概率统计总结*

易鹏

中山大学

内部版本号：V3.02.21(半正式版)

2020 年 8 月 4 日

*本笔记已经开源，可以免费下载，github 地址：https://github.com/Lovely-XPP/Notebook
gitee 分流地址：https://gitee.com/sysu_xpp/Notebook

目录

第 1 章　随机事件

1.1　随机事件

1.1.1　随机现象

确定性现象

在一定条件下必然出现的结果

随机现象

事先无法准确与之其结果的现象

1.1.2　随机现象的统计性规律

统计规律性

随机现象在大量重复出现时所表现出来的规律性.

随机试验

对随机现象的观察.

随机试验的特点

1. 可重复性
2. 可观察性
3. 随机性

1.1.3　样本空间

样本点

随机试验的每一个可能结果.

样本空间

样本点的全体.

1.1.4 随机事件

事件

实验结果具备的某一可观察的特征.

随机事件

在随机试验中可能发生也可能不发生.

必然事件

在试验中必然发生.

不可能事件

在试验中一定不发生.

基本事件

对应一个唯一的可能结果，即样本点.

1.1.5 事件的集合表示

1.1.6 事件建的关系和运算

事件的包含

A 发生必然导致 B 发生, 则称事件 B 包含事件 A, 记作 $B \supset A$ 或 $A \subset B$.

事件的相等

事件 A 包含事件 B，事件 B 也包含事件 A，则称事件 A 与 B 相等，记作 $A = B$.

事件的并（或和）

‘‘事件 A 与 B 至少有一个发生" 这一事件称为事件 A 和 B 的并（或和）, 记作 $A \cup B$ 或 $A + B$.

事件的交（或积）

‘‘事件 A 与 B 都发生" 这一事件称为事件 A 与 B 的交（或积）, 记作 $A \cap B$.

事件的差

‘‘事件 A 发生而 B 不发生" 这一事件称为事件 A 和 B 的差, 记作 $A - B$.

互不相容事件

若事件 A 与 B 不能同时发生, 也就是说 AB 时不可能事件, 即 $AB = \varnothing$, 则称事件 A 与 B 是不可能事件.

对立事件

‘‘事件 A 不发生" 这一事件称为事件 A 的对立事件, 记作 $\overline{A}$, 易见,$\overline{A} = \Omega - A$, 且 $\overline{(\overline{A})}$.

有限个事件的并与交

完备事件组

完备事件组设 $A_1, A_2, \cdots, A_n, \cdots$ 是有限或可数个事件, 如果其满足

(1) $A_i A_j = \varnothing, i \neq j, \quad i, j = 1, 2, \cdots$

(2) $\bigcup_i A_i = \Omega$

则称 $A_1, A_2, \cdots, A_n, \cdots$ 是一个完备事件组.

事件的关系与运算的文氏图

1.1.7 随机事件的运算律

求和运算

交换律

$$A \cup B = B \cup A \tag{1.1}$$

结合律

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C \tag{1.2}$$

求交运算

交换律

$$A \cap B = B \cap A \tag{1.3}$$

结合律

$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C \tag{1.4}$$

混合运算

第一分配律

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.5}$$

第二分配律

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{1.6}$$

求对立事件的运算

自反律

$$\overline{(\overline{A})} = A \tag{1.7}$$

求和及交事件的对立事件

第一对偶律

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \tag{1.8}$$

第二对偶律

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \tag{1.9}$$

1.2 随机事件的概率

1.2.1 概率及其频率解释

参见 P_9

1.2.2 从频率的性质看概率的性质

参见 P_{10}

1.2.3 概率的公理化定义

定义 1.2.1 概率公理化

设 Ω 是一个样本空间，定义在 Ω 的事件域 F 上的一个实值函数 $P(\cdot)$ 如果它满足下列三条公理：

1. $P(\Omega) = 1$
2. 对任意事件 A, 有 $P(A) \leq 0$
3. 对任意可数的两两不相容的事件 $A_1, A_2, \cdots, A_n, \cdots$, 有 $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

则称实值函数 $P(\cdot)$ 为 Ω 上的一个概率测度.

