第十二讲. 离散分布

12.1 伯努利分布

伯努利分布：如果随机变量 X 满足 $\text{Prob}(X=1)=p$ 且 $\text{Prob}(X=0)=1-p$，那么 X 就服从参数为 $p\in[0,1]$ 的**伯努利分布**. 我们把结果 1 看作**成功**，把结果 0 看作**失败**，并记 $X\sim\text{Bern}(p)$. X 也可以称为**二元标示随机变量**.

$$\mu_X = p$$

$$\text{Var}(X) = p\cdot(1-p)$$

12.2 二项分布

二项分布：设 n 是一个正整数，并设 $p\in[0,1]$. 如果随机变量 X 满足：

$$\text{Prob}(X=k)=\begin{cases}\binom{n}{k}p^k(1-p)^{n-k} & \text{若 } k\in\{0,1,\cdots,n\}\\ 0 & \text{其他.}\end{cases}$$

那么 X 就服从参数为 n 和 p 的**二项分布**，并记 $X\sim\text{Bin}(n,p)$. X 的均值是 np，方差是 $np(1-p)$.

均值、方差证明见 11 章.（微分恒等式）

可以看作重复 n 次的伯努利实验.

证法2（线性性质）：

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

$$E(X) = E(X_1 + X_2 + \cdots + X_n)$$

X_i是相互独立且同分布的.

$$\Rightarrow E(X) = n \cdot E(X_i) \qquad i = 1, 2, \cdots, n$$

$$X_i \sim Bern(p)$$

$$E(X) = np$$

$$Var(X) = Var(X_1 + X_2 + \cdots + X_n)$$

$$= n\,Var(X_i)$$

$$= n \cdot p \cdot (1-p)$$

得证.

12.3 多项分布

多项分布与多项式系数：设 n,k 是正整数且 $p_1,p_2,\cdots,p_n\in[0,1]$ 满足 $p_1+\cdots+p_n=1$. 设 $x_1,\cdots,x_n\in\{0,1,\cdots,n\}$ 满足 $x_1+\cdots+x_n=n$. 那么，相应的**多项式系数**就是

$$\binom{n}{x_1,x_2,\cdots,x_k}=\frac{n!}{x_1!x_2!\cdots x_k!},$$

且其余 x_i 的值都为 0. 仅当 $(x_1,\cdots,x_k)$ 满足上述条件时，参数为 n、k 和 $p_1,\cdots,p_k$ 的**多项分布**才不为 0，其概率密度函数为

$$\binom{n}{x_1,x_2,\cdots,x_k}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k},$$

记作 $X\sim\mathrm{Multinomial}(n,k,p_1,\cdots,p_k)$.

一个实验，重复 n 次，结果空间有 k 个元素，$0<i\le k$，第 i 种结果出现的概率是 p_i，次数是 x_i

计算可能的次序：

$$\binom{n}{x_1}\binom{n-x_1}{x_2}\cdot\cdots\cdot\binom{n-x_1-x_2-\cdots-x_{k-1}}{x_k}$$

$$=\frac{n!}{(n-x_1)!\cdot x_1!}\cdot\frac{(n-x_1)!}{(n-x_1-x_2)!\cdot x_2!}\cdot\cdots\cdot\frac{(n-\cdots-x_{k-1})!}{(n-\cdots-x_k)!\cdot x_k!}$$

$$=\frac{n!}{x_1!\cdot x_2!\cdot\cdots\cdot x_k!}$$

记作 $\binom{n}{x_1,x_2,\cdots x_k}$，多项式系数

法2：条件概率.

