概率论与数理统计

任课教师: 胡家信

清华大学数学科学系

2022 年 11 月 7 日

复习

- 庞加莱公式：对任意集合$A_1, A_2, \cdots, A_n$，下列等式成立：

$$P\left(\cup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j) - \sum_{j,k} P(A_j A_k) + \sum_{j,k,l} P(A_j A_k A_l)$$
$$- \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n),$$

其中求和的指标不同、取值从1到n，如

$$\begin{aligned}\sum_{j,k} P(A_j A_k) = P(A_1 A_2) + P(A_1 A_3) + \cdots + P(A_1 A_n) \\ + P(A_2 A_3) + P(A_2 A_4) + \cdots + P(A_2 A_n) \\ + \cdots \\ + P(A_{n-1} A_n).\end{aligned}$$

复习

例如

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = &\{P(A_1) + P(A_2) + P(A_3)\} \\ &- \{P(A_1A_2) + P(A_1A_3) + P(A_2A_3)\} \\ &+ P(A_1A_2A_3). \end{aligned}$$

若A_1, A_2, A_3两两互不相交，则

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3).$$

定理1：设X和Y是独立可积随机变量，则

$$E(XY) = E(X)E(Y).$$

证明：分两种情况：Ω可数和Ω不可数。

(1) Ω可数。令$\{x_j\}$为X所取的不同值的集合，$\{y_j\}$为Y 所取的不同值的集合，令$A_{jk} = \{\omega : X(\omega) = x_j, Y(\omega) = y_k\}$，则诸集合$A_{jk}$不相交, 且

$$\Omega = \sum_j \sum_k A_{jk}.$$

随机变量XY在A_{jk}上取值$x_j y_k$，但某些值可能一样，如$x_j = 2, y_k = 3$和$x_j = 3, y_k = 2$。

根据期望的定义，

$$E(XY)=\sum_j\sum_k x_j y_k P(A_{jk}).$$

由独立性，$P(A_{jk})=P(X=x_j)P(Y=y_k)$。所以

$$\begin{aligned}E(XY)&=\sum_j\sum_k x_j y_k P(X=x_j)P(Y=y_k)\\&=\left(\sum_j x_j P(X=x_j)\right)\left(\sum_k y_k P(Y=y_k)\right)=E(X)E(Y).\end{aligned}$$

注意这里假定两个级数均绝对收敛(为什么要此假设?)。

定理

(2) Ω不可数，设随机变量(X,Y)具有联合密度函数f。则

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}uvf(u,v)dudv.$$

由独立性，$f(u,v)=f_1(u)f_2(v)$，其中

$$f_1(u)=\int_{-\infty}^{\infty}f(u,v)dv,$$
$$f_2(v)=\int_{-\infty}^{\infty}f(u,v)du,$$

分别为X和Y的密度函数。所以

$$E(XY)=\left[\int_{-\infty}^{\infty}uf_1(u)du\right]\left[\int_{-\infty}^{\infty}vf_2(v)dv\right]=E(X)E(Y).$$

证毕。

力矩

对任意正整数r，数学期望

$$E(X^r)$$

称为随机变量X的**第r阶力矩**。注意：$r=1$为数学期望，$r=2$为重要情况。

定义：令$X^0=X-E(X)$，称X^0的二阶力矩为X的**方差**(variance)：

$$Var(X):=\sigma^2(X):=E\left((X^0)^2\right)=E\left((X-E(X))^2\right),$$

称$\sigma(X)=\sqrt{\sigma^2(X)}$为$X$的**偏差**(deviation)。

力矩

定理： 若$E(X^2)<\infty$，则$E(|X|)<\infty$，且

$$\sigma^2(X)=E(X^2)-[E(X)]^2; \tag{1}$$

从而$\{E(|X|)\}^2\le E(X^2)$。

证明： 由于$(|X|-1)^2=X^2-2|X|+1$，得

$$E(X^2-2|X|+1)\ge 0 \quad \Rightarrow E(X^2)+1\ge 2E(|X|).$$

故$E(|X|)<\infty$。注意$(X-E(X))^2=X^2-2E(X)\cdot X+[E(X)]^2$，两边同取数学期望得

$$\begin{aligned}\sigma^2(X) &= E\left((X-E(X))^2\right)=E(X^2)-2[E(X)]^2+[E(X)]^2\\ &= E(X^2)-[E(X)]^2,\end{aligned}$$

