概率论与数理统计

boatchanting

1 随机事件与概率

1.1 随机事件及其运算

事件的定义

在一次试验中，可能出现的结果称为基本事件，所有基本事件的集合称为样本空间，记作 Ω。

样本空间的子集称为事件。

随机事件

试验中每次可能发生或不发生的事件称为随机事件，用 A, B, C 等表示。

事件的运算

- 事件 A 和 B 的**并事件** $A \cup B$ 表示 A 或 B 发生。
- 事件 A 和 B 的**交事件** $A \cap B$ 表示 A 和 B 同时发生。
- 事件 A 的**补事件** $\bar{A}$ 表示 A 不发生。

1.2 概率的公理化定义与概率的性质

概率的公理化定义

设 Ω 为样本空间，其子集即为随机事件。概率是一个定义在事件上的函数 $P(\cdot)$，满足以下三条公理：

1. **非负性**：对任何事件 A，有 $P(A) \geq 0$。
2. **规范性**：$P(\Omega) = 1$。
3. **可加性**：若 A 和 B 为互不相容事件，则 $P(A \cup B) = P(A) + P(B)$。

概率的性质

对于任意两个事件 A 和 B，概率 P 满足以下性质：

- $P(\bar{A}) = 1 - P(A)$。
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$。
- 若 $A \subset B$，则 $P(A) \leq P(B)$。

1.3 条件概率与事件的独立性

条件概率

给定事件 B 且 $P(B) > 0$，事件 A 在 B 发生条件下的**条件概率**定义为：

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

事件的独立性

如果两个事件 A 和 B 满足 $P(A \cap B) = P(A) \cdot P(B)$，则称 A 和 B 是**独立事件**。

1.4 全概率公式与贝叶斯公式

全概率公式

设 $B_1, B_2, \ldots, B_n$ 为一组两两互不相容且穷尽样本空间 Ω 的事件，且 $P(B_i) > 0$。对于任意事件 A，有：

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

贝叶斯公式

在全概率公式的条件下，对于事件 A 和 B_i（$i = 1, 2, \ldots, n$），有：

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

2 离散型随机变量

2.1 随机变量

随机变量

设 Ω 为样本空间。若定义一个从样本空间到实数集合的映射 $X : \Omega \to \mathbb{R}$，则称 X 为一个**随机变量**。即每个样本点 $\omega \in \Omega$ 通过 X 映射为一个实数 $X(\omega)$。

2.2 一维离散型随机变量

离散型随机变量

如果随机变量 X 的取值是有限个或可列无限个，则称 X 为**离散型随机变量**。

概率质量函数

设 X 是一个离散型随机变量，且可能的取值为 $x_1, x_2, \ldots$。定义

$$p(x_i) = P(X = x_i), \quad i = 1, 2, \ldots$$

则函数 $p(x)$ 称为 X 的**概率质量函数**，满足

$$\sum_i p(x_i) = 1, \quad p(x_i) \geq 0.$$

2.3 离散型随机变量的数学期望与方差

数学期望

设 X 是一个离散型随机变量，其概率质量函数为 $p(x)$。则 X 的**数学期望** $E(X)$ 定义为：

$$E(X) = \sum_i x_i p(x_i).$$

数学期望可以理解为 X 的平均值或重心。

方差

设 X 的数学期望为 $E(X) = \mu$，则 X 的**方差** $\mathrm{Var}(X)$ 定义为：

$$\mathrm{Var}(X) = E\left[(X - \mu)^2\right] = \sum_i (x_i - \mu)^2 p(x_i).$$

方差表示随机变量取值的离散程度。

2.4 常用离散型随机变量及其分布

二项分布

若随机变量 X 表示 n 次独立重复试验中事件 A 发生的次数，每次试验中事件 A 发生的概率为 p，则 X 服从**二项分布**，记为 $X \sim B(n,p)$，其概率质量函数为：

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\ldots,n.$$

泊松分布

若随机变量 X 表示在单位时间内发生事件的次数，且事件发生服从平均为 λ 的泊松过程，则 X 服从**泊松分布**，记为 $X \sim \text{Poisson}(\lambda)$，其概率质量函数为：

