2.1 随机变量的定义

定义 2.1.1 **随机变量**是定义在样本空间 S 上的实数值函数, 通常用大写字母表示，例如 X

定义 2.1.2 取值于有限集合或由互不相同的实数组成的无穷序列 (即可数多个实数值) 的随机变量称为**离散随机变量**, 可取值与某区间中所有实数的随机变量成为连续随机变量.

定义 2.1.3 假设 X 是一个随机变量. 那么 X 的**分布**定义为所有如下形式的概率的集合：$P(X \in C)$ 对任意 $C \subset \mathbb{R}$. 其中 $\{X \in C\}$ 是 $\{\omega \in S : X(\omega) \in C\}$ 的简写 . ω 通常表示一个样本, 或是在随机过程中也成为一个轨道.

例题 2.1.1 假设一个硬币正面出现的概率为 p. 令 N 为投掷直到正面首次出现所需的次数. 假设投掷是相互独立的, 那么 N

是一个取值于 1,2,3,...... 的随机变量

- 每一次连续投掷是一次实验，这些实验组成了样本空间. 对应每一次实验 N 有唯一的取值, 即第一次正面在一次投掷序列中的位置, 因此它是一个随机变量. N 的分布为

$$P(N=1)=p$$

$$P(N=2)=(1-p)p$$

$$\vdots$$

$$P(N=n)=(1-p)^{n-1}p$$

$$\vdots$$

并且 $P(N\in C)=0$ 对任意 $C\backslash\{1,2,3,......\}$.

离散分布

定理 2.1.1 令 X 为一个随机变量，其取值为 $\{x_i\}_{i=1}^{\infty}$, $x_i \neq x_j$, $\forall i \neq j$. 那么其分布可以由如下式子来计算，对任意 $C \subset \mathbb{R}$,

$$P(X \in C) = \sum_{x_i \in C} P(X = x_i),$$

并且

$$\sum_{i} P(X = x_i) = 1.$$

■ 注意到 $\{X = x_i\} \cap \{X = x_j\} = \varnothing$, $\forall i \neq j$ 并运用概率公理.

对离散分布, $f(x) = P(X = x)$ 通常称为 X 的概率函数. 上面的定理说明对离散随机变量，概率函数即可确定其分布.

例题 2.1.2 连续投掷一个均匀硬币 10 次. 各次投掷相互独立. 令 X 为正面出现次数. 试确定 X 的概率函数.

■ 每一次实验包含 10 次投掷, 所有实验组成一个样本数目为 2^{10} 的样本空间. 根据题设，每一个样本具有概率 $\frac{1}{2^{10}}$. 容易看出 X 是定义在这些样本上的随机变量. 其取值范围是 $R = \{0, 1, ..., 10\}$. 对 $\forall x \in R$,

$$f(x) = P(X = x) = C_{10}^{x} \cdot \frac{1}{2^{10}}.$$

///

// Bernoulli 分布

定义 2.1.4 如果随机变量 Z 取值为 $\{0,1\}$, 并且 $P(Z=1)=p$, 那么我们称 Z 为具有参数 p 的 Bernoulli 随机变量. 其分布则称为具有参数 p 的 Bernoulli 分布.

二项分布

假设一台机器生产某产品, 所生产的产品以概率 p 出现次品. 假设机器生产各个产品的过程是相互独立的. 现对该机器生产的产品进行抽查, 每次检验抽取 n 个产品, 令 X 是在抽检的 n 个产品中次品的个数. 显然 X 是一个随机变量. 它的分布由以下概率函数确定:

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

定义 2.1.5 上面定义的随机变量 X 称为具有参数 n 和 p $(0<p<1)$ 的二项随机变量, 其分布则称为具有参数 n 和 p 的二项分布, 记为 $B(n,p)$.

二项分布 $B(n,p)$, $(n \geqslant 2,\ 0 < p < 1)$ 的概率 $P(X=k)$ 在 k 取什么值的时候达到最大？为此考察

$$\begin{aligned}\frac{P(X=k+1)}{P(X=k)} &= \frac{C_n^{k+1}p^{k+1}(1-p)^{n-k-1}}{C_n^k p^k (1-p)^{n-k}} \\ &= \frac{\dfrac{n!}{(k+1)!\,(n-k-1)!}p}{\dfrac{n!}{k!\,(n-k)!}(1-p)} \\ &= \frac{(n-k)\,p}{(k+1)(1-p)}.\end{aligned}$$

可见

$$\frac{P(X=k+1)}{P(X=k)} > 1$$

当且仅当

$$\frac{(n-k)\,p}{(k+1)\,(1-p)} > 1,$$

即

$$(n-k)\,p > (k+1)\,(1-p)\,,$$

化简得到

$$k < (n+1)\,p - 1.$$

令 $m = [(n+1)\,p]$ 为 $(n+1)\,p$ 的整数部分, 那么当 $(n+1)\,p$ 不是整数时, $P\,(X=k)$ 在 $k=m$ 处达到最大. 当 $(n+1)\,p$ 是整数时, $P\,(X=k)$ 在 $k=m-1, m$ 处达到同等最大值.

