松分布，且参数为各参数的和.

1.4 随机变量的特征函数

- 例1.4.10 设 X_1, X_2, \dots, X_n 相互独立，且 $X_k \sim N(\mu_k, \sigma_k^2)$, $k=1, 2, \dots, n$,
试用特征函数求随机变量 $\sum_{k=1}^n X_k$ 的概率分布.

解 由于 X_1, X_2, \dots, X_n 相互独立, 且 $X_k \sim N(\mu_k, \sigma_k^2)$, $k=1, 2, \dots, n$

故

$$\varphi_{X_k}(t) = e^{j\mu_k t - \frac{1}{2}\sigma_k^2 t^2}, k=1, 2, \dots, n$$

从而

$$\varphi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \varphi_{X_k}(t) = \prod_{k=1}^n [\exp(j\mu_k t - \frac{1}{2}\sigma_k^2 t^2)] = \exp[j(\sum_{k=1}^n \mu_k)t - \frac{1}{2}(\sum_{k=1}^n \sigma_k^2)t^2]$$

所以

$$\sum_{k=1}^n X_k \sim N(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2)$$

独立正态分布的和还是正态分布, 且参数为对应各参数的和.

1.4 随机变量的特征函数

- 定义1.4.3 设 $X = (X_1, X_2, \dots, X_n)$ 是 n 维随机变量，其联合分布函数为 $F(x) = F(x_1, x_2, \dots, x_n)$ ，则称

$$\begin{aligned}\varphi(t) &= \varphi(t_1, t_2, \dots, t_n) \stackrel{\text{def}}{=} E[e^{jtX^T}] = E[\exp(j \sum_{k=1}^n t_k X_k)] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[j(t_1 x_2 + t_2 x_1 + \cdots + t_n x_n)] dF(x_1, x_2, \dots, x_n) \\ t &= (t_1, t_2, t_n) \in R^n\end{aligned}$$

为 n 维随机变量 X 的特征函数.

1.4 随机变量的特征函数

- 若 $X = (X_1, X_2, \dots, X_n)$ 是 离散型随机变量, 其联合分布律为

$$P(X_1=x_1, X_2=x_2, \dots, X_n=x_n), \text{ 则}$$

$$\varphi(t) = \varphi(t_1, t_2, \dots, t_n)$$

$$= \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \exp[j(t_1 x_1 + t_2 x_2 + \cdots + t_n x_n)] P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

其中 \sum_{x_i} 是关于 X_i 的可能取值 x_i 求和.

- 若 $X = (X_1, X_2, \dots, X_n)$ 是 连续型随机变量, 其联合概率密度函

$$\text{数为 } f(x) = f(x_1, x_2, \dots, x_n), \text{ 则}$$

$$\varphi(t) = \varphi(t_1, t_2, \dots, t_n)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp[j(t_1 x_1 + t_2 x_2 + \cdots + t_n x_n)] f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

1.4 随机变量的特征函数

- n 维随机变量的特征函数具有下列性质：

$$(1) |\varphi(t_1, t_2, \dots, t_n)| \leq \varphi(0, 0, \dots, 0) = 1;$$

$$(2) \overline{\varphi(t_1, t_2, \dots, t_n)} = \varphi(-t_1, -t_2, \dots, -t_n);$$

(3) 设 $\varphi(t_1, t_2, \dots, t_n)$ 是 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的特征函数，
则随机变量 $Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ 的特征函数为

$$\underline{\varphi_Y(t) = \varphi(a_1 t, a_2 t, \dots, a_n t)}$$

$$(4) \varphi(t_1, t_2, \dots, t_n) \text{ 在 } R^n \text{ 上一致连续;}$$

1.4 随机变量的特征函数

(5) 设 $\varphi(t_1, t_2, \dots, t_n)$ 是 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的特征函数 $\varphi_{X_i}(t), i = 1, 2, \dots, n$ 是随机变量 X_1, X_2, \dots, X_n 相互独立的充要条件是

$$\varphi(t_1, t_2, \dots, t_n) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2)\cdots\varphi_{X_n}(t_n)$$

