第 1 讲 概率与随机变量

- 课程名：随机过程
- 任课老师：涂思铭
- 课程主页：http://39.107.124.3/teaching/stochastic2025/
- 教材：何书元. 随机过程，北京大学出版社
- 评分规则：平时作业 40% + 期末考试 60%

参考书籍

- 林元烈编著. 应用随机过程，清华大学出版社，2002.
- 张波，商豪，邓军编著. 应用随机过程，中国人民大学出版社，2023.
- 李增沪，张梅，何辉编著. 概率论（下册），高等教育出版社，2024.
- S. Ross. Introduction to Probability Models, thirteenth edition, Academic Press. (有中译本，人民邮电出版社)
- Richard Durrett. Essentials of Stochastic Processes, third edition, Springer. (有中译本，机械工业出版社)
- Robert G. Gallager. Stochastic Processes: Theory for Applications, Cambridge University Press.

本讲中，我们将回顾一些在概率与统计课程中已经介绍过，在本课程中经常会用到的概念. 通常把按照一定的想法去做的事情称为**试验**，把试验的可能结果称为**样本点**，称样本点的集合为**样本空间**. 对于一个特定的试验，以后总用 Ω 表示样本空间，用 ω 表示样本点，这时

$$\Omega = \{\omega | \omega\text{是试验的样本点}\}.$$

在概率论中，事件是样本空间 Ω 的子集. 在实际问题中人们往往并不需要关心 Ω 的所有子集，只要把关心的子集称为事件就够了. 但事件作为 Ω 的子集，必须满足以下三个条件:

(1) Ω 是事件;

(2) 若 A 是事件，则 A^c 是事件;

(3) 若 A_j 是事件，则 $\bigcup_{j=1}^{\infty} A_j$ 是事件.

对于事件 A，如果用 $P(A)$ 表示事件 A 发生的概率，则 $P(A)$ 满足以下条件：

(1) **非负性**：对任何事件 A，$P(A) \geq 0$；

(2) **完全性**：$P(\Omega) = 1$；

(3) **可列可加性**：对于互不相容的事件 $A_1, A_2, \cdots$，有

$$P(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j).$$

设 Ω 是一个样本空间，用 $\mathcal{F}$ 表示全体事件，P 表示一个概率，则称 $(\Omega, \mathcal{F}, P)$ 为一个**概率空间**.

下面是概率 P 的基本性质:

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

(2) **有限可加性**：如果 $A_1, A_2, \cdots, A_n$ 互不相容，则

$$P(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j).$$

(3) **单调性**：如果 $B \subset A$，则

$$P(A) - P(B) = P(A \backslash B) \geq 0.$$

(4) **加法公式**：$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(5) **次可加性**：$P(\bigcup_{j=1}^{n} A_j) \leq \sum_{j=1}^{n} P(A_j)$.

定义 1.1

设 A 和 B 是两个事件. 如果 $P(B)>0$，则在给定事件 B 发生的条件下事件 A 的**条件概率**为

$$P(A|B)=\frac{P(A\cap B)}{P(B)}. \tag{1.1}$$

定义 1.2

称事件 A 与 B **独立**，如果 $P(AB)=P(A)P(B)$.

定理 1.3

对已知的正概率事件 A，定义条件概率 $P_A(B) = P(B|A)$，$B \in \mathcal{F}$，则 P_A 是概率，$(\Omega, \mathcal{F}, P_A)$ 是概率空间. 当 $P(A \cap B) > 0$ 时，有

$$P_A(C|B) = P(C|A \cap B).$$

定理 1.4: 乘法公式

我们有

$$P(B_1B_2\cdots B_n) = P(B_1)P(B_2|B_1)\cdots P(B_n|B_1B_2\cdots B_{n-1}).$$

当 $P(A) > 0$ 时，我们有

$$P(B_1B_2\cdots B_n|A) = P(B_1|A)P(B_2|B_1A)\cdots P(B_n|B_1B_2\cdots B_{n-1}A).$$