于是，证明了式子（1）。

力矩

最后欲证$\{E(|X|)\}^2 \leq E(X^2)$。事实上，

$$(|X| - E(|X|))^2 = |X|^2 - 2E(|X|) \cdot |X| + [E(|X|)]^2$$

所以，两边同取数学期望得

$$\begin{aligned} 0 &\leq E\left((|X| - E(|X|))^2\right) \\ &= E(X^2) - 2[E(|X|)]^2 + [E(|X|)]^2 = E(X^2) - [E(|X|)]^2, \end{aligned}$$

所证成立。证毕。

方差的意义

方差的意义：$X^0 = X - E(X)$表示X与其均值的偏差，可正可负；也可考虑**绝对平均偏差**

$$E(|X^0|) = E(|X - E(X)|)$$

但实际计算此值困难。代之，考虑**偏差的平方，即方差**：

$$E(|X^0|^2) = E(|X - E(X)|^2).$$

开方后的值表示**随机变量与其均值之间的平均偏差**。该值愈小，随机变量愈集中在均值周围。

定理：设X, Y独立、方差均有限，则

$$\sigma^2(X+Y)=\sigma^2(X)+\sigma^2(Y).$$

更一般地，若$X_1,\cdots,X_n$独立，则

$$\sigma^2(X_1+\cdots+X_n)=\sigma^2(X_1)+\cdots+\sigma^2(X_n).$$

证明： 令

$$a = E(X), b = E(Y), X_1 = X - a, Y_1 = Y - b,$$

则$E(X_1) = E(Y_1) = 0$。由于$E(XY) = E(X)E(Y) = ab$，所以

$$E(X_1Y_1) = E\left((X-a)(Y-b)\right) = E\left(XY\right) - ab = ab - ab = 0.$$

（实际上，不难证明X_1和Y_1也独立）。于是，

$$\begin{aligned}\sigma^2(X+Y) &= E\{(X+Y-E(X+Y))^2\} = E\{(X_1+Y_1)^2\} \\ &= E\{X_1^2 + 2X_1Y_1 + Y_1^2\} = E(X_1^2) + 2E(X_1Y_1) + E(Y_1^2) \\ &= E(X_1^2) + E(Y_1^2) = \sigma^2(X) + \sigma^2(Y).\end{aligned}$$

最后，对任意n个独立随机变量，一样证明。证毕。

命题：设$X_1, \cdots, X_n$为任意随机变量，则

$$E\{(X_1+\cdots+X_n)^2\} = \sum_{j=1}^{n} E(X_j^2) + 2\sum_{1\le j<k\le n} E(X_jX_k).$$

事实上，利用

$$(X_1+\cdots+X_n)^2 = \sum_{j=1}^{n} X_j^2 + 2\sum_{1\le j<k\le n} X_jX_k,$$

两边取数学期望，即可。

Cauchy-Schwarz 不等式

Cauchy-Schwarz inequality：设X, Y为任意随机变量，则

$$[E(XY)]^2 \leq E(X^2)E(Y^2).$$

事实上，对任意实数λ,

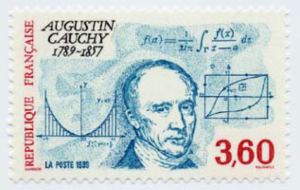
$$0 \leq E\{(X+\lambda Y)^2\} = E(X^2) + 2\lambda E(XY) + \lambda^2 E(Y^2)$$

注意$E(X^2) \geq 0$, 右边是λ的多项式，恒为非负，故判别式不大于零：

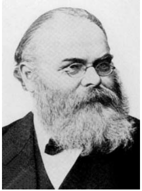
$$[2E(XY)]^2 - 4E(X^2)E(Y^2) \leq 0,$$

得证.

