博学 审问 慎思 明辨 笃行

概率论与数理统计

中山大學
SUN YAT-SEN UNIVERSITY

主讲老师：邝东阳， kuangdy@mail.sysu.edu.cn

助教：黄培， huangp89@mail2.sysu.edu.cn

范志隆， fanzhlong@mail2.sysu.edu.cn

Part 1
离 散 分 布

博学 | 审问 | 慎思 | 明辨 | 笃行

设 X 为一随机变量. 如果 X 只取有限个或可数个值, 则称 X 为一个 (一维) 离散型随机变量.

设 X 为一离散型随机变量, 其全部可能值为 $\{a_1, a_2, \ldots\}$. 则

$$p_i = P(X = a_i), \ \ i = 1, 2, \ldots \tag{2.1}$$

称为 X 的概率质量函数 (probability mass function, *pmf*) 或分布律.

概率分布表

可能值	a_1	a_2	$\cdots$	a_i	$\cdots$
概率	p_1	p_2	$\cdots$	p_i	$\cdots$

$$p_i \geq 0,\ \ i = 1, 2, \ldots.$$
$$\sum_i p_i = 1.$$

$$F(x) = P(X \leq x) = \sum_{i: p_i \leq x} P(X = a_i) = \sum_{i: p_i \leq x} p_i$$

例 2.2.2 设随机变量 X 只取非负整数值 $0,1,2,\cdots$，且 $P(X=k)=c3^{-k}$，其中 c 是未知参数，求 X 的概率分布.

解 根据命题 2.2.1，

$$1=\sum_{k=0}^{\infty}P(X=k)=\sum_{k=0}^{\infty}c3^{-k}=3c/2,$$

因而 $c=2/3$，而 X 有分布列

$$P(X=k)=2\times 3^{-(k+1)},\quad k=0,1,2,\cdots.$$

Bernoulli 分布

如果随机变量 X 只取值 0 和 1，有分布列

$$P(X=1)=p,\quad P(X=0)=q,$$

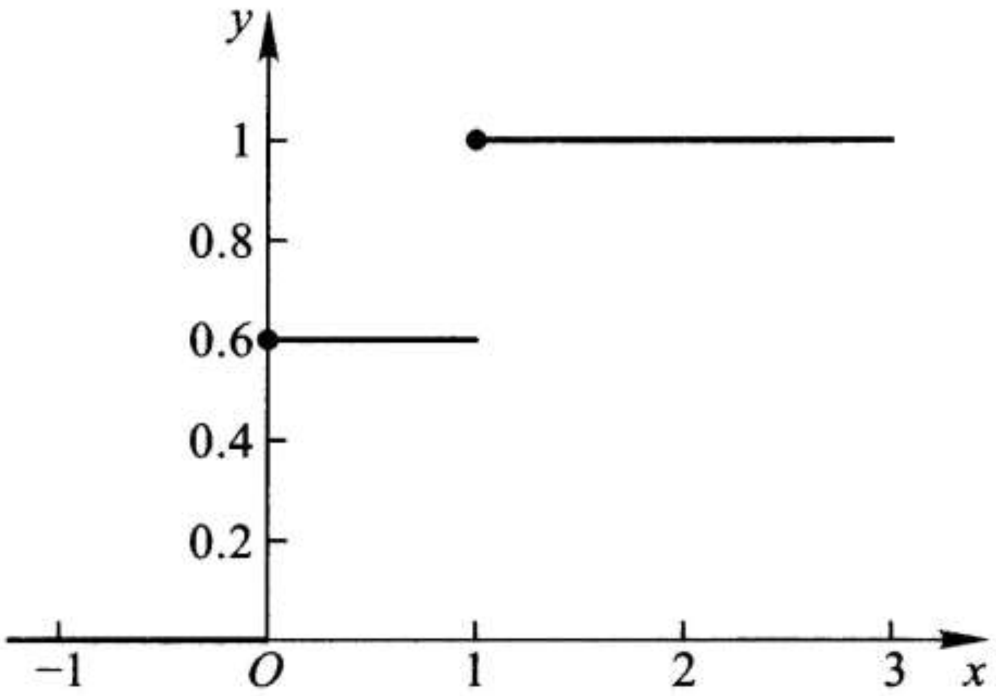
其中 $0\leqslant p\leqslant 1$，$p+q=1$，则称 X 服从参数为 p 的**伯努利**(Bernoulli)**分布**(亦有人称为 0 - 1 **分布或两点分布**)$B(1,p)$，记为 $X\sim B(1,p)$.