(6) 设 $\varphi(t_1, t_2, \dots, t_n)$ 是 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的特征函数，则 $k(1 \leq k < n)$ 维随机变量 (X_1, X_2, \dots, X_k) 的特征函数为

$$\varphi_{X_1, X_2, \dots, X_k}(t_1, t_2, \dots, t_k) = \varphi(t_1, t_2, \dots, t_k, 0, \dots, 0)$$

1.4 随机变量的特征函数

(7) 设 $\varphi(t_1, t_2, \dots, t_n)$ 是 n 维随机变量 $X = (X_1, X_2, \dots, X_n)$ 的特征函数, 如果 $E[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}]$ 存在, 则

$$E[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}] = j^{-\sum_{i=1}^n k_i} \left[\frac{\partial^{k_1+k_2+\cdots+k_n} \varphi(t_1, t_2, \dots, t_n)}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}} \right]_{t_1=t_2=\cdots=t_n=0}$$

- 特征函数要求随机变量 X 的取值范围为 $(-\infty, +\infty)$, 对只取非负值的随机变量, 用 *Laplace* 变换更为方便.

1.4 随机变量的特征函数（小结）

- 复随机变量
- 特征函数的定义、计算方法
 - 使用概率分布律（概率密度函数）求离散型（连续型）随机变量特征函数
- 常用随机变量的特征函数
 - 离散型：单点分布，二项分布，泊松分布
 - 连续型：均匀分布，正态分布，指数分布
- 特征函数的性质
 - 相互独立的随机变量之和的特征函数等于各自特征函数的乘积
 - 特征函数的 k 阶导数与 k 阶（原点）矩的关系
- 利用特征函数求各阶矩（期望，方差）
- 特征函数与分布函数的一一对应性
- n 维随机变量（向量）的特征函数

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.5 n 维正态随机变量

- 在概率论中，若 $(X_1, X_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，则 二维正态随机变量 (X_1, X_2) 的联合概率密度函数为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho\frac{x_1 - \mu_1}{\sigma_1}\frac{x_2 - \mu_2}{\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}\right]\right\}$$

其中， $\mu_1 = EX_1$, $\mu_2 = EX_2$, $\sigma_1^2 = DX_1$, $\sigma_2^2 = DX_2$, ρ 为随机变量 X_1, X_2 的相关系数。

1.5 n 维正态随机变量

- 复习一维情况:
- 正态分布（标准正态分布）的随机变量
- 正态分布的概率密度函数
- 正态分布的概率分布函数
- 正态分布的特征函数

1.5 n 维正态随机变量

- 下面用向量和矩阵的形式来表示二维正态分布的联合概率密度函数. 令

$$\mathbf{x} = (x_1, x_2), \quad \boldsymbol{\mu} = (\mu_1, \mu_2), \quad B = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

于是

$$|B| = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

$$B^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

1.5 n 维正态随机变量

所以

$$\frac{1}{1-\rho^2} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2 - \mu_2}{\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \}$$

$$= (\mathbf{x} - \boldsymbol{\mu}) \mathbf{B}^{-1} (\mathbf{x} - \boldsymbol{\mu})^T$$

于是

$$f(\mathbf{x}) = f(x_1, x_2) = \frac{1}{2\pi|B|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}) \mathbf{B}^{-1} (\mathbf{x} - \boldsymbol{\mu})^T\right]$$

1.5 n 维正态随机变量

- 定义 1.5.1 设 $X = (X_1, X_2, \dots, X_n)$ 是 n 维随机变量，如果其联合概率密度函数为

$$f(x) = f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |B|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)B^{-1}(x - \mu)^T\right]$$

其中 $x = (x_1, x_2, \dots, x_n)$, $\mu = (\mu_1, \mu_2, \dots, \mu_n) \stackrel{\text{def}}{=} (EX_1, EX_2, \dots, EX_n)$

$$B \stackrel{\text{def}}{=} \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{cov}(X_n, X_n) \end{bmatrix}$$

则称 $X = (X_1, X_2, \dots, X_n)$ 服从 μ 为均值向量、 B 为协方差矩阵的 n 维正态分布，记为 $X \sim N(\mu, B)$.

