第一章　预备知识与随机过程的基本概念

1.1 概　率

概率论的一个基本概念是随机试验. 一个试验 (或观察), 若它的结果预先无法确定, 则称之为**随机试验**, 简称为**试验**(experiment). 所有试验的可能结果组成的集合, 称为**样本空间**, 记作 Ω. Ω 中的元素则称为**样本点**, 用 ω 表示. 由 Ω 的某些样本点构成的子集合, 常用大写字母 A, B, C 等表示, 由 Ω 中的若干子集构成的集合称为集类, 用花写字母 $\mathcal{A}, \mathcal{B}, \mathcal{F}$ 等表示.

由于并不是在所有的 Ω 的子集上都能方便地定义概率, 一般只限制在满足一定条件的集类上研究概率性质, 为此引入 σ 域的概念:

定义 1.1.1　设 $\mathcal{F}$ 为由 Ω 的某些子集构成的非空集类, 若满足:

(1) 若 $A \in \mathcal{F}$, 则 $A^C \in \mathcal{F}$, A^C 是 A 的补集, 即 $A^C = \bar{A} = \Omega - A$;

(2) 若 $A_n \in \mathcal{F}, n \in \mathbb{N}$, 则 $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$.

则称 $\mathcal{F}$ 为σ **域**(σ 代数), 称 $(\Omega, \mathcal{F})$ 为**可测空间**.

容易验证, 若 $\mathcal{F}$ 为 σ 域, 则 $\mathcal{F}$ 对可列次交、并、差等运算封闭, 即 $\mathcal{F}$ 中的任何元素经可列次运算后仍属于 $\mathcal{F}$. 例: 集类 $\mathcal{F}_0 = \{\varnothing, \Omega\}$, $\mathcal{F}_1 = \{\varnothing, A, A^C, \Omega\}$ 及 $\mathcal{F}_2 = \{A\text{: } \forall A \subset \Omega\}$ 均是 σ 域, 但集类 $\mathcal{A} = \{\varnothing, A, \Omega\}$ 不是 σ 域.

通常最关心的是包含所要研究对象的最小 σ 域. 设 $\mathcal{A}$ 为由 Ω 的某些子集构成的集类. 一切包含 $\mathcal{A}$ 的 σ 域的交, 记为 $\sigma(\mathcal{A})$, 称 $\sigma(\mathcal{A})$ 为由 $\mathcal{A}$ 生成的 σ 域, 或称为包含 $\mathcal{A}$ 的最小 σ 域. 例:$\mathcal{A} = \{\varnothing, A, \Omega\}$, 则 $\sigma(\mathcal{A}) = \{\varnothing, A, A^C, \Omega\}$. 一维博雷尔(Borel)$\sigma$ 域: 包含 $\mathbb{R}$ 上所

有形如集合 $(-\infty, a]$ 的最小 σ 域称为一维博雷尔 σ 域, 记为 $\mathcal{B}$, 即 $\mathcal{B} = \sigma((-\infty, a], \forall a \in \mathbb{R})$.

定义 1.1.2 设 $(\Omega, \mathcal{F})$ 为可测空间, P 是一个定义在 $\mathcal{F}$ 上的集函数, 若满足:

(1) $P(A) \geqslant 0, \forall A \in \mathcal{F}$; (非负性)

(2) $P(\Omega) = 1$; (规一性)

(3) 若 $A_i \in \mathcal{F}, i = 1, 2, \cdots$, 且 $A_i A_j = \varnothing, \forall\, i \neq j$, 有

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i). \qquad \text{(可列可加性)}$$

则称 P 为可测空间 $(\Omega, \mathcal{F})$ 上的一个**概率测度**(probability measure), 简称**概率**(probability). 称 $(\Omega, \mathcal{F}, P)$ 为**概率空间**(probability space), 称 $\mathcal{F}$ 为事件域. 若 $A \in \mathcal{F}$, 则称 A 为**随机事件**(random event), 简称为**事件**, 称 $P(A)$ 为事件 A 的概率.

事件的概率刻画了事件出现可能性的大小. 概率的基本性质如下:

(1) $P(\varnothing) = 0, P(A^C) = 1 - P(A)$.

(2) 若 $A_1, A_2, \cdots, A_n$ 互不相容, 则

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i). \qquad \text{(有限可加性)}$$

(3) 对任意两个事件 A 及 B, 有

$$P(A \bigcup B) = P(A) + P(B) - P(AB),$$
$$P(A - B) = P(A) - P(AB).$$

(4) 若 $A \subset B$, 则 $P(A) \leqslant P(B)$.

(5) (若尔当 (Jordan) 公式) 对任意 $A_1, A_2, \cdots, A_n$ 有

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_iA_j) + \sum_{1\leqslant i<j<k\leqslant n} P(A_iA_jA_k) - \cdots + (-1)^{n+1}P(A_1A_2\cdots A_n).$$

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) \leqslant \sum_{i=1}^{n} P(A_i).$$

例 1　(1) $[0,1]$ 上的博雷尔概率空间. 设 $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, 即 $\mathcal{B}[0,1]$ 是 $\mathcal{B}$ 局限在 $[0,1]$ 上的博雷尔 σ- 域. 称 $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}[0,1])$ 为 $[0,1]$ 上的博雷尔可测空间. 设在可测空间 $([0,1], \mathcal{B}[0,1])$ 上定义一概率测度 P, 它满足：当 $\forall A = [a,b] \in \mathcal{B}[0,1]$ 时,$P(A) = b-a$, 称 $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}[0,1], P)$ 为 $[0,1]$ 上的博雷尔概率空间, 称 P 为 $[0,1]$ 上的博雷尔概率测度.

(2) 令 $B = [0,1]$ 上有理点全体, $\bar{B} = [0,1]$ 上无理点全体.

① 试证:$B \in \mathcal{F}, \bar{B} \in \mathcal{F}$;

② 用概率的定义与性质, 求证$P(B) = 0, \quad P(\bar{B}) = 1$.

证明　$\forall a \in [0,1]$, 单点集 $\{a\} = \bigcap\limits_{n=1}^{\infty} \left[a, a+\dfrac{1}{n}\right] \in \mathcal{B}[0,1] = \mathcal{F}$, 而 $B = 0 \bigcup \left\{\dfrac{m}{n}: 1 \leqslant m \leqslant n, n, m = \{1, 2, \cdots\}\right\}$ 是可列单点集的并, 故 $B \in \mathcal{F}$. 且 $\bar{B} = [0,1] - B \in \mathcal{F}$. 又 $\forall a \in [0,1], P(\{a\}) = 0$, 由完全可加性知 $P(B) = 0$, 而 $\bar{B} = [0,1] - B$, 故 $P(\bar{B}) = 1 - 0 = 1$.　□

概率的一个重要性质是它具有连续性. 为此先引入事件列的极限.

一事件列 $\{A_n, n \geqslant 1\}$ 称为单调**增序列**, 若 $A_n \subset A_{n+1}, n \geqslant 1$; 称为单调**减序列**, 若 $A_n \supset A_{n+1}, n \geqslant 1$. 如果 $\{A_n, n \geqslant 1\}$ 是单调增序列, 定义 $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{i=1}^{\infty} A_i$. 如果 $\{A_n, n \geqslant 1\}$ 是单调减序列, 定义 $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{i=1}^{\infty} A_i$.

连续性定理如下.

命题 1.1.1 若 $\{A_n, n \geqslant 1\}$ 是单调增序列 (或减序列), 则

$$\lim_{n\to\infty} P(A_n) = P\big(\lim_{n\to\infty} A_n\big).$$

证明 先设 $\{A_n, n \geqslant 1\}$ 为单调增序列, 令

$$B_1 = A_1, \quad B_n = A_n\Big(\bigcup_{i=1}^{n-1} A_i\Big)^C = A_n A_{n-1}^C, \quad n > 1.$$

容易验证 $\{B_n, n \geqslant 1\}$ 互不相容. 且有 $\bigcup\limits_{i=1}^{n} A_i = \bigcup\limits_{i=1}^{n} B_i, n \geqslant 1$ 及 $\bigcup\limits_{i=1}^{\infty} A_i = \bigcup\limits_{i=1}^{\infty} B_i$, 故

$$\begin{aligned}
P\big(\lim_{n\to\infty} A_n\big) &= P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = P\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \sum_{i=1}^{\infty} P(B_i) \qquad &&\text{(可列可加性)}\\
&= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} B_i\Big)\\
&= \lim_{n\to\infty} P\Big(\bigcup_{i=1}^{n} A_i\Big) = \lim_{n\to\infty} P(A_n). &&\Big(A_n = \bigcup_{i=1}^{n} A_i\Big)
\end{aligned}$$

若 $\{A_n, n \geqslant 1\}$ 为单调减序列, 则 $\{A_n^C, n \geqslant 1\}$ 为单调增序列, 于是

$$P\Big(\bigcup_{n=1}^{\infty} A_n^C\Big) = \lim_{n\to\infty} P(A_n^C),$$

由

$$\bigcup_{n=1}^{\infty} A_n^C = \Big(\bigcap_{n=1}^{\infty} A_n\Big)^C,$$

有

$$1-P\Big(\bigcap_{n=1}^{\infty}A_n\Big)=\lim_{n\to\infty}(1-P(A_n)),$$

即

$$P\Big(\bigcap_{n=1}^{\infty}A_n\Big)=\lim_{n\to\infty}P(A_n). \qquad \square$$

下面是著名的 Borel-Cantelli 引理.

