

概率论与数理统计

任课教师: 胡家信

清华大学数学科学系

2021年9月27日



上周内容总结

- 独立事件: $P(A \cap B) = P(A)P(B)$ 。
- 若干事件 $\{A_m\}_{m=1}^n$ 独立:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k})$$

对任意 $\{1, 2, \cdots, n\}$ 中的一组数 $\{i_1, i_2, \cdots, i_k\}$ 。

常见 $n = 3$: 三个事件 A_1, A_2, A_3 独立, 当且仅当下列均成立:

$$P(A_1 \cap A_2) = P(A_1) P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1) P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2) P(A_3)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3).$$

- 概率的计算: 抽样问题; 分配问题。

计算概率的更多例子

例子1: 标号为1至4的四张牌放在桌子上, 从4张牌中随机猜对至少1张牌的概率是多少?

提示:

- 仅猜对**1**张牌的概率;
- 仅猜对**2**张牌的概率;
- 仅猜对**3**张牌的概率;
- 仅猜对**4**张牌的概率。

以上事件的概率加起来即为所求。

讨论一下: 4张牌摆放的情况共有多少种? 仅猜对**第1**张牌的情况有多少种?

例1

解：4张牌有 $4! = 24$ 种不同的摆法。

(1) 先计算仅猜对**1**张牌的概率。

- 设某人仅猜对**第1**张牌，其余3张都猜错：

实际	abc	abc	abc	abc	abc	abc
猜测	abc	acb	bac	bca	cab	cba

此时，只有第4, 5两种情况均猜错，故总共有**2**种情况。

- 设某人仅猜对**第2**张牌，其余3张都猜错。同样，共有**2**种情况。
- 设某人仅猜对**第3**张牌或**第4**张牌，其余3张都猜错，分别有**2**种情况。

所以，仅猜对1张的情况共有 $2 * 4 = 8$ 种，故仅猜对**1**张牌的概率为

$$8/24 = 1/3.$$

例1

解：(2) 再计算仅猜对2张的概率。

设某人猜对第1, 2张牌, 其余2张都猜错:

实际	ab	ab
猜测	ab	ba

只有第2种情况猜错。同样, 设某人猜对第1, 3; 第1, 4; 第2, 3; 第2, 4; 第3, 4张牌, 其余2张都猜错, 均各出现1次。加起来, 仅猜对2张牌的情况共有 $6 * 1 = 6$ 种, 故仅猜对2张牌的概率

$$6/24 = 1/4.$$

例1

解：(3) 再计算猜对3张的概率。

若某人猜对第1, 2, 3张牌, 其余1张肯定也猜对, 故仅猜对3张的情况不存在, 有0种情况, 概率为0。

(4) 最后, 猜对4张牌的概率为 $1/24$ 。

答案：猜对至少1张牌的概率是

$$\frac{1}{3} + \frac{1}{4} + 0 + \frac{1}{24} = \frac{5}{8}.$$

例1

思考题：猜对至少2张的概率是多少？

例2

例：将 n 个球放在 n 个具有标号的箱子中，只有1只箱子是空的，问有多少种情况？

分两种情况：

- 球无标号。（试一试）
- 球有标号。

例2

(1) 球无标号。

- 从 n 个箱子中选1只箱子是空的，有 n 种情况。
- 剩下 $n-1$ 个箱子中，选1只箱子放2个球，有 $n-1$ 种情况。

故共有

$$n(n-1)$$

种情况。

例2

(2) 球有标号。

- 从 n 个箱子中选1只箱子是空的，有 n 种情况。
- 设第1个箱子是空的，剩下 $n-1$ 个箱子中，有1只箱子有2球。注意，两个球的组合有 $\binom{n}{2}$ 种情况，与其它 $n-2$ 个球一起，分别放在 $n-1$ 个箱子中，有 $(n-1)!$ 种放法，故有