1.5 n 维正态随机变量

- 定理1.5.1 设 $X \sim N(\mu, B)$, 则存在 n 阶正交矩阵 A , 使得

$$Y = (Y_1, Y_2, \dots, Y_n) = (X - \mu) A^T$$

是 n 维独立正态随机变量, 即 Y_1, Y_2, \dots, Y_n 相互独立, 且 $Y_k \sim N(0, d_k)$, 其中 $d_k > 0$ 是 B 的特征值, $k = 1, 2, \dots, n$.

- 定理1.5.2 设 $X \sim N(\mu, B)$, 则 X 的特征函数

$$\varphi(t) = \varphi(t_1, t_2, \dots, t_n) = \exp\left(j\mu t^T - \frac{1}{2}t B t^T\right)$$

1.5 n 维正态随机变量

- 定理1.5.3 (正态随机变量的性质)

设 $X = (X_1, X_2, \dots, X_n) \sim N(\boldsymbol{\mu}, \mathbf{B})$

(1) 若 l_1, l_2, \dots, l_n 是常数, 则 $Y = \sum_{k=1}^n l_k X_k$ 服从一维正态分布

$$N\left(\sum_{k=1}^n l_k \mu_k, \sum_{i=1}^n \sum_{k=1}^n l_i l_k \text{cov}(X_i, X_k)\right)$$

其中 $\mu_k = EX_k$, $k = 1, 2, \dots, n$.

(一维正态随机变量的线性组合仍然是正态随机变量)

1.5 n 维正态随机变量

(2) 若 $m < n$, 则 X 的 m 个分量构成的 m 维随机变量

$\tilde{X} = (X_1, X_2, \dots, X_n)$ 服从 m 维正态分布 $N(\tilde{\mu}, \tilde{B})$,

其中

$$\tilde{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$$

$$\tilde{B} = \begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_m) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_m) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_m, X_1) & \text{cov}(X_m, X_2) & \cdots & \text{cov}(X_m, X_m) \end{bmatrix}$$

(n 维正态随机变量的 m 维分量仍然是正态随机变量)

1.5 n 维正态随机变量

(3) 若 m 维随机变量 \mathbf{Y} 是 \mathbf{X} 的线性变换, 即 $\mathbf{Y} = \mathbf{X}\mathbf{C}$, 其中 \mathbf{C} 是 $n \times m$ 阶矩阵, 则 \mathbf{Y} 服从 m 维正态分布 $N(\mu\mathbf{C}, \mathbf{C}^T \mathbf{B} \mathbf{C})$.

(n 维正态随机变量的线性变换仍然是正态随机变量)

(4) X_1, X_2, \dots, X_n 相互独立的充要条件是 X_1, X_2, \dots, X_n 两两不相关.

(正态随机变量的相互独立性等价于两两独立性)

1.5 n 维正态随机变量（小结）

- n 维正态随机变量的定义
 - 联合概率密度函数的向量（均值向量）、矩阵（协方差矩阵）表示法
 - n 维正态随机变量由均值向量 μ 和协方差矩阵 B 唯一确定
- n 维正态随机变量的特征函数
- n 维正态随机变量的性质
 - 正态随机变量的线性变换仍然是正态随机变量
 - 正态随机变量的分量仍然是正态随机变量
 - 正态随机变量的相互独立性等价于两两独立性

第1章 概率论基础

- 1.1 概率空间
- 1.2 随机变量及分布
- 1.3 随机变量的数字特征
- 1.4 随机变量的特征函数
- 1.5 n 维正态随机变量
- 1.6 条件数学期望

1.6 条件数学期望

- 对于多个事件我们讨论了它们的条件概率，对于多个随机变量我们可以讨论它们的条件分布.
- 定义1.6.1 设 (X, Y) 是离散型二维随机变量，其联合分布律为 $P(X=x_i, Y=y_i)=p_{ij}$, $i, j=1, 2, \dots$, 如果 $P(Y = y_j) = p_{\cdot j} > 0$, 则称

$$\underline{P}_{i|j} \stackrel{\text{def}}{=} P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{\cdot j}}, i = 1, 2, \dots$$

为 (X, Y) 关于 X 在 $Y=y_j$ 的条件下的条件分布律(Conditional PMF).

