概率理论与数理统计

随机变量及其分布

随机变量

定义

设$(\Omega,\mathcal{F},\mathbf{P})$为一个概率空间. 随机变量$X$是定义在$\Omega$上的实值可测函数, 即$X:\Omega\to\mathbb{R}$, 对于$\mathcal{B}(\mathbb{R})$ 中任何集合B,

$$\{X\in B\}=\{\omega\in\Omega:X(\omega)\in B\}(=X^{-1}(B))\in\mathcal{F}.$$

随机变量

- 若随机变量仅取有限或可数个值, 则称其为离散随机变量,
- 若随机变量的取值充满数轴上的一个区间, 则称其为连续随机变量,
- 若$X_1,\cdots,X_n$都是空间$(\Omega,\mathcal{F},\mathbf{P})$上的随机变量, $f(\cdot)$为连续函数, 则以下都是随机变量:
  - X_1+X_2, $X_1\wedge X_2$, $f(X_1,\cdots,X_n)$,
  - $\inf_n X_n$, $\sup_n X_n$, $\lim_n\sup X_n$, $\lim_n\inf X_n$.

定义

设X为概率空间$(\Omega, \mathcal{F}, \mathbf{P})$上的一个随机变量. 随机变量$X$的分布测度$\mu_X$定义为: 对于$\mathbb{R}$ 中任何Borel集B,

$$\mu_X(B) = P(\{X \in B\}) = P(\{\omega \in \Omega : X(\omega) \in B\}) = P(X^{-1}(B)).$$

- 设$\Omega = \{$ 三次掷硬币$\}$, 随机变量$X =$ 出现正面的次数, 取值范围为$\{0,1,2,3\}$. 考虑$B \subset \{0,1,2,3\}$,

$$\mu_X(B) = P(\{X \in B\}) = \sum_{i \in B} P(X = i) = \frac{|\{\omega \in \Omega : X(\omega) \in B\}|}{8}$$

- 设$\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, P为$[0,1]$上的Lebesgue测度, 定义随机变量$X(\omega) = \omega$, $Y(\omega) = 1 - \omega$, $\omega \in [0,1]$, 则$0 \le a \le b \le 1$,

$$\mu_X([a,b]) = P(\{\omega \in \Omega : a \le X(\omega) \le b\}) = b - a = \mu_Y([a,b]).$$

离散型随机变量的分布律

设X的所有可能取值为$\{x_k\}_{1\le k<\infty}$，事件$\{X=x_k\}$的概率为

$$P(X=x_k)=p_k, k=1,2,\cdots$$

1. $p_k\ge 0,\ k=1,2,\cdots.$
2. $\sum_{k=1}^{\infty}p_k=1.$
3. 可列可加性.

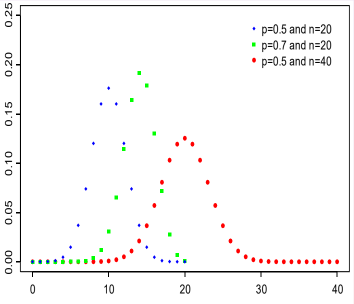
分布律-分布函数

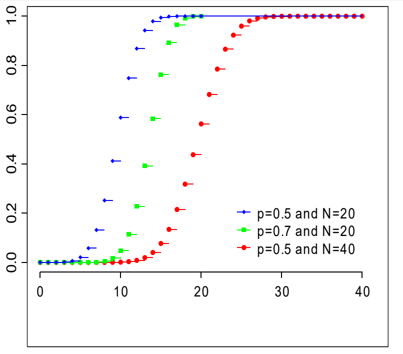
- 离散型随机变量X: 取值为$\{x_k\}_{1\le k<\infty}$，定义分布函数F_X为一个右连续非降阶梯函数.

$$F_X(x)=P(X\le x)=\sum_{j:x_j\le x}P(X=x_j),$$
$$p_k=F(x_k)-F(x_k-)=P(X=x_k),\quad k=0,1,2\cdots.$$

n重伯努利实验, 二项分布$X \sim B(n,p)$

- X为重复n次独立的伯努利实验中成功的次数，取值$0, 1, \cdots, n$,
- $P(X = k) = C_n^k p^k (1-p)^{n-k}$, $k = 0, 1, \cdots, n$.