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

几何分布

若随机变量 X 表示在一次成功前需要进行的独立重复试验次数，单次试验成功的概率为 p，则 X 服从**几何分布**，其概率质量函数为：

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

超几何分布

若随机变量 X 表示从包含 N 个个体的总体中抽取 n 个个体，其中有 M 个成功个体，则 X 表示成功个体数，且 X 服从**超几何分布**，其概率质量函数为：

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k=\max(0,n-(N-M)),\ldots,\min(n,M).$$

2.5 离散型随机变量函数的分布律

离散型随机变量函数的分布律

设 X 是离散型随机变量，函数 $g(X)$ 定义了 X 的一个新变量 $Y = g(X)$。若已知 X 的概率质量函数 $p(x)$，则 Y 的概率质量函数为：

$$P(Y = y) = \sum_{x:g(x)=y} p(x).$$

这种方法用于求离散型随机变量函数的分布。

3 连续型随机变量

3.1 随机变量的分布函数

分布函数

设 X 为随机变量，其分布函数（或称为累积分布函数）定义为：

$$F_X(x) = P(X \leq x).$$

分布函数 $F_X(x)$ 满足以下性质：

1. $0 \leq F_X(x) \leq 1$；
2. $F_X(x)$ 为非减函数；
3. $\lim_{x\to-\infty} F_X(x) = 0$ 且 $\lim_{x\to+\infty} F_X(x) = 1$。

3.2 连续型随机变量及其密度函数

连续型随机变量

若随机变量 X 的分布函数 $F_X(x)$ 是一个处处可导的连续函数，则称 X 为**连续型随机变量**。

概率密度函数

对于连续型随机变量 X，存在一个非负函数 $f_X(x)$，使得对于任意 $a \le b$，有

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx.$$

该函数 $f_X(x)$ 称为 X 的**概率密度函数**，且满足

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

3.3 连续型随机变量的数学期望与方差

数学期望

设 X 是一个连续型随机变量，概率密度函数为 $f_X(x)$。则 X 的**数学期望** $E(X)$ 定义为：

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx.$$

方差

设 X 的数学期望为 $E(X) = \mu$，则 X 的**方差** $\mathrm{Var}(X)$ 定义为：

$$\mathrm{Var}(X) = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\,dx.$$

3.4 常用连续型随机变量及其分布

均匀分布

若随机变量 X 在区间 $[a,b]$ 上均匀分布，则称 X 服从**均匀分布**，记为 $X \sim U(a,b)$，其概率密度函数为：

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

正态分布

若随机变量 X 的概率密度函数为

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

则称 X 服从**正态分布**，记为 $X \sim N(\mu, \sigma^2)$。其中，μ 为均值，σ^2 为方差。

指数分布

若随机变量 X 的概率密度函数为

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

则称 X 服从参数为 λ 的**指数分布**，记为 $X \sim \mathrm{Exp}(\lambda)$。

对数正态分布

若随机变量 $Y = \ln X$ 服从正态分布 $N(\mu, \sigma^2)$，则称 X 服从**对数正态分布**，记为 $X \sim \mathrm{Lognormal}(\mu, \sigma^2)$，其概率密度函数为：