命题 1.1.2　设 $\{A_n, n\geqslant 1\}$ 是一事件序列, 若 $\sum\limits_{i=1}^{\infty}P(A_i)<\infty$, 则

$$P\Big(\limsup_{i\to\infty}A_i\Big)=0,$$

其中 $\limsup\limits_{i\to\infty}A_i\triangleq\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{i=n}^{\infty}A_i$.

证明　易知 $\bigcup\limits_{i=n}^{\infty}A_i$ 是关于 n 的单调减序列, 故由命题 1.1.1 有

$$\begin{aligned}0\leqslant P\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}A_i\Big)&=P\Big(\lim_{n\to\infty}\bigcup_{i=n}^{\infty}A_i\Big)\\&=\lim_{n\to\infty}P\Big(\bigcup_{i=n}^{\infty}A_i\Big)\leqslant\lim_{n\to\infty}\sum_{i=n}^{\infty}P(A_i)=0.\end{aligned} \qquad \square$$

下面讨论事件间的一种重要关系, 即事件的独立性问题.

两个事件 $A,B\in\mathcal{F}$, 若满足 $P(AB)=P(A)P(B)$, 称 A 与 B 相互独立. 容易证明下列命题等价: ① A 与 B 独立; ② A 与 B^C 独立; ③ $P(A|B)=P(A)$; ④ $P(A|B^C)=P(A)$.

三个事件 $A,B,C\in\mathcal{F}$, 若满足

$$P(AB)=P(A)P(B),\quad P(AC)=P(A)P(C),\quad P(BC)=P(B)P(C)$$

及

$$P(ABC) = P(A)P(B)P(C),$$

称 A, B, C 相互独立. 请读者证明: 若 A, B, C 独立, 则 $A \bigcup B$ 与 C, AB 与 C, $A - B$ 与 C 相互独立.

n 个事件 $A_1, A_2, \cdots, A_n \in \mathcal{F}$, 若对其中任意 $k(2 \leqslant k \leqslant n)$ 个事件

$$A_{i_1}, A_{i_2}, \cdots, A_{i_k}(\text{其中} 1 \leqslant i_1 \leqslant i_2 \leqslant \cdots \leqslant i_k \leqslant n),$$

有 $P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$, 称 $A_1, A_2, \cdots, A_n$**相互独立**.

类似可证明: 若 $A_1, A_2, \cdots, A_n$ 相互独立, 取 $1 \leqslant m < n$, 记 $\mathcal{F}_1 = \sigma(A_k, 1 \leqslant k \leqslant m), \mathcal{F}_2 = \sigma(A_k, m+1 \leqslant k \leqslant n)$. 任取 $B_1 \in \mathcal{F}_1, B_2 \in \mathcal{F}_2$, 则 B_1 与 B_2 独立.

称事件序列 $\{A_n, n \geqslant 1\}$ 相互独立, 若任取其中有限个均相互独立.

命题 1.1.3 若 $\{A_n, n \geqslant 1\}$ 是相互独立的事件序列, 且 $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$, 则有

$$P\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\Big) = 1.$$

证明

$$P\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i\Big) = P\Big(\lim_{n\to\infty}\bigcup_{i=n}^{\infty} A_i\Big) = \lim_{n\to\infty} P\Big(\bigcup_{i=n}^{\infty} A_i\Big) = \lim_{n\to\infty}\Big[1 - P\Big(\bigcap_{i=n}^{\infty} A_i^C\Big)\Big].$$

但

$$
\begin{aligned}
P\Big(\bigcap_{i=n}^{\infty} A_i^C\Big) &= \prod_{i=n}^{\infty} P(A_i^C) && (\text{由独立性}) \\
&= \prod_{i=n}^{\infty}(1-P(A_i)) \leqslant \prod_{i=n}^{\infty} \mathrm{e}^{-P(A_i)} && (\text{由} 1-x \leqslant \mathrm{e}^{-x}, \quad x \geqslant 0) \\
&= \exp\Big(-\sum_{i=n}^{\infty} P(A_i)\Big) = 0. && \Big(\text{因为} \sum_{i=n}^{\infty} P(A_i) = \infty \text{对所有} n\Big)
\end{aligned}
$$

因此命题得证. □

1.2 随机变量、分布函数及数字特征

1. 随机变量与分布函数

考虑一样本空间 Ω, 记 $\mathbb{R}$ 为实数全体之集. 随机变量定义为:

定义 1.2.1 设 $(\Omega, \mathcal{F}, P)$ 是一概率空间,$X(\omega)$ 是定义在 Ω 上的单值实函数, 如果对 $\forall a \in \mathbb{R}$, 有 $\{\omega\text{：} X(\omega) \leqslant a\} \in \mathcal{F}$, 则称 $X(\omega)$ 为**随机变量**(random variable).

这里有几点说明：

(1) $\{\omega\text{：} X(\omega) \leqslant a\}$ 是指所有满足 $X(\omega) \leqslant a$ 的样本点 ω 的集合, 定义要求 $\{\omega\text{：} X(\omega) \leqslant a\}$ 是 $(\Omega, \mathcal{F}, P)$ 的一个事件, 因而可定义它的概率.

(2) 定义中 ω 为自变量, 为了书写方便, 简记 $\{\omega\text{：} X(\omega) \leqslant a\} = \{X \leqslant a\} = \{X \in (-\infty, a]\}$. 以下把 $X(\omega)$ 记为 X, 一般随机变量符号用大写字母 X, Y, Z 等表示.

(3) $X(\omega)$ 满足 $\{\omega\text{：} X(\omega) \leqslant a\} \in \mathcal{F}$, 则易证$\forall a, b \in \mathbb{R}, \{X > a\}, \{X < a\}, \{X = a\}, \{a < X \leqslant b\}, \{a \leqslant X < b\}, \{a < X < b\}, \{a \leqslant X \leqslant b\} \in \mathcal{F}$.

例 1 若 $(\Omega, \mathcal{F})$ 中的 $\mathcal{F} = \{\varnothing, A, A^C, \Omega\}$, $A \in \mathcal{F}$ 而 $A_1 \notin \mathcal{F}$, 容

易验证 A_1 的示性函数 $I_{A_1}(\omega)$ 使 $\{I_{A_1} \leqslant 1/2\} \notin \mathcal{F}$, 故 $I_{A_1}(\omega)$ 对 $\mathcal{F}$ 而言不满足随机变量的定义.

例 2 给定 $(\Omega, \mathcal{F})$, 设 $\{B_k\}(0 \leqslant k < \infty)$ 是 Ω 的一个划分, 即 $B_k B_l = \varnothing (k \neq l); \bigcup\limits_{k=1}^{\infty} B_k = \Omega$ 且 $B_k \in \mathcal{F}(0 \leqslant k < \infty)$, 定义 $X(\omega) = \sum\limits_{k=1}^{\infty} x_k I_{B_k}(\omega)$, 则容易验证 $X(\omega)$ 是随机变量.

可以证明$\forall B \in \mathcal{B}, \{\omega, X(\omega) \in B\} \in \mathcal{F}$ 等价于 $\forall a \in \mathbb{R}, \{\Omega, X(\omega) \leqslant a\} \in \mathcal{F}$. 参见 [21,22]. 简记 $X = X(\omega)$, 且记 $X^{-1}(B) = (\omega: X(\omega) \in B)$.

设 X 为 $(\Omega, \mathcal{F}, P)$ 上的随机变量, 对 $\forall x \in \mathbb{R}$, 定义

$$F(x) = P(X \leqslant x) = P(X \in (-\infty, x]),$$

称 $F(x)$ 为 X 的**分布函数**(distribution function).

若随机变量 X 的可能取值的全体是一可数集或有限集, 则称 X 是**离散型随机变量**.

对随机变量 X 的分布函数 $F(x)$, 若存在一非负函数 $f(x)$, 对 $\forall x \in \mathbb{R}$, 有

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

则称 $f(x)$ 为随机变量 X 的**概率密度函数**(probability density function). 若 $f(x)$ 连续, 则

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x),$$

即

$$\lim_{h \to 0} \frac{P(x < X \leqslant x + h)}{h} = f(x),$$

或

$$P(x < X \leqslant x + h) = f(x)h + o(h).$$

以上关系是以后用所谓“微元法”求概率密度函数的依据: 为求随机变量 X 的概率密度函数, 先求 X 落在一个小区域 $(x, x+h]$ 上的概率 $P(x < X \leqslant x+h)$, 然后令 $h \to 0$, 求其极限

$$\lim_{h \to 0} \frac{P(x < X \leqslant x+h)}{h},$$

即得 $f(x)$.

二维随机变量 (X, Y) 的**联合分布函数** (joint distribution function)$F(x, y)$ 定义为

$$F(x, y) = P(X \leqslant x, Y \leqslant y).$$

X 和 Y 的**边缘分布**定义为

$$F_X(x) = P(X \leqslant x) = \lim_{y \to +\infty} F(x, y) = F(x, +\infty),$$
$$F_Y(y) = P(Y \leqslant y) = \lim_{x \to +\infty} F(x, y) = F(+\infty, y).$$

若存在一非负函数 $f(x, y)$, 对 $\forall\, (x, y) \in \mathbb{R}^2$ 有

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

则称 $f(x, y)$ 为 (X, Y) 的**联合概率密度函数**.