Cauchy-Schwarz(柯西—施瓦茨)



Augustin Louis Cauchy (1789-1857)



Hermann Amandus Schwarz(1843-1921)

定义：设X, Y为任意方差有限的随机变量.

$$\begin{aligned} E(X^0Y^0) &= E\{(X-E(X))(Y-E(Y))\} \\ &= E\{XY - E(X)Y - XE(Y) + E(X)E(Y)\} \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

记$Cov(X, Y) = E(XY) - E(X)E(Y)$，称为$X, Y$的**协方差**。量

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)} = \frac{E(X^0Y^0)}{\sqrt{E\{(X^0)^2\}E\{(Y^0)^2\}}}$$

称为X, Y的**关联系数**。由Cauchy-Schwarz不等式,

$$-1 \le \rho(X, Y) \le 1.$$

若X, Y独立，则$Cov(X, Y) = 0$，故$\rho(X, Y) = 0$。反之，$\rho(X, Y) = 0$不一定意味X, Y独立(试举反例，课后思考题)。

课堂练习：设Bernoulli随机变量X如下：

$$X = \begin{cases} 1, & \text{概率为}p, \\ 0, & \text{概率为}q = 1 - p. \end{cases}$$

计算$E(X), \sigma^2(X)$.

协方差

答案:

- $E(X) = p$。
- $\sigma^2(X) = p - p^2 = p(1 - p)$。

课堂练习

课堂练习：设独立Bernoulli随机变量$X_1, \cdots, X_n$，令

$$S_n = X_1 + \cdots + X_n.$$

计算$E(S_n), \sigma^2(S_n)$。

协方差

答案:

- $E(S_n) = np$。
- $\sigma^2(S_n) = np(1-p)$。

多项式分布

- 二次多项式(binomial)：

$$(x_1+x_2)^n=\sum_{k=0}^{n}\binom{n}{k}x_1^k x_2^{n-k}=\sum_{k+j=n}\frac{n!}{k!j!}x_1^k x_2^j.$$

- 多次多项式(multinomial)：

$$(x_1+x_2+\cdots+x_r)^n=\sum_{k_1+\cdots+k_r=n}\frac{n!}{k_1!k_2!\cdots k_r!}x_1^{k_1}x_2^{k_2}\cdots x_r^{k_r}.$$

即$k_1\geq 0$，$\cdots$，$k_r\geq 0$为非负整数，且和为n：

$$k_1+\cdots+k_r=n.$$

例子

例：有一只箱子装r种不同颜色的球，每个颜色的球比率为

$$p_1 : p_2 : \cdots : p_r \quad (p_1 + p_2 + \cdots + p_r = 1).$$

每次放回、独立、一个一个地抽取n（$n \geq r$）个球，问每个颜色都出现的概率是多少?

注意：球无标号，只按颜色区分。如第1次抽红色球、第2次抽蓝色球的情况，与第1次抽蓝色球、第2次抽红色球的情况，不一样。

例子

比如，箱子装3种不同颜色的球，每次放回地独立抽4个球，问每个颜色都出现的概率是多少? 此时，$n = 4, r = 3$。记n_1为颜色为1的球的个数，n_2为颜色为2的球的个数，等等，则$n_1 + n_2 + n_3 = n = 4$。如$n_1 = 2, n_2 = n_3 = 1$，即抽取的4个球中颜色为1的球有2个、颜色为2，3的球各1个，共有多少种情况?