$$f_X(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

3.5 连续型随机变量函数的分布

连续型随机变量函数的分布

设 X 是连续型随机变量，其概率密度函数为 $f_X(x)$，且定义 $Y = g(X)$。若 g 是单调可导函数，则 Y 的概率密度函数为：

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

该公式用于求连续型随机变量的函数分布。

3.6 其他数字特征

矩

设 X 是一个随机变量，若 $E(X^k)$ 存在，则称 $E(X^k)$ 为 X 的 k 阶矩。

中心矩

设 X 的数学期望为 μ，若 $E((X-\mu)^k)$ 存在，则称 $E((X-\mu)^k)$ 为 X 的 k 阶中心矩。

偏度与峰度

- **偏度**（Skewness）描述分布的对称性，偏度为 0 表示分布关于均值对称，偏度大于 0 表示分布右偏，偏度小于 0 表示分布左偏。

$$\text{Skewness} = \frac{E[(X-\mu)^3]}{\sigma^3}.$$

- **峰度**（Kurtosis）描述分布的陡峭程度，正态分布的峰度为 3，峰度大于 3 表示分布更集中，峰度小于 3 表示分布更平坦。

$$\text{Kurtosis} = \frac{E[(X-\mu)^4]}{\sigma^4}.$$

4 随机向量

4.1 二维随机变量及其联合分布

二维离散型随机变量及联合分布函数

设 (X,Y) 是二维离散型随机变量，其联合分布函数定义为：

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

联合分布函数 $F_{X,Y}(x,y)$ 满足：

1. $\lim_{x\to-\infty, y\to-\infty} F_{X,Y}(x,y) = 0$；
2. $\lim_{x\to+\infty, y\to+\infty} F_{X,Y}(x,y) = 1$；
3. $F_{X,Y}(x,y)$ 是非减的。

二维离散型随机变量的联合分布律

设 (X,Y) 是二维离散型随机变量，取值为 (x_i, y_j)。则其联合分布律为：

$$P(X = x_i, Y = y_j) = p(x_i, y_j),$$

满足

$$\sum_i \sum_j p(x_i, y_j) = 1.$$

二维连续型随机变量及其联合密度函数

若随机变量 (X,Y) 的联合分布函数 $F_{X,Y}(x,y)$ 处处可导，则存在联合密度函数 $f_{X,Y}(x,y)$ 使得：

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x\,\partial y},$$

且满足

$$\iint_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = 1.$$

4.2 边缘分布、随机变量的独立性和条件分布

边缘分布函数

设 (X, Y) 是二维随机变量，若 $F_{X,Y}(x, y)$ 为其联合分布函数，则 X 和 Y 的边缘分布函数分别定义为：

$$F_X(x) = \lim_{y\to+\infty} F_{X,Y}(x, y), \quad F_Y(y) = \lim_{x\to+\infty} F_{X,Y}(x, y).$$

边缘分布律和边缘密度函数

对于二维离散型随机变量 (X, Y)，X 的边缘分布律为：

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j).$$

对于二维连续型随机变量 (X, Y)，X 的边缘密度函数为：

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dy.$$

随机变量的相互独立性

如果对于任意的 x 和 y，有 $P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$，则称 X 和 Y 相互独立。对于连续型随机变量，若 $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$，则 X 和 Y 相互独立。

4.3 二维随机变量函数的分布

二维离散型随机变量函数的分布

设 (X, Y) 是二维离散型随机变量，$Z = g(X, Y)$ 是其函数，则 Z 的概率分布可以通过 X 和 Y 的联合分布律求出。

二维连续型随机变量函数的分布

设 (X, Y) 是二维连续型随机变量，$Z = g(X, Y)$ 是其函数。若 g 可逆且可导，则 Z 的概率密度可以通过变量变换公式得到。

4.4 随机向量的数字特征

二维随机变量函数的数学期望

设 (X, Y) 是二维随机变量，若 $g(X, Y)$ 是一个函数，则其数学期望定义为：

$$E[g(X,Y)] = \begin{cases} \sum_i \sum_j g(x_i, y_j) p(x_i, y_j), & \text{离散型随机变量}, \\ \iint_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy, & \text{连续型随机变量}. \end{cases}$$

协方差及相关系数

设 X 和 Y 的数学期望分别为 $E(X) = \mu_X$ 和 $E(Y) = \mu_Y$，则 X 和 Y 的**协方差** $\mathrm{Cov}(X, Y)$ 定义为：

$$\mathrm{Cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

X 和 Y 的**相关系数** $\rho_{X,Y}$ 定义为：

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

期望向量和协方差矩阵

若 $(X_1, X_2, \dots, X_n)$ 为 n 维随机向量，则其**期望向量** μ 定义为：

$$\mu = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix}.$$

其**协方差矩阵** Σ 定义为：

$$\Sigma = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}.$$

多维正态分布

若随机向量 $X = (X_1, X_2, \dots, X_n)^T$ 服从多维正态分布，记为 $X \sim N(\mu, \Sigma)$，其中 μ 为期望向量，Σ 为协方差矩阵。则其概率密度函数为：

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right).$$