1.2.4 概率测度的性质

1. $P(\varnothing) = 0$
2. 有限可加性：$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$
3. $P(\overline{A}) = 1 - P(A)$
4. $P(A - B) = P(A) - P(AB) = P(B) - P(AB)$
5. $0 \leq P(A) \leq 1$
6. $P(A \cup B) = P(A) + P(B) - P(AB)$

1.3 古典概型与集合概型

1.3.1 古典概型

定义 1.3.1 古典概型

古典概型是满足下面两个假设条件的概率模型：

1. 随机试验只有有限个结果
2. 每一个可能记过发生的概率相同

所以，古典概型的概率测度可表述为：

$$P(A) = \frac{A\text{中的元素个数}}{\Omega\text{中的元素个数}} = \frac{\text{使}A\text{发生的基本事件数}}{\text{基本事件总数}} \tag{1.10}$$

1.3.2 几何概型

定义 1.3.2 几何概型

几何概型的概率测度可表述为

$$P(A) = \frac{S(A)}{S(\Omega)} \tag{1.11}$$

1.4 条件概率

1.4.1 条件概率的定义

定义 1.4.1 条件概率

给定概率空间 Ω, P，A, B 是其上的两个事件，且 $P(A) > 0$，则称 $P(B|A) = \dfrac{P(AB)}{P(A)}$ 为已知事件 A 发生的条件下，事件 B 发生的条件概率.

1.4.2 乘法公式

定理 1.4.1 乘法公式

乘法公式的两个形式：

$$P(AB) = P(A) \cdot P(B|A),\ P(A) > 0 \tag{1.12}$$

$$P(AB) = P(B) \cdot P(A|B),\ P(B) > 0 \tag{1.13}$$

1.4.3 全概率公式

定理 1.4.2 全概率公式

设 $\{A_i\}$ 是一列有限或可数无穷个两两不相容的非零概率事件，且 $\bigcup_i A_i = \Omega$，则对任意事件 $B, P(B) > 0$，有

$$P(B) = \sum_i P(A_i) \cdot P(B|A_i) \tag{1.14}$$

1.4.4 贝叶斯公式

定理 1.4.3 贝叶斯公式

设 $\{A_i\}$ 是一列有限或可数无穷个两两不相容的非零概率事件，且 $\bigcup_{i=1}^{\infty} A_i = \Omega$，则对任意事件 $B, P(B) > 0$，有

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_i) \cdot P(B|A_j)}{\sum\limits_j P(A_j) \cdot P(B|A_j)} \tag{1.15}$$

1.5 事件的独立性

1.5.1 两个事件的独立性

两个事件的独立性

如果 $P(AB) = P(A)P(B)$，则称 A 与 B 相互独立，简称 A 与 B 独立.

有限个事件的独立性

(1) 如果有 $n(n \le 2)$ 个事件：$A_i, A_2, \cdots, A_n$ 中任意两个使劲按均相互独立，即对任意 $1 \le i \le j \le n$，均有 $P(A_iA_j) = P(A_i)P(A_j)$，则称 n 个事件 $A_i, A_2, \cdots, A_n$ 两两独立.

(2) 设 $A_i, A_2, \cdots, A_n$ 为 $n(n \le 2)$ 个事件，如果对其中任何 $k(2 \le k \le n)$ 个事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ $(1 \le i_1 < i_2 < \cdots < i_k \le n)$，均有 $P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$，则称事件 $A_i, A_2, \cdots, A_n$ 为 $n(n \le 2)$ 相互独立.

1.5.2 相互独立性的性质

定理 1.5.1 相互独立性的性质

1. 如果 n 个事件 $A_1, A_2, , A_n$ 相互独立，则将其中任何 $m(1 \le m \le n)$ 个事件改为相应的对立事件，形成的新的 n 个事件仍然相互独立.

2. 如果 n 个事件 $A_1, A_2, , A_n$ 相互独立，则有

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} P\left(\overline{A_i}\right) = 1 - \prod_{i=1}^{n} \left[1 - P\left(A_i\right)\right] \tag{1.16}$$

1.5.3 伯努利概型

定义 1.5.1 伯努利概型

只有两个可能的结果的试验称为伯努利试验，一个伯努利试验独立重复 n 次形成的试验序列称为 n 重伯努利试验.

定理 1.5.2 伯努利定理

在一次试验中，事件 A 发生的概率为 $p(0 < p < 1)$，则在 n 重伯努利试验中，事件 A 恰好发生 k 次的概率 $b(k; n, p)$ 为

$$b(k; n, p) = C_n^k\, p^k\, q^{n-k} \tag{1.17}$$

其中，$q = 1 - p$.

