2. 随机变量及其分布

- 随机变量建立了样本空间与实数的联系
- 离散型、连续型、其它类型随机变量.

2.1 随机变量及其分布函数.

定义2.1 设Ω样本空间. $\mathcal{F}$ σ-域, $X:\Omega\to\mathbb{R}$实函数. 若$\forall x\in\mathbb{R}$, $\{\omega\in\Omega: X(\omega)\le x\}\in\mathcal{F}$, 则称$X$为随机变量. (可测性).

记号: $\{X\le x\}\triangleq\{\omega: X(\omega)\le x\}$.

例 掷硬币. $\Omega=\{正, 反\}$. 定义 $X:\Omega\to\mathbb{R}$

$$X(正)=1,\qquad X(反)=0.$$

若 $\mathcal{F}=\{\phi, \Omega, \{正\}, 反\}$, 则$X$为随机变量;
若 $\mathcal{F}=\{\phi, \Omega\}$, 则$X$不是随机变量.

注: (1) 若X为随机变量, 则 $\{X<x\}\in\mathcal{F}$, $\forall x\in\mathbb{R}$.

$$\because \{X<x\}=\bigcup_{n=1}^{\infty}\{X\le x-\tfrac{1}{n}\}.$$

(2) 若X为随机变量, 则 $\forall a<b$, $\{a<X\le b\}$, $\{X=a\}\in\mathcal{F}$.

定义2.2 设X为随机变量. 定义

$$F_X(x)=P(X\le x),\quad \forall x\in\mathbb{R}.$$

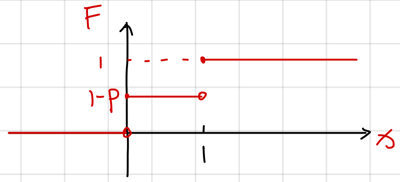
称F_X为X的分布函数.

记号: $F\triangleq F_X$.

例 $A\in\mathcal{F}$. 定义 $I_A(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \omega\notin A.\end{cases}$ 称I_A为A的示性函数.

则I_A为随机变量. $\{I_A\le x\}=\begin{cases}\phi, & x<0\\ \bar{A}, & 0\le x<1\\ \Omega, & x\geq 1\end{cases}$

$$F(x)=\begin{cases}0, & x<0,\\ 1-p, & 0\le x<1,\\ 1, & x\ge 1.\end{cases}\qquad p=P(A).$$



定理2.1 设F为X的分布函数. 则F有如下性质:

(1) 单调非降性: $\forall\, x_1\le x_2,\ F(x_1)\le F(x_2)$.

(2) 右连续性: $\lim\limits_{x\downarrow a} F(x)=F(a)$.

(3) 规范性: $F(-\infty)=0,\quad F(+\infty)=1$.

证明: 见板书.

注 若F满足定理2.1中三个性质, 则F为某个随机变量分布函数.

注. 设X分布函数为F. 则

(1) $P(X>b)=1-F(b)$. (2) $P(a<X\le b)=F(b)-F(a)$.

(3) $P(X<b)=F(b^-):=\lim\limits_{x\uparrow b}F(x)$

(4) $P(X=b)=F(b)-F(b^-)$.

注 (1) 概率问题 ⇒ 随机变量 ⇒ 分布函数.

(2) 随机变量 $\begin{cases}\text{离散型: 取值有限个或可列无限个}\\ \text{连续型}\\ \text{其它类型.}\end{cases}$

例*(Dimer problem) n 个不稳定的分子排成一列 $m_1, m_2, \cdots, m_n$. 从 $n-1$ 条边中随机选一条边, 该边连接的两个分子结合成一个稳定的二元聚合物; 依次类推, 直到所有不稳定分子不相邻. 令 P_n 表示最后 m_1 保持孤立的概率. 证 $P_n=\sum\limits_{r=0}^{n-1}\frac{(-1)^r}{r!}$.

提示: 假设第一次选择边 (m_r, m_{r+1}), $r\ge 2$, 则 $P_n=P_{r-1}$;

$$\therefore\ P_n=\sum_{r=2}^{n-1}\frac{1}{n-1}P_{r-1}\Rightarrow n P_{n+1}-(n-1)P_n=P_{n-1}$$

$$\Rightarrow P_{n+1}-P_n=\frac{-1}{n}(P_n-P_{n-1})=\frac{(-1)^{n-1}}{n!}(P_2-P_1)=\frac{(-1)^n}{n!}$$

$$\Rightarrow P_n=\sum_{r=2}^{n-1}(P_{r+1}-P_r)=\sum_{r=2}^{n-1}\frac{(-1)^r}{r!}=\sum_{r=0}^{n-1}\frac{(-1)^r}{r!}\qquad\square$$

2.2 离散型随机变量 (取值有限个或可列无限个)

2.2.1 分布列

设 $X: \Omega \to \{x_1, x_2, \cdots\}$ 记

$$P_i = P(X=x_i), \quad i=1,2,\cdots \qquad (2.1)$$

称 (2.1) 为 X 的概率分布或分布列.

性质. (1) 非负性: $P_i \geq 0$. (2) 规范性: $\sum_i P_i = 1$.

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

注: 分布函数与分布列相互决定

$$F(x) = \sum_{i \leq x} p_i, \qquad p_i = F(x_i) - F(x_i^-).$$

但分布列更直观.

2.2.2 常见的离散型分布

(1) 单点分布

$$P(X=c) = 1, \quad c \text{ 为常数}.$$

(2) 两点分布 (伯努利分布)

X	0	1
P	$1-p$	p

$0<p<1$ 掷不均匀硬币

例. 示性函数 I_A 服从参数为 $P(A)$ 的两点分布. (只关心事件A是否发生).