$$\binom{n}{2}(n-1)!$$

种情况。

答案：总共有

$$n \times \binom{n}{2}(n-1)! = n(n-1) \frac{n!}{2}$$

种情况。

随机变量（在可数集上）

设 Ω 是一个可数（有限或者无限）样品空间

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}.$$

定义（随机变量）：定义域为 Ω 上的数值函数 X ，称为一个**随机变量**，记为 $X(\omega)$ 。

特点：

- X 是一个数值函数；
- 定义域是样品空间。

随机变量（在可数集上）

命题：设 X, Y 是可数样品空间 Ω 上的随机变量，则

$$X + Y, X - Y, XY, \frac{X}{Y} (Y \neq 0),$$

和 $aX + bY$ ，均是 Ω 上的随机变量，其中 a, b 是两个数。

证明：显然，

$$\omega \rightarrow X(\omega) + Y(\omega)$$

是 Ω 上的数值函数，故是一个随机变量。其它同理，证毕。

注意：在微积分中，函数 $f + g$ 定义在 $x \in D \subset \mathbb{R}$ （实数域）上，而这里 ω 定义在样品空间 Ω ，不一定属于实数域，如 $\Omega = \{\text{正面、反面}\}$ 。

随机变量（在可数集上）

命题： 设 φ 是一个普通二元函数， X, Y 是可数样品空间 Ω 上的随机变量，则

$$\omega \rightarrow \varphi(X(\omega), Y(\omega))$$

也是 Ω 上的随机变量。

证明： 显然。特例： $\varphi(x, y) = x + y$ 。

同理，若 X_i 是 Ω 上的随机变量，则

$$\omega \rightarrow X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)$$

也是 Ω 上的随机变量。

随机变量（例子）

例：设出版社出版1000（含）册以下图书的成本为3元，出版1001至5000（含）册图书的成本为2元，出版5000册以上图书的成本为1元，每册图书市场上卖5元。写出出版社的利润 X 。

注意：出版社要预测图书销售量，分批印刷图书，但至少印刷**1000**册图书。例如，可能开始印刷1000册，然后看销售量，再追加印刷500册等等，以达到利润最大化；也可能开始印刷1500册，但这里假设前面印刷的第1 - 1000册中每本的成本和后面的成本不一样。

随机变量（例子）

解：令 $\Omega = \{1, 2, \dots\}$ 为销售量，是不固定的（样本空间）。利润 X 为销售量 $\omega \in \Omega$ 的函数，是一个随机变量（实值函数）。

$$X(\omega) = \begin{cases} 5\omega - 3 * 1000, & \text{若 } \omega \leq 1000, \\ (5 - 3) * 1000 + 3(\omega - 1000), & \text{若 } 1000 < \omega \leq 5000, \\ 14000 + 4(\omega - 5000), & \text{若 } \omega > 5000. \end{cases}$$

注意： $14000 = (5 - 3) * 1000 + (5 - 2) * 4000$ 。

什么时候亏本？ $\{5\omega - 3000 < 0\} = \{\omega < 600\}$ 。

什么时候盈利10000元？

$$\{2000 + 3(\omega - 1000) > 10000\} \cup \{\omega > 5000\} = \{\omega \geq 3667\}.$$

注：实际上， X 是定义在 Ω 上的一个分段函数。

分布

设 A 是一个由某些实数构成的集合，定义

$$P(X \in A) = P(\{\omega : X(\omega) \in A\}).$$

注意: 集合 $\{\omega : X(\omega) \in A\}$ 是一个可数集, 属于 P 的定义域 (因为这里假设 Ω 是一个可数样本空间, 不可数样品空间后面讨论)。

例: $A = \{x\}$ 只含一点 x , 记

$$P(X = x) = P(X \in \{x\}).$$

定义(分布): 称函数

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

为随机变量 X 的分布函数。显然,

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

数学期望

设 X 是定义在可数样本空间 Ω 上的随机变量，其数学期望 $E(X)$ 定义为

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) \quad (\text{设级数收敛})$$

即取 X 每点的值，乘以该点的概率，然后相加，得到数学期望 $E(X)$ 。

设 $\Omega = \{\omega_1, \omega_2, \dots\}$ ，则

$$E(X) = \sum_{n=1}^{\infty} X(\omega_n)P(\{\omega_n\}) = \sum_n a_n P(X(\omega) = a_n).$$

数学期望

例：设 $\Omega = \{\omega_1, \omega_2, \dots, \omega_7\}$ 是某块土地分成7份的集合，其所占面积及价格如下：

5%	10%	10%	10%	15%	20%	30%
800元	900元	1000元	1200元	800元	900元	800元

设 $X(\omega)$ 表示某块地 ω 的价格，则其数学期望 $E(X)$ 为

$$\begin{aligned} E(X) &= (800) \frac{5}{100} + (900) \frac{10}{100} + (1000) \frac{10}{100} + (1200) \frac{10}{100} \\ &\quad + (800) \frac{15}{100} + (900) \frac{20}{100} + (800) \frac{30}{100} = 890, \end{aligned}$$