在伯努利试验序列中，设每次试验中事件 A 发生的概率为 p，“事件 A 在第 k 次试验中才首次发生”$(k1)$ 这一事件的概率为

$$g(k, p) = p\, q^{k-1} \tag{1.18}$$

第 2 章　随机变量的分布与数字特征

2.1　随机变量及其分布

2.1.1　随机变量的概念

定义 2.1.1　随机变量

定义在概率空间 (Ω, P) 上，取值为实数的函数 $X = X(\omega)(\omega \in \Omega)$ 称为 (Ω, P) 上的一个随机变量.

2.1.2　离散型随机变量的概率分布

定义 2.1.2　离散型随机变量

设 X 是定义在概率空间 (Ω, P) 上的一个随机变量，如果 X 的全部可能取值只有有限个或可数无穷个，则称 X 是一个离散型随机变量.

定义 2.1.3　离散型随机变量的概率分布

设 X 是离散型随机变量，其全部可能取值为 $\{x_i, i = 1, 2, \cdots, \}$，记 $p(x_i) = PX = x_i, i = 1, 2, \cdots$，则称 $\{p(x_i), i = 1, 2, \cdots\}$ 为 X 的概率分布. .

2.1.3　分布函数

定义 2.1.4　分布函数

设 X 是一随机变量，则称函数 $F(x) = P\{X \le x\}, x \in (-\infty, +\infty)$ 为随机变量 X 的分布函数，记作 $X \sim F(x)$.

分布函数的性质

1. 单调性：若 $x_1 < x_2$，则 $F(x_2)F(x_1)$.
2. 归零性与归一性

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 \tag{2.1}$$

$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1 \tag{2.2}$$

3. 右连续性

$$F(x + 0) = F(x) \tag{2.3}$$

2.1.4　离散型随机变量的分布函数

一般不同题不同，通常是离散的点.

2.1.5　连续性随机变量及其概率密度

定义 2.1.5　概率密度

一个随机变量 X 称为连续型随机变量，如果存在一个非负可积函数 $f(x)$，使得 X 的分布函数

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} \mathrm{d}t \tag{2.4}$$

称 $f(x)$ 为 X 的概率密度函数，简称密度函数.

密度函数的性质　密度函数有以下两个性质：

$$f(x) \geq 0, x \in (-\infty, \infty) \tag{2.5}$$

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1 \tag{2.6}$$

2.2　随机变量的数字特征

2.2.1　数学期望

离散型随机变量的数学期望

若离散型随机变量 X 的可能值为 $x_i(i = 1, 2, \cdots)$, 其概率分布为 $P\{X = x_i\} = p_i, i = 1, 2, \cdots$, 如果 $\sum_{i=1}^{\infty} |x_i|\, p_i < \infty$，则称

$$E(X) = \sum_{i=1}^{\infty} x_i p_i \tag{2.7}$$

为随机变量 X 的数学期望（简称期望），也称为 X 的均值，记作 EX 或 $E(X)$.

连续型随机变量的数学期望

若 X 为连续型随机变量，$f(x)$ 为其密度函数，如果 $\int_{-\infty}^{+\infty} |x|\, f(x)\,\mathrm{d}x < \infty$，称

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x \tag{2.8}$$

为随机变量 X 的数学期望（简称期望），也称为 X 的均值，记作 EX 或 $E(X)$.

随机变量函数的数学期望

1. 若 X 为离散型随机变量，其概率分布为 $P\{X = x_i\} = p_i, i = 1, 2, \cdots$，如果 $\int_{-\infty}^{+\infty} |g(x_i)|\, f(x)\,\mathrm{d}x < \infty$ 时，则 $Eg(x)$ 存在，且

$$Eg(x) = \sum_{i=1}^{\infty} g(x_i) p_i \tag{2.9}$$

2. 若 X 为连续型随机变量，$f(x)$ 为其密度函数，如果 $\int_{-\infty}^{+\infty} |g(x)|\, f(x)\,\mathrm{d}x < \infty$，则 $Eg(x)$ 存在，且

$$E(X) = \int_{-\infty}^{+\infty} g(x) f(x)\,\mathrm{d}x \tag{2.10}$$

2.2.2 数学期望的性质

1. 对任意常数 a，有

$$Ea = a \tag{2.11}$$

2. 设 α_1, α_2 为任意实数，$g_1(x), g_2(x)$ 为任意实函数，如果 $Eg_1(X), Eg_2(X)$ 均存在，则

$$E[\,\alpha_1\, g_1(x) + \alpha_2\, g_2(x)\,] = \alpha_1 Eg_1(x) + \alpha_2 Eg_2(x) \tag{2.12}$$

