Lecture 6 Binomial, Hypergeometric and Normal Distribution (optional)

1. 设随机变量 $X(\omega):\Omega\to\{1,2,3,\cdots,n,\cdots\}$，称 X 服从**几何分布**，若

$$P\{X=k\}=(1-p)^{k-1}p\ (k=1,2,\cdots,n,\cdots,\ 0<p<1).$$

同时记 $X\sim \mathrm{Ge}(p)$. 对这种随机变量的取值有可列个整数的情形，定义 $\mathrm{E}(X)=\sum_{k=1}^{\infty}x_kp_k=\lim_{n\to\infty}\sum_{k=1}^{n}x_kp_k$.

(1) 验证这是一个合法的分布列，即 $\sum_{k=1}^{\infty}P\{X=k\}=\lim_{n\to\infty}\sum_{k=1}^{n}P\{X=k\}=1$;

先验证几何级数 $\sum_{k=1}^{\infty}a_1q^{k-1}=\lim_{n\to\infty}\sum_{k=1}^{n}a_1q^{k-1}=a_1/(1-q)\ (|q|<1)$.

$$\lim_{n\to\infty}\frac{a_1[1-q^n]}{1-q}=\frac{a_1}{1-q}-\frac{1}{1-q}\lim_{n\to\infty}q^n=\frac{a_1}{1-q}$$

$$\lim_{n\to\infty}\sum_{k=1}^{n}(1-p)^{k-1}p=p\sum_{k=1}^{\infty}(1-p)^{k-1}=p\cdot\frac{1}{1-(1-p)}=\frac{p}{p}=1.$$

(2) 先待定 a,b 满足 $k(1-p)^{k-1}=[a(k+1)+b](1-p)^k-[ak+b](1-p)^{k-1}$，再证明 $\mathrm{E}(X)=1/p$;

$$\Big[(1-p)a(k+1)+b(1-p)-ak-b\Big](1-p)^{k-1}$$

$$\Big[\big((1-p)a-a\big)k+(1-p)(a+b)-b\Big]$$

$$\Rightarrow\begin{cases}-pa=1 & \Rightarrow a=-\frac{1}{p}\\ (1-p)a+(-p)b=0 & \Rightarrow b=\frac{1-p}{p}\cdot-\frac{1}{p}=\frac{-1+p}{p^2}=\frac{p-1}{p^2}\end{cases}$$

$$E(X)=\lim_{n\to\infty}p\sum_{k=1}^{n}k(1-p)^{k-1}=\lim_{n\to\infty}p\Big\{\big[a(n+1)+b\big](1-p)^n-(a+b)\Big\}$$

同级增长（同阶无穷）

$$\lim_{n\to\infty}\boxed{n\cdot\frac{1}{2^n}}=\lim_{n\to\infty}\frac{0}{2^n}=0$$

消掉

$$=-(a+b)p.=\cancel{-}p.\frac{\cancel{p-1}-p}{p^2}=1\cdot\frac{p}{p^2}=\boxed{\frac{1}{p}}.$$

(3) 用 $D(X)=E(X^2)-E^2(X)$ 来证明 $D(X)=(1-p)/p^2$;

$$E(X^2)=\sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = \lim_{n\to\infty} p\left[\{a(k+1)^2+b(k+1)+c\}(1-p)^n-(a+b+c)\right]$$

$$[a(k+1)^2+b(k+1)+c](1-p)^k-[ak^2+bk+c](1-p)^{k-1} = -(a+b+c)\cdot p.$$

$$=\left[(1-p)a(k+1)^2+(1-p)b(k+1)+c(1-p)-ak^2-bk-c\right](1-p)^{k-1}$$

$$\left\{[(1-p)a-a]k^2+[2(1-p)a+(1-p)b-b]k+(1-p)a+(1-p)b\overset{-pc}{\boxed{-c+c(1-p)}}\right\}$$

$$\Rightarrow \begin{cases} -pa=1 \\ (1-p)\cdot 2a+(-p)b=0 \\ (1-p)(a+b)-pc=0 \end{cases}$$

$a=-\frac{1}{p}$

$b=\frac{1-p}{p}\cdot\frac{-2}{p}=\frac{2p-2}{p^2}$

$c=(1-p)\cdot\frac{-p+2p-2}{p^2}\cdot\frac{1}{p}$

$=\frac{(1-p)(p-2)}{p^3}$

$$E(X^2)=-(a+b+c)\cdot p=-\frac{\{-p^2+2p^2-2p+(p-2-p^2+2p)\}}{p^3}\cdot p$$

$$=\frac{2-p}{p^2}$$

$$D(X)=E(X^2)-E^2(X)=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}$$ $\square$.