假设有A、B、C三结果，先记作A与$\bar{A}$.
概率分别为P_1、P_2、P_3.

$$Pr(A发生k_1次) = P_1^{k_1}\cdot(1-P_1)^{n-k_1}\cdot\binom{n}{k_1}$$

已知A不发生，B的概率 $= \dfrac{P_2}{1-P_1}$

$C = \dfrac{P_3}{1-P_1}$

则 $Pr(A\ k_1次,\ B\ k_2次,\ C\ k_3次)$

$$= P_1^{k_1}\binom{n}{k_1}\cdot(1-P_1)^{n-k_1}\cdot\binom{n-k_1}{k_2}\cdot\left(\frac{P_2}{1-P_1}\right)^{k_2}\cdot\left(\frac{P_3}{1-P_1}\right)^{k_3}$$

$$= P_1^{k_1}\cdot\frac{n!}{\cancel{(n-k_1)!}\cdot k_1!}\cdot\frac{\cancel{(n-k_1)!}}{(n-k_1-k_2)!\cdot k_2!}\cdot\frac{P_2^{k_2}\cdot P_3^{k_3}}{(1-P_1)^{k_2}\cdot(1-P_1)^{k_3}}\cdot(1-P_1)^{n-k_1}$$

$$= P_1^{k_1}\cdot\frac{n!}{k_3!\cdot k_1!\cdot k_2!}\cdot P_2^{k_2}\cdot P_3^{k_3}$$

$(k_1+k_2+k_3=n)$

由数学归纳法可以得到一般性结论.

例. ①, ②, ③三种巧克力在罐子中.
拿①的概率是0.45, ②是0.3, ③是0.25

问: 有放回地取6块, ①②③分别有3, 2, 1块的概率?

$$\Pr(X_1=3, X_2=2, X_3=1)$$

$$=\binom{6}{1,2,3}\cdot(0.45)^3\cdot(0.3)^2\cdot(0.25)^1$$

$$=\frac{6!}{1!\cdot 2!\cdot 3!}\cdot \cdots$$

$$\doteq 0.123$$

12.4 几何分布

几何分布: 设 $p \in [0,1]$. 如果随机变量 X 满足

$$\mathrm{Prob}(X=n)=\begin{cases} p(1-p)^{n-1} & \text{若 } n \in \{1,2,3,\cdots\} \\ 0 & \text{其他,} \end{cases}$$

∞

那么 X 就服从参数为 p 的几何分布, 并记作 $X \sim \mathrm{Geom}(p)$. 它的均值为 $1/p$, 方差为 $\frac{1-p}{p^2}$.

随机变量X表示首次成功时已经完成的实验次数.

均值：$\mu_X = \sum_{n=1}^{\infty} np\cdot(1-p)^{n-1}$

$$= p\cdot\sum_{n=1}^{\infty} n\cdot(1-p)^{n-1}$$

$$\sum_{n=1}^{\infty}(1-p)^n = \frac{(1-p)\cdot[1-(1-p)^n]}{1-(1-p)}$$

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

$$\Rightarrow -\sum_{n=1}^{\infty} n\cdot(1-p)^{n-1} = -\frac{1}{p^2}$$

故 $\mu_X = p\cdot\frac{1}{p^2} = \frac{1}{p}$

方法2：无记忆过程

$$\mu_X = 1\cdot p + (\mu_X+1)\cdot(1-p)$$

$$\Rightarrow \mu_X = \frac{1}{p}$$

方差：

由 $1 = \sum_{n=1}^{\infty} p \cdot (1-p)^{n-1}$

$$p \cdot \frac{d}{dp} 1 = p \cdot \frac{d}{dp} \cdot \sum_{n=1}^{\infty} p(1-p)^{n-1}$$

$$0 = p \cdot \sum_{n=1}^{\infty} \left[(1-p)^{n-1} - p \cdot (n-1) \cdot (1-p)^{n-2}\right]$$

$$0 = \sum_{n=1}^{\infty} p(1-p)^{n-1} - \sum_{n=1}^{\infty} p^2 (n-1)(1-p)^{n-2}$$