3. 如果 EX 存在，则对任意实数 a，有

$$E(x+a) = EX + a \tag{2.13}$$

2.2.3 方差

离散型随机变量的方差

若 X 为离散型随机变量，其概率分布为 $P\{X = x_i\} = p_i, i = 1, 2, \cdots$，则

$$DX = E(X - EX)^2 = \sum_i (x_i - EX)^2 p_i \tag{2.14}$$

连续型随机变量的方差

若 X 为连续型随机变量，$f(x)$ 为其密度函数，则

$$DX = E(X - EX)^2 == \int_{-\infty}^{+\infty} (X - EX)^2 f(x)\,\mathrm{d}x \tag{2.15}$$

2.2.4 方差的性质

1. 对任意常数 a，有

$$Da = 0 \tag{2.16}$$

2. 对任意常数 a，有

$$D(x+a) = DX \tag{2.17}$$

3. 对任意常数 a，有

$$D(aX) = a^2 DX \tag{2.18}$$

4. 可以用以下公式简化方差的计算

$$DX = EX^2 - (EX)^2 \tag{2.19}$$

2.2.5 随机变量的矩与切比雪夫不等式

定义 2.2.1　k 阶原点矩

X 为一随机变量，k 为正整数，如果 EX^k 存在（即 $E|X|^k < -\infty$），则称 $E|X|^k$ 为 X 的 k 阶原点矩，称 $E|X|^k$ 为 X 的 k 阶绝对矩.

定义 2.2.2　k 阶原点矩

X 为一随机变量，k 为正整数，如果 EX^k 存在，则称 $E\,(E - EX)^k$ 为 X 的 k 阶中心矩，称 $E\,|E - EX|^k$ 为 X 的 k 阶绝对中心矩.

定理 2.2.1 马尔可夫不等式

设 X 的 k 阶矩存在（k 为正整数），即 $E|X|^k < \infty$，则对任意 $\varepsilon > 0$ 有

$$P\{|X-EX| \geq \varepsilon\} \leq \frac{E|X|^k}{\varepsilon^k} \tag{2.20}$$

定理 2.2.2 切比雪夫不等式

设 X 的方差存在，则对任意 $\varepsilon > 0$ 有

$$P\{\,|X-EX| \geq \varepsilon\,\} \leq \frac{DX}{\varepsilon^k} \tag{2.21}$$

2.3 常用的离散型分布

表 2.1: 常用的离散型分布及其数字特征

分布类型	分布律	数学期望	方差
退化分布	$P\{X=a\}=1$	$E(X)=a$	$D(X)=0$
两点分布	$P\{X=x_1\}=p$ $P\{X=x_2\}=1-p$	$E(X)=px_1+(1-p)x_2$	$D(X)=p(1-p)(x_1-x_2)^2$
$(0-1)$ 分布	$P\{X=1\}=p$ $P\{X=0\}=1-p$	$E(X)=p$	$D(X)=p(1-p)$
均匀分布	$P\{X=x_i\}=\frac{1}{n}$ $i=1,2,\cdots,n$	$E(X)=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$	$D(X)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$
二项分布 [1] $X\sim b(n,p)$	$P\{X=k\}=C_n^k\,p^k(1-p)^{n-k}$ $i=0,1,\cdots,n$	$E(X)=np$	$D(X)=np(1-p)$
几何分布 [2]	$P\{X=k\}=q^{k-1}p, k\leq 1$ $q^k p \overset{\text{def}}{=\!=} g(k,p)$	$E(X)=\frac{1}{p}$	$D(X)=\frac{p}{q^2}$
超几何分布	$P\{X=k\}=\frac{C_{N_1}^k C_{N_2}^{n-k}}{C_N^n}$ $0\leq k\leq n, N=N_1+N_2$	$E(X)=n\cdot\frac{N_1}{N}$	$D(X)=n\cdot\frac{N_1}{N}\cdot\frac{N_2}{N}\cdot\frac{N-n}{N-1}$
泊松分布 $X\sim P(\lambda)$	$P\{X=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$ $k=0,1,\cdots,\lambda>0$	$E(X)=\lambda$	$D(X)=\lambda$

[1] 二项分布中 $C_n^k\,p^k(1-p)^{n-k} \overset{\text{def}}{=\!=} b(k;n,p)$

[2] 几何分布的统计意义为: 在独立重复试验中，事件 A 发生的概率为 p，设 X 为直到 A 发生为止所进行的试验的次数.