(4) 几何分布有一个重要的性质：**无记忆性**。也就是对 $m, n \in \mathbb{N}^*$ 有

$$\boxed{P\{X = m+n \mid X > n\} = P\{X = m\}. \rightsquigarrow P\{X > m+n \mid X > n\} = P\{X > m\}.}$$

为了证明上述性质你需要一点准备工作，先验证几何级数 $\sum_{k=1}^{\infty} a_1 q^{n-1} = \lim_{n\to\infty} \sum_{k=1}^{n} a_1 q^{n-1} = a_1/(1-q)$ $(|q| < 1)$.

这一点在你求解 $P\{X > n\} = \sum_{k=n+1}^{\infty} P\{X = k\}$ 时会使用到。

首次成功　　1失败. 2成功　　1.2失败 3成功

$$\xi(x) = 1 \cdot p + 2 \cdot (1-p)^1 \cdot p + 3 (1-p)^2 \cdot p + \cdots$$

$$= 1 \cdot p + (1-p)\left[(1+1) \cdot p + (1+2) \cdot p \cdot (1-p) + (1+3) \cdot p (1-p)^2 + \cdots\right]$$

$$\left[p + (1-p)p + (1-p)^2 p + \cdots\right] \qquad 1 \cdot p + 2 \cdot p(1-p) + 3 \cdot p(1-p)^2 + \cdots$$

$= 1$ 全部分布列　　　　$\xi(x)$

2成功 | 1失败　　3成功, 2失败 | 1失败

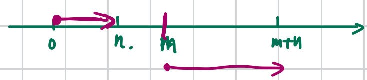
$$P\{X=2 \mid X>1\} = P\{X=1\} \qquad P\{X=3 \mid X>1\} = P\{X=2\}$$

$$= 1 \cdot p + (1-p)\left[1 + \xi(x)\right]$$

$$\forall m, n \in N^* \; \sum_{n=1}^{\infty} P\{X = m+n \mid \boxed{X > m}\} = \sum_{n=1}^{\infty} P\{X = n\}$$

再做n次到第n次才成功　　前m次全失败　　直接实验n次，在第n次首次成功

$$P\{X > m+n \mid X > m\} = P\{X > n\}$$



$$\frac{P\{X = m+n, X > m\}}{P\{X > m\}} = \frac{P\{X = m+n\}}{P\{X > m\}} = \frac{(1-p)^{m+n-1} \cdot \cancel{p}}{\sum_{k=m+1}^{\infty} (1-p)^{k-1} \cancel{p}} = \frac{(1-p)^{m+n-1}}{\frac{\cancel{(1-p)^m}}{1-(1-p)}} = (1-p)^{n-1} p$$

$$1-(1-p) = p \qquad = P\{X = n\}$$

2. 设 $X_1(\omega), X_2(\omega)$ 都是定义在样本空间 Ω 上的随机变量，且 $X_i : \Omega \to \{1,2,\cdots,n\}$. 我们称

$$P\{X_1 = i, X_2 = j\} = p_{ij}\ (1 \le i,j \le n,\ p_{ij} \ge 0).$$

(2) 我们称 X_1, X_2 这两个随机变量是独立的，若

$$P\{X_1 = i, X_2 = j\} = P\{X_1 = i\} \cdot P\{X_2 = j\}\ (\forall\, 1 \le i,j \le n).$$

请验证在 X_1, X_2 独立的情况下有，$E(X_1X_2) = E(X_1)E(X_2)$.

$$E(X_1X_2) = \sum_{j=1}^{n}\sum_{i=1}^{n} ij\, P\{X_1=i, X_2=j\}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} ij\, P\{X_1=i\}\, P\{X_2=j\}$$
$$= \sum_{i=1}^{n} i\, P\{X_1=i\} \sum_{j=1}^{n} j\, P\{X_2=j\}$$
$$= \left(\sum_{i=1}^{n} i\, P\{X_1=i\}\right)\left(\sum_{j=1}^{n} j\, P\{X_2=j\}\right) = E(X_1)E(X_2)$$