1.6 条件数学期望

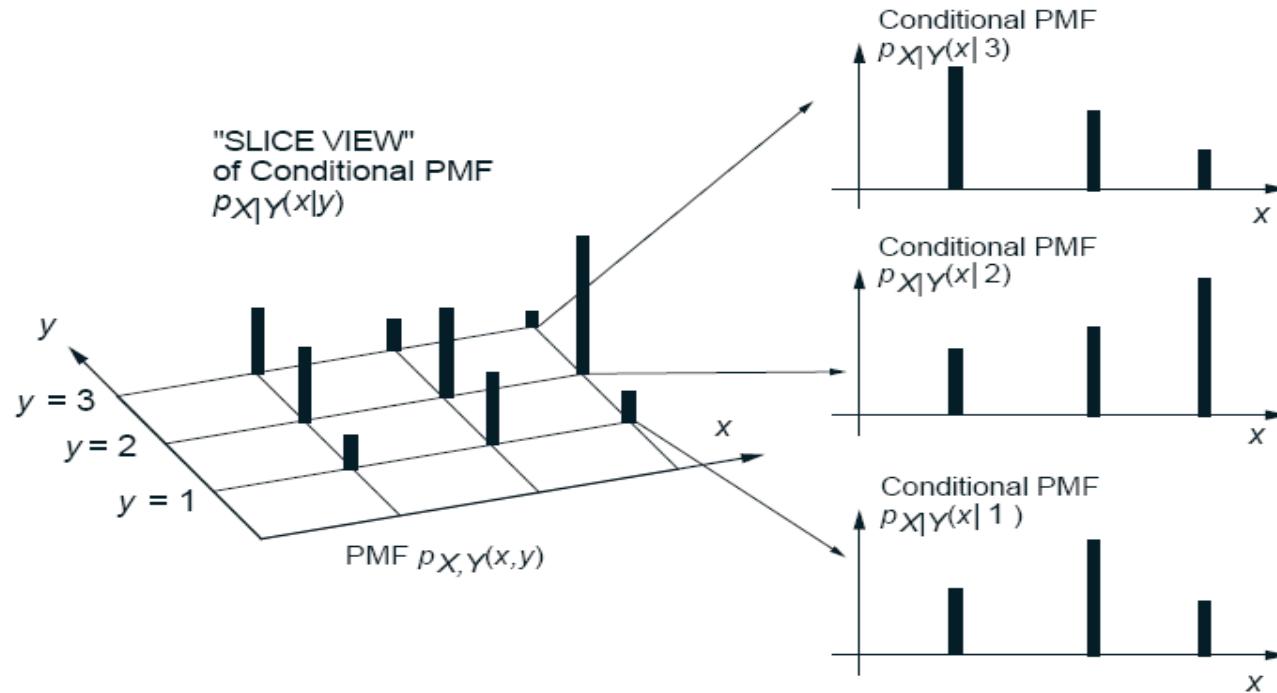


Figure 2.13: Visualization of the conditional PMF $p_{X|Y}(x|y)$. For each y , we view the joint PMF along the slice $Y = y$ and renormalize so that

$$\sum_x p_{X|Y}(x|y) = 1.$$

1.6 条件数学期望

- 如果 $P(X = x_i) \stackrel{\text{def}}{=} p_{i.} > 0$, 则称

$$p_{j|i} \stackrel{\text{def}}{=} P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_{i.}}, j = 1, 2, \dots$$

为 (X, Y) 关于 Y 在 $X=x_i$ 的条件下的 条件分布律.

- 称 $\underline{F_{X|Y}(x | y) \stackrel{\text{def}}{=} P(X \leq x | Y = y_j)} = \sum_{x_i \leq x} \frac{p_{ij}}{p_{.j}}$, $-\infty < x < +\infty$

为 (X, Y) 关于 X 在 $Y=y_j$ 的条件下的 条件分布函数

(Conditional CDF).

- 称 $F_{Y|X}(y | x) \stackrel{\text{def}}{=} P(Y \leq y | X = x_j) = \sum_{y_i \leq y} \frac{p_{ij}}{p_{i.}}$, $-\infty < y < +\infty$

为 (X, Y) 关于 Y 在 $X=x_i$ 的条件下的 条件分布函数.