$$\begin{array}{cc} 0 & 1 \\ 1-p & p \end{array}$$

例 2.2.4 设射击一次，命中率为 0.4. 若用 $X=1$ 来表示"中"，用 $X=0$ 来表示"不中"，则

$$P(X=1)=0.4,\quad P(X=0)=0.6.$$

$$F(x)=\begin{cases}0, & x<0,\\ 0.6, & 0\leqslant x<1,\\ 1, & x\geqslant 1.\end{cases}$$



二项分布

设某事件 A 在一次试验中发生的概率为 p. 现把试验独立地重复 n 次. 以 X 记 A 在这 n 次试验中发生的次数，则 X 取值 $0,1,...,n$，且有

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\cdots,n.$$

称 X 服从二项分布，记为 $X\sim B(n,p)$.

一个变量服从二项分布有两个条件：

- 各次试验的条件是稳定的，这保证了事件 A 的概率 p 在各次试验中保持不变
- 各次试验的独立性

例2.3.1. 按规定, 某种型号的电子元件的使用寿命超过1500小时的为一级品. 已知某一大批产品的一级品率为0.2, 现随机地抽查20件产品, 问这20件产品中恰有$k(k=0,1,\cdots,20)$件一级品的概率是多少。

解: 这是不放回抽样，但是由于产品总数很大，而抽取的20件相对于总数来说很小，故我们可以视为是由放回的抽样。我们检查一个产品相当于作一次试验，抽查20件产品相对于做20次独立的贝努里试验，若以X表示20件中的一级品个数，则X服从二项分布$B(20,0.2)$，所以

$$P(X=k)=C_{20}^{k}0.2^{k}0.8^{20-k},\quad k=0,1,\cdots,20.$$

例2.3.2. 某人进行射击，设每次射击的命中率为0.02。现独立的重复射击400次，求至少击中两次的概率。

解：设X表示射击400次中的击中的次数，则$X \sim B(400, 0.02)$。所以

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.98^{400} - 400 * 0.02 * 0.98^{399} = 0.9972.$$

本例说明小概率事件在试验次数足够多时必然发生。

Poisson 分布

如果随机变量 X 只取非负整数值 0,1,2,…,且分布列为

$$P(X=k)=e^{-\lambda}\lambda^k/k!,\quad k=0,1,2,\cdots,$$

其中 $\lambda>0$,则称 X 服从(参数为 λ 的)**泊松分布** $P(\lambda)$,记为 $X\sim P(\lambda)$.

Poisson分布通常用来对“一段连续的时间目标事件发生的次数”进行建模

Poisson 分布

定理 2.2.2 对任意正整数 k 和正实数 λ,

$$\lim_{n\to\infty} C_n^k (\lambda/n)^k (1-\lambda/n)^{n-k} = e^{-\lambda}\lambda^k/k!.$$

证* $\lim_{n\to\infty} C_n^k (\lambda/n)^k (1-\lambda/n)^{n-k} = \lim_{n\to\infty}\left(\frac{n!}{(n-k)!\ k!}\cdot\frac{\lambda^k}{n^k}\cdot(1-\lambda/n)^{n-k}\right)$

$$=\frac{\lambda^k}{k!}\lim_{n\to\infty}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\cdots\cdot\frac{n-k+1}{n}\right)\lim_{n\to\infty}(1-\lambda/n)^{-k}\lim_{n\to\infty}(1-\lambda/n)^n$$

$$=e^{-\lambda}\lambda^k/k!.$$

由上面的定理可以得到近似式:当 n 很大,$\lambda=np$ 不大时,

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

Poisson 分布

例 2.2.6 设当次射击打中目标的概率等于 0.001，如果射击 5000 次，试求至少两次打中目标的概率 p.

