

概率论与数理统计

随机变量与分布函数

量化，并更加有效的描述随机事件：

$$\{X \in I\} := \{\omega \in \Omega : X(\omega) \in I\}, \quad I \subset \mathbb{R}$$

随机变量 X 的分布函数定义为：

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]), \quad x \in \mathbb{R}$$

分布函数

$$(-\infty, x] = \bigcap_n^{\infty} \left(-\infty, x + \frac{1}{n} \right]$$

$$\begin{aligned} F(x) &= P(X \leq x) = P\left(X \in \bigcap_n^{\infty} \left(-\infty, x + \frac{1}{n} \right]\right) \\ &= P\left(\bigcap_n^{\infty} \left\{ X \in \left(-\infty, x + \frac{1}{n} \right] \right\}\right) \\ &= \lim_n P\left(X \in \left(-\infty, x + \frac{1}{n} \right]\right) \\ &= \lim_n F\left(x + \frac{1}{n}\right) \end{aligned}$$

分布函数

$$(-\infty, x) = \bigcup_n \left(-\infty, x - \frac{1}{n} \right]$$

$$P(X < x) = \lim_n P \left(X \in \left(-\infty, x - \frac{1}{n} \right] \right)$$

$$= \lim_n F \left(x - \frac{1}{n} \right)$$

$$(-\infty, \infty) = \bigcup_n (-\infty, n] \quad \emptyset = \bigcap_n (-\infty, -n]$$

$$F(-\infty) = 0 \quad F(\infty) = 1$$

二项分布

Bernoulli分布：投掷正面出现概率为p的硬币1次，正面出现次数X的分布

$$X \sim \text{Bernoulli}(p), 0 \leq p \leq 1 :$$

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

二项分布：投掷正面出现概率为p的硬币n次，正面出现次数X的分布

$$X \sim \text{Bin}(n, p), n \geq 1, 0 \leq p \leq 1 :$$

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}, k = 0, 1, 2, \dots$$

二项分布最大概率点

定理 2.1 设 $n \geq 2, 0 < p < 1, m = [(n+1)p]$ (不超过 $(n+1)p$ 的最大整数),

$$p_n(k) = C_n^k p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n),$$

则有下列结论:

(1) 当 $(n+1)p$ 不是整数时,

$$\begin{aligned} p_n(0) &< p_n(1) < \cdots < p_n(m-1) < p_n(m) \\ &> p_n(m+1) > \cdots > p_n(n); \end{aligned} \tag{2.4}$$

(2) 当 $(n+1)p$ 是整数时,

$$\begin{aligned} p_n(0) &< p_n(1) < \cdots < p_n(m-1) = p_n(m) \\ &> p_n(m+1) > \cdots > p_n(n). \end{aligned} \tag{2.5}$$

证明 显然

$$\frac{p_n(k+1)}{p_n(k)} = \frac{n-k}{k+1} \cdot \frac{p}{1-p},$$

又 $\frac{n-k}{k+1} \cdot \frac{p}{1-p} > 1$ 的充要条件是 $k < (n+1)p - 1$, 于是有下列结论:

$$\text{当 } k < (n+1)p - 1 \text{ 时, } p_n(k+1) > p_n(k); \quad (2.6)$$

$$\text{当 } k > (n+1)p - 1 \text{ 时, } p_n(k+1) < p_n(k); \quad (2.7)$$

$$\text{当 } k = (n+1)p - 1 \text{ 时, } p_n(k+1) = p_n(k). \quad (2.8)$$

$(n+1)p$ 不是整数:

$$[(n+1)p] - 1 < (n+1)p - 1 < [(n+1)p] < (n+1)p$$

$(n+1)p$ 是整数:

$$[(n+1)p] - 1 = (n+1)p - 1 < [(n+1)p] = (n+1)p$$

二项分布Poisson逼近

定理 7.2 如果 $0 < p_n < 1$ 且 $\lim_{n \rightarrow \infty} np_n = \lambda > 0$, 则

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (7.4)$$

证明 根据排列组合公式,有

$$\begin{aligned} C_n^k p_n^k (1 - p_n)^{n-k} &= \frac{n(n-1)\cdots(n-k+1)}{k!} p_n^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} (np_n)^k \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1 - p_n)^{n-k}. \end{aligned}$$