1.6 条件数学期望

- 对于连续型二维随机变量，由于对于任意的 x, y , $P(X=x)=0, P(Y=y)=0$, 因此就不能直接用条件概率公式引入条件分布函数了. 我们用极限的方法来处理.
- 给定 y , 设对于任意固定的正数 ε , $P(y-\varepsilon < Y \leq y+\varepsilon) > 0$ 且若对于任意的 x , 有

$$P(X \leq x | y - \varepsilon < Y \leq y + \varepsilon) = \frac{P(X \leq x, y - \varepsilon < Y \leq y + \varepsilon)}{P(y - \varepsilon < Y \leq y + \varepsilon)}$$

上式给出了在条件 $y - \varepsilon < Y \leq y + \varepsilon$ 下 X 的条件分布函数.

1.6 条件数学期望

- 定义1.6.2 给定 y , 设对于任意固定的正数 ε ,

$P(y-\varepsilon < Y \leq y+\varepsilon) > 0$, 且若对于任意实数 x , 极限

$$\lim_{\varepsilon \rightarrow 0^+} P(X \leq x | y - \varepsilon < Y \leq y + \varepsilon) = \lim_{\varepsilon \rightarrow 0^+} \frac{P(X \leq x, y - \varepsilon < Y \leq y + \varepsilon)}{P(y - \varepsilon < Y \leq y + \varepsilon)}$$

存在, 则称此极限为 (X, Y) 关于 X 在条件 $Y=y$ 下的条件分布函数, 记为 $P(X \leq x | Y=y)$ 或 $F_{X/Y}(x/y)$.

- (X, Y) 关于 X 在条件 $Y=y$ 下的条件概率密度函数

(Conditional PDF) 为

$$f_{X|Y}(x | y) \stackrel{\text{def}}{=} \frac{f(x, y)}{f_Y(y)}$$

1.6 条件数学期望

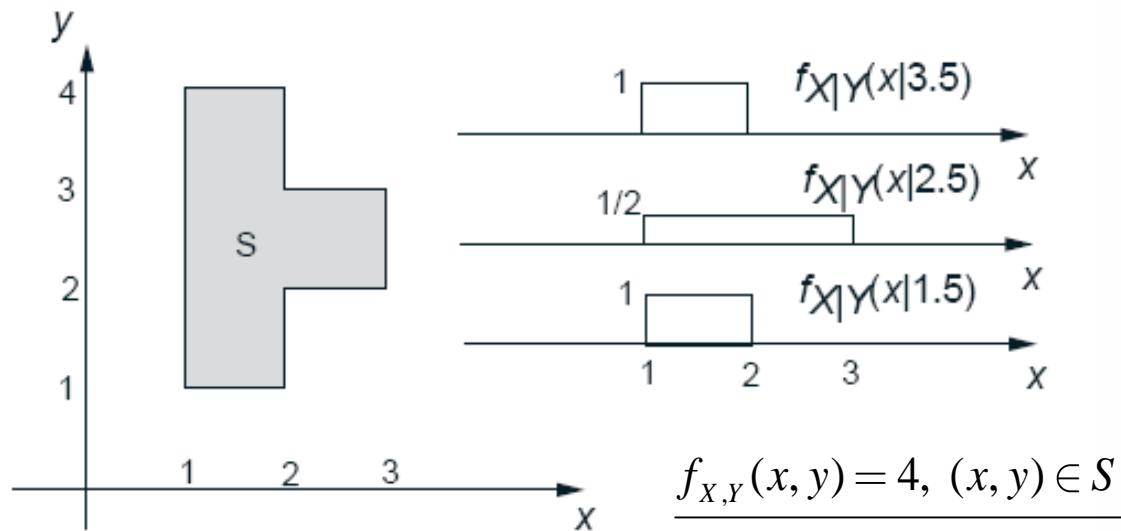


Figure 3.18: Visualization of the conditional PDF $f_{X|Y}(x|y)$. Let X, Y have a joint PDF which is uniform on the set S . For each fixed y , we consider the joint PDF along the slice $Y = y$ and normalize it so that it integrates to 1.