$\square$

(3) 利用 $D(X) = E[(X - E(X))^2]$ 验证在 X_1, X_2 独立的情况下有

$$D(X_1 \pm X_2) = D(X_1) + D(X_2).$$

$$D(X_1+X_2) = E\left[\left(X_1+X_2-E(X_1)-E(X_2)\right)^2\right]$$
$$= E\left\{\left[(X_1-E(X_1))+(X_2-E(X_2))\right]^2\right\}$$
$$= E\left\{(X_1-E(X_1))^2 + 2(X_1-E(X_1))(X_2-E(X_2)) + (X_2-E(X_2))^2\right\}$$
$$= \underbrace{E\left[(X_1-E(X_1))^2\right]}_{D(X_1)} + \underbrace{E\left[(X_2-E(X_2))^2\right]}_{D(X_2)} + 2E\left[(X_1-E(X_1))(X_2-E(X_2))\right] \overset{?}{=} 0$$

Note: X_1, X_2 $\boxed{独立} \Rightarrow Cor(X_1, X_2) = 0$

Goal: X_1, X_2 独立

$\boxed{X_1, X_2\ 相关性}$

$$2E\left[X_1X_2 - E(X_1)X_2 - E(X_2)X_1 + E(X_1)E(X_2)\right]$$
$$= 2\left\{E(X_1X_2) - 2E(X_1)E(X_2) + E(X_1)E(X_2)\right\}$$
$$= 2\left\{E(X_1X_2) - E(X_1)E(X_2)\right\} = 0.$$

$\square$

def. 协方差 X_1, X_2 相关性

$$Cor(X_1, X_2) \overset{def}{=} E\left[(X_1-E(X_1))(X_2-E(X_2))\right]$$
$$= E(X_1X_2) - E(X_1)E(X_2).$$

二项分布

正 → 1　　反 → 0

X	0	1
P	$1-p$	p

$p=\frac{1}{2}$.

$A\in\Omega$　　$X(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \omega\notin A.\end{cases}$

两点分布 (示性函数)

$E(X)=p$　　$D(X)=E(X^2)-E^2(X)=(1-p)p.$

有放回摸球

a 红　b 黄　　$P(红)=\frac{a}{a+b}.$

二项分布　　n重伯努利试验. $\begin{cases}1^\circ & \text{每个伯努利试验只有两种可能取值，重复做}n\text{次}\\ 2^\circ & n\text{重伯努利试验相互独立.}\end{cases}$

$X=X_1+X_2+\cdots+X_n.$　　$X_i \underbrace{\ \text{i.i.d.}\ }_{\text{独立同分布}}$ 同一个两点分布

X_i 取值只有 0 和 1.

X的可能取值：$0,1,2,\cdots,n.$

分布列　★ $P\{X=k\}=C_n^k p^k(1-p)^{n-k}$　$(k=0,1,2,\cdots,n)$

$\sum_{k=0}^{n} C_n^k p^k(1-p)^{n-k}=(p+1-p)^n=1.$ 分布列✓

$X\sim B(n,p)$　n：伯努利试验重复n次；p：每次试验概率 = 随机试验概率 = E(两点分布)

数字特征.

$$E(X)=\sum_{k=0}^{n} k C_n^k p^k(1-p)^{n-k}$$
$$=\sum_{k=1}^{n} n C_{n-1}^{k-1} p^k(1-p)^{[n-1]-(k-1)}$$
$$=np\sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1}(1-p)^{[n-1]-(k-1)}$$
$$=np(p+1-p)^{n-1}=np.$$

$$E(X)=E(X_1+\cdots+X_n)$$
$$=E(X_1)+\cdots+E(X_n)$$
$$=p+\cdots+p=np.$$

$D(X)=E(X^2)-E^2(X)$

$$E(X^2)=\sum_{k=0}^{n} k^2 C_n^k p^k(1-p)^{n-k}\qquad k^2=k(k-1)+k$$
$$=\sum_{k=2}^{n} n(n-1)C_{n-2}^{k-2}p^{k-2}\cdot p^2\cdot(1-p)^{[n-2]-(k-2)}+E(X)$$
$$=n(n-1)p^2\cdot(p+1-p)^{n-2}+E(X)$$
$$=n(n-1)p^2+np.$$

$$D(X)=n(n-1)p^2+np-n^2p^2$$
$$=\cancel{n^2p^2-n^2p^2}-np^2+np=n(1-p)p.$$

$X_1=1\quad X_2=0\qquad A_1\quad \bar{A_2}$ 独立

$X_1=0\quad X_2=1 \Leftarrow \bar{A_1}\quad A_2$

$X_1=0,\ X_2=0\qquad \bar{A_1}\cdot\bar{A_2}$

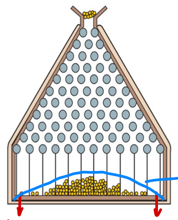
$D(X_1+\cdots+X_n)=D(X_1)+\cdots+D(X_n)=np(1-p)$

$\boxed{A_1\ A_2}$ — A_n 独立.