称随机变量 X 与 Y **相互独立** (independent), 若对 $\forall\, (x, y) \in \mathbb{R}^2$, 有

$$F(x, y) = F_X(x) F_Y(y).$$

n 维随机向量 $X_1, X_2, \cdots, X_n$ 的**联合分布函数**定义为

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n).$$

若对 $\forall\, (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ 有 $F(x_1, x_2, \cdots, x_n) = F_1(x_1) F_2(x_2) \cdots F_n(x_n)$, 则称 $X_1, X_2, \cdots, X_n$ **相互独立**. 这里 $F_i(x_i) = P(X_i \leqslant x_i)$.

可以证明, 若 X, Y, Z 相互独立, 则 $X \pm Y$ 与 Z 独立,$X \cdot Y$ 与 Z 独立,$X/Y (Y \neq 0)$ 与 Z 独立, 更一般有 $g_1(X, Y)$ 与 $g_2(Z)$ 独立 (其中 $g_1(X, Y), g_2(Z)$ 可以是逐段单调函数或逐段连续函数).

2. 黎曼－斯蒂尔切斯积分

为以后表示简便, 这里我们引出黎曼－斯蒂尔切斯 (Riemann-Stieltjes)
积分.

设 $F(x)$ 为 $(-\infty, +\infty)$ 上的单调不减右连续函数$g(x)$ 为 $(-\infty, +\infty)$ 上的单值实函数,$\forall\, a < b$.

定义 1.2.2　任取分点 $a = x_0 < x_1 < x_2 < \cdots < x_{i-1} < x_i < \cdots < x_n = b, \forall\, u_i \in [x_{i-1}, x_i]$, 作积分和式

$$\sum_{i=1}^{n} g(u_i) \Delta F(x_i) = \sum_{i=1}^{n} g(u_i)[F(x_i) - F(x_{i-1})].$$

令 $\lambda = \max\limits_{1 \leqslant i \leqslant n} \Delta x_i = \max\limits_{1 \leqslant i \leqslant n} (x_i - x_{i-1})$, 若极限

$$J(a, b) = \lim_{\lambda \to 0} \sum_{i=1}^{n} g(u_i) \Delta F(x_i)$$

存在, 则记

$$J(a, b) = \int_a^b g(x)\, \mathrm{d}F(x) \qquad \left(\text{或} \int_a^b g(x) F(\mathrm{d}x)\right),$$

称极限 $J(a, b)$ 为 $g(x)$ 关于 $F(x)$ 在 $[a, b]$ 上的 R-S 积分.

注　(1) $\lambda \to 0$ 意味着 $n \to \infty$, 且最大子区间的长度趋于 0. (2) 当取 $F(x) = x$ 时,R-S 积分化为原来的黎曼 (Riemann) 积分, 所以 R-S 积分是黎曼积分的推广.

当 $a \to -\infty, b \to +\infty$ 时, 若极限

$$J(-\infty, +\infty) = \lim_{b \to \infty, a \to -\infty} \int_a^b g(x)\, \mathrm{d}F(x)$$

存在, 则称

$$J(-\infty,+\infty)=\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}F(x) \qquad \left(\text{或}\int_{-\infty}^{+\infty} g(x)F(\mathrm{d}x)\right)$$

为 $g(x)$ 关于 $F(x)$ 在 $(-\infty,+\infty)$ 上的 R-S 积分.

R-S 积分的基本性质:

(1) 当 $a<c_1<\cdots<c_n<b$ 时

$$\int_a^b g(x)\,\mathrm{d}F(x)=\sum_{i=0}^{n}\int_{c_i}^{c_{i+1}} g(x)\,\mathrm{d}F(x) \quad (a=c_0, b=c_{n+1});$$

(2)

$$\int_a^b \sum_{i=1}^{n} g_i(x)\,\mathrm{d}F(x)=\sum_{i=1}^{n}\int_a^b g_i(x)\,\mathrm{d}F(x);$$

(3) 若 $g(x)\geqslant 0$, 且 $a<b$, 则

$$\int_a^b g(x)\,\mathrm{d}F(x)\geqslant 0;$$

(4) 若 $F_1(x),F_2(x)$ 为两个分布函数,c_1,c_2 为常数.$c_1,c_2>0$, 则

$$\int_a^b g(x)\,\mathrm{d}[c_1F_1(x)+c_2F_2(x)]=c_1\int_a^b g(x)\,\mathrm{d}F_1(x)+c_2\int_a^b g(x)\,\mathrm{d}F_2(x).$$

几个特例:

设 $F(x)$ 为 X 的分布函数.

(1) 若 $g(x)=1$, 则

$$\int_a^b \mathrm{d}F(x)=F(b)-F(a)=P(a<X\leqslant b).$$

(2) 若 X 为离散型随机变量, 即 $P(X=c_i)=p_i$, $i\in\{1,2,\cdots\}$, 则

$$F(x)=\sum_{c_i\leqslant x} p_i$$

是一跳跃型分布函数, 即 $F(x)$ 的变化只在 $c_1, c_2, \cdots$ 这些点且其跃度为 p_i, 则 R-S 积分

$$\int_{-\infty}^{\infty} g(x)\,\mathrm{d}F(x) = \sum_{n=1}^{\infty} g(c_n)[F(c_n+0) - F(c_n-0)] = \sum_{n=1}^{\infty} g(c_n)p_n$$

化成了一个级数.

3. 数字特征

(1) 随机变量的数学期望 (expectation or mean)

定义 1.2.3 设 X 为随机变量,$F(x)$ 为 X 的分布函数, 若 $\int_{-\infty}^{+\infty} |x|\,\mathrm{d}F(x)$ 存在, 则称

$$EX = \int_{-\infty}^{\infty} x\,\mathrm{d}F(x)$$

为随机变量 X 的**数学期望**(或称为 X 的均值).

性质 (1) 若 $(c_i, i = 1, 2, \cdots, n)$ 为常数,$X_i (i = 1, 2, \cdots, n)$ 为随机变量, 则

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

性质 (2) 设 $g(x)$ 为 x 的函数,$F_X(x)$ 为 X 的分布函数, 若 $E[g(X)]$ 存在, 则

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}F_X(x).$$

当 X 为离散型随机变量, 即 $P(X = x_n) = p_n (n \in \mathbb{N})$ 时, 则

$$EX = \sum_{n=1}^{\infty} x_n p_n,$$

即 EX 是 X 所有可能取值的加权平均.

当 X 为连续随机变量, 且有概率密度函数 $f(x)$ 时, 则

$$EX = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x.$$

(2) 方差 (variance)

定义 1.2.4 令 $DX \triangleq E(X-EX)^2 = EX^2-(EX)^2$, 称 DX 为随机变量 X 的**方差**(有时记 $DX = \operatorname{var} X = \sigma_X^2$).

DX 刻画了 X 取值的集中或分散程度.

(3) 协方差 (covariance)

定义 1.2.5 两个随机变量 (X,Y), 称

$$\operatorname{cov}(X,Y) \triangleq E[(X-EX)(Y-EY)] = E(XY)-(EX)(EY)$$

为 (X,Y) 的**协方差**.

若 X,Y 独立, 则 $E(XY)=EX\,EY$, 从而得 $\operatorname{cov}(X,Y)=0$. 于是, 若 $\operatorname{cov}(X,Y)\neq 0$, 则 X,Y 不独立. 因此 $\operatorname{cov}(X,Y)\neq 0$ 刻画了 X,Y 取值存在某种统计上的线性相关关系.

(4) 相关系数 (correlation coefficient)

定义 1.2.6 若 $0<DX=\sigma_X^2<\infty, 0<DY=\sigma_Y^2<\infty$, 称

$$\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}$$

为 (X,Y) 的**相关系数**.

$\rho(X,Y)$ 刻画了 X,Y 之间线性关系的密切程度, 若 $\rho=0$, 称 X,Y 不

相关.

主要性质:

① $$D\Big(\sum_{i=1}^n a_iX_i\Big)=\sum_{i=1}^n a_i^2DX_i+2\sum_{i<j}a_ia_j\operatorname{cov}(X_i,X_j);$$

② 若 $X_1,X_2,\cdots,X_n$ 相互独立, 则 $\operatorname{cov}(X_i,X_j)=0,\ j\neq i$, 即 X_i,X_j 不相关;

③ 若 $X_1,X_2,\cdots,X_n$ 两两不相关, 则

$$D\Big(\sum_{i=1}^n X_i\Big)=\sum_{i=1}^n DX_i;$$

④ 施瓦茨 (Schwarz) 不等式, 若随机变量 X, Y 的二阶矩存在, 则

$$|E(XY)|^2 \leqslant E(X^2)E(Y^2);$$

特别是

$$|\operatorname{cov}(X,Y)|^2 \leqslant \sigma_X^2\sigma_Y^2,$$
$$|\rho(X,Y)| \leqslant 1.$$

⑤ $\rho = \pm 1$, 当且仅当

$$P\Big(\frac{Y-EY}{\sqrt{DY}} = \pm\frac{X-EX}{\sqrt{DX}}\Big) = 1,$$

即 $\rho = \pm 1$, (X,Y) 以概率 1 取值在直线 $Y - EY = \pm\sqrt{DY}(X - EX)/\sqrt{DX}$ 上.

(5) 矩 (moment)

定义 1.2.7 记

$$E(X^k) = \int_{-\infty}^{+\infty} x^k \,\mathrm{d}F_X(x), \quad (k \geqslant 1),$$

称 $E(X^k)$ 为随机变量 X 的k **阶矩**$(k \in \mathbb{N})$.