定理 2.3.1 泊松定理

在 n 重伯努利试验中，事件 A 在每次试验中发生的概率为 p_n（注意这与试验的次数 n 有关），如果 $n\to\infty$

时,$np_n \to \lambda(\lambda > 0$且为常数), 则对任意给定的正整数 k, 都有

$$\lim_{n\to\infty} b(k; n, p_n) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \tag{2.22}$$

2.4 常用的连续型分布

表 2.2: 常用的连续型分布及其数字特征

分布类型	分布函数	密度函数	数学期望	方差
均匀分布 $X \sim U(a,b)$	$F(x)=\begin{cases}0,\ x\le b\\ \frac{x-a}{b-a}, x\in[a,b]\\ 1,\ x\ge a\end{cases}$	$f(x)=\begin{cases}\frac{1}{b-a},\ a\le x\le b\\ 0,\ \text{others}\end{cases}$	$E(X)=\frac{a+b}{2}$	$D(X)=\frac{(b-a)^2}{12}$
指数分布 $X \sim e[\lambda]$	$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x},\ x\ge 0\\ 0,\ x<0\end{cases}$	$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, x\ge 0\\ 0,\ x<0\end{cases}$	$E(X)=\frac{1}{\lambda}$	$D(X)=\frac{1}{\lambda^2}$
正态分布 $X \sim N(\mu,\sigma)$	$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$E(X)=\mu$	$D(X)=\sigma^2$
标准正态分布 $X \sim N(0,1)$	$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t$	$f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$	$E(X)=0$	$D(X)=1$

定理 2.4.1 标准正态分布的性质

设 $X \sim N(0,1)$, 设标准正态分布的分布函数为 $\Phi_0(x)$, 密度函数为 $\varphi_0(x)$, 则

$$\varphi_0(-x) = \varphi_0(x) \tag{2.23}$$

$$\Phi_0(x) + \Phi_0(-x) = 1 \tag{2.24}$$

定理 2.4.2 正态分布的性质

设 $X \sim N(\mu, \sigma^2)$, 标准正态分布的分布函数为 $\Phi_0(x)$, 密度函数为 $\varphi_0(x)$, 正态分布的分布函数为 $\Phi(x)$, 密度函数为 $\varphi(x)$, 则

1. 设 $Y = ax + b, a, b$ 为常数，且 $a \neq 0$，那么

$$Y \sim N(a\mu + b, a^2\sigma^2) \tag{2.25}$$

2. 正态分布的标准化 ξ

$$\xi = \frac{X-\mu}{\sigma} \sim N(0,1) \tag{2.26}$$

3. $X \sim N(\mu, \sigma^2)$ 的充要条件是存在一个随机变量 $\xi \sim N(0,1)$, 使得 $X = \sigma\xi + \mu$

4. 由正态分布的标准化,

$$\Phi(x) = \Phi_0\left(\frac{X-\mu}{\sigma}\right) \tag{2.27}$$

$$\varphi(x) = \frac{1}{\sigma}\varphi_0\left(\frac{X-\mu}{\sigma}\right) \tag{2.28}$$

2.5 随机变量函数的分布

随机变量的函数

若随机变量 Y 满足 $Y = g(x)$ 的形式, 则称随机变量 Y 是随机变量 X 的函数.

离散型随机变量的分布

连续型随机变量的分布

已知 X 的分布函数 $F_X(x)$ 或密度函数 $f_X(x)$, 为求 $Y = g(X)$ 的分布函数, 则

$$\begin{aligned} F_Y(x) &= P\{Y \le x\} \\ &= P\{g(X) \le x\} \\ &= P\{x \in C_x\}, \quad C_x = \{\, t | g(t) \le x \,\} \end{aligned} \tag{2.29}$$

χ^2 分布和对数正态分布

参见书 P_{72}.

第 3 章　随机向量

3.1　随机向量的分布

3.1.1　随机向量及其分布函数

定义 3.1.1　随机向量

设 $X_1, X_2, \cdots, X_n$ 是定义在概率空间 Ω, P 上的 n 个随机变量, 则称 $(X_1, X_2, \cdots, X_n)$ 是 (Ω, P) 上的一个 n 维随机向量.

