随机变量与分布函数

随机变量

定义

- 随机变量：对概率空间 $(\Omega,\ \mathscr{F},\ P)$ ，定义单值函数 $X \to \mathbb{R}$ 。若 $\forall x \in \mathbb{R}$ ，均有 $\{\omega | X(\omega) \le x\} \in \mathscr{F}$ ，则称 X 是随机变量。常把 $X(\omega)$ 简记作 X 。

- 离散型 \ 非离散型随机变量：
  - 随机变量的可能取值是有限个或可列个，则称为离散型随机变量
  - 若随机变量的可能取值“连续”，则称为连续性随机变量，连续性随机变量是一种重要的非离散型随机变量。

分布函数

定义

设 X 为随机变量，x 是任意实数，称函数

$$F(x) = P(X \le x),\ -\infty < x < +\infty$$

为 X 的分布函数。由此定义，X 落在区间 $(a, b]$ 里的概率可用分布函数

$$\begin{aligned} P(a < X \le b) &= P(X \le b) - P(X \le a) \\ &= F(b) - F(a) \end{aligned}$$

来计算。

性质

- $F(x)$ 单调不减，即

$$\forall x_1 < x_2,\ F(x_1) \le F(x_2)$$

- $0 \le F(x) \le 1$ 且

$$lim_{x \to +\infty} F(x) = 1,\ lim_{x \to -\infty} F(x) = 0$$

- $F(x)$ 右连续，即

$$F(x+0) \triangleq lim_{t \to x+0} F(t) = F(x)$$

- 设 $a < b$ ，则：
  - $P(a < X \le b) = F(b) - F(a)$
  - $P(X > a) = 1 - F(a)$
  - $P(x < a) = F(a-0)$
  - $P(X = a) = F(a) - F(a-0)$

离散型随机变量

- 定义：离散型随机变量的概率分布（分布律）：它在其所有取值 x_k 上的概率列。即：

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

- 离散型随机变量的分布函数：

$$\begin{aligned} F(x) &= P(X \le x) = P(\cup_{x_k \le x}(X = x_k)) \\ &= \sum_{x_k \le x} P(X = x_k) = \sum_{x_k \le x} p_k \end{aligned}$$

- 常见分布：
  首先定义多重伯努利试验：
  1. 试验可以多次重复
  2. 每次试验的结果与其他次试验无关
  3. 每次试验只有两个可能的结果：$A,\ \bar{A}$，且 $P(A)=p$

  - 0-1 分布：

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

  - 二项分布:

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

  - 几何分布: 描述多次伯努利试验中，A 首次发生时的试验次数。$P(A)=p$

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

  - Pascal 分布：在多重伯努利试验中，事件 A 第 r 次发生时的试验次数记为随机变量 X，已知 $P(A)=p,\ q=1-p$

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

  - 超几何分布：

  $$P(X=k)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n},\ \ k=0,1,\cdots,min(n,M)$$

  - Poisson 分布：

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

  记作 $X\sim\pi(\lambda)$ 或 $P(\lambda)$

  NOTE: 二项分布是超几何分布当样本总量 $N\rightarrow+\infty$ 时的极限；Poisson 分布是二项分布当 $N\rightarrow+\infty$ 时的极限，此时 $\lambda=N*p$。

连续型随机变量

- 定义：连续型随机变量的概率密度：设 X 是随机变量，若对于任意实数 x 存在一个非负可积函数 $f(x)$，使得 X 的分布函数：

  $$F(x)=\int_{-\infty}^{x}f(t)dt,\ \ -\infty<x<+\infty$$

  则称 X 是连续型随机变量，$f(x)$ 是它的概率密度函数，简称为密度函数或概率密度。

- 概率密度的性质：
  1. $f(x)\geq 0$
  2. $\int_{-\infty}^{+\infty}f(x)dx=F(+\infty)=1$
  3. 取任一常数的概率为 0
  4. $P(X<b)=P(X\leq b)=F(b)$
  5. $P(X>a)=P(X\geq a)=1-F(a)$
  6. **连续型随机变量的密度不唯一，可以任意改变密度在有限个点上的值**

- 求法：
  1. 已知概率密度函数，求分布函数：积分

  $$F(x)=\int_{-\infty}^{x}f(t)dt$$

  2. 已知分布函数，求概率密度函数：求导

  $$f(x)=F'(x)$$

- 常见分布：
  - 均匀分布：

$$f(x)=\begin{cases}\dfrac{1}{b-a} & ,\ a<x<b \\ 0 & ,\ otherwise\end{cases}$$

记作 $X\sim U(a,b)$

- 指数分布：

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,\ x\geq 0 \\ 0 & ,\ otherwise\end{cases}$$

称 X 为服从参数为 λ 的指数分布，记作 $X\sim E(\lambda)$

- 正态分布（高斯分布）：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$$

记作 $X\sim N(\mu,\sigma^2)$

- Γ 分布：

$$f(x)=\begin{cases}\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & ,\ x>0 \\ 0 & ,\ x\leq 0\end{cases}$$

记作 $X\sim\Gamma(\alpha,\beta)$ 其中

$$\Gamma(x)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$$

具有性质

1. $\Gamma(\alpha+1)=\alpha*\Gamma(\alpha)$
2. $\Gamma(0.5)=\pi^{1/2},\ \ \Gamma(1)=1$
3. 对于正整数 n，$\Gamma(n+1)=n!$