

概率论与数理逻辑笔记整理V1.0

陈鸿峥

2018.08 *

目录

1	基本概念	1
1.1	事件与概率	1
1.2	条件概率	2
2	随机变量及其分布	3
2.1	常见的离散分布	4
2.2	常见的连续分布	5
3	大数定律	6
4	统计	6

1 基本概念

1.1 事件与概率

命题 1. 事件的基本运算

1. 分配律: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. 德摩根律: $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$

注意概率是对一个集合的函数, 有如下定义.

定义 1 (概率). E 为随机试验, S 是它的样本空间, 称 $\mathbb{P}(A)$ 为 E 中某一事件 A 的概率, 若集合函数 $\mathbb{P}(\cdot)$ 满足:

1. 非负性: 对每一个事件 A , $\mathbb{P}(A) \geq 0$
2. 规范性: 必然事件 B , $\mathbb{P}(B) = 1$
3. 可列可加性: A_1, A_2, \dots 两两不相容, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

*Build 20180803

由定义可得一些基本性质：

1. $\mathbb{P}(\emptyset) = 0, \mathbb{P}(A) \leq 1$
2. 有限可加性： A_1, A_2, \dots 两两不相容, $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$
3. 若 $A \subset B$, 则 $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A), \mathbb{P}(B) \geq \mathbb{P}(A)$
4. 逆事件概率： $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
5. 容斥原理： $\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{k=1}^n \left((-1)^{k-1} \sum_{\substack{I \subset \{1, \dots, n\} \\ |I|=k}} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) \right)$

定理 1 (实际推断原理). 概率很小的事件在一次试验中实际上几乎是不发生的

1.2 条件概率

定义 2 (条件概率). 设 A, B 为两个事件, 且 $\mathbb{P}(A) > 0$, 则称

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

为在事件 A 发生的条件下事件 B 发生的条件概率

定理 2 (乘法公式). 设 A_1, A_2, \dots, A_n 为 n 个事件, $n \geq 2$, 且 $\mathbb{P}(A_1 A_2 \cdots A_{n-1}) > 0$, 则

$$\mathbb{P}(A_1 A_2 \cdots A_n) = \mathbb{P}(A_n | A_1 A_2 \cdots A_{n-1}) \mathbb{P}(A_{n-1} | A_1 A_2 \cdots A_{n-2}) \cdots \mathbb{P}(A_2 | A_1) \mathbb{P}(A_1)$$

定义 3 (划分). 两两交为空, 所有并为全集

定理 3 (全概率公式). 设试验 E 的样本空间为 S , A 为 E 的事件, B_1, B_2, \dots, B_n 为 S 的一个划分, 且 $\mathbb{P}(B_i) > 0$, 则

$$\mathbb{P}(A) = \mathbb{P}(AB_1) + \mathbb{P}(AB_2) + \cdots + \mathbb{P}(AB_n) = \mathbb{P}(A | B_1) \mathbb{P}(B_1) + \mathbb{P}(A | B_2) + \cdots + \mathbb{P}(A | B_n) \mathbb{P}(B_n)$$

定理 4 (贝叶斯(Bayes)公式). 设试验 E 的样本空间为 S , A 为 E 的事件, B_1, B_2, \dots, B_n 为 S 的一个划分, 且 $\mathbb{P}(A) > 0, \mathbb{P}(B_i) > 0$, 则

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

特别地, 当 $n = 2$ 时有

$$\mathbb{P}(B | A) = \frac{\mathbb{P}(A | B) \mathbb{P}(B)}{\mathbb{P}(A | B) \mathbb{P}(B) + \mathbb{P}(A | \overline{B}) \mathbb{P}(\overline{B})}$$

定义 4 (独立性). 对于事件 A_1, \dots, A_n ,

- 若 $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \mathbb{P}(A_j), \forall i, j$, 则称 A_1, \dots, A_n 两两(pairwise)独立

- 若 $\mathbb{P}\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in I} \mathbb{P}(A_j), \forall I \in 2^{[n]}$, 其中 $2^{[n]}$ 为 $\{A_i\}_{i=1}^n$ 的所有子集, 则称 A_1, \dots, A_n 相互(mutually)独立

区分以下两个概念

1. A, B 对立(exclusive) $\Leftrightarrow \mathbb{P}(A \cap B) = 0$, 即不相交(disjoint)
2. A, B 独立(independent) $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, 即不相关(unrelated)

2 随机变量及其分布

定义 5 (概率质量函数(PMF)). 对于离散随机变量 X , $f_X(k) = \mathbb{P}(X = k)$

定义 6 (概率密度函数(PDF)). 对于连续随机变量 X , $f_X(k) = \mathbb{P}(X \in A) = \int_A f_X(x) dx$