定义 3.1.2　联合分布函数

设 $(X_1, X_2, \cdots, X_n)$ 是 (Ω, P) 上的一个 n 维随机向量, 则称 n 元函数

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\} \tag{3.1}$$

为随机向量 $(X_1, X_2, \cdots, X_n)$ 的分布函数或 n 个随机变量 $X_1, X_2, \cdots, X_n$ 的联合分布函数.
其中 $\{X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n\}$ 表示 $X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n$ 的交事件.

二维联合分布函数的性质

1. $0 \le F(x, y) \le 1$.
2. $F(x, y)$ 关于 x 和 y 均单调递增, 且右连续.
3. 归零性和归一性

$$F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0$$

$$F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0$$

$$F(-\infty, -\infty) = \lim_{(x,y) \to (-\infty, -\infty)} F(x, y) = 0$$

$$F(+\infty, +\infty) = \lim_{(x,y) \to (+\infty, +\infty)} F(x, y) = 1$$

定义 3.1.3　边缘分布函数

随机向量中分量各自的概率分布称为边缘分布函数, 其表达式为

$$\begin{aligned} F_X(x) &= P\{X \le x\} = P\{X \le x, Y < +\infty\} = F(x, +\infty) \\ F_Y(y) &= P\{Y \le y\} = P\{X < +\infty, Y \le y\} = F(+\infty, y) \end{aligned} \tag{3.2}$$

3.1.2 离散型随机向量的概率分布

离散型随机向量的概率分布

随机向量 (X, Y) 的概率分布 (X 和 Y 的联合概率分布) 为

$$P\{X = x_i, Y = y_i\} = p_{ij}, \ i, j = 1, 2, \cdots \tag{3.3}$$

离散型随机向量的概率分布的性质

1. $p_{ij} = 0, \ i, j = 1, 2 \cdots$
2. $\sum_i \sum_j p_{ij} = 1$

3.1.3 连续型随机向量的概率密度函数

连续型随机向量的概率密度函数

若 (X, Y) 是二维连续型随机变量,$f(x, y)$ 为 (X, Y) 的概率密度函数 (X 与 Y 的联合密度函数), 必满足

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s, t) \ \mathrm{d}s\mathrm{d}t \tag{3.4}$$

离散型随机向量的概率分布的性质

1. $f(x, y) > 0$
2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \ \mathrm{d}x\mathrm{d}y = 1$
3. 若 D 是平面上的一个区域, 则

$$P\{(X, Y) \in D\} = \iint_D f(x, y) \ \mathrm{d}x\mathrm{d}y$$

边缘密度函数

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) \, \mathrm{d}y \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) \, \mathrm{d}x \end{aligned} \tag{3.5}$$

例 3.1.1 二元正态分布

二元正态分布的概率密度函数如下:

$$\varphi(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]} \tag{3.6}$$

记做 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

3.2 条件分布与随机变量的独立性

3.2.1 条件分布与独立性的一般概念

定义 3.2.1 条件分布与独立性的一般概念

我们记条件概率 $P\{X \le x|A\}$ 为 $F(x|A), -\infty < x < +\infty$ 为在 A 发生的条件下 X 的条件分布函数

$$F(X|Y \le y) = \frac{P\{X \le x, Y \le y\}}{P\{Y \le y\}} = \frac{F(x,y)}{F_Y(y)} \tag{3.7}$$

若 $F(x,y) = F_X(x) \cdot F_Y(y)$, 则称随机变量 X 和 Y 相互独立.

3.2.2 离散型随机变量的条件概率分布与独立性

定义 3.2.2 离散型随机变量的条件概率分布与独立性

已知 $Y = y_i$ 的条件下 X 的条件概率分布为

$$P\{X = x_i|Y = y_i\} = \frac{P\{X \le x, Y \le y\}}{P\{Y \le y\}} = \frac{p_{ij}}{p_j^y} = p_{i|j},\ i = 1, 2, \cdots \tag{3.8}$$