注意,

$$\begin{aligned} (1 - p_n)^{n-k} &= \exp\{(n-k)\ln(1-p_n)\} \\ &= \exp\left\{(n-k)p_n \cdot \frac{1}{p_n}\ln(1-p_n)\right\}, \end{aligned}$$

$$\lim_{n \rightarrow \infty} p_n = 0, \quad \lim_{n \rightarrow \infty} np_n = \lambda, \quad \lim_{n \rightarrow \infty} \frac{1}{p_n} \ln(1-p_n) = -1,$$

故(7.4)式成立. \square

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

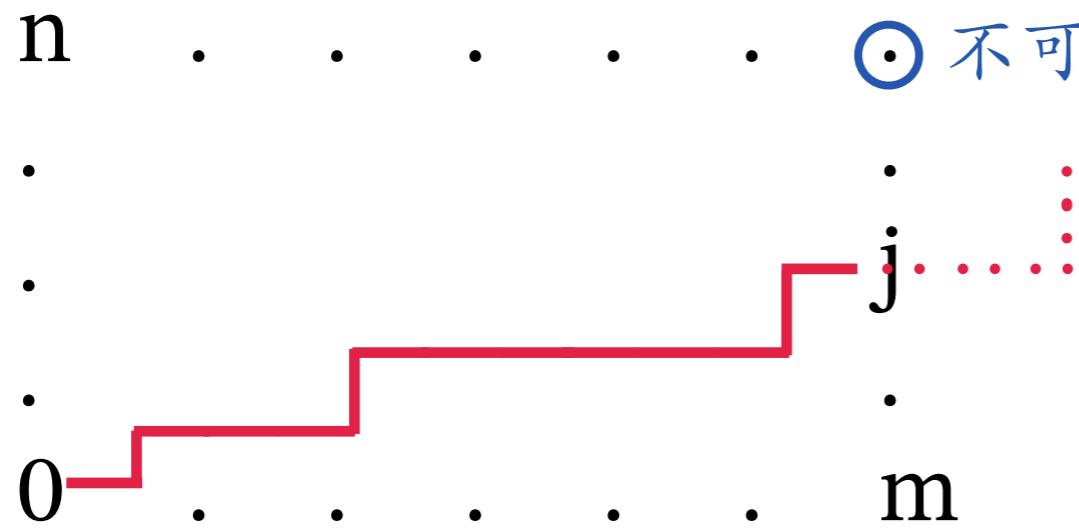
得分问题（解答）

甲、乙两人各出同样的赌注，用掷硬币作为博奕手段。每掷一次，若正面朝上，甲得1分乙不得分。反之，乙得1分，甲不得分。谁先得到规定分数就赢得全部赌注。当进行到甲还差2分乙还差3分，就分别达到规定分数时，发生了意外使赌局不能进行下去，问如何公平分配赌注？

$P(m, n)$: 甲乙距离分数线分别为 m, n 时，甲最终赢得比赛

$$P(m, n) = \sum_{k=m}^{m+n-1} \binom{m+n-1}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{m+n-1-k}$$

$$P(m, n) = \frac{1}{2} \cdot P(m-1, n) + \frac{1}{2} \cdot P(m, n-1)$$



甲需赢得 $m+n-1$ 步
中的至少 m 步

$P(3, 0)$	$P(2, 0)$	$P(1, 0)$	*
$P(3, 1)$	$P(2, 1)$	$P(1, 1)$	$P(0, 1)$
$P(3, 2)$	$P(2, 2)$	$P(1, 2)$	$P(0, 2)$
$P(3, 3)$	$P(2, 3)$	$P(1, 3)$	$P(0, 3)$