定理 1.5: 全概率公式

如果事件 $A_1, A_2, \cdots$ 互不相容且概率均为正，则当 $B \subset \bigcup_{j=1}^{\infty} A_j$ 或者 $\bigcup_{j=1}^{\infty} A_j = \Omega$ 时，有

$$P(B) = \sum_{j=1}^{\infty} P(A_j) P(B|A_j),$$

$$P(B|A) = \sum_{j=1}^{\infty} P(A_j|A) P(B|AA_j), \text{当 } P(A) > 0 \text{ 时.}$$

定理 1.6

设 $\{B_1, B_2, \cdots\} \subseteq \mathscr{F}$ 为 Ω 的一个划分, 而 $A \in \mathscr{F}$ 满足 $P(A) > 0$. 则对任意 $n \geq 1$ 有

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{P(A)} = \frac{P(B_n)P(A|B_n)}{\sum_{k=1}^{\infty} P(B_k)P(A|B_k)}.$$

定理 1.7: 概率的连续性

如果 $A_1 \subset A_2 \subset \cdots$, $B_1 \supset B_2 \supset \cdots$，则有

$$P(\bigcup_{j=1}^{\infty} A_j) = \lim_{n\to\infty} P(A_n), \quad P(\bigcap_{j=1}^{\infty} B_j) = \lim_{n\to\infty} P(B_n).$$

我们记

$$P(A_n \text{ i.o.}) = P(\text{有无穷多个 } A_j \text{ 发生}).$$

定理 1.8: Borel-Cantelli 引理

(1) 如果 $\sum_{j=1}^{\infty} P(A_j) < \infty$，则

$$P(A_n \text{ i.o.}) = 0.$$

(2) 如果 $\sum_{j=1}^{\infty} P(A_j) = \infty$ 且 $A_1, A_2, \cdots$ 相互独立, 则 $P(A_n \text{ i.o.}) = 1$.

推论 1.9: 0-1 律

若 $A_1, A_2, \cdots$ 相互独立, 则 $P(A_n \text{ i.o.})$ 非 0 即 1.

随机变量 X 是定义在样本空间 Ω 上的函数，使得对于 $\mathbb{R}=(-\infty,\infty)$ 的子集 A，$\{X\in A\}$ 是事件.
对于随机变量 X，称 $F(t)=P(X\le t)$ 为 X 的**分布函数**. 分布函数是单调不减的右连续函数. 用 $F(t-)$ 表示 F 在 t 的左极限，有

$$P(X=t)=F(t)-F(t-),\ t\in(-\infty,\infty).$$

称 $\overline{F}(t)=P(X>t)$ 为 X 的**生存函数**. 于是

$$\overline{F}(t)=1-F(t)=P(X>t).$$

如果 $F(t)$ 是 X 的分布函数，若非负函数 $f(s)$ 使得对所有的 t，有

$$F(t) = \int_{-\infty}^{t} f(s)ds,$$

则称 $f(s)$ 为 F 或 X 的**密度函数**，称 X 是**连续型随机变量**. 这时对于 $(-\infty, \infty)$ 的子集 A，有

$$P(X \in A) = \int_A f(s)ds.$$

如果 $X_1, X_2, \cdots, X_n$ 都是随机变量，则称 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 是**随机向量**. 我们称 $\mathbb{R}^n$ 上的 n 元函数

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

是 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 的**分布函数**. 如果

$$F(x_1, \cdots, x_n) = F_{X_1}(x_2) F_{X_2}(x_2) \cdots F_{X_n}(x_n),$$

其中

$$F_{X_i}(x_i) = \lim_{x_j \to \infty, j \neq i} F(x_1, \cdots, x_n).$$

则称这 n 个随机变量是**独立**的.