X 与 Y 相互独立的充要条件是

$$p_{ij} = p_i^X \cdot p_j^Y,\ i, j = 1, 2, \cdots \tag{3.9}$$

3.2.3 连续型随机变量的条件概率密度函数与独立性

定义 3.2.3 连续型随机变量的条件概率密度函数与独立性

已知 $Y = y$ 的条件下 X 的条件概率密度函数记为 $f_{X|Y}(x|y)$, 由 $F_{X|Y}(x|y) = \int_{-\infty}^{x} \frac{f(u,v)}{f_Y(y)}\,\mathrm{d}u$ 可知其表达式为

$$f_{X|Y}(x|y) = \frac{x, y}{f_Y(y)} \tag{3.10}$$

进而可以得到密度函数的乘法公式

$$\begin{aligned} f(x,y) &= f_X(x) \cdot f_{Y|X}(x,y) \\ &= f_Y(y) \cdot f_{X|Y}(x,y) \end{aligned} \tag{3.11}$$

X 与 Y 相互独立的充要条件是

$$f(x,y) = f_X(x) \cdot f_Y(y) \tag{3.12}$$

3.3 随机向量的函数的分布与数学期望

离散型随机向量的函数的分布

设随机变量 $Z = g(X,Y)$, 则 Z 的概率分布为

$$F_Z(z) = P\{Z = z_k\} = P\{g(X,Y) = z_k\} = \sum_{g(x_i,y_j)=z_k} P\{X = x_i, Y = y_i\},\ k = 1, 2, \cdots \tag{3.13}$$

连续型随机向量的函数的分布

设随机变量 $Z = g(X, Y)$, 则 Z 的概率分布为

$$F_Z(z) = P\{Z \le z\} = P\{g(X,Y) \le z\} = P\{(X,Y) \in D_z\} = \iint\limits_{D_z} f(x,y)\,\mathrm{d}x\mathrm{d}y \tag{3.14}$$

其中,$D_z = \{(x,y)|g(x,y) \le z\}$. 特别地, 如果 $Z = X + Y$, 且 X 和 Y 是相互独立的随机变量, 则有

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)\,\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} f_X(z-y)\cdot f_Y(y)\,\mathrm{d}y \end{aligned} \tag{3.15}$$

上式分别称为函数 $f_X(x)$ 和 $f_Y(y)$ 的卷积.

3.3.1 随机向量的函数的数学期望

二维离散型随机向量的数学期望

设 (X, Y) 是二维离散型随机变量, 所以其数学期望为

$$EZ = Eg(X,Y) = \sum_{i,j} g(x_i, y_i)p_{ij} \tag{3.16}$$

二维连续型随机向量的数学期望

设 (X, Y) 是二维连续型随机变量, 所以其数学期望为

$$EZ = Eg(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,\mathrm{d}x\mathrm{d}y \tag{3.17}$$

数学期望的性质

1. $E(X+Y) = EX + EY$
2. X, Y 为任意两个相互独立的随机变量, 则 $EXY = EX \cdot EY$

3.4 随机向量的数字特征

3.4.1 协方差

定义 3.4.1 协方差

随机变量 X 与 Y 的协方差为

$$\begin{aligned} cov(X,Y) &= E[(X-EX)(Y-EY)] \\ &= E(XY) - EX\cdot EY \end{aligned} \tag{3.18}$$

协方差的性质

1. $cov(X, X) = DX$.
2. $cov(X, Y) = cov(Y, X)$.

3. $cov(aX,bY)=ab\cdot cov(X,Y)$, a,b 为任意常数.

4. $cov(C,X)=0$, C 为任意常数.

5. $cov(X_1+X_2Y)=cov(X_1,Y)+cov(X_2,Y)$.

6. X,Y 为任意两个相互独立的随机变量, 则 $cov(X,Y)=0$.

定理 3.4.1　方差与协方差

方差与协方差的关系为

$$D(X+Y)=D(X)+D(Y)+2cov(X,Y) \tag{3.19}$$

特别地, 如果 X 与 Y 相互独立, 则

$$D(X+Y)=D(X)+D(Y) \tag{3.20}$$

3.4.2　协方差矩阵

参见课本 P_{107}

3.4.3　相关系数

定义 3.4.2　相关系数

相关系数是研究变量之间线性相关程度的量, 用来度量两个变量间的线性关系, 其数学表达式为

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sqrt{DX}\cdot\sqrt{DY}},\quad |\rho_{X,Y}|\le 1 \tag{3.21}$$

$|\rho_{X,Y}|=1$ 的充要条件是 X 与 Y 具有线性关系, 即存在常数 $a\neq 0$ 及常数 b, 使得 $P\{Y=ax+b\}=1$.

当 $a>0$ 时,$\rho_{X,Y}=1$.

当 $a<0$ 时,$\rho_{X,Y}=-1$.

