随机变量及其概率分布

随机变量

随机变量用于将随机试验的结果数值化

设 Ω 是随机试验的样本空间，若对每个 $e\in\Omega$ ，都对应一个唯一的实数 $X(e)$ ，称单值实函数 $X(e)$ 为随机变量

对于任意实数 x ，都对应概率空间中的一个事件 $\{\omega:\omega\in\Omega,X(\omega)\leqslant x\}$ ，即 $X\leqslant x$

随机变量的取值与概率之间的对应描述称为 X 的概率分布

设 X 是一个随机变量，x 是任意实数，称函数

$$F(x)=P(X\leqslant x)(-\infty\leqslant x\leqslant\infty)$$

为随机变量 X 的分布函数（CDF, Cumulative Distribution Function）

CDF 有性质

- 单调：$\forall x_1,x_2\in\mathbb{R},x_1<x_2,F(x_1)\leqslant F(x_2)$
- 有界：$0\leqslant F(x)\leqslant 1,F(-\infty)=0,F(\infty)=1$
- 右连续：$F(x_0+0)=\lim_{x\to x_0+0}=F(x_0)$

一个函数是 CDF $\iff$ 满足上述性质

对任意实数 $a,b(a<b)$ 有

$$\begin{aligned}&P(a<X\leqslant b)=P(X\leqslant b)-P(x\leqslant a)=F(b)-F(a)\\&P(X>b)=1-P(X\leqslant b)=1-F(b)\\&P(X\geqslant b)=1-P(X<b)=1-F(b-0)\end{aligned}$$

离散型随机变量

定义

若随机变量 X 的取值为有限个或可列无限个，则称 X 为离散型随机变量（discrete random variable)

设离散型随机变量所有可能取值为 $x_1,x_2,\dots,x_n,\dots$ 则

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

称为 X 的**分布律**，可使用表格形式给出

分布律具有性质

- $p_k\geqslant 0,k=1,2,\dots$
- $\sum_k p_k=1$

> 分布律即**概率质量函数**（PMF, Probability Mass Function）
>
> 设 X 为离散型随机变量，其 PMF $f_X(x)$ 的定义为
>
> $$f_X(x)=P(X=x)=P(\{s\in S:X(s)=x\})$$

常见离散型随机变量

0-1 分布（Bernoulli Distribution）

随机变量可能取值只有 0，1，且

$$P(X=1)=p,P(X=0)=1-p=q$$

二项分布（Binomial Distribution）

若随机变量的分布律满足

$$P(x=k)=p_k=C_n^kp^k(1-p)^{n-k}$$

则称 X 服从参数为 n,p 的二项分布，记为 $X\sim B(n,p)$

> n 重伯努利试验中设 A 发生的概率为 p，A 发生的次数为 X，则 $X\sim B(n,p)$

对于二项分布的概率最大值

- 当 $(n+1)p$ 为整数时，p_k 在 $(n+1)p-1,(n+1)p$ 达到最大
- 当 $(n+1)p$ 不为整数时，p_k 在 $\lfloor(n+1)p\rfloor$ 达到最大

泊松分布（Poisson Distribution）

若随机变量 X 的分布律满足

$$p_k=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,1,\dots$$

其中 $\lambda>0$ 为常数，则称 X 服从参数为 λ 的泊松分布，记为 $X\sim P(\lambda)$

泊松分布常用于描述大量试验时稀有事件出现频数的概率。

> 泊松分布可以用来近似二项分布，当二项分布的参数 n,p 满足 n 很大而 p 很小时，设 $\lambda = np$ ，则有
>
> $$\lim_{n\to\infty} C_n^k p^k (1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$

几何分布（Geometry Distribution）

若随机变量 X 的分布律为

$$p_k = (1-p)^{k-1}p, k = 1, 2, \dots$$

其中 $0 < p < 1$ 为常数，则称 X 服从参数为 p 的几何分布，记为 $X \sim g(p)$

几何分布的含义是 n 重伯努利试验中 A 首次出现时所需的试验次数。

几何分布有**无记忆性**，即

$$P(X = s + t | X > t) = P(X = s)$$

离散型随机变量的分布函数

离散型随机变量的分布函数是一个分段函数，且每一段是左闭右开的区间。下一段相比上一段只增加一个有概率分布的点。

更正式的定义为

$$F(X) = P(X \leqslant x) = \sum_{x_k \leqslant x} P(X = x_k) = \sum_{x_k \leqslant x} p_k$$

连续型随机变量

定义

设随机变量 X 的分布函数为 $F(x)$ ，若存在非负可积函数 $p(x)$ ，对任意实数 x 有

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

则称 X 为连续型随机变量（continuous random variable），称 $p(x)$ 为 X 的概率密度函数（PDF, Probability Density Function）

PDF 有性质

- $p(x) \geqslant 0$
- $\int_{-\infty}^{+\infty} p(x)dx = 1$
- 由于 $p(x)$ 为可积函数，$F(x)$ 为连续函数

- 对任意实数 $a,b(a<b)$ 有

$$P(a<X\leqslant b)=F(b)-F(a)=\int_a^b p(x)dx$$

由于连续型随机变量有特性，即 X 在任意一点 x_0 处取值的概率为 0

> 对任意 $\Delta x>0$ 有
>
> $$0\leqslant P(X=x_0)\leqslant P(x_0-\Delta x<X\leqslant x_0)=F(x_0)-F(x_0-\Delta x)$$
>
> 分布函数 $F(x)$ 为连续函数，在 $\Delta x\to 0$ 时上式趋向于 0，故 $P(X=x_0)=0$