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

$$\frac{d}{dp} \cdot 0 = -\frac{d}{dp}\left(\sum_{n=1}^{\infty} n p^2 (1-p)^{n-2} - \frac{p}{1-p}\right)$$

$$0 = \sum_{n=1}^{\infty}\left[-n \cdot 2p \cdot (1-p)^{n-2} + np^2 (n-2)(1-p)^{n-3}\right] + \frac{1}{(1-p)^2}$$

$$0 = -2 \cdot E(x) \cdot \frac{1}{1-p} + \sum_{n=1}^{\infty}\left[n^2 p^2 (1-p)^{n-3} - 2np^2(1-p)^{n-3}\right] + \frac{1}{(1-p)^2}$$

$$0 = -2E(x) \cdot \frac{1}{1-p} + \frac{p}{(1-p)^2} \cdot E(x^2) - \frac{2p}{(1-p)^2} E(x) + \frac{1}{(1-p)^2}$$

$$= -\frac{2}{p} + \frac{p}{(1-p)} E(X^2) - \frac{1}{(1-p)}$$

$$E(X^2) = \frac{2-p}{p^2}$$

$$\Rightarrow Var(X) = E(X^2) - E^2(X)$$

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

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

12.5 负二项分布

负二项分布：设 r 是一个正整数，并且 $p \in [0,1]$．假设抛掷一枚硬币获得成功（出现正面）的概率是 p．随机变量 X 表示当恰好出现 r 次失败（出现反面）时已经成功的次数．X 服从参数为 r 和 p 的**负二项分布**．它的概率密度函数为

$$\mathrm{Prob}(X=k) = \begin{cases} \binom{k+r-1}{k} p^k (1-p)^r & \text{若 } k \in \{0,1,2,\cdots\} \\ 0 & \text{其他}, \end{cases}$$

记作 $X \sim \mathrm{NegBin}(r,p)$．它的均值为 $\frac{pr}{1-p}$，方差为 $\frac{pr}{(1-p)^2}$．

当 $r=1$ 时.

$$\text{Prob}(X=k) = \begin{cases} p^k \cdot (1-p) & k=0,1,\cdots \\ 0 & \text{其它} \end{cases}$$

和几何分布很像.

负二项分布就是由几何分布推广而来.

均值:

$$1 = \sum_{k=1}^{\infty} \binom{k+r-1}{k} p^k (1-p)^r$$

$$p \cdot \frac{d}{dp} 1 = p \cdot \frac{d}{dp} \sum_{k=1}^{\infty} \binom{k+r-1}{k} p^k (1-p)^r$$

$$0 = p \cdot \sum_{k=1}^{\infty} \binom{k+r-1}{k} [k \cdot p^{k-1} (1-p)^r + p^k \cdot r \cdot (1-p)^{r-1} \cdot (-1)]$$

$$p \cdot \sum_{k=1}^{\infty} \binom{k+r-1}{k} p^k \cdot r \cdot (1-p)^{r-1} = \sum_{k=1}^{\infty} \binom{k+r-1}{k} \cdot k \cdot p^k \cdot (1-p)^r$$

$$\frac{pr}{1-p} \cdot 1 = \mu_X \quad ①$$

$$\Rightarrow \mu_X = \frac{pr}{1-p}$$

方差：

① 两边作用 $p\cdot\frac{d}{dp}$.

$$p\cdot\frac{d}{dp}\cdot r\cdot\left(-1+\frac{1}{1-p}\right)=p\cdot\frac{d}{dp}\mu_x$$

$$pr\cdot\frac{1}{(1-p)^2}=p\left[\sum_{k=1}^{\infty}\binom{k+r-1}{k}k^2p^{k-1}(1-p)^r-\sum_{k=1}^{\infty}\binom{k+r-1}{k}k\cdot p^k\cdot r\cdot(1-p)^{r-1}\right]$$

$$\frac{pr}{(1-p)^2}=E(X^2)-pr\frac{\mu_x}{1-p}$$

$$E(X^2)=\frac{pr}{(1-p)^2}+p\cdot\frac{pr^2}{(1-p)^2}$$

$$\Rightarrow Var(X)=\frac{pr}{(1-p)^2}+\frac{p^2r^2}{(1-p)^2}-\frac{p^2r^2}{(1-p)^2}$$

