

1

可能性为 0 的时间不一定时不可能事件，设想：在圆内找随机找到任意一个点的概率都为 0，但是不是不可能事件，
给定一个随机试验，对于这个试验的样本空间的描述不止一种。

2 条件概率与事件的独立性

定义：设 A, B 为两个事件，且 $P(A) > 0$ 则称

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

为事件 A 发生条件下 B 发生的条件概率

注意 $P(AB)$ 和 $P(B|A)$ 的差别。并且注意这两种情况的不同措辞。

计算方法：如果 A 发生之后，样本空间能够重新很清楚的得出，那么就用古典概型来计算。否则使用公式来进行计算（公式中的所有值都是依照于原来的样本空间来计算的）。条件概率同样适用于概率的性质，是已知原因寻求结果。

$$P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

上面的式子就是概率的乘法定理，用来计算复杂事件的概率，下面时推广。

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 \cdots A_{n-1})$$

波利亚罐子模型：两种球，取球时会加入某数量个同色球。这是一个简化的传染病模型。注意在这个模型中复杂事件中使用惩罚概率。

在描述事件的时候要尽量做到周全完善，比如「所取两件中至少有一件是不合格品」而不是「所取的两件中有一件是不合格品」。

2.1 事件的独立性

当 A, B 是随机事件的时候， $P(A) > 0$,

$$P(B) = P(B|A)$$

则称事件 B 独立于事件 A 。

独立性是相互的，及 A 独立于 B 则 B 也独立与 A ，以此为基础定义 A 和 B 相互独立，当且仅当以下条件满足：

$$P(AB) = P(A) \cdot P(B)$$

注意当 $P(A) = 0$ 的时候，则一定 A 和 B 相互独立。判断 A, B 事件是否相互独立，一定要按照定义进行计算。相互独立的本质是一个事件的发生对于另外一个事件发生的概率没有任何影响。

如果 A, B 相互独立，则 \bar{A} 与 B, A 与 \bar{B}, \bar{A} 和 \bar{B} 相互独立，可以通过这个互相判断是否事件是否相互独立。

注意「 A, B 相互独立」和「 A, B 互不相容」的区别：互不相容是集合关系，表示交集为空。相互独立表示的是一个数值相等的概率关系。

$$\text{if } P(A) > 0, P(B) > 0, AB = \phi$$

$$P(AB) = 0, P(A) \cdot P(B) \neq 0$$

2.1.1 多个事件独立性

n 个事件要相互独立，则要求所有的任意两个事件之间都相互独立。

2.2 全概率公式和 Bayes 公式

2.2.1 分割

定义：设 A_1, A_2, \dots, A_n 满足

$$A_i A_j = \phi, P(A_i) > 0, i \neq j$$

2.2.2 全概率公式

公式：设 A_1, A_2, \dots, A_n 是一个分割， B 为任意一个事件，则

$$B = \bigcup_{i=1}^n A_i B$$

显然 $A_i B$ 互不相容，于是

$$P(B) = P\left(\bigcup_{i=1}^n A_i B\right) = \sum_{i=1}^n P(A_i B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i)$$

全概率公式是一个概率加法定理和乘法定理的结合，可以将事件进行分解成简单概率的总和。全概率事件的实质是不同事件 A_1, A_2, \dots, A_n 导致了 B 事件发生的概率的综合。

在知道概率递推公式的情况下，可以使用数学归纳法取证明。如使用数学归纳法计算抽签模型的概率的时候：

设定事件 A_i 为「在第 i 次抽取到白球」，而数学归纳法假定 $P(A_i) = \frac{a}{a+b}$

$$P(A_{i+1}) = P(A_{i+1}|A_i) + P(A_{i+1}|\bar{A}_i)$$

$$P(A_{i+1}|A_i) = P(A_2|A_1), \quad P(A_{i+1}|\bar{A}_i) = P(A_2|\bar{A}_1)$$

2.3 Bayes 公式

设 A_1, A_2, \dots, A_n 是一个分割， B 为任意一个事件， $P(B) > 0$ ，则

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i) \cdot P(B|A_i)}$$

内涵：全概率公式的逆用，相当于分割其中的某一个事件相比于所有事件对于事件 B 的贡献的占比。表示结果一定的情况下，求导致原因的的概率。不用执着于直接看出来是 Bayes 公式，在计算条件概率的时候会自然想到全概率公式。在公式之中，我们认为 $P(A_i)$ 是验前概率， $P(A_i|B)$ 是验后概率，是对于新信息出现时对于概率的修正（例子：用 Bayes 公式解释狼来了）。

重要的例子：习题集 P29, 13; P27, 7;