1.6 条件数学期望

- 类似地， (X, Y) 关于 Y 在条件 $X=x$ 下的条件分布函数为

$$\underline{P(Y \leq y | X=x)} \quad \text{或} \quad \underline{F_{X/Y}(y/x)}.$$

- (X, Y) 关于 X 在条件 $Y=y$ 下的条件概率密度函数为

$$\underline{f_{Y|X}(y | x) \stackrel{\text{def}}{=} \frac{f(x, y)}{f_X(x)}}$$

- 条件分布的概念完全可推广到 n 维随机变量的情形

1.6 条件数学期望

- 定义1.6.3 设 (X, Y) 是二维随机变量， $F_{X/Y}(x/y)$ ， $F_{Y/X}(y/x)$ 分别是 X 和 Y 的条件分布函数，则称

$$E(X | y) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x dF_{X|Y}(x | y)$$

为 X 在条件 $Y=y$ 下的条件数学期望 (Conditional Expectation). 称

$$E(Y | x) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} y dF_{Y|X}(y | x)$$

为 Y 在条件 $X=x$ 下的条件数学期望.

1.6 条件数学期望

- 若 X, Y 是 离散型随机变量, 其可能取值分别是 x_1, x_2, \dots 和 y_1, y_2, \dots , 则

$$\underline{E(X | y) = \sum_i x_i p_{i|j} = \sum_i x_i P(X = x_i | Y = y_j)}$$

$$E(Y | x) = \sum_j y_j p_{j|i} = \sum_j y_j P(Y = y_j | X = x_i)$$

- 若 X, Y 是 连续型随机变量, 则

$$\underline{E(X | y) = \int_{-\infty}^{+\infty} xf_{X|Y}(x | y)dx}$$

$$E(Y | x) = \int_{-\infty}^{+\infty} yf_{Y|X}(y | x)dy$$

1.6 条件数学期望

- 由于 $E(X|y)$ 是随机变量 Y 可能取值 y 的函数，因此 $E(X|Y)$ 是随机变量 Y 的函数，称为 X 在条件 Y 下的条件数学期望。
- 类似地，称 随机变量 X 的函数 $E(Y|X)$ 为 Y 在条件 X 下的条件数学期望。

1.6 条件数学期望

- 定义1.6.4 设 $X=(X_1, X_2, \dots, X_n)$ 是 n 维随机变量，

$F_{X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 为 X_i 的条件分布函数，则称

$$\begin{aligned} & E(X_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \\ &= \int_{-\infty}^{+\infty} x dF_{X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \end{aligned}$$

为 X_i 在条件 $\underline{X_1=x_1, \dots, X_{i-1}=x_{i-1}, X_{i+1}=x_{i+1}, \dots, X_n=x_n}$ 下的条件数学期望.

称 $E(X_i/X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ 为 X_i 在条件 $\underline{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}$ 下的条件数学期望.

1.6 条件数学期望

- 定理1.6.1 $E(E(X_i/X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)) = EX_i$ (全期望公式)

证明 仅就二维随机变量的情形加以证明. 即设 (X, Y) 是二维随机变量, 则 $E(E(X/Y)) = EX$.

(1) 若 (X, Y) 是离散型随机变量, 则

$$\begin{aligned} E(E(X | Y)) &= \sum_j E(X | Y = y_j) P(Y = y_j) \\ &= \sum_j \left(\sum_i x_i P(X = x_i | Y = y_j) \right) P(Y = y_j) \\ &= \sum_j \sum_i x_i \underbrace{\left(P(X = x_i | Y = y_j) P(Y = y_j) \right)}_{\text{——}} \\ &= \sum_i \sum_j x_i p_{ij} \quad \text{——} \\ &= EX \end{aligned}$$

1.6 条件数学期望

- (2) 若 (X, Y) 是 连续型随机变量, 则

$$\begin{aligned} E(E(X | Y)) &= \int_{-\infty}^{+\infty} E(X | y) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x f_{X|Y}(x | y) dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_Y(y) \underline{f_{X|Y}(x | y)} dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \underline{f(x, y)} dx dy \\ &= EX \end{aligned}$$

$$EX = \sum_j E(X | Y = y_j) P(Y = y_j) \quad EX = \int_{-\infty}^{+\infty} E(X | y) f_Y(y) dy$$