4. 常用随机变量的分布

(1) 离散型随机变量

① 二项分布

设 $0 \leqslant p < 1, n \geqslant 1$, 若 X 的分布律为

$$P(X=k) = \mathrm{C}_n^k p^k (1-p)^{n-k}, \quad 0 \leqslant k \leqslant n,$$

称 X 是参数为 (n,p) 的**二项分布**(binomial with parameters (n,p)), 简记为 $X \sim B(n,p)$.

② 泊松 (Poisson) 分布

设 $\lambda > 0$, 若 X 的分布律为

$$P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad k=0,1,2,\cdots,$$

称 X 是参数为 λ 的**泊松分布**(Poisson with parameter λ), 简记为 $X \sim P(\lambda)$.

③ 几何分布

设 $0<p<1$, 若 X 的分布律为

$$P(X=k)=(1-p)^{k-1}p, \quad k=1,2,\cdots,$$

称 X 是参数为 p 的**几何分布**(geometric distribution), 简记为 $X \sim G(p)$.

(2) 具有概率密度的随机变量

① 均匀分布

设 $a<b$, 若 X 的概率密度函数为

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

称 X 是 (a,b) 上的**均匀分布**(uniform over (a,b)), 简记为 $X \sim U(a,b)$.

② 正态分布

设 $\mu \in \mathbb{R}, \sigma>0$, 若 X 的概率密度函数为

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

称 X 是参数为 (μ,σ^2) 的**正态分布**(normal with parameters (μ,σ^2)), 简记为 $X \sim N(\mu,\sigma^2)$.

③ 指数分布

设 $\lambda > 0$, 若 X 的概率密度函数为

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

称 X 是参数为 λ 的**指数分布**(exponential with parameter λ), 简记为 $X \sim E(\lambda)$.

5. 连续型随机变量的事件示性函数的线性组合表示

(1) 设 $X(\omega)$ 为非负随机变量, $P(X < \infty) = 1$, 令

$$X_n(\omega) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} I_{\left\{\frac{k}{2^n} \leqslant X < \frac{k+1}{2^n}\right\}}(\omega) + nI_{\{X \geqslant n\}}(\omega),$$

则 $X_n(\omega)$ 是随机变量, 满足 $\forall \omega \in \Omega, n \in \mathbb{N}, X_n(\omega) \leqslant X_{n+1}(\omega)$(记为 $X_n(\omega) \uparrow$), 且当 $0 \leqslant X(\omega) < n$ 时, $|X_n(\omega) - X(\omega)| < 1/2^n$; 当 $X(\omega) = n$ 时, $X_n(\omega) = n$. 故 $\forall \omega \in \Omega, \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)$.

(2) 设 $X(\omega)$ 为一般的随机变量, 令 $X^+ = X \vee 0 = \max(X, 0), X^- = -(X \wedge 0) = -\min(X, 0)$, 显然 $X^+, X^- \geqslant 0$ 由上面的结论, 对 X^+, X^- 存在 $X_n^+ \uparrow X^+, X_n^- \uparrow X^-$. 令 $X_n = X_n^+ - X_n^-$, 则 $X_n \uparrow X$.

1.3 矩母函数、特征函数和拉普拉斯变换

1. 矩母函数 (moment generating function)

定义 1.3.1 随机变量 X 的**矩母函数**定义为

$$\psi(t) \triangleq E(\mathrm{e}^{tX}) = \int_{-\infty}^{+\infty} \mathrm{e}^{tx}\, \mathrm{d}F_X(x),$$

如果上式右边积分存在.

显然, 如 X 的 k 阶矩存在, 则

$$E(X^k) = \psi^{(k)}(0).$$

矩母 (生成) 函数由此得名. 可以证明矩母函数与分布函数是一一对应的.

对于取值非负整数的随机变量 X, 即 $P(X=k)=p_k\geqslant 0(k\geqslant 0), \sum\limits_{k=0}^{\infty}p_k=1$, 则 X 的矩母函数记为

$$g(s)=E(s^X)=\sum_{k=0}^{\infty}p_ks^k,\quad 0\leqslant s\leqslant 1.$$

显然$p_k=g^{(k)}(0)/k!$, 且有 $E[X(X-1)(X-2)\cdots(X-k+1)]=g^{(k)}(1)$. 若 X_1,X_2 相互独立, 其矩母函数分别记为 $g_1(s),g_2(s)$, 则不难证明 X_1+X_2 的矩母函数为

$$g_{X_1+X_2}(s)=g_1(s)g_2(s).$$

对于数列 $\{a_n,n\geqslant 0\}$, 如

$$A(s)=\sum_{n=0}^{\infty}a_ns^n,\quad |s|\leqslant 1,$$

则称 $A(s)$ 为 $\{a_n,n\geqslant 0\}$ 的**母函数**.

母函数的几个重要性质：

(1) $E(X)=g'(1)$.

(2) $D(X)=g''(1)+g'(x)-[g'(1)]^2$.

(3) $g'(1)=\sum\limits_{k=1}^{\infty}kp_k$.

2. 特征函数 (characteristic function)

定义 1.3.2 记

$$\phi(t)\stackrel{\triangle}{=}E\{\exp(\mathrm{i}tX)\}=\int_{-\infty}^{+\infty}\exp(\mathrm{i}tx)\,\mathrm{d}F_X(x),$$

其中 $\mathrm{i}=\sqrt{-1}$, $-\infty<t<+\infty$. 称 $\phi(t)$ 是随机变量 X 的**特征函数**.

$\phi(t)$ 的几个重要性质:

(1) $\phi(0)=1$, $|\phi(t)|\leqslant\phi(0)$, $\phi(-t)=\overline{\phi(t)}$, 且 $\phi(t)$ 在 $(-\infty,+\infty)$ 上一致连续.

(2) $\phi(t)$ 具有非负定性, 即对任给 n 个实数 t_i 及复数 $\lambda_i(1\leqslant i\leqslant n)$ 有

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\phi(t_i-t_j)\lambda_i\overline{\lambda_j}\geqslant 0.$$

(3) 若 X 与 Y 相互独立, 则

$$\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t).$$

(4) 随机变量 X, 若 EX^n 存在, 则当 $k\leqslant n$ 时

$$\phi^{(k)}(0)=i^kE(X^k).$$

(5) 随机变量的分布函数与特征函数有一一对应的关系, 即给定 $F(x)$ 可唯一决定 $\phi(t)$; 反之, 给定 $\phi(t)$ 可唯一决定 $F(x)$ (唯一性定理).

上述性质的证明详见参考文献 [20].

3. 拉普拉斯—斯蒂尔切斯变换

定义 1.3.3 设非负随机变量 X, 分布函数 $F_X(x), s=a+b\mathrm{i}$, 这里 $a>0, b$ 是实数, 称

$$\widehat{F}_X(s)=\int_0^{+\infty}\exp(-sx)\,\mathrm{d}F_X(x)$$

为 $F_X(x)$ 的拉普拉斯—斯蒂尔切斯变换 (Laplace-Stieltjes transform, 简记 L-S 变换), 或称随机变量 X 的 L-S 变换.

注 $\widehat{F}_X(s)$ 与 $F_X(x)$ 也有一一对应关系, 且对 $X_1, X_2 \geqslant 0$ 相互独立, 有

$$\widehat{F}_{X_1+X_2}(s) = \widehat{F}_{X_1}(s)\widehat{F}_{X_2}(s).$$

1.4 条件数学期望

条件数学期望是随机过程中最基本最重要的概念之一. 为了直观地对此概念有正确的理解. 我们先从离散型随机变量入手, 再讨论连续型随机变量情形, 然后推广到多元随机变量的情形.

1. 离散型随机变量的情形

设两随机事件 A, B, 若 $P(B) > 0$, 称 $P(A|B) = P(AB)/P(B)$ 为事件 B 发生时, 事件 A 的**条件概率**(若 $P(B) = 0$, 则 $P(A|B)$ 没定义或规定为 0). 设 (X, Y) 为两个离散型随机变量, 其联合分布律为 $P(X = x_i, Y = y_j) = p_{ij} \geqslant 0$, $\sum\limits_i \sum\limits_j p_{ij} = 1$, 若 $P(Y = y_j) = \sum\limits_i p_{ij} \triangleq p_{\cdot j} > 0$, 称

$$P(X = x_i|Y = y_j) = P(X = x_i, Y = y_j)/P(Y = y_j) = p_{ij}/p_{\cdot j}$$

为给定 $Y = y_j$ 时X 的**条件分布律**. 称

$$E(X|Y = y_j) \triangleq \sum_i x_i P(X = x_i|Y = y_j)$$

为给定 $Y = y_j$ 时X 的**条件数学期望**.

比较 (无条件) 数学期望 $EX = \sum\limits_i x_i P(X = x_i)$ 与条件数学期望 $E(X|Y = y_j)$ 的异同EX 是对所有 $\omega \in \varOmega, X(\omega)$ 取值全体的加权平均; 而 $E(X|Y = y_j)$ 是局限在 $\omega \in \{\omega\colon Y(\omega) = y_j\} \triangleq B_j$ 时, $X(\omega)$ 取值局部 ($\omega \in B_j$) 的加权平均. 这是因为: 记 $B_j = \{\omega\colon Y = y_j\}, A_i = \{\omega\colon X = x_i\}$, 于是整个样本空间 $\varOmega$ 按 Y 的不同取值分为

$B_1, \cdots, B_j, \cdots$ 等互不相容的事件 $\left(\Omega = \sum_j B_j\right)$. 而 Ω 又按 X 不同取值分为 $A_1, \cdots, A_i, \cdots$ 等互不相容的事件 $\left(\Omega = \sum_i A_i\right)$.