例子

事实上，列出所有情况如下：

$$
\begin{array}{lll}
(1,1,2,3) & (1,1,3,2) & \text{（第1、2次为颜色1的球）} \\
(1,2,1,3) & (1,3,1,2) & \text{（第1、3次为颜色1的球）} \\
(1,2,3,1) & (1,3,2,1) & \text{（第1、4次为颜色1的球）} \\
(2,1,1,3) & (3,1,1,2) & \text{（第2、3次为颜色1的球）} \\
(2,1,3,1) & (3,1,2,1) & \text{（第2、4次为颜色1的球）} \\
(2,3,1,1) & (3,2,1,1) & \text{（第3、4次为颜色1的球）}.
\end{array}
$$

情况有$12 = \frac{4!}{2!1!1!}$。该事件（即“抽取的4个球中颜色为1的球有2个、颜色为2，3的球各1个”）发生的概率为$12p_1^2p_2p_3$（其它情形一样处理）。

欲求的概率是多少？（对颜色再排列组合即可，见下面）

例子

解：设$X_1, \cdots, X_n$是具有和下面随机变量X相同分布的n个独立随机变量：

$$X = \begin{cases} 1, & 概率为p_1, \\ 2, & 概率为p_2, \\ \vdots & \\ r, & 概率为p_r. \end{cases}$$

例子

引入新的随机变量N_j 如下：

$$N_j\text{表示}(X_1,\cdots,X_n)\text{中取值为}j\text{随机变量}X\text{的个数}.$$

每个N_j从$\{0,1,\cdots,n\}$中取值，但$N_1,\cdots,N_r$不独立，因为

$$N_1+\cdots+N_r=n.$$

而事件$\{N_1=n_1,\cdots,N_r=n_r\}$表示在抽取的$n$个球中，颜色为1的球有$n_1$个，颜色为2的球有$n_2$个，颜色为3的球有$n_3$个，等等。所以，

$$P(N_1=n_1,\cdots,N_r=n_r)=\frac{n!}{n_1!\cdots n_r!}p_1^{n_1}\cdots p_r^{n_r}.$$

从而，所求概率为

$$P(A)=\sum_{\substack{n_1+\cdots+n_r=n\\ n_1\geq 1,\cdots,n_r\geq 1}}\frac{n!}{n_1!\cdots n_r!}p_1^{n_1}\cdots p_r^{n_r}.$$ 解毕。

课后讨论题：

1 （加分题，0.2分，网络学堂上交）化简下列计算

$$\sum_{\substack{n_1+n_2+n_3+n_4=15,\\ n_1\ge 1, n_2\ge 1, n_3\ge 1, n_4\ge 1}} \frac{15!}{n_1!n_2!n_3!n_4!}\left(\frac{1}{4}\right)^{15}.$$

2 计算（答案见下面）

- 随机变量$(N_1,\cdots,N_r)$的**边际分布**.
- 数学期望$E(N_j)$和方差$\sigma^2(N_j)$.

课后概念：什么是边际分布？

答案

答案：计算随机变量$(N_1, \cdots, N_r)$的**边际分布**(直接计算)：

$$P(N_1 = n_1) = \frac{n!}{n_1!(n-n_1)!} p_1^{n_1}(1-p_1)^{n-n_1}.$$

数学期望

$$\begin{aligned} E(N_j) &= \sum_{m=0}^{n} mP(N_j = m) \\ &= \sum_{m=1}^{n} m \frac{n!}{m!(n-m)!} p_j^m (1-p_j)^{n-m} = np_j. \end{aligned}$$

和方差

$$\sigma^2(N_j) = np_j(1-p_j), \quad 1 \le j \le r.$$

生成函数或母函数

设X是随机变量，**取值非负整数**，概率决定如下

$$P(X=j)=a_j,\quad j=0,1,2,\cdots$$

注意$\sum_{j=0} a_j=1$。引入函数

$$g(z)=a_0+a_1z+a_2z^2+\cdots=\sum_{j=0}^{\infty}a_jz^j,$$

称此函数为数列$\{a_j\}_{j\geq 0}$或者随机变量X的**生成函数（或母函数）**，变量z可取复数值。当$|z|\leq 1$时，级数收敛，因为

$$|g(z)|\leq\sum_{j=0}^{\infty}a_j|z|^j\leq\sum_{j=0}^{\infty}a_j=1,\quad |z|\leq 1.$$