$\rho_{X,Y}=0$ 表明 X 与 Y 之间不存在线性联系,X 与 Y 不相关等价下列的任何一个条件:

1. $cov(X,Y)=0$.
2. $E(XY)=EX\cdot EY$.
3. $D(X+Y)=DX+DY$.

3.4.4　条件数学期望

离散性随机变量的条件数学期望

设 (X,Y) 为离散性随机变量, 则 X 在 $Y=y_i$ 条件下的条件数学期望记为 $E[X|Y=y_i]$, 其表达式为

$$E[X|Y=y_i]=\sum_i |x_i|\,p_{i|j} \tag{3.22}$$

连续性随机变量的条件数学期望

设 (X,Y) 为连续性随机变量, 则 X 在 $Y=y_i$ 条件下的条件数学期望记为 $E[X|Y=y_i]$, 其表达式为

$$E[X|Y=y_i]=\int_{-\infty}^{+\infty} x\,f_{X|Y}(x|y)\,\mathrm{d}x \tag{3.23}$$

条件数学期望的性质

1. $E[C|Y] = C, C$ 为任意常数.
2. $E[(k_1X_1 + k_2X_2)|Y] = k_1E[X_1|Y] + k_2E[X_2|Y], k_1, k_2$ 是常数.
3. X, Y 为任意两个相互独立的随机变量, 则 $E[X|Y] = EX$.
4. $g(x)$ 是一个任意函数, 则 $E[g(Y) \cdot X|Y] = g(Y)E[X|Y]$, 特别地, 有 $E[g(Y)|Y] = g(Y)$.
5. 全期望公式：$E(\,E[X|Y]\,) = EX$.

3.4.5 条件期望的预测含义

3.5 大数定律与中心极限定理

3.5.1 依概率收敛

定理 3.5.1 依概率收敛

设 $X, X_1, X_2, \cdots, X_n, \cdots$ 是一列随机变量, 如果对任意的 $\varepsilon > 0$, 恒有

$$\lim_{n\to\infty} \{\,|X_n - X| > \varepsilon\,\} = 0 \tag{3.24}$$

则称 $\{X_n\}$ 依概率收敛到 X, 记做 $X_n \xrightarrow{P} X$ 或 $P - \lim\limits_{n\to\infty} X_n = X$.

3.5.2 大数定律

定理 3.5.2 伯努利大数定律

设 μ_n 是 n 重伯努利试验中事件 A 的发生的次数, 已知在每次试验中 A 发生的概率为 $p(0 < p < 1)$, 则对任意的 $\varepsilon > 0$ 都有

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| > \varepsilon\right\} = 0 \tag{3.25}$$

即

$$\frac{\mu_n}{n} \xrightarrow{P} p \quad \text{或} \quad P - \lim_{n\to\infty} \frac{\mu_n}{n} = p$$

定理 3.5.3 切比雪夫大数定律

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是一列两两不相关的随机变量, 它们的数学期望 $E\xi_i$ 和方差 $D\xi_i$ 均存在, 且方差有界, 即存在常数 C, 使得 $D\xi_i \le C(i = 1, 2, \cdots)$, 则对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}E\xi_i < \varepsilon\right\} = 1 \tag{3.26}$$

定理 3.5.4 辛钦大数定律

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是一列相互独立同分布的随机变量, 且数学期望存在. 记 $E\xi_i = \mu$, 则有

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu < \varepsilon\right\} = 1 \tag{3.27}$$

3.5.3 中心极限定律

定理 3.5.5 林德伯格-列维定律

设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 是一列相互独立同分布的随机变量, 且 $E\xi_i = \mu, D(\xi_i) = \sigma^2 > 0, i = 1, 2, \cdots$, 则有

$$\lim_{n\to\infty} P\left\{ \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n}\xi_i - n\mu\right) \leq x \right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \tag{3.28}$$

或者写成

$$\frac{\sum_{i=1}^{n}\xi_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) \tag{3.29}$$

定理 3.5.6 棣莫弗-拉普拉斯中心极限定理

设 $X_n \sim b(n, p), 0 < p < 1$, 则

$$\lim_{n\to\infty} P\left\{ \frac{x_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \tag{3.30}$$

或者写成

$$\frac{x_n - np}{\sqrt{np(1-p)}} \sim N(0, 1) \tag{3.31}$$

索引

Q

S

T

W

X

Y

Z