ECE 313 Additional Note

概率质量函数（PMF）与典型离散分布例子

概率质量函数（PMF, Probability Mass Function）用于描述**离散随机变量**取某个具体值的概率。

设 X 是一个定义在可数样本空间 $S \subseteq \mathbb{R}$ 上的离散随机变量，则其概率质量函数为：

$$f_X(x) = \begin{cases} \Pr(X = x), & x \in S \\ 0, & x \in \mathbb{R} \setminus S \end{cases}$$

PMF 的定义覆盖整个实数域，对于不属于状态空间 S 的值，概率为 0。

伯努利分布（Bernoulli Distribution）$X \sim \mathrm{Bern}(p)$

场景

一次试验只有两个结果：成功（1）或失败（0）。

参数

- p：成功的概率
- $1 - p$：失败的概率

表达式

$$P_X(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

期望

$$E[X] = p$$

方差

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

示例

抛一枚硬币（正面为 1，反面为 0）：

- 如果硬币公平，则 $p = 0.5$，有：

$$P_X(x) = \begin{cases} \frac{1}{2}, & x \in \{0, 1\} \\ 0, & \text{otherwise} \end{cases}$$

二项分布（Binomial Distribution）$X \sim \mathrm{Bin}(n, p)$

场景

进行 n 次相同的伯努利试验，每次成功概率为 p，记录成功次数。

参数

- n：实验次数
- p：每次实验成功的概率

表达式

若 $X \sim \mathrm{Bin}(n, p)$，则：

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

期望

$$E[Y] = np$$

方差

$$\mathrm{Var}(Y) = np(1-p)$$

示例

抛 4 次硬币，每次成功概率为 $p = \frac{1}{4}$：

- $\Pr(X = 0) = \binom{4}{0} p^0 (1-p)^4$
- $\Pr(X = 1) = \binom{4}{1} p^1 (1-p)^3$
- $\Pr(X = 2) = \binom{4}{2} p^2 (1-p)^2$

几何分布（Geometric Distribution）$X \sim \mathrm{Geo}(p)$

场景

在一系列伯努利试验中，直到第一次成功所需的试验次数。

参数

- p：每次成功的概率

表达式

$$\Pr(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \ldots$$

期望

$$E[X] = \frac{1}{p}$$

方差

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

示例

假设某项任务每次完成的概率为 $p = 0.2$，则第一次成功发生在第 3 次的概率为：

$$\Pr(X = 3) = (1 - 0.2)^2 \cdot 0.2 = 0.64 \cdot 0.2 = 0.128$$

超几何分布（Hyper-geometric Distribution）$X \sim \mathrm{HypGeo}(N, K, n)$

场景

超几何分布描述的是：**从有限总体中不放回地抽取样本**时，某种类型元素出现的次数的分布。

例如：

- 总体中有 N 个元素，其中有 K 个是"成功"类型（比如红球），$N-K$ 个是"失败"类型（比如蓝球）。
- 从中抽取 n 个元素，**不放回。**
- 记 X 为抽到的"成功"类型（红球）个数，则 X 就服从超几何分布。

参数

- N：总体大小（如所有球的总数）
- K：总体中"成功"元素的数量
- n：抽样次数（样本量）
- k：感兴趣的成功次数

概率质量函数（PMF）

$$\Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \quad \max(0, n - (N - K)) \le k \le \min(K, n)$$

其中：

- $\binom{K}{k}$ 表示从 K 个成功中选 k 个的方法数。
- $\binom{N-K}{n-k}$ 表示从失败的 $N-K$ 个中选剩下的 $n-k$ 个。
- $\binom{N}{n}$ 是从 N 个总体中抽取 n 个样本的总方法数。

期望

$$E[X] = n \cdot \frac{K}{N}$$

方差（应该是不要求的，我后面的课程还没看，先放在这）

$$\mathrm{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$$

示例

假设有一个装有 20 个球的盒子：

- 其中有 7 个是红球（成功），13 个是蓝球（失败）
- 从中不放回地抽取 5 个球

问：抽到恰好 2 个红球的概率是多少？

代入公式得：

$$\Pr(X=2) = \frac{\binom{7}{2}\binom{13}{3}}{\binom{20}{5}} = \frac{21 \cdot 286}{15504} \approx 0.387$$

超几何分布 vs 二项分布

特点	超几何分布	二项分布
抽样方式	不放回	放回（或独立）
总体规模	有限且固定（N）	无需考虑总体
试验独立性	不独立（抽一个影响另一个）	独立
应用场景	抽牌、彩票、无放回抽签等	抛硬币、重复实验等

应用场景举例

- 从一堆零件中随机抽样检测不良品数量；
- 抽牌时计算抽到特定花色的概率；
- 抽奖游戏中中奖概率分析。

负二项分布（Negative Binomial Distribution） $X \sim \mathrm{NegBin}(r,p)$

场景

表示直到第 r 次成功所需的试验总次数（从 r 开始计数），每次试验成功概率为 p。

概率质量函数 (PMF)

$$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r},\quad k=r,r+1,r+2,\dots$$

期望

$$E[X]=\frac{r}{p}$$

方差

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

总结

分布类型	变量取值	参数	概率质量函数表达式
伯努利分布	$\{0,1\}$	p	$\Pr(X=x)=p^x(1-p)^{1-x}$
二项分布	$0,1,\dots,n$	n,p	$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$
几何分布	$1,2,3,\dots$	p	$\Pr(X=k)=(1-p)^{k-1}p$
负二项分布	$r,r+1,r+2,\dots$	r,p	$\Pr(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$
超几何分布	$0,1,\dots,\min(n,K)$	N,K,n	$\Pr(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

- **说明**:
  - 负二项分布的变量取值从 r 开始（第 (r) 次成功所需的试验次数），参数 r 是目标成功的次数，p 是单次成功概率。
  - 超几何分布的参数 N 是总体大小，K 是总体中成功的个数，n 是抽样次数，变量取值受限于 $\min(n,K)$