可求导

$$g'(z) = a_1 + 2a_2 z + \cdots = \sum_{n=1}^{\infty} n a_n z^{n-1},$$

$$g''(z) = 2a_2 + 6a_3 z \cdots = \sum_{n=2}^{\infty} n(n-1) a_n z^{n-2},$$

$$\vdots$$

$$g^{(j)}(z) = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1) a_n z^{n-j}.$$

令$z = 0$，得$g^{(j)}(0) = j! a_j$，或$a_j = \frac{1}{j!} g^{(j)}(0)$，故从$g$可得$a_j$的值，**生成函数**和**分布**两者一一对应。

定理： 取非负整数值的随机变量的概率分布由生成函数唯一决定。（证明如上，略。）

令$z=1$，得

$$g'(1)=\sum_{n=1}^{\infty} na_n = E(X),$$

$$g''(1)=\sum_{n=2}^{\infty} n(n-1)a_n = \sum_{n=0}^{\infty} n^2 a_n - \sum_{n=0}^{\infty} na_n = E(X^2)-E(X).$$

故

$$E(X)=g'(1),\quad E(X^2)=g'(1)+g''(1),$$

$$\sigma^2(X)=E(X^2)-(E(X))^2=g'(1)+g''(1)-(g'(1))^2.$$

生成函数

设Y是另一个随机变量，取值非负整数，概率决定如下

$$P(Y=j)=b_j,\quad j=0,1,2,\cdots$$

生成函数为$h(z)=\sum_{j=0}^{\infty}b_jz^j$。问：若$X$和$Y$独立，求$X+Y$的概率分布？事实上，

$$g(z)h(z)=\left(\sum_{j=0}^{\infty}a_jz^j\right)\left(\sum_{k=0}^{\infty}b_kz^k\right)=\sum_j\sum_k a_jb_kz^{j+k}.$$

令$g(z)h(z)=\sum_{i=0}^{\infty}c_iz^i$得

$$c_i=\sum_{j+k=i}a_jb_k=\sum_{j=0}^{i}a_jb_{i-j}.$$

数列$\{c_i\}$

$$c_i = \sum_{j=0}^{i} a_j b_{i-j},$$

称为两个数列$\{a_j\}, \{b_j\}$的**卷积**(convolution)。

若X和Y独立，则

$$\begin{aligned} c_i &= \sum_{j=0}^{i} P(X=j)P(Y=i-j) \\ &= \sum_{j=0}^{i} P(X=j, Y=i-j) = P(X+Y=i). \end{aligned}$$

生成函数

另一方面,

$$P(X+Y=i)=\sum_{j=0}^{\infty}P(X=j)P(X+Y=i|X=j)$$
$$=\sum_{j=0}^{\infty}P(X=j)P(Y=i-j|X=j)$$
$$=\sum_{j=0}^{i}P(X=j)P(Y=i-j).$$

即$P(X+Y=i)=c_i$, 所以$\{c_i, i\geq 0\}$是随机变量$X+Y$的**概率分布**（注意：c_i是X, Y的生成函数乘积中关于z^i项的系数）。

定理：若随机变量$X_1,\cdots,X_n$独立，生成函数分别为$g_1,\cdots,g_n$，则其和$X_1+\cdots+X_n$的生成函数g为各个随机变量生成函数的乘积，

即$g=g_1\cdots g_n$。

例：设Bernoulli随机变量$X_1, \cdots, X_n$独立，求和$X_1 + \cdots + X_n$的生成函数。

解：因为$a_0 = P(X = 0) = q, a_1 = P(X = 1) = p$，每个随机变量$X_1, \cdots, X_n$的生成函数为

$$g(z) = a_0 + a_1 z = q + pz.$$

于是，和$X_1 + \cdots + X_n$的生成函数为

$$g(z)^n = (q + pz)^n = \sum_{k=0}^{n} \binom{n}{k} q^{n-k} p^k z^k.$$

证毕。

课堂练习题：利用上面结论，试求和$S_n = X_1 + \cdots + X_n$的概率分布。

解：事实上，由定义

$$g(z)^n = \sum_{k=0}^{\infty} P(S_n = k) z^k = \sum_{k=0}^{n} \binom{n}{k} q^{n-k} p^k z^k.$$