当 $A_iB_j = \varnothing$ 时,$P(X = x_i, Y = y_j) = 0,\ P(X = x_i \big| Y = y_j) = 0$, 于是

$$E(X\big|Y = y_j) = \sum_i x_i P(X = x_i\big|Y = y_j) = \sum_{i:\ A_iB_j \neq \varnothing} x_i P(X = x_i\big|Y = y_j).$$

因此 $E(X\big|Y = y_j)$ 是 $\omega \in B_j$ 时 $X(\omega)$ 的局部加权平均.

显然,$E(X\big|Y = y_1), \cdots, E(X\big|Y = y_j), \cdots$, 依赖于 $Y = y_j$, 即依赖于 $\omega \in B_j = \{\omega\text{: } Y = y_j\}$, 这样, 从全局样本空间 Ω 及对 $\omega \in \Omega$ 可以变化的观点看, 有必要引进一个新的随机变量, 记为 $E(X\big|Y)$. 对于这个随机变量 $E(X\big|Y)$, 当 $\omega \in B_j$ 时 (即 $Y = y_j$ 时) 它的取值为 $E(X\big|Y = y_j)$, 称随机变量 $E(X\big|Y)$ 为随机变量 X 关于随机变量 Y 的条件数学期望.

为给出 $E(X\big|Y)$ 的确切定义及表示式, 引进事件的示性函数如下: 记

$$I_{B_j}(\omega) = \begin{cases} 1, & \omega \in B_j = \{\omega\text{: } Y(\omega) = y_j\}, \\ 0, & \omega \notin B_j = \{\omega\text{: } Y(\omega) = y_j\}. \end{cases}$$

显然 $I_{B_j}(\omega) = 1 \leftrightarrow Y(\omega) = y_j$ 发生, 亦记 $I_{B_j}(\omega) = I_{(Y=y_j)}(\omega)$.

定义 1.4.1　记

$$E(X\big|Y) \triangleq \sum_j I_{(Y=y_j)}(\omega) E(X\big|Y = y_j), \tag{1.4.1}$$

称 $E(X\big|Y)$ 为 X 关于 Y 的条件数学期望.

$E(X\big|Y)$ 的定义包含如下的直观意义:

(1) 随机变量 $E(X|Y)$ 是随机变量 Y 的函数, 当 $\omega \in \{\omega: Y = y_j\}$ 时, $E(X|Y)$ 的取值为 $E(X|Y = y_j)$. 事实上, 它是局部平均 $\{E(X|Y = y_j), j \in \mathbb{N}\}$ 的统一表达式.

(2) 当 $E(X|Y = y_j) \neq E(X|Y = y_k)(j \neq k)$ 时,$P[E(X|Y) = E(X|Y = y_j)] = P(Y = y_j)$; 否则, 令 $D_j = \{k: E(X|Y = y_k) = E(X|Y = y_j)\}$, 则

$$P\{E(X|Y) = E(X|Y = y_j)\} = \sum_{k \in D_j} P(Y = y_k).$$

(3) 由于随机变量 $E(X|Y)$ 是随机变量 Y 的函数, 故它的数学期望应为

$$E(E(X|Y)) = \sum_j E(X|Y = y_j)P(Y = y_j).$$

例 1 随机变量 (X, Y) 的联合分布律如表 1.1.

表 1.1

$Y \backslash X$	1	2	3	$p_{\cdot j}$
1	2/27	4/27	1/27	7/27
2	5/27	7/27	3/27	15/27
3	1/27	2/27	2/27	5/27
$p_{i\cdot}$	8/27	13/27	6/27	

试求 $E(X|Y)$ 的分布律,EX 及 $E(E(X|Y))$.

解 为求 $E(X|Y = j)$, 先求 $P(X = i, Y = j), i, j = 1, 2, 3$. 当 $Y = 1$ 时, 有

$$P(X = 1|Y = 1) = P(X = 1, Y = 1)/P(Y = 1) = \frac{2/27}{7/27} = \frac{2}{7},$$

同理 $P(X=2\big|Y=1)=4/7, P(X=3\big|Y=1)=1/7$. 故

$$E(X\big|Y=1)=\sum_{i=1}^{3}iP(X=i\big|Y=1)=1\times\frac{2}{7}+2\times\frac{4}{7}+3\times\frac{1}{7}=\frac{13}{7}.$$

类似地有

$$E(X\big|Y=2)=\sum_{i=1}^{3}iP(X=i\big|Y=2)=1\times\frac{5}{15}+2\times\frac{7}{15}+3\times\frac{3}{15}=\frac{28}{15},$$

$$E(X\big|Y=3)=\sum_{i=1}^{3}iP(X=i\big|Y=3)=1\times\frac{1}{5}+2\times\frac{2}{5}+3\times\frac{2}{5}=\frac{11}{5}.$$

又 $P\{E(X\big|Y)=E(X\big|Y=j)\}=P(Y=j), j=1,2,3$. 故 $E(X\big|Y)$ 的分布律列表如下: 于是

$E(X\big\|Y=j)$	$13/7$	$28/15$	$11/5$
$P\{E(X\big\|Y)=E(X\big\|Y=j)\}=P(Y=j)$	$7/27$	$15/27$	$5/27$

$$E(E(X\big|Y))=\frac{13}{7}\times\frac{7}{27}+\frac{28}{15}\times\frac{15}{27}+\frac{11}{5}\times\frac{5}{27}=\frac{52}{27},$$

而

$$EX=\sum_{i=1}^{3}iP(X=i)=1\times\frac{8}{27}+2\times\frac{13}{27}+3\times\frac{6}{27}=\frac{52}{27},$$

得

$$EX=E(E(X\big|Y))=\frac{52}{27}.$$ □

自然要问, 一般情形下, 是否 $EX=E(E(X\big|Y))$? 它的直观意义又是什么? 试证明:

(1) 在 (X,Y) 为离散随机变量且 $E|X|<\infty$ 时,$EX=E(E(X\big|Y))$;

(2) $E(X\big|X)=X$.

2. 连续随机变量 $(\boldsymbol{X},\boldsymbol{Y})$ 的情形

设 (X,Y) 的联合概率密度函数 (jointly probability density function) 为 $f(x,y)$, Y 的概率密度函数为 $f_Y(y)=\displaystyle\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}x$, 设 $f_Y(y)>0, E|X|<\infty$, 给定 $Y=y, X$ 的条件概率密度函数为

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

条件分布函数为

$$F_{X|Y=y}(x|y)=P(X\leqslant x\big|Y=y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}\,\mathrm{d}u,$$

条件数学期望为

$$E(X|Y=y)=\int_{-\infty}^{+\infty}xf_{X|Y=y}(x|y)\,\mathrm{d}x=\int_{-\infty}^{+\infty}x\frac{f(x,y)}{f_Y(y)}\,\mathrm{d}x. \quad (1.4.2)$$

令 $D\in\mathcal{B}$, 考虑 $Y\in D$ 下, 若 $P(Y\in D)>0, X$ 的条件分布函数为

$$F(x\big|D)=P\{X\leqslant x\big|Y\in D\}=\frac{P(X\leqslant x,Y\in D)}{P(Y\in D)}=\frac{\displaystyle\int_{-\infty}^{x}\Big(\int_{y\in D}f(x,y)\,\mathrm{d}y\Big)\,\mathrm{d}x}{\displaystyle\int_{y\in D}f_Y(y)\,\mathrm{d}y}.$$

在 $Y\in D$ 下,X 的条件概率密度函数为

$$f_{X|D}(x\big|D)=\frac{\displaystyle\int_{y\in D}f(x,y)\,\mathrm{d}y}{P(Y\in D)}. \quad (1.4.3)$$

于是在 $Y\in D$ 下,X 的条件数学期望定义为

$$E(X\big|Y\in D)\triangleq\int_{-\infty}^{+\infty}xf_{X|D}(x\big|D)\,\mathrm{d}x=\int_{y\in D}\Big(\int_{-\infty}^{+\infty}xf(x,y)\,\mathrm{d}x\Big)\,\mathrm{d}y/P(Y\in D).$$

由上式定义, 有

$$
\begin{aligned}
E(X\big|Y\in D) &= \int_{y\in D}\Big(\int_{-\infty}^{+\infty} x\frac{f(x,y)}{f_Y(y)}\,\mathrm{d}x\Big)f_Y(y)\,\mathrm{d}y/P(Y\in D)\\
&= \frac{1}{P(Y\in D)}\int_{y\in D} E(X\big|Y=y)f_Y(y)\,\mathrm{d}y.
\end{aligned}
$$

显然, 条件数学期望 $E(X\big|Y=y)$ 是 y 的函数. 这样, 从整个样本空间 Ω 及从 $\omega\in\Omega$ 可以变化的宏观上看, 可以且有必要定义一个随机变量 $E(X\big|Y)$, 使其在 $Y=y$ 时, $E(X\big|Y)$ 的取值为 $E(X\big|Y=y)$.