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

均值法2：无记忆过程（用于证明）

由归纳得到，$\mu_{p,r}=\frac{pr}{1-p}$

$$\mu_{p,r+1} = p\cdot(\mu_{p,r+1}+1) + (1-p)\cdot\mu_{p,r}$$

代入，得 $\mu_{p,r+1} = p\,\mu_{p,r+1} + p + (1-p)\cdot\dfrac{pr}{1-p}$

$$(1-p)\mu_{p,r+1} = p\cdot(1+r)$$

$$\Rightarrow \mu_{p,r+1} = \frac{p\cdot(1+r)}{1-p}.$$

得证.

例、某机器故障率为 2%，累积故障 5 次后宣布报废. 问：该机器预计能用几次？

设 X 是能用的次数.

$$X \sim NegBin(5, 0.98)$$

$$\mu_X = \frac{pr}{1-p}$$

$$= \frac{0.98\times 5}{0.02}$$

$$= 245$$

12.6 泊松分布

泊松分布：设 $\lambda > 0$. 如果**随机变量** X 满足

$$\mathrm{Prob}(X = n) = \begin{cases} \lambda^n e^{-\lambda}/n! & \text{若 } n \in \{0, 1, 2, \cdots\} \\ 0 & \text{其他.} \end{cases}$$

那么 X 就服从参数为 λ 的**泊松分布**, 并记作 $X \sim \mathrm{Pois}(\lambda)$. X 的均值和方差都是 λ.

可以被定义为参数为 $n \cdot p$ 的二项分布极限. 其中 $n \to \infty$. $np_n \to \lambda$

首先, 证明是 pdf:

$$\sum_{n=0}^{\infty} \lambda^n e^{-\lambda}/n!$$

$$= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$$

由 $e^x = \sum_{n=0}^{\infty} \frac{x^n}{x!}$

原式 $= e^{-\lambda} \cdot e^{\lambda} = 1$.

得证.

均值:

$$\mu_X = \sum_{n=0}^{\infty} n \cdot \lambda^n \cdot \frac{e^{-\lambda}}{n!}$$

$$1 = \sum_{n=0}^{\infty} \lambda^n \frac{e^{-\lambda}}{n!}$$

$$\lambda \cdot \frac{d}{d\lambda} \cdot 1 = \lambda \cdot \sum \frac{d}{d\lambda} \lambda^n \cdot \frac{e^{-\lambda}}{n!}$$

$$0 = \lambda \cdot \sum_{n=0}^{\infty} \Big[\frac{1}{n!} \cdot n \cdot \lambda^{n-1} \cdot e^{-\lambda} +$$

$$\frac{1}{n!} \cdot \lambda^n \cdot e^{-\lambda} \cdot (-1) \Big]$$

$$0 = \sum_{n=0}^{\infty} n \cdot \frac{\lambda^n \cdot e^{-\lambda}}{n!} - \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \cdot \lambda$$

$$\mu_x = \lambda \quad ②$$

方差：

② 等式两边同时作用 $\lambda \cdot \frac{d}{d\lambda}$

$$\lambda = \lambda \cdot \sum_{n=0}^{\infty} \Big(n^2 \cdot \frac{\lambda^{n-1}}{n!} e^{-\lambda} - \frac{n \cdot \lambda^n \cdot e^{-\lambda}}{n!} \Big)$$

$$\lambda = E(x^2) - \lambda \cdot \mu_x$$

$$E(x^2) = \lambda + \lambda \cdot \lambda$$

$$= \lambda + \lambda^2$$

$$Var(x) = E(x^2) - E^2(x)$$

$$= \lambda .$$

例. 今日进入邮局的顾客数服从 $\lambda=\frac{1}{3}$ 的泊松分布. 今天至少有一位顾客进入邮局的概率是:

设 X 为进入邮局的人数.

$$X \sim Pois(\tfrac{1}{3})$$

$$Pr(X=0) = (\tfrac{1}{3})^0 \cdot \frac{e^{-\frac{1}{3}}}{0!}$$

$$= e^{-\frac{1}{3}}$$

$$Pr(X \geq 1) = 1-e^{-\frac{1}{3}} \doteq 0.283$$

12.7 离散均匀分布.