定义 7 (分布函数/累积密度函数(CDF)). $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(z) dz$

定义 8 (期望).

$$\mathbb{E}(X) = \sum_k k f_X(k)$$

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

期望具有线性性, 即

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y), \mathbb{E}(cX) = c\mathbb{E}(X)$$

定义 9 (方差).

$$\text{Var}(X) = \sigma^2 = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$$

由方差定义和期望的线性性有

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

注意方差并不线性

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

若 X, Y 独立, 则 $\text{Cov}(X, Y) = 0$

定理 5. 若 X_1, \dots, X_n 独立, 则

$$\begin{aligned}\mathbb{E}\left(\prod_{i=1}^n X_i\right) &= \prod_{i=1}^n \mathbb{E}(X_i) \\ \text{Var}\left(\prod_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{Var}(X_i) \\ \text{Cov}(X_i, X_j) &= 0, i \neq j\end{aligned}$$

2.1 常见的离散分布

1. 伯努利分布 **Bernoulli(p)** (二项分布的特殊情形)

$$f_X(k) = \begin{cases} 1-p & k=0 \\ p & k=1 \end{cases}$$
$$\mathbb{E}(X) = p$$

2. 二项分布 **Binomial(n,p)**

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, 0 \leq k \leq n$$
$$\mathbb{E}(X) = n \cdot p$$

e.g. 扔 n 次硬币扔到 k 次正面 (做实验 n 次, 记录成功的次数)

3. 几何分布 **Geometric(p)** (负二项分布的特殊情形)

$$f_X(k) = (1-p)^{k-1} \cdot p, k \geq 1$$
$$\mathbb{E}(X) = \frac{1-p}{p}$$

e.g. 扔 k 次反面直至扔到正面 (做实验直到你成功, 记录失败的次数)

4. 负二项分布 **NegativeBinomial(t,p)**

$$f_X(k) = \binom{k+t-1}{t-1} p^t (1-p)^{k-t}, k \geq t$$
$$\mathbb{E}(X) = t \cdot \frac{1-p}{p}$$

e.g. 扔 k 次反面直到有 t 个正面 (做实验直到你获得 t 次成功, 记录失败次数)

5. 超几何分布 **HyperGeometric(N,n,M)**

$$f_X(k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}(X) = n \frac{M}{N}$$

e.g. M 个产品中有 N 个次品，检查 n 次得到 k 个次品

6. 泊松分布 **Poisson(λ)**

$$f_X(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$\mathbb{E}(X) = \lambda$$

$X \sim B(n, p)$, 若 $p = \frac{\lambda}{n}$, 且 n 非常大, 则

$$\begin{aligned} \mathbb{P}(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 + \frac{-\lambda}{n}\right)^{n-k} \\ &\approx \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

注意泊松分布具有无记忆性(memoryless), 即

$$\mathbb{P}(X \geq a \mid X \geq b) = \mathbb{P}(X \geq a - b)$$

2.2 常见的连续分布

1. 指数分布 **Exponential(λ)**

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$$

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

2. 正态分布 **Normal(μ, σ)**

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\mathbb{E}(X) = \mu$$

算平方

$$\begin{aligned}
 & \left(\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right)^2 \\
 &= \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy \\
 &= \int_0^{2\pi} \int_0^{+\infty} r e^{-\frac{r^2}{2}} dr d\theta \\
 &= \int_0^{2\pi} \int_0^{+\infty} e^{-u} du d\theta \\
 &= \int_0^{2\pi} 1 d\theta \\
 &= 2\pi
 \end{aligned}$$

3 大数定律

定理 6 (切比雪夫(Chebyshev)不等式).

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{\lambda^2}$$

定理 7 (弱大数定律). 若 X_1, X_2, \dots 为独立随机变量且同等分布, $\mathbb{E}(X_i) = \mu, \text{Var}(X_i) = \sigma^2$, 则

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| > \varepsilon\right) \rightarrow 0, \forall \varepsilon > 0$$

定理 8 (强大数定律). 若 X_1, X_2, \dots 为独立随机变量且同等分布, $\mathbb{E}(X_i) = \mu, \text{Var}(X_i) = \sigma^2$, 则

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mu\right) = 1$$

4 统计

定义 10 (估计). X_1, X_2, \dots, X_n 为独立随机变量, 从有参数 $\mu, \sigma, \theta, \dots$ 的分布 f 中得到, 对参数 θ 的估计是函数 $T(X_1, \dots, X_n)$, 称 T 是期望(*expected*)估计, 若

$$\mathbb{E}(T(X_1, \dots, X_n)) = \theta$$

合适(*probable*)的估计, 若

$$\mathbb{P}(|T(X_1, \dots, X_n) - \theta| > \varepsilon) \rightarrow 0, n \rightarrow \infty$$