如果 $X_1, X_2, \cdots, X_n$ 都是随机变量，则随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 是定义在 Ω 上的一个多元函数. 对每个 $\omega \in \Omega$,

$$\boldsymbol{X}(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$$

是实数向量，称为 $\boldsymbol{X}$ 的**一次观测或一次实现**.

对于随机向量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$，如果有 $\mathbb{R}^n$ 上的非负函数 $f(\boldsymbol{x}) = f(X_1, x_2, \cdots, x_n)$，使得对 $\mathbb{R}^n$ 的任何子立方体

$$D = \{(x_1, x_2, x_3, \cdots, x_n) | a_i < x_i \le b_i, 1 \le i \le n\},$$

有

$$P(\boldsymbol{x} \in D) = \int_D f(\boldsymbol{x}) dx_1 \cdots dx_n,$$

则称 $\boldsymbol{X}$ 是**连续型随机向量**，称 $f(\boldsymbol{x})$ 是 $\boldsymbol{X}$ 的**联合密度**. 这时，可以证明对于 $\mathbb{R}^n$ 中的任何区域 D，上式成立.

定理 2.1

设 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 有联合分布函数 $F(\boldsymbol{x}) = F(x_1, x_2, \cdots, x_n)$. $F(\boldsymbol{x})$ 在 $\mathbb{R}^n$ 的开区域 D 中有连续的 n 阶混合偏导数. 定义

$$f(\boldsymbol{x}) = \begin{cases} \frac{\partial^n F(\boldsymbol{x})}{\partial x_n \cdots \partial x_2 \partial x_1}, & \boldsymbol{x} \in D, \\ 0, & \text{其它}. \end{cases}$$

若下面的条件 (a), (b) 之一成立:

(a) $P(\boldsymbol{X} \in D) = 1$;

(b) $\int_D f(\boldsymbol{x}) dx_1 dx_2 \cdots dx_n = 1$.

则 $f(\boldsymbol{x})$ 是 $\boldsymbol{X}$ 的联合密度.

对于随机向量 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$, $\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_m)$. 定义

$$(\boldsymbol{X},\boldsymbol{Y})=(X_1,X_2,\cdots,X_n,Y_1,Y_2,\cdots,Y_m).$$

以后用 $\boldsymbol{X}\sim f(\boldsymbol{x})$ 表示 $\boldsymbol{X}$ 有联合密度 $f(\boldsymbol{x})$，用 $\boldsymbol{Y}\sim g(\boldsymbol{y})$ 表示 $\boldsymbol{Y}$ 有联合密度 $g(\boldsymbol{y})$.

定理 2.2

设 $\boldsymbol{X} \sim f(\boldsymbol{x})$，则

(1) X_j 有密度函数

$$f_j(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x}) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_n,$$

并且 $X_1, X_2, \cdots, X_n$ 相互独立的充分必要条件是

$$f(\boldsymbol{x}) = \prod_{j=1}^{n} f_j(x_j), \boldsymbol{x} \in \mathbb{R}^n;$$

(2) 当 $\boldsymbol{Y} \sim g(\boldsymbol{y})$ 时，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 相互独立的充分必要条件是

$$(\boldsymbol{X}, \boldsymbol{Y}) \sim f(\boldsymbol{x}) g(\boldsymbol{y});$$

(3) 当 $\boldsymbol{X}, \boldsymbol{Y}$ 都是离散随机向量时，$\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 相互独立的充分必要条件是对所有的 $\boldsymbol{x}$，$\boldsymbol{y}$, 有

$$P(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y}) = P(\boldsymbol{X} = \boldsymbol{x}) P(\boldsymbol{Y} = \boldsymbol{y}).$$

定理 2.3: Fubini 定理

设 D 是 $\mathbb{R}^n$ 上的区域，$g(x_1, x_2, \cdots, x_n)$ 是 D 上的非负函数或是满足条件

$$\int_D |g(x, x_2, \ldots, x_n)| dx_1 x_2 \cdots dx_n < \infty$$

的函数，则对区域 D 上的 n 重积分

$$\int_D g(x, x_2, \ldots, x_n) dx_1 x_2 \cdots dx_n$$

可以进行累次积分计算，且积分的次序可以交换.