- Poisson 分布 $X \in \{0, 1, ...\}$, 参数 $\lambda > 0$,

$$P\,(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, k = 0, 1, ...,$$

记为 $Poisson\,(\lambda)$

例如单位时间内放射性物质释放的 α 粒子数量可以用 Poisson 分布模拟.

Poisson 分布 $Poisson\,(\lambda)$, $(\lambda > 0)$ 的概率 $P\,(X = k)$ 在 k 取什么值的时候达到最大？为此考察

$$\frac{P\left(X = k + 1\right)}{P\left(X = k\right)} = \frac{\dfrac{\lambda^{k+1}}{(k+1)!}e^{-\lambda}}{\dfrac{\lambda^{k}}{k!}e^{-\lambda}} = \frac{\lambda}{k+1}.$$

可见

$$\frac{P\left(X = k + 1\right)}{P\left(X = k\right)} > 1$$

当且仅当

$$\frac{\lambda}{k+1} > 1,$$

即

$$k < \lambda - 1,$$

令 $m = [\lambda]$ 为 λ 的整数部分, 那么当 λ 不是整数时, $P(X = k)$ 在 $k = m$ 处达到最大. 当 λ 是整数时, $P(X = k)$ 在 $k = m - 1$, m 处达到同等最大值.

二项分布的 Poisson 逼近

定理 2.1.2 若 $\forall n, 0 < p_n < 1$, 且 (p_n 简记为 p)

$$\lim_{n\to\infty} np = \lambda > 0.$$

那么对固定的 k,

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

- 首先展开

$$\begin{aligned}
&C_n^k p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\
&= \frac{1}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} (np)^k (1-p)^{n-k}
\end{aligned}$$

$$= \frac{1}{k!} 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (np)^k (1-p)^{n-k}.$$

由假设当 $n \to \infty$ 时 $p \to 0$, 又对固定的 k,

$$1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \to 1,$$

$$(np)^k \to \lambda^k,$$

$$(1-p)^{n-k} \to e^{-\lambda}.$$

因此

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

例题 2.1.3 每次打靶命中率为 $p = 0.001$. 射击 5000 次后, 求至少命中两次的概率.

■ 命中数目 X 为随机变量. 令 $n = 5000$, 根据二项分布, 要求的概率为

$$
\begin{aligned}
& 1 - P(X=0) - P(X=1) \\
& = 1 - (1-0.001)^{5000} - C_{5000}^{1} 0.001 (1-0.001)^{4999} \\
& \approx 0.9598.
\end{aligned}
$$

当 n 和 k 较大时, $P(X=k) = C_n^k p^k (1-p)^{n-k}$ 较难计算, 然而根据 Poisson 逼近就有

$$
P(X=k) \approx \frac{\lambda^k}{k!} e^{-\lambda},
$$

其中 $\lambda = 5000 \cdot 0.001 = 5$. 例如

$$P(X=0) = (1-0.001)^{5000} \approx e^{-5},$$
$$P(X=1) = C_{5000}^{1} 0.001 (1-0.001)^{4999} \approx 5e^{-5}.$$

超几何分布

设一批产品共有 N 个, 其中 D 个为不合格产品, 其余为合格产品, 现任取 n 个, $1 \leqslant n \leqslant N$. 则这 n 个产品中的不合格产品数 X 为一随机变量. 这里 $0 \leqslant D \leqslant N$, $1 \leqslant n \leqslant N$ 那么

$$P(X=k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n},\ k \in \{0, 1, \ldots, \min(n, D)\}.$$

约定: $i > j \geqslant 0$, $C_j^i = 0$, $C_0^0 = 1$.

超几何分布与二项分布

定理 2.1.3 设超几何分布中, D 是 N 的函数, 且满足

$$\lim_{N\to\infty}\frac{D}{N}=p.$$

其中 $0<p<1$. 固定 n. 那么对任意 $k\geqslant 0$,

$$\lim_{N\to\infty}\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}=C_n^k p^k(1-p)^{n-k}.$$

- 由假设 $N\to\infty$ 时, $D\to\infty$. 不妨认为 $n<D<N$.

$$\frac{C_D^k C_{N-D}^{n-k}}{C_N^n}$$