$P\{X_1=i, X_2=j\}=P\{X_1=i\}P\{X_2=j\}$ ✓

$i,j=0,1.$

4. 下图是一块高尔顿板的示意图。在一块木板上钉着与排列相互平行但相互错开的圆柱形小木钉，小木钉之间留有适当的空隙作为通道，前面挡有一块玻璃。将小球从顶端放入，小球下落的过程中，每次碰到小木钉后都会等概率地向左或向右落下，最后落入底部的格子中。格子从左到右分别编号为 0，1，2，…，10，用 X 表示小球最后落入格子的编号，求 X 的分布列。



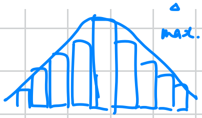
$X=0$ — $X=10$

向右的次数为X

$P\{X=k\}=\boxed{C_{10}^{k}}\left(\frac{1}{2}\right)^{10}$

$C_{10}^{0}\qquad C_{10}^{4}\quad C_{10}^{5}\quad C_{10}^{6}\ \cdots\ C_{10}^{10}$

max.



max.

$X\sim B(n,\frac{1}{2})$ $n\to\infty$

$X_1+\cdots+X_n$ $X_i \overset{iid}{\sim}$ 两点.

中心极限定理: $\sum_{k=1}^{n} X_k \longrightarrow$ 正态分布.

X独立同分布 分布.

近似连续

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

正态分布曲线.

5. 甲、乙两人选手进行象棋比赛，如果每局比赛甲获胜的概率为 0.6，乙获胜的概率为 0.4，那么采用 3 局 2 胜制还是采用 5 局 3 胜制对甲更有利？

[i导一]　✓✓　×✓✓　✓×✓

✓✓✓　ab^3

$C_3^1\ \boxed{\times\ \times\ \times}$ ✓ $C_3^1\ 0.6^3\cdot0.4$

$\boxed{C_4^2}$ ✓ $C_4^2\cdot0.4^2\cdot0.6^3$

[i导二] =二项分布

局限性: 胜了后比赛停了

假定赛满5局　$P\{X\geq3\}$?

✓ ✓ ✓

$P\{X\geq3\}$

$P\{X\geq3\}$

8. 一个数轴上的质点在随机外力的作用下，从原点 O 出发，每隔 1 秒随机向左或向右移动一个单位，设向右移动的概率为 p $(0 < p < 1)$，移动 n 次后位于位置 X_n.

(1) $p = 1/2$ 时，求 $P\{X_6 = 4\}$； $C_6^5\left(\frac{1}{2}\right)^6$

(2) 求 $E(X_n)$； 向右移动 Y 次 左 $n-Y$ $X_n = Y-(n-Y) = 2Y-n$ $E(X_n) = 2\cdot np - n = n(2p-1)$

(3) 移动 n 次后质点最有可能位于哪个位置？即求使得 $P\{X_n = x\}$ 最大的 x.

$x = [(n+1)p]$ 取整，$(n+1)p \notin \mathbb{Z}$.

$(n+1)p-1$ 或 $(n+1)p$，$(n+1)p \in \mathbb{Z}$.

$P\{X_n = x\} = C_n^x \cdot p^x \cdot (1-p)^{n-x} = f(x)$ 单调性.

$$\frac{f(x)}{f(x+1)} = \frac{C_n^x \cdot p^x \cdot (1-p)^{n-x}}{C_n^{x+1} p^{x+1} \cdot (1-p)^{n-x-1}} = \frac{C_n^x}{C_n^{x+1}} \cdot \frac{1-p}{p} = \frac{\frac{n!}{(n-x)!\,x!}}{\frac{n!}{(n-x-1)!\,(x+1)!}} \cdot \frac{1-p}{p} = \frac{(x+1)(1-p)}{(n-x)p} \geq 1 \; (<)$$

$(x+1)(1-p) \geq (n-x)p$

$= x[1-p+p] + (1-p) - np$

$= x - [np+p-1]$

$= x - [(n+1)p-1] \geq 0 \; (<)$

$= [(n+1)p]$ 取整 $x+1$

$\boxed{取 k = [(n+1)p-1] 取整}$ $+1$

当 $x - k \leq 0$ 时，

$\frac{f(k)}{f(k+1)} < 1. \Rightarrow f(k) < f(k+1)$

N^*

$\Rightarrow f(1) < \cdots < f(k) < \boxed{f(k+1)} > f(k+2) > \cdots$

$(n+1)p-1 \notin \mathbb{Z}$

当 $(n+1)p-1 \in \mathbb{Z}$ 时 取 $=$.