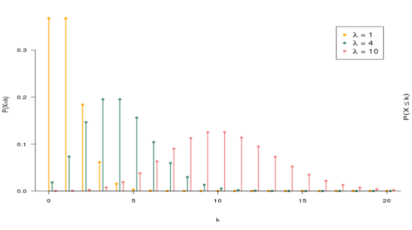
Poisson分布$\pi(\lambda)$

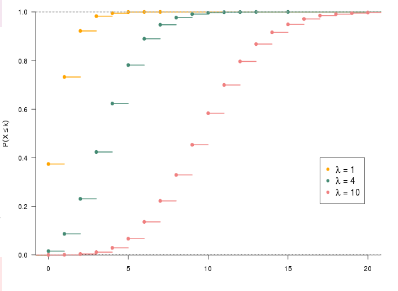
Poisson分布$\pi(\lambda)$: 随机变量取值为$0, 1, 2, \cdots$

- $P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0, 1, 2, \cdots$,
- 二项分布逼近Poisson分布: 设n为任意正整数, $np_n = \lambda$, 对于任意固定非负整数k,

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

- 下图三个参数分别为$\lambda = 1, 2, 5$.

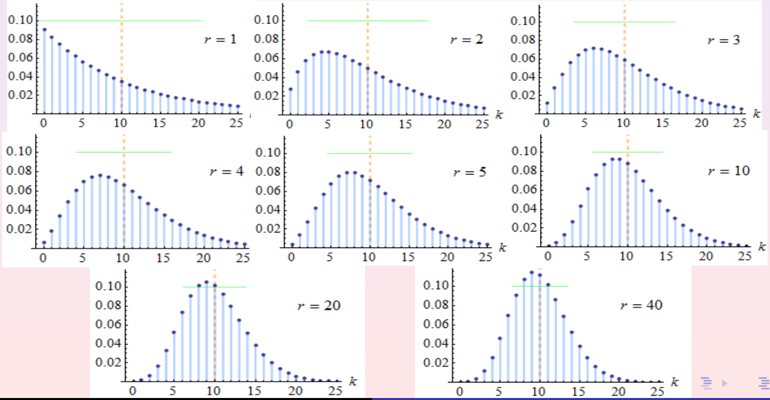




Pascal分布/负二项分布

Pascal分布/负二项分布X

- 固定r, X为多次重复独立的伯努利实验，当r次成功出现时失败的次数，取值于0,1,$\cdots$,
- $P(X=k)=C_{r-1+k}^{r-1}p^r(1-p)^k$, $k=0,1,2,\cdots$,



分布函数

如果随机变量X的取值空间是$\mathbb{R}$或其子集，集合$\{X \leq x\} \in \mathcal{F}$ 可测，定义函数

$$F_X(x) = P(X \leq x)$$

即$F(x)$给出X小于或等于给定数值x的概率, 称为X的分布函数, 满足

1. 单调性: F_X是单调非降函数，$F(x_1) \leq F(x_2)$, $x_1 \leq x_2$,
2. 有界性: $0 \leq F(x) \leq 1$, $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = 1$.
3. 右连续性: $\forall x_0 \in \mathbb{R}$, $\lim_{x\to x_0+} F_X(x) = F(x_0)$,

定理

- 由$(\mathbb{R}, \mathcal{B}(\mathbb{R})$上的概率$\mathbf{P}$可以诱导出一个分布函数
$$F(x) = \mathbf{P}((-\infty, x])$$
- 由分布函数F可以确定一个在$(\mathbb{R}, \mathcal{B}(\mathbb{R})$ 上的概率$\mathbf{P}$.
- 若函数F满足定义中的1)-3), 则F为一个在$(\mathbb{R}, \mathcal{B}(\mathbb{R})$上(唯一)确定概率的分布函数.