该数值数值表示这块土地的平均价格。

数学期望

若 X 是一个随机变量, 则 $\varphi(X)$ 也是一个随机变量, 其数学期望 $E(\varphi(X))$ 为

$$E(\varphi(X)) = \sum_{\omega \in \Omega} \varphi(X(\omega))P(\{\omega\}).$$

特别地, $\varphi(x) = x^r$,

$$E(\varphi(X)) = \sum_{\omega \in \Omega} [X(\omega)]^r P(\{\omega\}),$$

称为 X 的第 r 阶力矩。

数学期望的历史

源于17世纪一名法国贵族关于"赌注分配问题" (见课件1, Pascal和Fermat赌注分配)。1654年, Fermat给Pascal的一封私人信件解决了该问题, 但当时人们不知。3年以后, 即1657年, 德国科学家Huygens (14 April 1629 – 8 July 1695) 公布他关于"赌注分配问题"的结果, 解法与Fermat的解法本质一样。

160年后, 即1814年, Laplace 使用"数学期望"一词, 为大家接受; 因为, 英语"Expectation", 德文"Erwartungswert", 法文"Espérance mathématique" 首个字母均含 E 。

赌金分配问题

(赌金分配问题) Fermat和Pascal坐在巴黎咖啡馆，玩作简单的游戏，抛硬币。若正面，Fermat得1分，若反面，Pascal的1分。先得10分的人赢全部赌注160法郎（每人各出80法郎）。但奇怪的事情发生了，Fermat的朋友病了，他要在8 : 7的好形势下赶回Toulouse。问全部赌注160法郎如何分？

答案：Fermat应得赌金为

$$160 \times \frac{11}{16} = 110 \text{ (法郎).}$$

赌金分配问题

关键思想：赌金分配不完全依赖过去赢多少盘，而很依赖未来要赢多少盘。比如，赌20盘决定胜负比赛在17：15的情况下中断，和赌10盘决定胜负比赛在7：5的情况下中断，两种情况赌金分配比例应该是一样的。

赌金分配问题

现选手A需赢 r 盘胜，选手B需赢 s 盘胜，则只需再赌 $r + s - 1$ 盘就可决定胜负（为什么？）。选手A和选手B合理的赌金分配比例为

$$\sum_{k=r}^{r+s-1} \binom{r+s-1}{k} : \sum_{k=s}^{r+s-1} \binom{r+s-1}{k}.$$

注意到

$$\begin{aligned} \sum_{k=r}^{r+s-1} \binom{r+s-1}{k} &= \sum_{m=0}^{s-1} \binom{r+s-1}{r+s-1-m} (\text{令 } k = r+s-1-m) \\ &= \sum_{m=0}^{s-1} \binom{r+s-1}{m}. \end{aligned}$$

注意：最后一个式子是对对手未能赢 s 盘（即输掉）的所有情况。

赌金分配问题

例如： $r = 2$ (Fermat) 和 $s = 3$ (Pascal)

$$\frac{\sum_{m=0}^2 \binom{4}{k}}{\sum_{k=0}^1 \binom{4}{k}} = \frac{\binom{4}{0} + \binom{4}{1} + \binom{4}{2}}{\binom{4}{0} + \binom{4}{1}} = \frac{1 + 4 + 6}{1 + 4} = \frac{11}{5},$$

即Fermat 应得全部赌注的 $\frac{11}{16}$ ，而Pascal应得全部赌注的 $\frac{5}{16}$ 。

数学期望

例：设 L 是一个正整数，随机变量 X 满足

$$P(X = n) = \frac{1}{L}, 1 \leq n \leq L.$$

则其数学期望 $E(X)$ 为

$$E(X) = \sum_{n=1}^L n \frac{1}{L} = \frac{1}{L} \cdot \frac{L(L+1)}{2} = \frac{L+1}{2}.$$

该随机变量 X 称为集合 $\{1, 2, \dots, L\}$ 的均匀分布。

数学期望

例：抛一枚均匀硬币直至正面出现，设 X 是直至正面出现所抛次数，求 $E(X)$ （想一想，什么意思？）。

猜一猜： $E(X) = ?$

数学期望

解： $X = n$ 意味前 $n - 1$ 次出现反面，第 n 次出现正面。

$$P(X = n) = \left(\frac{1}{2}\right)^{n-1} \cdot \frac{1}{2} = \frac{1}{2^n}, \quad n \geq 1.$$