所以，比较z^k的系数，得

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k},\ 0 \le k \le n,$$

$$P(S_n = k) = 0, k > n.$$

证毕。

例：设等时随机变量$T_1, \cdots, T_n$独立，求$S_n = T_1 + \cdots + T_n$的生成函数。

解：回忆**等时随机变量**T的分布：$P(X = j) = q^{j-1}p, j \geq 1$（如抛硬币，前面$j-1$次出现反面、第$j$次出现正面的概率）。每个随机变量$X_1, \cdots, X_n$的生成函数为

$$g(z) = \sum_{j=1}^{\infty} q^{j-1} p z^j = \frac{p}{q} \sum_{j=1}^{\infty} (qz)^j = \frac{p}{q} \cdot \frac{qz}{1-qz} = \frac{pz}{1-qz}.$$

故和S_n的生成函数为

$$g(z)^n = \left(\frac{pz}{1-qz} \right)^n.$$

证毕。

注意，对函数$f(x)=(1-x)^{-n}$进行Taylor展开，则和的生成函数为

$$g(z)^n=\left(\frac{pz}{1-qz}\right)^n=(pz)^n(1-qz)^{-n}=\sum_{k=n}^{\infty}\binom{k-1}{n-1}p^nq^{k-n}z^k.$$

所以得

$$P(S_n=n+j)=\binom{n+j-1}{n-1}p^nq^j,\quad j\geq 0.$$

意义：它表明在$n+j$次投币中，"前$n+j-1$次中有$n-1$次出现正面、j次出现反面、而最后一次（即第$n+j$次）出现正面"这一事件（即在$n+j$次投币中，n次正面出现的等时）的概率。

题目讲解（男孩、女孩问题）

题目： Consider families with three children. What is the probability that all are boys **given that**：

1. the first is a boy;
2. there is at least one boy.

求条件概率（猜一猜）。

题目讲解（解法一）

解法一：两个问题的样品空间相同，均为

$$\Omega = \{bbb, bbg, bgb, gbb, ggb, gbg, bgg, ggg\}$$

样品空间中每个事件发生的概率均为$\frac{1}{8}$。

令$A = \{bbb\} = \{$三个均是男孩$\}$，$B_1 = \{$老大是男孩$\}$，$B_2 = \{$至少有一个是男孩$\}$。欲求概率$P(A|B_1)$和概率$P(A|B_2)$。事实上

$$P(B_1) = P(\{bbb\}) + P(\{bbg\}) + P(\{bgb\}) + P(\{bgg\}) = 4 \times \frac{1}{8}.$$

$$P(B_2) = 1 - P(\{ggg\}) = 7 \times \frac{1}{8}.$$

所以，**已知老大是男孩，三个全是男孩**的概率为

$$P(A|B_1) = \frac{P(AB_1)}{P(B_1)} = \frac{P(A)}{P(B_1)} = \frac{\frac{1}{8}}{4 \times \frac{1}{8}} = \frac{1}{4}.$$

已知至少有一个是男孩，三个全是男孩的概率为

$$P(A|B_2) = \frac{P(AB_2)}{P(B_2)} = \frac{P(A)}{P(B_2)} = \frac{\frac{1}{8}}{7 \times \frac{1}{8}} = \frac{1}{7}.$$