引理 2.4

设 $\boldsymbol{S}=(S_1,S_2,\cdots,S_n)$ 有联合密度 $g(\boldsymbol{x})$，$\boldsymbol{X}=(u_1(\boldsymbol{S}),u_2(\boldsymbol{S}),\cdots,u_n(\boldsymbol{S}))$ 是 $\boldsymbol{S}$ 的函数，D 是 $\mathbb{R}^n$ 的区域使得 $P(\boldsymbol{X}\in D)=1$. 如果有 D 上的 n 维向量值函数 $\boldsymbol{s}(\boldsymbol{x})$，使得

(a) 对 $\boldsymbol{x}\in D$，有 $\{\boldsymbol{X}=\boldsymbol{x}\}=\{\boldsymbol{S}=\boldsymbol{s}(\boldsymbol{x})\}$；

(b) $\boldsymbol{s}(\boldsymbol{x})$ 是 D 到其值域的可逆映射，偏导数连续，雅可比行列式的绝对值

$$\left|\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{x}}\right|\neq 0,\ \boldsymbol{x}\in D,\ i=1,2,\cdots,m.$$

则 $\boldsymbol{X}$ 有联合密度

$$f(\boldsymbol{x})=\begin{cases}(g(\boldsymbol{s}(x))\left|\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{x}}\right|, & x\in D,\\ 0, & x\notin D.\end{cases}$$

定理 2.5

设数列 $\{a_j\}$ 绝对可和：$\sum_{j=0}^{\infty}|a_j| < \infty$，函数列 $\{h_j(s)\}$ 一致有界：$\sup_{a\le s\le b}|h_j(s)| \le M$. 对于 $c \in [a,b]$，如果 $\lim_{s\in(a,b),s\to c} h_j(s) = h_j$，则

$$\lim_{s\in(a,b),s\to c} \sum_{j=0}^{\infty} a_j h_j(s) = \sum_{j=0}^{\infty} a_j h_j.$$

在统计学中，我们把所要调查对象的全体叫做总体，把总体中的成员叫做个体. 当我们关心总体的某个指标时，就称这个指标为参数.

当 $y_1, y_2, \cdots, y_N$ 是总体的全部个体时，总体均值是

$$\mu = \frac{y_1 + y_2 + \cdots + y_N}{N},$$

总体方差为

$$\sigma^2 = \frac{(y_1 - \mu)^2 + (y_2 - \mu)^2 + \cdots + (y_N - \mu)^2}{N}.$$

总体标准差是总体方差的开平方 $\sigma = \sqrt{\sigma^2}$. 总体均值、总体方差和总体标准差都是参数.

当 X 是从总体中随机抽样得到的个体时，X 是随机变量，X 的分布就是总体的分布. 如果对总体进行有放回的抽样，则得到独立同分布的，且和 X 同分布的随机变量 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本.

在进行统计分析时，为了强调 $X_1, X_2, \cdots, X_n$ 是随机变量，也称 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的随机变量.

例 4.1

如果 X 只取值 0 或 1，概率分布是

$$P(X=1)=p,\ P(X=0)=q,\ p+q=1,$$

则称 X 服从两点分布，或 *Bernoulli* 分布. 记做 $X\sim\mathcal{B}(1,p)$.

例 4.2

设某试验成功的概率为 p，$q = 1 - p$. 将该试验独立重复 n 次时，用 X 表示成功的次数，则 X 的概率分布为

$$P(X = k) = C_n^k p^k q^{n-k}, \; k = 0, 1, \cdots, n.$$

此时称 X 服从二项分布，记做 $X \sim \mathcal{B}(n, p)$.