解 设打中目标的次数为 X，则 X 是随机变量，$X \sim B(5000, 0.001)$. 于是

$$p = P(X \geqslant 2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - C_{5000}^{0} 0.001^{0} \times (1-0.001)^{5000} - C_{5000}^{1} 0.001 \times (1-0.001)^{4999} = 0.959640.$$

而用泊松近似律(2.2.3)求近似值. 这时 $\lambda = np = 5000 \times 0.001 = 5$，所以

$$C_{5000}^{0} 0.001^{0} \times (1-0.001)^{5000} \approx e^{-5},$$

$$C_{5000}^{1} 0.001 \times (1-0.001)^{4999} \approx 5e^{-5},$$

从而

$$p \approx 1 - e^{-5} - 5e^{-5} = 0.959572.$$

Poisson 分布

例2.3.6. 假设一块放射性物质在单位时间内发射出的α粒子数ξ服从参数为λ的Poisson分布。而每个发射出来的α粒子被记录下来的概率是p，就是说有$q=1-p$的概率被记数器漏记。如果各粒子是否被记数器记录是相互独立的，试求记录下来的α粒子数η的分布。

解：以事件$\{\xi=n\}, n=0,1,2,\cdots$为划分，则由全概率公式有

对比课本例2.2.8

$$\begin{aligned}
P(\eta=k) &= \sum_{n=0}^{\infty} P(\eta=k|\xi=n)P(\xi=n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \sum_{n=k}^{\infty} \frac{(\lambda q)^{n-k}}{k!(n-k)!} e^{-\lambda} (\lambda p)^k = \frac{(\lambda p)^k}{k!} e^{-\lambda p}, \quad k=0,1,2,\cdots .\#
\end{aligned}$$

Pascal分布

在可列重贝努里试验中，若以 X_r 表示第 r 次成功发生时的试验次数，则 X_r 的分布律为

$$\begin{aligned}
P(X_r=k) &= P(\{\text{前}k-1\text{次恰有}r-1\text{次成功且第}k\text{次成功}\}) \\
&= P(\{\text{前}k-1\text{次恰有}r-1\text{次成功}\})P(\{\text{第}k\text{次成功}\}) \\
&= C_{k-1}^{r-1}p^{r-1}q^{k-r}\cdot p \\
&= C_{k-1}^{r-1}p^{r}q^{k-r}, \quad k=r,r+1,\cdots.
\end{aligned}$$

称此概率分布为 Pascal 分布。

几何分布

如果随机变量 X 只取正整数值 $1,2,\cdots$ 且分布列为

$$P(X=k)=q^{k-1}p,\quad k=1,2,\cdots.$$

称此分布为几何分布. 记为 $X\sim G(p)$.

若以X表示在可列重贝努里试验中结果A出现时的试验次数，

X表示首次成功时的试验次数

几何分布

定理 2.3.1. 以所有正整数为取值集合的随机变量ξ服从几何分布$G(p)$，当且仅当对任何正整数m和n,都有

$$P(\xi > m + n \mid \xi > m) = P(\xi > n).$$

这个性质称为几何分布的无记忆性(*memoryless property*).

几何分布是唯一的具有无记忆性的取值集合为正整数集的离散型分布.

几何分布

证: 设随机变量ξ服从几何分布$G(p)$,写$q=1-p$,那么对任何非负整数k,都有

$$P(\xi > k) = \sum_{j=k+1}^{\infty} P(\xi = j) = p \sum_{j=k+1}^{\infty} q^{j-1} = q^k .$$

所以对任何正整数m和n,都有

$$P(\xi > m+n \,|\, \xi > m) = \frac{P(\xi > m+n,\ \xi > m)}{P(\xi > m)}$$

$$= \frac{P(\xi > m+n)}{P(\xi > m)} = \frac{q^{m+n}}{q^n} = q^n = P(\xi > n).$$

超几何分布

设有产品 N 个,其中有 M 个是次品,从这些产品中任取 n 个. 以 X 记取出的产品中的次品数,则 X 是离散型随机变量,取值 $s,s+1,\cdots,t$,其中 $s=\max\{0,n-(N-M)\}$, $t=\min(M,n)$. 不难知道 X 有分布列

$$P(X=m)=\mathrm{C}_M^m\mathrm{C}_{N-M}^{n-m}/\mathrm{C}_N^n,\quad m=s,s+1,\cdots,t.$$