则数学期望 $E(X)$ 为

$$E(X) = \sum_{n=1}^{\infty} n \frac{1}{2^n} = ?$$

数学期望

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n=0}^{\infty} x^n, \quad |x| < 1.$$

两边关于 x 微分, 得

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots = \sum_{n=0}^{\infty} (n+1)x^n, \quad |x| < 1.$$

两边乘以 x , 得

$$\sum_{n=0}^{\infty} (n+1)x^{n+1} = \sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}.$$

令 $x = \frac{1}{2}$ 得,

$$\sum_{n=1}^{\infty} \frac{n}{2^n} = 2.$$

则数学期望 $E(X) = 2$.

数学期望

例：抛一枚不均匀硬币直至正面出现，正面出现的概率为 p ，反面出现的概率为 $q = 1 - p$ 。设 X 是直至正面出现所抛次数，求 $E(X)$ 。

猜一猜： $E(X) = ?$ 如 $p = \frac{1}{3}$ ， $p = \frac{1}{4}$ 。

数学期望

解: $X = n$ 意味前 $n - 1$ 次出现反面, 第 n 次出现正面, 则

$$P(X = n) = q^{n-1} \cdot p, \quad n \geq 1.$$

于是, 数学期望 $E(X)$ 为

$$E(X) = \sum_{n=1}^{\infty} np(1-p)^{n-1} = \frac{1}{p}.$$

此处利用上面推出的公式:

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2} \quad (x < 1).$$

随机变量 X 称为正面出现的"等时"(waiting time)。

数学期望

例：一枚均匀硬币抛 n 次，设 S_n 是正面出现的次数，求 $E(S_n)$ 。

猜一猜： $E(S_n) = ?$ ，即正面平均出现多少次？

数学期望

解：在 n 次抛硬币中出现 k 次正面的概率为

$$P(S_n = k) = \binom{n}{k} \left(\frac{1}{2}\right)^n, \quad 0 \leq k \leq n.$$

则数学期望 $E(X)$ 为

$$E(X) = \sum_{k=0}^n k \binom{n}{k} \left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n \sum_{k=0}^n k \binom{n}{k} = ?$$

数学期望

$$(1+x)^n = \sum_{k=0}^n \binom{n}{k} x^k, \quad x \geq 0.$$

两边关于 x 微分, 得

$$n(1+x)^{n-1} = \sum_{k=1}^n k \binom{n}{k} x^{k-1}$$

令 $x = 1$ 得,

$$n2^{n-1} = \sum_{k=1}^n k \binom{n}{k},$$

则数学期望

$$E(X) = \sum_{k=1}^n \binom{n}{k} \frac{k}{2^n} = \frac{1}{2^n} \sum_{k=1}^n k \binom{n}{k} = \frac{1}{2^n} \cdot n2^{n-1} = \frac{n}{2}.$$

想一想, 合理吗?

课堂练习(Bernoulli): 一枚不均匀硬币抛 n 次, 正面出现的概率为 p , 反面出现的概率为 $q = 1 - p$ 。设 S_n 是正面出现的次数, 求 $E(S_n)$ 。

答案：数学期望 $E(S_n)$ 为

$$E(S_n) = np.$$

数学期望的性质

- $E(c) = c$;
- $E(X + Y) = E(X) + E(Y)$;
- $E(aX) = aE(X)$;
- 若 $X \leq Y$ 则 $E(X) \leq E(Y)$ 。

具有密度的随机变量(不可数样品空间)

- 以上考虑定义在可数样品空间上的随机变量，不够。
- 现考虑定义在不可数样品空间上的随机变量。现实中，不可能对不可数样品空间上的每个子集赋予一个概率，太大。
- 需对某些子集赋予一个概率，这些子集构成一个 σ -代数。
- 避免复杂的数学理论，将讨论具密度的随机变量（下周内容）。

作业

第3次作业(钟开来书): P. 71-73: 第19, 20, 21, 25, 28题。
P. 109-110: 第1, 2, 3, 4, 11, 15题。

作业: (1)在作业第1页左上角醒目学号、姓名。
(2)将作业扫描成单个、**PDF**文件 (不要压缩、不要Word格式)

预习内容: 条件概率、**Bayes**公式