0	0	0	*
$P(3, 1)$	$P(2, 1)$	$P(1, 1)$	1
$P(3, 2)$	$P(2, 2)$	$P(1, 2)$	1
$P(3, 3)$	$P(2, 3)$	$P(1, 3)$	1

Poisson分布

Poisson分布：单位时间发射粒子平均数目 $\lambda > 0$.
在给定的单位时间内，粒子数量 X 的分布

$X \sim Poisson(\lambda), \lambda > 0 :$

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

$$0, \quad \frac{1}{n}, \quad \frac{2}{n}, \quad \frac{3}{n}, \quad \dots, \quad 1$$

$1/n$ 的时长内出现粒子的概率近似为 $p = \lambda/n$

$$np \approx \lambda, n \rightarrow \infty$$

Poisson分布最大概率点

定理 2.2 设 X 服从泊松分布, $p_k = P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ($\lambda > 0$; $k=0, 1, 2, \dots$), 则有下列结论:

(1) 当 λ 不是整数时,

$$p_0 < p_1 < \cdots < p_{[\lambda]} > p_{[\lambda]+1} > \cdots$$

(这里 $[\lambda]$ 是不超过 λ 的最大整数);

(2) 当 λ 是整数时,

$$p_0 < p_1 < \cdots < p_{\lambda-1} = p_\lambda > p_{\lambda+1} > \cdots.$$

$$\frac{p_n(k+1)}{p_n(k)} = \frac{\lambda}{k+1} > 1 \iff k < \lambda - 1$$

超几何分布

超几何分布：一罐子中有 N 个小球，其中 K 个涂有颜色，从中取出 n 个，有颜色的小球数 X 的分布

$X \sim Hypergeometric(N, K, n), N \geq K, n \geq 1 :$

$$P(X = k) = \begin{cases} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, & \max(0, K - N + n) \leq k \leq \min(n, K) \\ 0, & \text{otherwise} \end{cases}$$

- $n=1$, Bernoulli(p), $p=K/N$
- 二项分布 $K/N \approx p, N \rightarrow \infty$

超几何分布二项逼近

定理 2.3 设超几何分布(2.10)中 D 是 N 的函数, 即 $D=D(N)$ 且

$$\lim_{N \rightarrow \infty} \frac{D(N)}{N} = p \quad (0 < p < 1),$$

则

$$\lim_{N \rightarrow \infty} \frac{\binom{D(N)}{k} \binom{N-D(N)}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k} \quad (2.11)$$

(对任何固定的 $k \geq 0$).

证明 既然 $0 < p < 1$, 故 N 充分大时, $n < D(N) < N$ (注意, n 是固定的!) 以下简记 $D(N)$ 为 D . 易知

$$\frac{C_D^k C_{N-D}^{n-k}}{C_N^n} = \frac{D!}{k!(D-k)!} \cdot \frac{(N-D)!}{(n-k)!(N-D-n+k)!}$$

$$\cdot \frac{n!(N-n)!}{N!}$$

$$= \frac{n!}{k!(n-k)!} \cdot \frac{D(D-1)\cdots(D-k+1)}{N^k}$$

$$\cdot \frac{(N-D)(N-D-1)\cdots(N-D-n+k+1)}{N^{n-k}}$$

$$\cdot \frac{N^n}{N(N-1)\cdots(N-n+1)}$$

$$= C_n^k \left[\prod_{i=1}^k \frac{D-i+1}{N} \right] \left[\prod_{i=1}^{n-k} \frac{N-D-i+1}{N} \right]$$

$$\cdot \left[\prod_{i=1}^n \frac{N}{N-i+1} \right]$$

$$\rightarrow C_n^k p^k (1-p)^{n-k} \quad (N \rightarrow \infty). \quad \square$$

超几何分布

$$\sum_{0 \leq k \leq \min(n, K)} P(X = k) = \sum_{0 \leq k \leq \min(n, K)} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = 1$$

回顾：**Vandermonde等式**

$$\binom{m+n}{k} = \sum_{j=0}^k \binom{m}{j} \cdot \binom{n}{k-j}$$

注意n与K的对称性

$$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\binom{n}{k} \binom{N-n}{K-k}}{\binom{N}{K}}$$