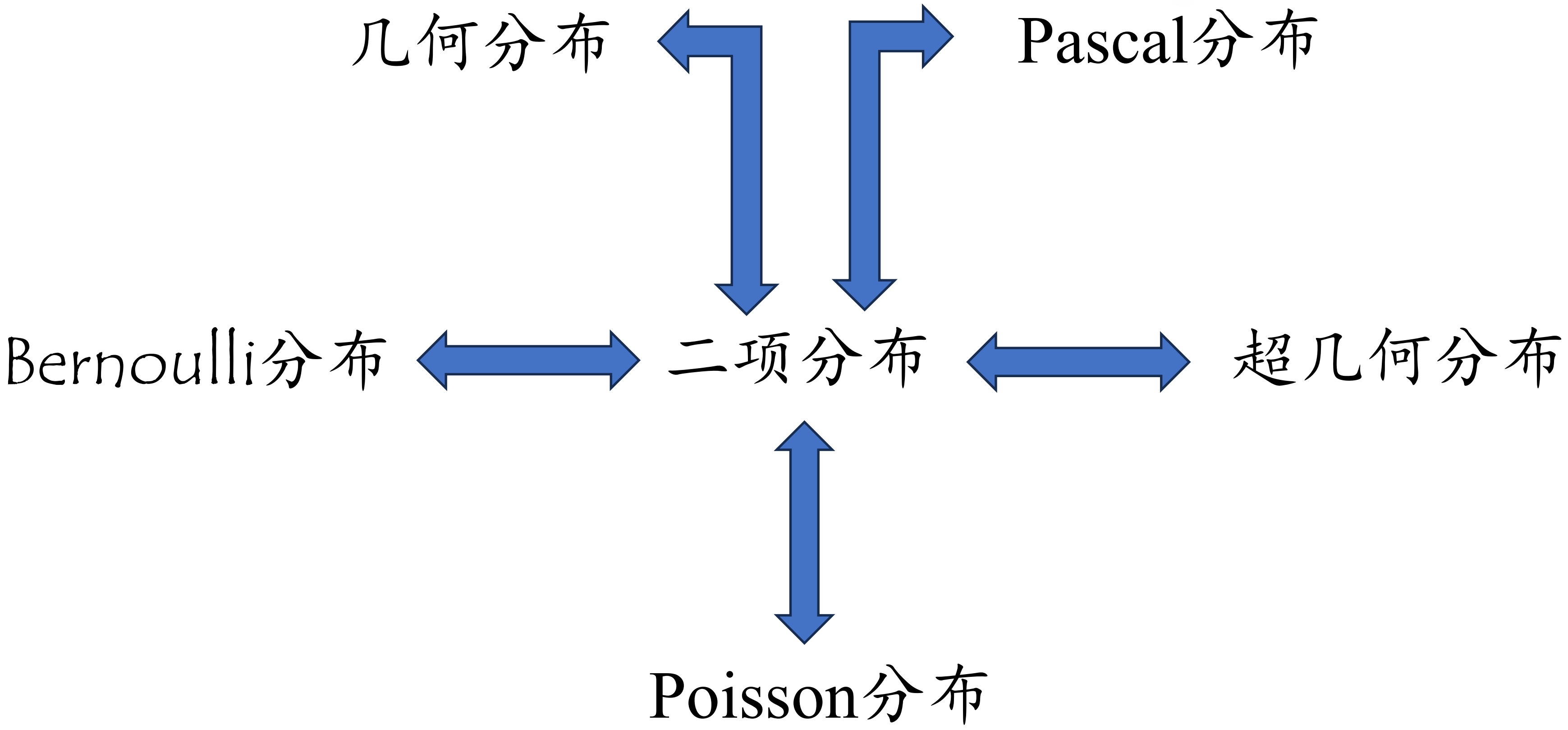
这个 X 的概率分布称为**超几何分布**.

超几何分布

定理 2.2.3 设当 $N\to\infty$ 时，$M/N\to p$，而 n,m 不变. 则当 $N\to\infty$ 时，

$$C_M^m C_{N-M}^{n-m}/C_N^n \to C_n^m p^m (1-p)^{n-m}.$$

二项分布用于对有放回抽样，超几何分布用于无放回抽样，当供抽样的总体容量N很大时，后者可以用前者逼近。


几何分布
Pascal分布
Bernoulli分布
二项分布
超几何分布
Poisson分布