解毕。

解法二：通过缩小样品空间，将条件概率化为无条件概率（我们将利用点数法求概率）。

(1) 已知老大是男孩，其样品空间为$\widetilde{\Omega} = \{b_{bb}, b_{bg}, b_{gb}, b_{gg}\}$，样品空间仅含4个点。令$A = \{$三个均是男孩$\}$，则$A = \{b_{bb}\}$。于是，

$$P(A) = \frac{1}{4}.$$

解法二

(2) 已知至少有一个是男孩，其样品空间为

$$\widetilde{\Omega} = \{bbb, bbg, bgb, gbb, ggb, gbg, bgg\},$$

样品空间含7个点。令$B = \{\text{三个均是男孩}\}$，则$B = \{bbb\}$。于是，

$$P(B) = \frac{1}{7}.$$

解毕。

题目讲解（第158页第5题）

题目（课堂讨论）： 考虑有三个孩子的家庭，生男生女概率一样。

(1) 随机选取一个家庭，发现老大是男孩，问另两个孩子是女孩的概率？（猜一猜）

(2) 从一个家庭随机选取一个孩子（不一定是老大），发现是男孩，问另两个孩子是女孩的概率？（猜一猜）

关键点：如何确定样品空间。

注意：第1问是孩子确定、家庭随机（有很多这样的家庭）；而第2问是家庭确定、孩子随机（有3个孩子）。

答案

解：(1). 样品空间为

$$\Omega = \{bbb, bbg, bgb, gbb, ggb, gbg, bgg, ggg\}$$

样品空间中每个事件发生的概率均为$\frac{1}{8}$。令$A = \{$老大是男孩$\}$，$B = \{$另两个孩子是女孩$\}$。欲求$P(B|A)$? 事实上

$$P(A) = P(\{bbb\}) + P(\{bbg\}) + P(\{bgb\}) + P(\{bgg\}) = 4 \times \frac{1}{8}.$$

$$P(AB) = P(\{bgg\}) = \frac{1}{8}.$$

所以，**已知老大是男孩，另外两个是女孩**的概率为

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{1}{8}}{4 \times \frac{1}{8}} = \frac{1}{4}.$$

答案

解：(2). 此时选取的男孩不一定是老大，与前一问不一样。但可以从选取的男孩出发（利用Poisson关于序列抽样思想），然后考虑其它两个孩子的可能情形，样品空间为

$$\widetilde{\Omega} = \{b_{bb}, b_{bg}, b_{gb}, b_{gg}\}$$

该样本空间的总点数为4，其中另外两个是女孩的情形为b_{gg}，只有一种情形。所以，**已知选取的是男孩，另外两个是女孩**的概率为

$$P\big(\{b_{gg}\}\big) = \frac{1}{4}. \quad \text{解毕。}$$

注意：第（2）小问的样品空间**不是**

$$\widetilde{\Omega} = \{bbb, bbg, bgb, gbb, ggb, gbg, bgg\}$$

（它是“已知至少有一个男孩”的样品空间，不是“有一个男孩”的样品空间）

题目讲解（第158页第7题）

题目（课堂讨论）： 考虑有三个孩子的家庭，生男生女概率一样，抽样是从**这种家庭的孩子中随机选取2个小孩**。若2个孩子均是女孩，问另一个孩子是男孩的概率？（猜一猜）

答案

解：样品空间为

$$\widetilde{\Omega} = \{gg_b, gg_g\}$$

该样本空间的总点数为2，其中另外一个是男孩的情形为gg_b，只有一种情形。故所求概率为

$$P(\{gg_b\}) = \frac{1}{2}.$$

解毕。

思考题：为什么样品空间不能写成

$$\Omega = \{ggb, gbg, bgg, ggg\}$$

（事实上，该空间表示 "已知至少有两个是女孩"，和"2个孩子（同时）均是女孩"的情况不同）

题目： 设$A_j, 1 \le j \le 5$独立，则

(1) $(A_1 \cup A_2)A_3$，$A_4^c \cup A_5^c$独立；(2) $A_1 \cap A_2, A_3 \cap A_4$，$A_5^c$独立。

关键思想：（1）在独立性中只有**交运算**，故尽量处理**交运算**为宜；（2）充分运用**补运算**，以及事实：A, B独立，当且仅当A, B^c独立、或者A^c, B独立、或者A^c, B^c独立（**为什么？**）