几何分布

几何分布：投掷正面出现概率为 p 的硬币直到正面出现，总投掷次数 W 的分布

$W \sim Geometric(p), 0 < p < 1 :$

$$P(W = k) = (1 - p)^{k-1} \cdot p, k = 1, 2, 3, \dots$$

等待时间 W : $P(W > k) = (1 - p)^k$

无记忆性:

$$P(W > k + n \mid W > n) = P(W > k) = (1 - p)^k, k, n \geq 1,$$

负二项分布

负二项分布：投掷正面出现概率为 p 的硬币直到正面出现 r 次，总投掷次数 W_r 的分布

$W_r \sim NBin(r, p), r \geq 1, 0 < p < 1 :$

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

$$W_r = \underbrace{W + W + \cdots + W}_{r \text{ (independent) copies of } W}$$

在 k 个小球的 $k-1$ 个间隙中插入 $r-1$ 个搁板将小球分为 r 组

负二项分布(附)

Binomial 展开: $|x| < 1, \alpha \in \mathbb{C}$

$$\begin{aligned}(1+x)^\alpha &= \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \\&= 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} x^3 + \dots,\end{aligned}$$

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}.$$

负二项分布(附)

$$\begin{aligned} P(W_r = r+j) &= \binom{r+j-1}{r-1} \cdot p^r \cdot (1-p)^j \\ &= \binom{r+j-1}{j} \cdot p^r \cdot (1-p)^j, j = 0, 1, 2, \dots \end{aligned}$$

投掷k次，r次正面，j次反面

$$\begin{aligned} p^{-r} &= (1 + (p - 1))^{-r} = \sum_{j=0}^{\infty} \binom{-r}{j} (p - 1)^j \\ &= \sum_{j=0}^{\infty} \frac{(-r)(-r-1) \cdots (-r-j+1)}{j!} (p - 1)^j \end{aligned}$$

负二项分布(附)

$$p^{-r} = (1 + (p - 1))^{-r} = \sum_{j=0}^{\infty} \binom{-r}{j} (p - 1)^j$$

$$= \sum_{j=0}^{\infty} \frac{(-r)(-r - 1) \cdots (-r - j + 1)}{j!} (p - 1)^j$$

← "负二项"

$$= \sum_{j=0}^{\infty} \frac{(r + j - 1) \cdots (r + 1)(r)}{j!} (1 - p)^j$$

$$= \sum_{j=0}^{\infty} \binom{r + j - 1}{j} (1 - p)^j$$

$$\Rightarrow \sum_{j=0}^{\infty} \binom{r + j - 1}{j} p^r (1 - p)^j = 1$$

指数分布

$$T \sim Exp(\lambda), \lambda > 0$$

$$p(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

任意时段内，粒子出现率与时间长成比例 $\lambda > 0$

固定 $t > 0$: $0, \delta, 2\delta, 3\delta, \dots, t = k\delta$

$$P(T > t) = P(T > k\delta) = P\left(T > \left(\frac{t}{\delta}\right)\delta\right)$$

$$\approx (1 - \lambda\delta)^{t/\delta} \rightarrow e^{-\lambda t}, k \rightarrow \infty$$

指数分布

例 3 设某人打一次电话所用的时间 ζ 服从参数为 $1/10$ (单位: min) 的指数分布, 当你走进电话室需要打电话时, 某人恰好在你前面打电话. 求以下几个事件的概率: $P(\zeta > T_{in} + 10 \mid \zeta > T_{in}) = P(\zeta > 10)$

- (1) 你需要等待 10min 以上;
- (2) 你需要等待 10~20min.

$$P(T_{in} + 20 > \zeta > T_{in} + 10 \mid \zeta > T_{in}) = P(20 > \zeta > 10)$$

解 用 ζ 表示某人的通话时间, 也就是你的等待时间, 则 ζ 的分布密度函数

$$f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

所以要求的概率分别为

$$(1) P(\zeta > 10) = \int_{10}^{+\infty} \frac{1}{10} e^{-\frac{x}{10}} dx = e^{-1} \approx 0.368 ;$$

Galton板到正态分布