2.3.1 贝努利模型

n 次相互独立的试验：对于随机试验 E ，如 E_1, E_2, \dots, E_n ，如哦任何一次试验结果的发生都不依赖于其他试验的结果。注意试验独立性和事件独立性的区别。

n 次重复独立试验：将随机试验 E 重复进行 n 次，如哦每次试验样本空间相同，有关事件的概率保持不变，且各次试验相互独立。

贝努利模型：设随机试验满足

1. 在相同条件下进行 n 次重复试验
2. 每次试验两种可能结果， A 和 \bar{A}
3. 在每次实验中 A 发生的概率不变 $P(A) = p$
4. 各次试验相互独立

则成为 n 重贝努利试验。

定理：对贝努利模型，每次试验 A 发生的概率是 p ，则 n 次试验中 A 恰好发生 K 次的概率为

$$B(k, n, p) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$$

$$\sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = 1$$

例题：书 p29, 例 19, 例 20;

小概率事件：在一次实验中发生的概率很小的事件，称为小概率事件。一般取在 0.01 或者 0.0 以下的值。

小概率原理：小概率事件在一次实验中几乎不可能发生，而某一概率很小的事件，居然在某一次实验发生了，那么要怀疑原来概率的正确性。

2.3.2 章末习题

P31, 9;11

数学归纳法的一般形式和使用数学归纳法证明抽签模型类似，在证明 $n+1$ 次的概率时，可以化为在第一次实验已经展开的情况下的 $n+1$ 的概率，这时候相对于第一次已经发生时 $n+1$ 时就被视为第 n 次，可以套入假设。

对于 n 次试验后的概率讨论，可以先从 1 开是找找规律。

习题集 P29 6, 13, 24

3 () 〰

3.1 随机变量及其分布

在实际问题中，随机的试验的结果可以用数量来表示，由此产生了随机变量的概念。有些试验，试验结果或直接为数值，或者直接数值化。这里引入实值函数的概念：从样本空间映射到实轴上。

定义：设随机试验的样本空间为 $\Omega = \{\omega\}$ ，如哦对于每一个 ω

对于随机变量的研究时包含了所有相对应的随机试验的研究。

随机变量的分布函数：设 X 是一个随机变量， x 是一个任意实数，概率 $P(X \leq x)$ 是 x 的函数，则称此函数为随机变量 X 的随机函数，记为 $F(x)$ ，即

$$F(x) = P(X \leq x)$$

使用 $X \leq x$ 是他能够比较好的表述各种其他的区间，而且可以只拥有一个自变量。对于其他区间：

$$P(X > x) = 1 - F(x)$$

$$P(a < X \leq b) = F(b) - F(a)$$

性质；

$$0 \leq F(x) \leq 1, \forall x \in R$$

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$$

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$$

用于解函数中的待定常数

$$x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$$

$$\lim_{x \rightarrow x_0^+} F(x) = F(x_0), \forall x_0 \in R$$

随机变量可以分为离散型随机变量或者连续型随机变量。

定义 1 若随机变量 X 的所有值为 x_1, x_2, \dots, x_k

$$P(X = x_k) = p_k, k = 1, 2, 3, \dots$$

则称上式为离散型随机变量 X 的概率分布或者分布律

$$p_k > 0$$

$$\sum_{k=1}^{\infty} p_k = 1$$

离散型随机变量 x 的概率分布和分布函数的关系

$$F(x) = \int_{-\infty}^x p(X = x)$$

3.2 常用离散型随机变量的分布

3.2.1 0-1 分布

$$P(X = k) = p^k(1-p)^{1-k}, k = 0, 1$$

3.2.2 二项分布

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \dots, n$$

记作:

$$X \sim B(n, p)$$

二项分布在 $n = 1$ 时退化为 0-1 分布。如果 $X \sim B(n, p)$ 则 $X = \sum_{i=1}^n X_i$, 其中它们互相独立, 且 $X \sim B(1, p)$

3.2.3 泊松分布

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

则称 X 服从参数为 λ 的泊松分布：

$$X \sim P(\lambda)$$

泊松分布最开始是对于二项分布的一个近似。若对于二项分布中的 n 充分大, p 充分小时, 则 $\lambda = np$ 。

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

3.2.4 几何分布

$$P(X = k) = (1-p)^{k-1} p, k = 1, 2, 3, \dots$$

记作：

$$X \sim G(p)$$

一般用于独立重复试验 A 首次发生时的概率。

3.3 连续型随机变量及其分布

定义：设变量 X 的分布函数为 $F(x)$, 如存在非负可积函数 $p(x)$, 满足：

$$F(x) = \int_{-\infty}^x p(t) dt$$

则称 X 为连续型随机变量, 其中 $p(x)$ 为概率密度函数, 简称密度函数。由上面的式子可以看出, 分布函数在整个区间上可导 (积分上限函数一定可导), 所以分布函数一定可积。概率密度函数的区间面积表示概率。