例 4.3

甲向一个目标独立重复射击，每次击中目标的概率是 $p = 1 - q > 0$. 用 X 表示其首次击中目标的射击次数，则 X 的概率分布为

$$P(X = k) = q^{k-1}p,\ \ k = 1, 2, \cdots, \tag{4.1}$$

这时称 X 服从几何分布，记为 $X \sim \text{Geom}(p)$.

例 4.4

设 $X_1, X_2, \cdots, X_i$ 独立同分布，有共同的几何分布 (4.1). 将 $S_i = X_1 + X_2 + \cdots + X_i$ 视为第 i 次击中目标时的射击次数，称其服从负二项分布：

$$P(S_i = j) = C_{j-1}^{i-1} q^{j-i} p^i, \ j \geq i.$$

例 4.5

如果随机变量 X 有概率分布

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

则称 X 服从参数是 λ 的泊松分布，简记为 $X\sim P(\lambda)$，这里 λ 是正常数.

例 4.6

设 X 是取值于 $[a,b]$ 上的连续型随机变量，我们称 X 服从 $[a,b]$ 上的均匀分布，记为 $X \sim U[a,b]$，如果其概率密度函数为

$$f(x) = (b-a)^{-1}, a \leq x \leq b.$$

例 4.7

称连续型随机变量 X 服从参数为 λ 的指数分布，记做 $X \sim \mathcal{E}(\lambda)$，如果 X 有密度函数

$$f(t) = \lambda e^{-\lambda t}, \ t > 0,$$

例 4.8

称连续型随机变量 X 服从均值 μ 和方差 σ^2 的正态分布，如果其概率密度函数为

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

我们记为 $X \sim N(\mu, \sigma^2)$. 特别地，如果 $X \sim N(0,1)$，则称 X 服从**标准正态分布**. 设 $\boldsymbol{\mu}$ 是 n 维常数列向量，B 是 $n \times m$ 常数矩阵，$\epsilon_1, \epsilon_2, \cdots, \epsilon_m$ 是来自总体 $N(0,1)$ 的随机变量，$\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_m)^T$. 如果

$$\boldsymbol{X} = \boldsymbol{\mu} + B\boldsymbol{\epsilon},$$

则称 $\boldsymbol{X}$ 服从 n 元正态分布，记做 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$. 其中 $\Sigma = BB^T$ 是 $\boldsymbol{X}$ 的协方差矩阵 (我们将在下一讲中回顾).

命题 4.9

$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^T \sim N(\boldsymbol{\mu}, \Sigma)$ 的充分必要条件是对任何常数 $a_1, a_2, \cdots, a_n$，线性组合 $\sum_{j=1} a_j X_j$ 服从正态分布.

命题 4.10

当 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$ 时，其分量 $X_1, X_2, \cdots, X_n$ 相互独立的充分必要条件是他们互不相关，即 $\mathrm{Cov}(X_i, X_j) = 0$，$j \neq i$.

定理 5.1: 马尔可夫不等式

对随机变量 X 和常数 $\epsilon > 0$，有

$$P(|X| \geq \epsilon) \leq \frac{1}{\epsilon^{\alpha}} E[|X|^{\alpha}], \ \alpha > 0.$$

推论 5.2: 切比雪夫不等式

对随机变量 X 和常数 $\epsilon > 0$，有

$$P(|X - E[X]| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathrm{Var}(X).$$

证明.

在马尔可夫不等式中取 $\alpha = 2$ 并用 $X - E[X]$ 代替 X 可得. □

定理 5.3

设 $E[X^2] < \infty$，$E[Y^2] < \infty$，则有

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}.$$

并且上面不等式中的等号成立的充分必要条件是有不全为零的常数 a, b，使得 $aX + bY = 0$ a.s..

定理 5.4

若 f 是凸函数，则只要期望存在，就有

$$E[f(X)] \geq f(E[X]).$$