定义 1.4.2　设 (X,Y) 具有联合概率密度函数 $f(x,y), Y$ 的概率密度函数为 $f_Y(y)>0, E|X|<\infty$, 若随机变量 $E(X\big|Y)$ 满足:

(1) $E(X\big|Y)$ 是随机变量 Y 的函数, 当 $Y=y$ 时, 它的取值为 $E(X\big|Y=y)$;

(2) 对任意 $D\in\mathcal{B}$, 有

$$
E[E(X\big|Y)\big|Y\in D]=E(X\big|Y\in D). \tag{1.4.4}
$$

称随机变量 $E(X\big|Y)$ 为 X 关于 Y 的条件数学期望.

从 (1), 由于 $E(X\big|Y)$ 是随机变量 Y 的函数, 故它的数学期望应为

$$
E[E(X\big|Y)]=\int_{-\infty}^{+\infty} E(X\big|Y=y)f_Y(y)\,\mathrm{d}y.
$$

而从 (2), 当取 $D=\mathbb{R}=(-\infty,+\infty)$ 时

$$
\begin{aligned}
EX &= E\{X\big|Y\in(-\infty,+\infty)\}=E\{E(X\big|Y)\big|Y\in(-\infty,\infty)\}=E[E(X\big|Y)]\\
&= \int_{-\infty}^{+\infty} E(X\big|Y=y)f_Y(y)\,\mathrm{d}y.
\end{aligned}
$$

例 2　(X,Y) 是二维正态分布, 即 $(X,Y)\sim N(\mu_1,\mu_2,\rho,\sigma_1^2,\sigma_2^2)$,

则联合概率密度函数为

$$\begin{aligned}f(x,y)=&\frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\Big\{-\frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_1)^2}{\sigma_1^2}-\\&2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\Big]\Big\},\end{aligned}$$

则

$$\begin{aligned}f_{Y|X=x}(y\big|x)&=\frac{f(x,y)}{f_X(x)}\\&=\frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\Big\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\Big[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\Big]^2\Big\}\\&\sim N\Big(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\Big).\end{aligned}$$

故

$$E(Y\big|X=x)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1);$$

$$E(Y\big|X)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1). \tag{1.4.5}$$

此时 $E(Y\big|X)$ 是 X 的线性函数, 这是正态分布的重要性质. □

3. 一般随机变量的情形

设 (X,Y) 为一般随机变量, 其联合分布函数为 $P(X\leqslant x,Y\leqslant y)$. 以下假设 $E|X|<\infty$, 分两种情况讨论.

定义 1.4.3 设 $D\in\mathcal{B},P(Y\in D)>0$. $\forall x\in\mathbb{R}$, 称 $P(X\leqslant x\big|Y\in D)=P(X\leqslant x,Y\in D)/P(Y\in D)$ 为 X 关于事件 $\{Y\in D\}$ 的条件分布函数.容易证明, 若 X 与 Y 独立, 则对 $\forall x\in\mathbb{R}$, $\forall D\in\mathcal{B},P(X\leqslant x\big|Y\in D)=P(X\leqslant x)$. 称 $E(X\big|Y\in D)=\int_{\mathbb{R}}x\,\mathrm{d}P(X\leqslant x\big|Y\in D)$ 为 X 关于 $\{Y\in D\}$ 的条件数学期望.

在许多问题中常常需要考虑 D 为单点集 $\{y\}$ 的情形. 若 $P(Y=y)>0$, 这时定义条件分布同上. 问题是当 $P(Y=y)=0$ 时, 如何定义 $P(X\leqslant x\big|Y=y)$.

定义 1.4.4　设 $(x,y)\in\mathbb{R}^2$, 对充分小的 $h>0$, 有 $P(y<Y\leqslant y+h)>0$. 若 $P(X\leqslant x\big|Y=y)\triangleq\lim\limits_{h\to 0}P(X\leqslant x\big|y<Y\leqslant y+h)$ 存在, 则称 $P(X\leqslant x\big|Y=y)$ 为 X 关于 $\{Y=y\}$ 的条件分布函数, 称 $E(X\big|Y=y)=\displaystyle\int_{\mathbb{R}}x\,\mathrm{d}P(X\leqslant x\big|Y=y)$ 为 X 关于 $\{Y=y\}$ 的条件数学期望.

若随机变量 $E(X\big|Y)$ 满足:

(1) $E(X\big|Y)$ 是随机变量 Y 的函数, 当 $Y=y$ 时, 它的取值为 $E(X\big|Y=y)$.

(2) 对于 $\forall D\in\mathcal{B}, E\{E(X\big|Y)\big|Y\in D\}=E(X\big|Y\in D)$.

称随机变量 $E(X\big|Y)$ 为 X 关于 Y 的条件数学期望.

从该定义中的 (1) 知 $E(X\big|Y)$ 是随机变量 Y 的函数, 则它的数学期望为

$$E[E(X\big|Y)]=\int_{\mathbb{R}}E(X\big|Y=y)\,\mathrm{d}P(Y\leqslant y).$$

但是从 (2) 知, 当取 $D=\mathbb{R}=(-\infty,+\infty)$ 时, 有

$$EX=E(X\big|Y\in(-\infty,+\infty))=E\{E(X\big|Y)\big|Y\in(-\infty,+\infty)\}=E\{E(X\big|Y)\},$$

故有 $EX=E\{E(X\big|Y)\}$, 即

$$EX=\int_{\mathbb{R}}E(X\big|Y=y)\,\mathrm{d}P(Y\leqslant y). \tag{1.4.6}$$

上式可看作是数学期望形式的全概率公式.

4. 条件概率与条件分布函数

设随机变量 (X,Y) 及任一随机事件 $B\in\mathcal{F}$, 记

$$I_B(\omega)=\begin{cases}1, & \omega\in B,\\ 0, & \omega\notin B,\end{cases}$$

即 I_B 是 B 的示性函数. 显然

$$P(B) = E(I_B(\omega)).$$

称

$$E(I_B(\omega)\big|Y) \triangleq P(B\big|Y)$$

为事件 B 关于随机变量 Y 的条件概率. 此时 $P(B\big|Y)$ 是随机变量且是 Y 的函数, 对于 $\forall\, x \in \mathbb{R}$, 取 $B = (\omega\colon X \leqslant x)$, 称

$$F(x\big|Y) \triangleq P(X \leqslant x\big|Y) = E(I_{(X\leqslant x)}\big|Y) \tag{1.4.7}$$

为 X 关于 Y 的条件分布函数.

于是有关条件概率, 条件分布函数均可用条件数学期望的概念及性质来处理.

5. 条件数学期望的基本性质

两个随机变量 Z_1, Z_2, 如果 $P(Z_1 = Z_2) = 1$, 称 Z_1, Z_2 几乎处处 (或称几乎必然) 相等, 记作 $Z_1 = Z_2$ a.s..

设 $X, Y, X_i(1 \leqslant i \leqslant n)$ 为随机变量,$g(x), h(y)$ 为一般函数, 且 $E\big|X\big|, E\big|X_i\big| < \infty(1 \leqslant i \leqslant n), E\big|g(X)h(Y)\big| < \infty, E\big|g(X)\big| < \infty$. 则有

(1)

$$E(E(X\big|Y)) = EX. \tag{1.4.8}$$

(2)

$$E\Big(\sum_{i=1}^{n} \alpha_i X_i\big|Y\Big) = \sum_{i=1}^{n} \alpha_i E(X_i\big|Y) \qquad \text{a.s.,} \tag{1.4.9}$$

其中 $\alpha_i(1 \leqslant i \leqslant n)$ 为常数.

(3)

$$E[(g(X)h(Y)\big|Y] = h(Y)E(g(X)\big|Y) \qquad \text{a.s.}, \tag{1.4.10}$$

特别地

$$E(X\big|X) = X \qquad \text{a.s.},$$

$$E[g(X)h(Y)] = E[h(Y)E(g(X)\big|Y)]. \tag{1.4.11}$$

(4) 如 X, Y 相互独立, 则

$$E(X\big|Y) = EX.$$

下面仅证 (1.4.11) 式, 其他留作练习.

设 $(X, Y) \sim f(x, y)$, 则

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{+\infty}\Big[\int_{-\infty}^{+\infty} g(x)\frac{f(x,y)}{f_Y(y)}\,\mathrm{d}x\Big]h(y)f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} E(g(X)\big|Y=y)h(y)f_Y(y)\,\mathrm{d}y \\ &= E[h(Y)E(g(X)\big|Y)]. \end{aligned}$$

由上面倒数第二式, 有

$$E(g(X)h(Y)) = \int_{-\infty}^{+\infty} E(g(X)\big|Y=y)h(y)f_Y(y)\,\mathrm{d}y. \tag{1.4.12}$$

特别地, 取 $g(X) = I_A(\omega), h(y) \equiv 1$ 时, 得

$$P(A) = \int_{-\infty}^{+\infty} P(A\big|Y=y)f_Y(y)\,\mathrm{d}y.$$

(1.4.12) 式是全概率公式的推广.