离散均匀分布: 设 $\{a_1, a_2, \cdots, a_n\}$ 是一个有限集. 如果 X 满足

$$\text{Prob}(X=a) = \begin{cases} 1/n & \text{若 } a \in \{a_1, a_2, \cdots, a_n\} \\ 0 & \text{其他.} \end{cases}$$

那么 X 就是一个服从**离散均匀分布**的**随机变量**. 最重要的一种情况出现在上述集合为 $\{a, a+1, a+2, \cdots, a+n-1\}$ 时. 此时, X 的均值为 $a+\frac{n-1}{2}$, 方差为 $\frac{n^2-1}{12}$.

只关注重要情况:

当 $a=0$ 时.

$$\mu_{x_0} = \frac{0+1+2+\cdots+n-1}{n} = \frac{n\cdot(n-1)}{2n} = \frac{n-1}{2}$$

在一般情况下:

$$\mu_x = a + \mu_{x_0} = a + \frac{n-1}{2}$$

方差:

$a=0$ 时:

$$prob(x=k) = \begin{cases} \frac{1}{n} & k \in \{0,1,2,\cdots,n-1\} \\ 0 & \end{cases}$$

$$Var(x) = \frac{(0-\mu_x)^2 + (1-\mu_x)^2 + \cdots + (n-1-\mu_x)^2}{n}$$

连续平方和公式:

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

$$Var(x) = \frac{1}{n}\left(\sum_{k=1}^{\mu_x} k^2 - \sum_{k=1}^{\mu_x+1-n} k^2\right)$$

$$= \frac{\mu_x(\mu_x+1)(2\mu_x+1)}{6n} - \frac{(\mu_x+1-n)\cdots}{6n}$$

$$= \frac{n^2-1}{12}$$

例、德国坦克问题

最大序列号 = N

最小序列号 = 1

采样次数 = k

k 次采样中的最大序列号 = m

目标：已知 k、m，估计 N.

记 M 为观测到的最大序列.

$$P_r(M=m) = \frac{\binom{m-1}{k-1}}{\binom{m}{k}} \qquad m \in [k, n]$$

有 $$\sum_{m=k}^{N} \frac{\binom{m-1}{k-1}}{\binom{m}{k}} = 1$$

$$E(M) = \sum_{m=k}^{N} m\, P_r(M=m)$$

$$= \sum_{m=k}^{N} \frac{\frac{m!}{(k-1)!(m-k)!}}{\frac{N!}{k!(N-k)!}}$$

(重写代数表达式)

$$= \sum_{m=k}^{N} \frac{\frac{m!}{(k-1)!(m-k)!} \cdot \frac{k}{k}}{\frac{N!}{k!(N-k)!} \cdot \frac{N+1}{N+1} \cdot \frac{k+1}{k+1}}$$

$$= \frac{k}{k+1}(N+1) \sum_{m=k}^{N} \frac{\binom{m}{k}}{\binom{N+1}{k+1}}$$

$$= \frac{k}{k+1}(N+1) \sum_{u=k+1} \frac{\binom{u-1}{k+1-1}}{\binom{N+1}{k+1}}$$

$$= \frac{k}{k+1}(N+1) \cdot 1$$

$$\Rightarrow N = \left(1+\frac{1}{k}\right) \cdot E(M) - 1$$

$$= \left(1+\frac{1}{k}\right) m - 1$$

由此通过观测值 m 反推测 N.

Q: m 只是单次观测值，如何可以代替多次实验后才能得到的 $E(M)$?