1.6 条件数学期望

- 例1.6.1 一矿工被困在矿井中，要到达安全地带，有三个通道可供选择。他从第一个通道出去要走3个小时可到达安全地带，从第二个通道出去要走5个小时又返回原处，从第三个通道出去要走7个小时也返回原处。设在任一时刻都等可能地选中其中一个通道，试问他到达安全地带平均要花多长时间。

解 设 X 表示矿工到达安全地带所需时间， Y 表示他选定的通道，则

$$\begin{aligned} EX &= E(E(X|Y)) \\ &= E(X|Y=1)P(Y=1)+E(X|Y=2)P(Y=2)+E(X|Y=3)P(Y=3) \\ &= \frac{1}{3}\times 3 + \frac{1}{3}(5+EX) + \frac{1}{3}\times(7+EX) \end{aligned}$$

所以， $EX = 15$.

1.6 条件数学期望

- 例1.6.2 设某日进入某商店的顾客人数是随机变量 N , X_i 表示第*i*个顾客所花的钱数, X_1, X_2, \dots 是相互独立同分布的随机变量, 且与 N 相互独立, 是求该日商店一天营业额的均值.

解

$$\begin{aligned} E\left(\sum_{i=1}^N X_i\right) &= E(E(\sum_{i=1}^N X_i) | N) = \sum_{n=1}^{\infty} E(\sum_{i=1}^N X_i | N = n) P(N = n) \\ &= \sum_{n=1}^{\infty} E(\sum_{i=1}^n X_i | N = n) P(N = n) = \sum_{n=1}^{\infty} E(\sum_{i=1}^n X_i) P(N = n) \\ &= \sum_{n=1}^{\infty} n EX_1 P(N = n) = (EX_1) \sum_{n=1}^{\infty} n P(N = n) \\ &= EX_1 EN \end{aligned}$$

1.6 条件数学期望

- **定理1.6.2** 设 X_1, X_2, \dots, X_n 相互独立，则

$$E(X_i/X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) = EX_i$$

- **定理1.6.3** 设 $X=(X_1, X_2, \dots, X_n)$ 是 n 维随机变量，

$g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 是连续函数，则

$$\begin{aligned} & E(X_i g(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) / X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \\ &= g(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) E(X_i / X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \end{aligned}$$

- **定理1.6.4** 设 $X=(X_1, X_2, \dots, X_n)$ 是 n 维随机变量， $k < n-1$ 则

$$E(E(X_n / X_1, \dots, X_{n-1}) / X_1, \dots, X_k) = E(X_n / X_1, \dots, X_k)$$

1.6 条件数学期望 (小结)

- （离散随机变量）条件分布函数、条件分布律

$$\underline{F_{X|Y}(x|y)} \stackrel{\text{def}}{=} P(X \leq x | Y = y_j) = \sum_{x_i \leq x} \frac{p_{ij}}{P_{\cdot j}}$$
$$\underline{P_{i|j}} \stackrel{\text{def}}{=} P(X = x_i | Y = y_j) = \frac{p_{ij}}{P_{\cdot j}}$$

- （连续随机变量）条件分布函数、条件概率密度函数

$$\underline{F_{X|Y}(x|y)} \stackrel{\text{def}}{=} P(X \leq x | Y = y) = \lim_{\varepsilon \rightarrow 0^+} P(X \leq x | y - \varepsilon < Y \leq y + \varepsilon)$$

$$= \lim_{\varepsilon \rightarrow 0^+} \frac{P(X \leq x, y - \varepsilon < Y \leq y + \varepsilon)}{P(y - \varepsilon < Y \leq y + \varepsilon)}$$

$$\underline{f_{Y|X}(y|x)} \stackrel{\text{def}}{=} \frac{f(x,y)}{f_X(x)}$$

1.6 条件数学期望 (小结)

- 条件数学期望

$$E(X | y) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} x dF_{X|Y}(x | y) \quad E(X | Y)$$

- 全期望公式

$$E(E(X_i | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)) = EX_i$$

$$E(E(X | Y)) = EX$$

$$EX = \sum_j E(X | Y = y_j) P(Y = y_j) \quad EX = \int_{-\infty}^{+\infty} E(X | y) f_Y(y) dy$$

The End