6. 对多元随机变量的条件数学期望

(1) 离散型随机变量

设三个随机变量 (X, Y, Z), 其中 (Y, Z) 为离散随机变量, 称随机变量 $E(X|Y, Z)$ 是 X 关于 Y, Z 的条件数学期望, 若它满足:

① $E(X|Y, Z)$ 是 (Y, Z) 的二元函数, 当 $Y = y_j;\ Z = z_k$ 时, $E(X|Y, Z)$ 的取值为 $E(X|Y = y_j;\ Z = z_k)$;

② 对任意 $D_j \in \mathbb{R}^1$, $D_k \in \mathbb{R}^1$, 有

$$E[E(X|Y, Z)|Y \in D_j, Z \in D_k] = E(X|Y \in D_j,\ Z \in D_k),$$

用示性函数表示, 即

$$E(X|Y, Z) \triangleq \sum_j \sum_k I_{(Y=y_j, Z=z_k)}(\omega) E(X|Y = y_j,\ Z = z_k).$$

当 $E|X| < \infty$ 时, 请读者证明

$$E[E(X|Y, Z)|Y] = E(X|Y) = E[E(X|Y)|Y, Z]. \tag{1.4.13}$$

(2) 连续型随机变量

如 (X, Y, Z) 为连续型随机变量, 联合概率密度函数为 $f(x, y, z)$; (Y, Z) 的联合概率密度函数为 $f_{Y,Z}(y, z)$; X 关于 $Y = y, Z = z$ 的条件概率密度函数为

$$f_{X|(Y,Z)=(y,z)}(x|y, z) = f(x, y, z)/f_{Y,Z}(y, z),$$

设 $E|X| < \infty, f_{Y,Z}(y, z) > 0$, 若随机变量 $E(X|Y, Z)$ 满足:

① $E(X|Y, Z)$ 是 Y, Z 的函数, 当 $Y = y$, $Z = z$ 时, 它们取值为

$$E(X|Y = y,\ Z = z).$$

② 对任意,$D_1 \in \mathbb{R}^1$, $D_2 \in \mathbb{R}^1$, 有

$$E\{(E(X|Y,Z)|Y \in D_1, Z \in D_2)\} = E(X|Y \in D_1, Z \in D_2).$$

称 $E(X|Y,Z)$ 为 X 关于 (Y,Z) 的条件数学期望.

(3) n 元随机变量

对离散随机变量 $\{X, Y_k, 1 \leqslant k \leqslant n\}$ 的情况, 称

$$\begin{aligned} E(X|Y_1,\cdots,Y_n) \triangleq \sum_{j_1}\cdots\sum_{j_n} I_{(Y_k=j_k,1\leqslant k\leqslant n)}(\omega)\times \\ E(X|Y_1=j_1,Y_2=j_2,\cdots,Y_n=j_n) \end{aligned}$$

为 X 关于 $(Y_1,\cdots,Y_n)$ 的条件数学期望. 类似可定义一般多元随机变量作为条件的条件数学期望.

若 $E|X|, E|X_i| < \infty, 1 \leqslant i \leqslant 2, E|g(Y_1,\cdots,Y_n)| < \infty$, 则类似有:

① $E(\alpha_1 X_1 + \alpha_2 X_2|Y_1,\cdots,Y_n) = \sum_{i=1}^{2} \alpha_i E(X_i|Y_1,\cdots,Y_n)$ a.s.;

② $E[g(Y_1,\cdots,Y_n)X|Y_1,\cdots,Y_n] = g(Y_1,\cdots,Y_n)E(X|Y_1,\cdots,Y_n)$ a.s.;

③ 如 X 与 $Y_1,\cdots,Y_n$ 独立, 则 $E(X|Y_1,\cdots,Y_n) = EX$ a.s..

④ $EX = E[E(X|Y_1,\cdots,Y_n)]$ a.s..

⑤ $\forall 1 \leqslant m < n$

$$\begin{aligned} E(X|Y_1,\cdots,Y_m) &= E[E(X|Y_1,\cdots,Y_n)|Y_1,\cdots,Y_m] \\ &= E[E(X|Y_1,\cdots,Y_m)|Y_1,\cdots,Y_n] \quad \text{a.s..} \end{aligned}$$

证明从略.

7. 条件乘法公式与条件独立性

(1) 条件概率的乘法公式

设 A, B 为两个随机事件, 由条件概率的定义可知 $P(AB) = P(A)P(B|A)$. 与上面的概率乘法公式类似, 条件概率的乘法公式如下:

命题 1.4.1 设 A,B,C 为 3 个随机事件, 则

$$P(BC\big|A)=P(B\big|A)P(C\big|AB). \tag{1.4.14}$$

证明 按条件概率的定义, 当 $P(AB)>0$ 时

$$P(BC\big|A)=\frac{P(ABC)}{P(A)}=\frac{P(AB)}{P(A)}\frac{P(ABC)}{P(AB)}=P(B\big|A)P(C\big|AB).$$

因此按对条件概率的等式的有关约定(1.4.14) 式成立. □

(2) 条件独立性

当两个随机事件 A,B 独立时, 有 $P(AB)=P(A)P(B)$, 即 $P(A\big|B)=P(A)$. 同样, 与上面的独立性概念类似, 条件独立性的定义如下：

定义 1.4.5 设 A,B,C 为 3 个随机事件, 称事件 A,C 关于事件 B**条件独立**, 若满足

$$P(C\big|AB)=P(C\big|B). \tag{1.4.15}$$

对于条件独立性有如下结论:

命题 1.4.2 设 A,B,C 为 3 个随机事件, 则事件 A,C 关于事件 B 条件独立的充要条件为

$$P(AC\big|B)=P(A\big|B)P(C\big|B). \tag{1.4.16}$$

证明 必要性. 由命题 1.4.1, 则

$$P(AC\big|B)=P(A\big|B)P(C\big|AB).$$

当 $P(AB)=0$ 时,$P(A\big|B)=0$, 因此 (1.4.16) 式两边均为 0, 而当 $P(AB)>0$ 时, 将 (1.4.15) 式代入上式即得 (1.4.16) 式.

充分性. 只需考虑 $P(AB)>0$ 的情形, 由命题 1.4.1 及 (1.4.16) 式可得

$$P(C\big|AB)=\frac{P(AC\big|B)}{P(A\big|B)}=P(C\big|B). \qquad \square$$

本书有关初等概率论的内容, 读者可参考文献 [22].

1.5 随机过程的概念

在概率论中, 研究了随机变量, n 维随机向量. 在极限定理中, 涉及到了无穷多个随机变量, 但局限在它们之间是相互独立的情形. 将上述情形加以推广, 即研究一族无穷多个、相互有关的随机变量, 这就是随机过程.

1. 概念

设对每一个参数 $t \in T, X(t,\omega)$ 是一随机变量, 称随机变量族 $X_T = \{X(t,\omega), t \in T\}$ 为一**随机过程**(stochastic process) 或称随机函数. 其中 $T \subset \mathbb{R}$ 是一实数集, 称为指标集.

用映射来表示 X_T,

$$X(t,\omega)\colon\ T \times \Omega \to \mathbb{R}\,,$$

即 $X(\cdot,\cdot)$ 是定义在 $T \times \Omega$ 上的二元单值函数, 固定 $t \in T$, $X(t,\cdot)$ 是定义在样本空间 Ω 上的函数, 即为一随机变量. 对于 $\omega \in \Omega$, $X(\cdot,\omega)$ (t 在 T 中顺序变化) 是参数 $t \in T$ 的一般函数, 通常称 $X(\cdot,\omega)$ 为样本函数, 或称随机过程的一个实现, 或说是一条轨道. 记号 $X(t,\omega)$ 有时也写为 $X_t(\omega)$ 或简记为 $X(t)$ 或 X_t.

参数 $t \in T$ 一般表示时间或空间. 参数集 T 常用的有 3 种:(1) $T_1 = \mathbb{N}_0 = \{0, 1, 2, \cdots\}$, (2) $T_2 = \mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$, (3) $T_3 = [a, b]$, 其中 a 可以取 $-\infty$ 或 0, b 可以取 $+\infty$. 当 T 取可列集 (T_1 或 T_2) 时, 通常称 X_T 为随机序列.

X_T 的取值也可以是复数,$\mathbb{R}^n$ 或更一般的抽象空间. $X_t (t \in T)$ 可能取值的全体所构成的集合称为**状态空间**, 记作 S. S 中的元素称为状态.

2. 例子

(1) 质点在直线上的随机游动. 设一质点在时刻 $t=0$ 时处于位置 a (整数), 以后每隔单位时间, 分别以概率 p 及 $q=1-p$ 向正的或负的方向随机移动一个单位, 记 X_n 为质点在时刻 $t=n$ 的位置. 固定 n, X_n 是随机变量. 考虑不同的 n 时,$\{X_n, n \geqslant 0\}$ 是一随机序列.

(2) 考虑某 “服务站” 在 $[0,t]$ 内来的 “顾客” 数, 记为 $N(t)$, 固定 $t, N(t)$ 就是一随机变量. 因此 $\{N(t), t \geqslant 0\}$ 是一随机过程. 这里的 “顾客” 可以是电话的 “呼唤”, 通信设备中的 “信号”, 一个系统的 “更换设备”, 放散性物质衰变的 “粒子” 等.

(3) 在外界是随机载荷条件下, 某零件 t 时的应力 $X(t)$ 是随机的, 故 $\{X(t), t \in T\}$ 是一随机过程$X(t)$ 亦可表示某电路中的电压、设备的温度、河流的流量 (或水位), 以及气体的压力等等.