$\boxed{(n+1)p-1 \in \mathbb{Z}}$

当 $x \geq k+1$ 时，$x - [(n+1)p-1] \geq 0$

$f(k+1) > f(k+2) \cdots \to f(n)$

超几何分布（有放回模型）

引例 一批产品 N 件（5），其中有 M 件次品（4）。不放回从中抽 n 件（3）。

$X \overset{def}{=}$ 抽出的次品数。

- X 的可能取值：$k = m, m+1, \cdots, r$

  $m = \max\{0, \underbrace{n-(N-M)}_{至少要抽次品}\}$，$r = \min\{n, M\}$

- X 的分布列

  n件产品给次品选择位置

$$P\{X=k\} = \frac{C_M^k\, C_{N-M}^{n-k}}{C_N^n} = \frac{C_n^k\, A_M^k\, A_{N-M}^{n-k}}{A_N^n}$$

不考虑顺序　　　考虑

$$= \frac{\dfrac{M!}{(M-k)!\,k!}\cdot\dfrac{(N-M)!}{(N-M-n+k)!\,(n-k)!}}{\dfrac{N!}{(N-n)!\,n!}} = \frac{\dfrac{n!}{(n-k)!\,k!}\cdot\dfrac{M!}{(M-k)!}\cdot\dfrac{(N-M)!}{(N-M-n+k)!}}{\dfrac{N!}{(N-n)!}}$$

- X 的 $E(X)$、$D(X)$

分类加法：$\sum_{k=m}^{r} C_M^k C_{N-M}^{n-k} = C_N^n$

合法分布列：

$$\frac{\sum_{k=m}^{r} C_M^k C_{N-M}^{n-k}}{C_N^n} = \frac{C_N^n}{C_N^n} = 1.$$

$$E(X) = \frac{\sum_{k=m}^{r} \boxed{k\, C_M^k}\, C_{N-M}^{n-k}}{C_N^n} = \frac{M \sum_{k=m}^{r} C_{M-1}^{k-1} C_{N-M}^{n-k}}{C_N^n} = M\cdot\frac{C_{N-1}^{n-1}}{C_N^n} = M\cdot\frac{\dfrac{(N-1)!}{(N-n)!\,(n-1)!}}{\dfrac{N!}{(N-n)!\,n!}}$$

$$= M\cdot\frac{n}{N} = n\cdot\boxed{\frac{M}{N}}$$

概率 = 第1次抽出次品率

记 $p = \frac{M}{N} \Rightarrow E(X) = np$.

$$E(X^2)=\frac{\sum_{k=m}^{r} k(k-1)\, C_M^k\, C_{N-M}^{n-k}}{C_N^n}+E(X)$$

$$k(k-1)C_M^k = M(M-1)\,C_{M-2}^{k-2}$$

打和 $M(M-1)\,C_{N-2}^{n-2}$

$$E(X^2) = n(n-1)\cdot\frac{C_{N-2}^{n-2}}{C_N^n}+n\cdot\frac{M}{N}$$

$$= M(M-1)\cdot\frac{\dfrac{(N-2)!}{\cancel{(N-n)!}\,\cancel{(n-2)!}}}{\dfrac{\cancel{N!}}{\cancel{(N-n)!}\,\cancel{n!}}}\qquad \frac{\cancel{N!}}{(N-2)!}=N(N-1),\quad \frac{\cancel{n!}}{(n-2)!}=n(n-1)$$

$$=\frac{n(n-1)M(M-1)}{N(N-1)}+\frac{nM}{N}$$

$$D(X)=\frac{n(n-1)M(M-1)}{N(N-1)}+\frac{nM}{N}-\frac{n^2M^2}{N^2}=\frac{Nn(n-1)M(M-1)+N(N-1)nM-(N-1)n^2M^2}{N^2(N-1)}$$

$$=\frac{nM(N-M)(N-n)}{N^2(N-1)}$$

$$=n\,\frac{M}{N}\cdot\underbrace{\left(1-\frac{M}{N}\right)}_{=\frac{N-M}{N}}\cdot\boxed{\frac{N-n}{N-1}}\ ☆$$

超几何 $\underline{\underline{D(X)}}$ 小 $=np(1-p)\cdot\boxed{\frac{N-n}{N-1}}$ ☆ $\underline{N-n} < N-1$

二项 $D(X)=np(1-p)$