描述了伯努利试验.

(3) 二项分布　　　　　　　　　　　　　　掷 n 次不均匀硬币

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

其中 $0<p<1$. 称 X 服从参数为 (n,p) 的二项分布, 记为 $X\sim B(n,p)$.

n 重伯努利试验: 独立重复 n 次伯努利试验.

例1. 在 n 重伯努利试验中, 记 $p=P(A)$, X 为事件 A 发生的次数,

则 $X\sim B(n,p)$.

定理 2.2　设 $X\sim B(n,p)$, 则

$$P(X=[(n+1)p]) = \max_{0\le k\le n} P(X=k).$$

若 $(n+1)p$ 为整数, 则 $P(X=(n+1)p) = P(X=(n+1)p-1)$.

称 $[(n+1)p]$ 为 $B(n,p)$ 的最可能出现次数. $P(X=[(n+1)p])$ 为中心项.

证
$$\frac{P(X=k+1)}{P(X=k)} = \frac{C_n^{k+1} p^{k+1} (1-p)^{n-k-1}}{C_n^k p^k (1-p)^{n-k}} = \frac{n-k}{k+1}\frac{p}{1-p} \geq 1$$

当且仅当 $k\le (n+1)p-1$, 即

$$P(X=1) < P(X\le 2) < \cdots < P([(n+1)p]-1) \le P([(n+1)p])$$

$$P([(n+1)p]) > P([(n+1)p]+1) > \cdots > P(X=n).$$ □.

(4) 泊松分布

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

称 X 服从参数为 λ 的泊松分布, 记为 $X\sim P(\lambda)$.

定理 2.3　若 $X \sim B(n, \frac{\lambda}{n})$, $\lambda > 0$. 则

$$\lim_{n\to\infty} P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \forall k=0,1,2,\cdots$$

(可用泊松分布近似二项分布).

证.　$P(X=k) = C_n^k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$

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

$$\longrightarrow \frac{\lambda^k}{k!}e^{-\lambda}, \quad n\to\infty. \qquad \square$$

实际应用：进入商场的顾客数，书中印刷错误的个数，……

中心项：$\dfrac{P(X=k)}{P(X=k-1)} = \dfrac{\lambda}{k}$

$P(X=k) \geq P(X=k-1) \iff \lambda \geq k$

$\therefore$ $P(X=[\lambda])$ 为中心项.

例. 顾客数 $N \sim P(1500)$. 女性顾客占 $p=70\%$. 求女性顾客数 N_F 的分布列.

解. 记 $\lambda = 1500$. 对 $k=0,1,\cdots$,

$$P(N_F=k) = \sum_{n=k}^{\infty} P(N=n)P(N_F=k \mid N=n)$$

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

$$= \sum_{n=k}^{\infty}\frac{\lambda^n}{k!(n-k)!}e^{-\lambda}p^k(1-p)^{n-k}$$

$$= \frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{n=k}^{\infty}\frac{[\lambda(1-p)]^{n-k}}{(n-k)!}$$

$$= \frac{(\lambda p)^k}{k!}e^{-\lambda p}.$$

$\therefore \quad N_F \sim P(\lambda p) = P(1050)$. $\square$

(泊松分布的分解).

(5) 超几何分布.

$M \le N$, $n \le N$.

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k=0,\cdots,n.$$

称 X 服从超几何分布.

例. 共 N 个球, M 个白球. 抓 n 个球, 令 X 表示抓到的白球数. 则 X 服从超几何分布.

二项分布近似超几何分布 $\quad p=\frac{M}{N}$. N 很大, n 很小时,

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = \frac{M\cdot\cdots(M-k+1)\cdot(N-M)\cdots(N-M-n+k+1)}{N\cdots(N-k+1)\cdot(N-k)\cdots(N-n+1)} \frac{n!}{k!(n-k)!}$$

$$\approx p^k(1-p)^{n-k} C_n^k.$$

(6) 几何分布.

$0<p<1$.

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

例. 独立重复试验, 设事件 A 发生概率为 p. 令 X 表示事件 A 首次发生时所需试验次数. 则 X 服从几何分布.

定理 2.4. (几何分布的无记忆性) X 取值 $\{1,2,\cdots\}$. 则

X 服从几何分布 $\Leftrightarrow$ $P(X>m+n \mid X>m) = P(X>n)$, $\forall\, m,n \in \mathbb{Z}_+$.

$\mathbb{Z}_+ := \{1,2,\cdots\}$.

证. "$\Rightarrow$" $P(X>m+n \mid X>m) = \dfrac{P(X>m+n,\, X>m)}{P(X>m)}$

$$= \frac{P(X>m+n)}{P(X>m)} = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = P(X>n).$$

"$\Leftarrow$" 令 $g(n) = P(X>n)$. 则 $g(m+n) = g(m)g(n)$

$\Rightarrow$ $g(n) = g(1)^n$. 令 $p = 1-g(1)$.

$\Rightarrow$ $P(X>n) = (1-p)^n$, $P(X=n) = p(1-p)^{n-1}$, $n\geq 1$. $\blacksquare$

例 独立重复试验. A 发生概率为 p. 记 $X=$ A 恰好发生 r 次时总试验次数, $Y = X - r$. 求 X, Y 分布.

解: $\forall\, k \geq r$,

$$P(X=k) = C_{k-1}^{r-1}\, p^r (1-p)^{k-r}$$ 帕斯卡分布.

$\forall\, k \geq 0$,

$$P(Y=k) = P(X=k+r) = C_{k+r-1}^{r-1}\, p^r (1-p)^k$$

负二项分布. $\square$