存在连续型随机变量既不是离散型随机变量, 也不是连续型随机变量。

$$\int_{-\infty}^{\infty} p(t) dt = 1$$

$$P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p(t) dt$$

$$P(X = a) = \lim_{h \rightarrow 0} \int_a^{a+h} p(t) dt = 0$$

判断一个分布是不是密度函数:

$$\int_{-\infty}^{\infty} p(t) dt = 1, \quad p(x) \geq 0 \text{ for } \forall x \in R$$

3.4 常用连续型随机变量的分布

3.4.1 均匀分布

$$p(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b) \end{cases}$$
$$X \sim U(a, b)$$

均匀分布中相同长度区间的概率一样，对应于之前的几何概型。

3.4.2 指数分布

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$
$$X \sim EXP(\lambda)$$

λ 越大，靠近原点的分布就越密集。拥有无记忆性

$$\begin{aligned} P(X > s+t | X > s) &= \frac{P(X > s+t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= P(X > t) \end{aligned}$$

无记忆性可以理解为：电子元件的寿命在符合指数分布的情况下，以上式子可以解释为在已经工作了 s 小时的情况下，再工作 t 小时的概率和新的电子原件工作 t 小时的概率一样，和已经正常工作的时间无关。几何概型中也类似满足这恶样的无记忆性。

3.4.3 正态分布

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in R, \sigma > 0$$
$$X \sim N(\mu, \sigma^2)$$

标准正态分布:

$$X \sim N(0, 1) \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$p(x)$ 关于 $x = \mu$ 对称, 顶点为 $(\mu, \sqrt{2\pi}\sigma)$; μ 会导致曲线平移, 被陈伟位置参数; σ 反映了取值分散性, 被称为形状参数。

正态分布于标准正态分布之间的关系, 可以将任意的正态分布转化为标准正态分布

$$p(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$$
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$X \sim N(\mu, \sigma^2), \quad Y = \frac{X-\mu}{\sigma} \Rightarrow Y \sim N(0, 1)$$

证明：

$$\begin{aligned}F_Y(y) &= P(Y \leq y) \\&= P\left(\frac{X - \mu}{\sigma} \leq y\right) \\&= P(X \leq \sigma y + \mu) \\&= F(\sigma y + \mu) \\&= \Phi(y)\end{aligned}$$

以上可以推出， $X \sim N(\mu, \sigma^2)$ ：

$$\begin{aligned}P(a < X \leq b) &= F(b) - F(a) \\&= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)\end{aligned}$$

标准正态分布的性质：

$$\Phi(-x) = 1 - \Phi(x)$$

这个时候负值就可由正值得到

3 σ 原则

$$P(|x - \mu| < \sigma) = 2\Phi(1) - 1 = 0.6826$$

$$P(|x - \mu| < 2\sigma) = 2\Phi(2) - 1 = 0.9546$$

$$P(|x - \mu| < 3\sigma) = 2\Phi(3) - 1 = 0.9974$$

所以我们常常只研究 $(\mu - 3\sigma, \mu + 3\sigma)$ 范围内的概率

3.5 随机变量函数的分布

$$P(X = x_k) = p_k, \quad Y = f(X)$$

Y 一定是一个离散型随机变量，确定 y 的取值然后的带等价的随机变量分布。

$$P(Y = y_j) = P(f(X) = y_j, X \in D)$$

计算的时候应该先算更加简单的分布；连续型随机变量可能会映射为离散型随机变量。连续型随机变量要先算出 Y 的分布函数，然后再算出概率密度函数。

当 $Y = f(X)$ 可以求出反函数的 $X = h(Y)$ ，则可以先求出密度函数：

$$p_Y(y) = p_X(h(y)) \cdot |h'(y)|$$

其中 $y \in (\alpha, \beta)$, $\alpha = \min(f(-\infty), f(+\infty))$, $\beta = \max(f(-\infty), f(+\infty))$

$$X \sim N(\mu, \sigma^2), \quad Y = aX + b \Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

3.5.1 习题课

书:p62 2; p64 34, 41; 判别一个函数是不是分布函数：单调不减， $F(-\infty) = 0, F(+\infty) = 1$ ，右连续，

习题集：p52, 6, 26, 27, 3, 12, 16, 27, 28, 31, 8, 27;

4 作业

p65, 42, 44, 48, 54