(4) 考虑某输入输出系统, 例如最简单的 R-C 电路, 设输入端有一个干扰信号电压, 记为 $\xi(t)$, 记 $Q(t)$ 为 t 时电路的电量, 则它满足

$$R\frac{\mathrm{d}Q(t)}{\mathrm{d}t}+\frac{1}{C}Q(t)=\xi(t).$$

由于 $\{\xi(t), t \in T\}$ 是一随机过程, 容易理解 $\{Q(t), t \in T\}$ 也是一随机过程, 上式是一个最简单的随机微分方程.

3. 随机过程的数字特征及有限维分布函数族

设 $\{X(t), t \in T\}$ 是一随机过程. 为了刻画它的概率特征, 通常用到随机过程的均值函数、方差函数、协方差函数 (相关函数) 以及有限维分布函数族及特征函数族等概念.

(1) 均值函数. 随机过程 $\{X(t), t \in T\}$ 的**均值函数**定义为 (以下均假定右端存在)

$$m(t) \triangleq E(X(t)).$$

(2) 方差函数. 随机过程 $\{X(t), t \in T\}$ 的**方差函数**定义为

$$D(t) \triangleq E\{(X(t) - m(t))^2\}.$$

(3) 协方差函数. 随机过程 $\{X(t), t \in T\}$ 的**协方差函数**定义为

$$R(s,t) \triangleq \operatorname{cov}(X(s), X(t)).$$

(4) 相关函数. 随机过程 $\{X(t), t \in T\}$ 的**相关函数**定义为

$$\rho(s,t) \triangleq \frac{\operatorname{cov}(X(s), X(t))}{\sqrt{D(t)D(s)}}.$$

(5) 有限维分布族. 设 $t_i \in T, 1 \leqslant i \leqslant n$ (n 为任意正整数), 记

$$\begin{aligned}&F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n)\\&= P(X(t_1) \leqslant x_1, X(t_2) \leqslant x_2, \cdots, X(t_n) \leqslant x_n),\end{aligned}$$

其全体

$$\{F(t_1, t_2, \cdots, t_n, x_1, x_2, \cdots, x_n), t_1, t_2, \cdots, t_n \in T, n \geqslant 1\}$$

称为随机过程的**有限维分布族**. 它具有以下两个性质:

① 对称性　对 $(1, 2, \cdots, n)$ 的任一排列 $(j_1, j_2, \cdots, j_n)$, 有

$$F(t_{j_1}, t_{j_2}, \cdots, t_{j_n}; x_{j_1}, x_{j_2}, \cdots, x_{j_n}) = F(t_1, t_2, \cdots, t_n; x_1, x_2, \cdots, x_n).$$

② 相容性　对 $m < n$, 有

$$\begin{aligned}&F(t_1, \cdots, t_m, t_{m+1}, \cdots, t_n; x_1, \cdots, x_m, \infty, \cdots, \infty)\\&\qquad = F(t_1, \cdots, t_m; x_1, \cdots, x_m).\end{aligned}$$

一个随机过程的概率特性完全由其有限维分布族决定.

(6) 特征函数. 记

$$\begin{aligned}\phi(t_1,t_2,\cdots,t_n;\theta_1,\cdots,\theta_n) =& E\{\exp\{\mathrm{i}[\theta_1X(t_1)+\cdots+\theta_nX(t_n)]\}\}\\ =& \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp\{\mathrm{i}[\theta_1x_1+\cdots+\theta_nx_n]\}\times\\ & F(t_1,\cdots,t_n;\,\mathrm{d}x_1,\cdots,\,\mathrm{d}x_n),\end{aligned}$$

称 $\{\phi(t_1,\cdots,t_n;\theta_1,\cdots,\theta_n),n\geqslant 1,t_1,\cdots,t_n\in T\}$ 为随机过程 $\{X(t),t\in T\}$ 的有限维特征函数族.

1.6 随机过程的分类

设 $X_T=\{X(t),t\in T\}$ 为随机过程, 按其概率特征, 分类如下.

1. 独立增量过程

对 $t_1<t_2<\cdots<t_n$, $t_i\in T$, $1\leqslant i\leqslant n$, 若增量

$$X(t_1),X(t_2)-X(t_1),X(t_3)-X(t_2),\cdots,X(t_n)-X(t_{n-1})$$

相互独立, 则称 $\{X(t),t\in T\}$ 为**独立增量过程**(process with independent increments). 若对一切 $0\leqslant s<t$, 增量 $X(t)-X(s)$ 的分布只依赖于 $t-s$, 则称 X_T 有平稳增量. 有平稳增量的独立增量过程简称为独立平稳增量过程.

常见的泊松 (Poisson) 过程和维纳 (Wiener) 过程 (或称布朗运动 (Brownian motion)) 就是两个最简单也是最重要的独立平稳增量过程.

2. 马尔可夫过程

粗略地说, 一随机过程, 若已知现在的 t 状态 X_t, 那么将来状态 $X_u(u>t)$ 取值 (或取某些状态) 的概率与过去状态 $X_s(s<t)$ 取值无关, 或更简单地说, 已知现在, 将来与过去无关 (条件独立), 则称此性

质为马尔可夫性 (无后效性或简称为马氏性). 具有这种马尔可夫性的过程称为马尔可夫过程. 精确定义为:

随机过程 $\{X_t, t \in T\}$, 若对任意 $t_1 < t_2 \cdots < t_n < t, x_i, 1 \leqslant i \leqslant n$, 及 $A \subset \mathbb{R}$, 总有

$$P(X_t \in A \big| X_{t_1} = x_1, X_{t_2} = x_2, \cdots, X_{t_n} = x_n) = P(X_t \in A \big| X_{t_n} = x_n),$$

则称此过程为**马尔可夫过程**(Markov process), 简称马氏过程.

称 $P(s, x; t, A) = P(X_t \in A \big| X_s = x)(s < t)$ 为转移概率函数 (transition probability function).

X_t 的取值全体构成的集合记为 S, 称为状态空间. 对于马尔可夫过程 $X_T = \{X_t, t \in T\}$, 当 $S = \{1, 2, 3, \cdots\}$ 为可列无限集或有限集时, 通常称为**马尔可夫链**(Markov chain), 简称马氏链.

样本函数是连续的马尔可夫过程 $\{X_t, t \in [0, \infty]\}$ 称为扩散 (diffusion) 过程. 泊松过程是一个最简单连续时间马尔可夫链, 而布朗运动则是一个最简单的扩散过程.

3. 平稳过程及二阶矩过程

(1) 宽平稳过程 (或协方差平稳过程)

一随机过程 X_T, 若对 $\forall \tau, t \in T, D(X(t))$ 存在且

$$E(X(t)) = m,\ \operatorname{cov}(X_t, X_{t+\tau}) = R(\tau)$$

仅依赖 τ, 则称 X_T 为**宽平稳过程**(wide sense stationary process), 即它的协方差不随时间推移而改变.

(2) 二阶矩过程

一随机过程 X_T, 若对 $\forall t \in T, DX_t$ 存在, 则称为**二阶矩过程** (finite second moments process).

(3) 严平稳过程

一随机过程 X_T, 若对 $\forall\, t_1, t_2, \cdots, t_n \in T$, 及 $h > 0$, $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ 与 $(X_{t_1+h}, X_{t_2+h}, \cdots, X_{t_n+h})$ 有相同的联合分布, 则称该过程为**严平稳过程**(strictly stationary process). 严平稳过程的一切有限维分布对时间的推移保持不变. 特别地,$X(t), X(s)$ 的二维分布只依赖于 $t-s$.

尽管从实际应用的角度来看, 要求追溯到无穷的过去似乎有点不现实, 但为数学讨论方便, 平稳过程 (包括宽、严两种情况) 的指标集应取为 $(-\infty, +\infty)$ 或全体整数 (离散时间情形).

4. 鞅

若对 $\forall\, t \in T, E|X(t)| < \infty$, 且对 $\forall\, t_1 < t_2 < \cdots < t_n < t_{n+1}$, 有

$$E(X(t_{n+1})|X(t_1), X(t_2), \cdots, X(t_n)) = X(t_n) \qquad \text{a.s.},$$

则称 $\{X(t), t \in T\}$ 为**鞅**(martingales).

近十多年, 鞅在现代科技中有越来越广泛的应用.

5. 更新过程

设 $(X_k, k \geqslant 1)$ 为独立同分布的正的随机变量序列, 对 $\forall\, t > 0$, 令 $S_0 = 0, S_n = \sum\limits_{k=1}^{n} X_k$, 并定义

$$N(t) = \max\{n\colon n \geqslant 0, S_n \leqslant t\},$$

称 $\{N(t), t \geqslant 0\}$ 为**更新过程**(renewal process).

$N(t)$ 可以解释为 $[0,t]$ 内更换零件的个数或系统来的信号 (粒子) 数, 或 “服务站” 来的 “顾客数” 等.

6. 点过程 (或称计数过程)

一个随机过程 $\{N(A), A \subset T\}$ 是一**点过程**(point process), 若 $N(A)$ 表示在集合 A 中 “事件” 发生的总数, 即它满足:

(1) 对 $\forall\, A \subset T, N(A)$ 是一取值非负整数的随机变量 $(N(\varnothing) = 0)$;

(2) 对 $\forall\, A_1, A_2 \subset T$, 若 $A_1 A_2 = \varnothing$, 则对每一个样本有 $N(A_1 \bigcup A_2) = N(A_1) + N(A_2)$.

注 参数集 T 可以是 $\mathbb{R}^n$, 也可以是任意一抽象非空集.

泊松过程是简单的点过程.