解答：（1）只需证明$(A_1 \cup A_2)A_3$与A_4A_5独立（**为什么？**）。事实上

$$
\begin{aligned}
& P\left((A_1 \cup A_2)A_3 \cap A_4A_5\right) = P\left((A_1 \cup A_2)A_3A_4A_5\right) \\
& = P\left(A_1A_3A_4A_5 \cup A_2A_3A_4A_5\right) \\
& = P\left(A_1A_3A_4A_5\right) + P\left(A_2A_3A_4A_5\right) - P\left(A_1A_2A_3A_4A_5\right)
\end{aligned}
$$

利用独立性，我们有

$$
\begin{aligned}
P\left((A_1 \cup A_2)A_3 \cap A_4A_5\right) &= P\left(A_1A_3A_4A_5\right) + P\left(A_2A_3A_4A_5\right) \\
&\quad - P\left(A_1A_2A_3A_4A_5\right) \\
&= P(A_4)P(A_5)\left\{P(A_3)\left[P(A_1) + P(A_2) - P(A_1A_2)\right]\right\} \\
&= P(A_4A_5) \cdot P((A_1 \cup A_2)A_3).
\end{aligned}
$$

2 只需证$A_1 \cap A_2, A_3 \cap A_4, A_5$独立（为什么？），即需证四种交集的概率等于概率相乘（略）。证毕。

题目讲解（老虎机问题, P.159. 17题）

题目：赌场有三种数目相同的老虎机，赔率分别为$\frac{1}{3},\frac{1}{2},\frac{2}{3}$。已知一个老虎机在四局中陪了两局，问它再次陪付的概率是多少？

解：利用"分割征服"公式。符号：设U_1, U_2, U_3 "分别表示三种老虎机"，A"四局中赔了两局"，B"第五局赔付"。

- 由于三种老虎机数目相同，选取均等，
  故$P(U_1) = P(U_2) = P(U_3) = 1/3$。
- 计算知$P(A|U_1) = \binom{4}{2}\cdot\left(\frac{1}{3}\right)^2\cdot\left(\frac{2}{3}\right)^2 = \frac{8}{27}$ (为什么)。同理，
  $P(A|U_2) = \binom{4}{2}\cdot\left(\frac{1}{2}\right)^4 = \frac{3}{8}$, 以及$P(A|U_3) = \binom{4}{2}\cdot\left(\frac{2}{3}\right)^2\cdot\left(\frac{1}{3}\right)^2 = \frac{8}{27}$。
- 计算$P(A)$如下：

$$P(A) = P(U_1)P(A|U_1) + P(U_2)P(A|U_2) + P(U_3)P(A|U_3) = \frac{209}{648}.$$

- 于是，

$$P(U_1|A)=\frac{P(AU_1)}{P(A)}=\frac{P(U_1)P(A|U_1)}{P(A)}=\frac{1/3\times 8/27}{209/648}=\frac{64}{209},$$

$$P(U_2|A)=\frac{P(AU_2)}{P(A)}=\frac{P(U_2)P(A|U_2)}{P(A)}=\frac{81}{209},$$

$$P(U_3|A)=\frac{P(AU_3)}{P(A)}=\frac{P(U_3)P(A|U_3)}{P(A)}=\frac{64}{209}.$$

- 由题意可知，

$P((B|A)|(U_1|A))=\frac{1}{3}$, $P((B|A)|(U_2|A))=\frac{1}{2}$, $P((B|A)|(U_3|A))=\frac{2}{3}$。

- 故所求概率为

$$P(B|A)=P(U_1|A)\cdot P((B|A)|(U_1|A))+P(U_2|A)\cdot P((B|A)|(U_2|A))$$
$$+P(U_3|A)\cdot P((B|A)|(U_3|A))=\frac{64}{209}\cdot\frac{1}{3}+\frac{81}{209}\cdot\frac{1}{2}+\frac{64}{209}\cdot\frac{2}{3}=\frac{1}{2}.\blacksquare$$

作业

第7次作业(钟开来书)：

P. 197-200：第18，20，21，22，23，24题，

第29，33，37，38题.

预习内容：Bernoulli、二次项、Poisson分布等