根据该性质，连续型随机变量在某个区间取值的概率与区间的开闭无关，故

$$\begin{aligned}P(a<X\leqslant b)&=P(a<X<b)=P(a\leqslant X\leqslant b)=P(a\leqslant X<b)\\&=F(b)-F(a)=\int_a^b p(x)dx\end{aligned}$$

- 若 $p(x)$ 在点 x 连续，则分布函数在 $F(x)$ 可导，且 $p(x)=F(x)$
- 若令 $a=x,b=x+\Delta x$ ，有

$$P(x<X\leqslant x+\Delta x)=F(x+\Delta x)-F(x)=\int_x^{x+\Delta x}p(x)dx\approx p(x)\Delta x$$

即点 x 处密度函数值越大，在该点附近取值的概率就越大

常见连续型随机变量

均匀分布（Uniform Distribution）

若随机变量 X 的 PDF 为

$$p(x)=\begin{cases}\frac{1}{b-a} & a<x<b\\ 0 & \text{otherwise}\end{cases}$$

则称 X 在区间 $[a,b]$ 上服从均匀分布，记为 $X\sim U[a,b]$

X 的 CDF 为

$$F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\leqslant x<b\\ 1 & x\geqslant b\end{cases}$$

均匀分布取值的概率仅与区间长度有关，与区间位置无关

指数分布（Exponential Distribution）

若随机变量 X 的 PDF 为

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0\\ 0 & x<0\end{cases}$$

其中 $\lambda>0$ 为常数，则称 X 服从参数为 λ 的指数分布，记为 $X\sim E(\lambda)$

X 的 CDF 为

$$F(x)=\begin{cases}1-e^{-\lambda x} & x\geqslant 0\\ 0 & \text{otherwise}\end{cases}$$

指数分布常用于描述各种寿命的分布

指数分布具有**无记忆性**，即

$$P(X>s+t|X>t)=P(X>s)$$

正态分布（Normal Distribution/Gaussian Distribution）☆

若随机变量 X 的 PDF 为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

其中 μ,σ 为常数，则称 X 服从参数为 μ,σ^2 的正态分布，记为 $X\sim N(\mu,\sigma^2)$

正态分布的 CDF 为

$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

正态分布的密度函数与分布函数的图形有以下特点

- 曲线 $p(x)$ 关于 $x=\mu$ 对称，且对任意 $b>0$ 有 $P(X\leqslant\mu-b)=P(X\geqslant\mu+b)$
- $x=\mu\pm\sigma$ 为 $p(x)$ 的拐点
- 当 $x\to\pm\infty$ 时，$p(x)\to 0$ ，即渐近线为 x 轴
- $x=\mu$ 时 $p(x)$ 取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$ ，由于 σ^2 为方差，故 σ 越大图形越平缓，σ 越小图形越陡峭
- 固定 σ 改变 μ 时曲线形状不变，而是横向平移

当 $\mu=0,\sigma=1$ 时，$X\sim N(0,1)$ ，称 X 服从标准正态分布，特别地，将其 PDF 与 CDF 记为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$
$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt$$

对于其 CDF，有 $\Phi(-x)=1-\Phi(x)$

> Proof.
>
> $$\Phi(-x)=\int_{-\infty}^{-x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt$$
>
> 做积分变换，令 $u=-t, dt=-du$
>
> $$\begin{aligned}\Phi(-x)&=\int_{\infty}^{x}-\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,du\\&=\int_{x}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,du\\&=1-\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,du\\&=1-\Phi(x)\end{aligned}$$

这一性质对于所有 x 均成立

☆ 若随机变量 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

> Proof.
>
> $$\begin{aligned}P(Z\leqslant x)&=P(\frac{X-\mu}{\sigma}\leqslant x)\\&=P(X\leqslant \sigma x+\mu)\\&=\int_{-\infty}^{\sigma x+\mu}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt\end{aligned}$$
>
> 两边同时求导，利用积分变限函数求导性质
>
> $$(P(Z\leqslant x))'=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2}}\sigma=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\varphi(x)$$
>
> 即 Z 的 PDF 为 $\varphi(x)$，根据定义，$Z\sim N(0,1)$

根据以上重要性质，可得对于 $X\sim N(\mu,\sigma^2)$

$$F(x)=P(X\leqslant x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$$
$$P(a<X<b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

3σ 法则：$X\sim N(\mu,\sigma^2)$ 在取值时落入 $[\mu-3\sigma,\mu+3\sigma]$ 的概率超过 99.7%

随机变量函数的分布

设 X 是随机变量，$y=g(x)$ 是普通实函数，则令随机变量 Y 在 X 取 x 时取 $g(x)$ ，记为 $Y=g(X)$ ，Y 作为一个随机变量，也有自己的概率分布

离散型随机变量

设 X 是离散型随机变量，其分布律为

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

则对于 X 的函数 $Y=g(x)$ ，当 X 取 x_k 时 Y 取 $y_k=g(x_k)$ ，Y 也是离散型随机变量

当 $y_1,y_2,\ldots,y_k,\ldots$ 取值各不相同时，有

$$P(Y=y_k)=P(Y=g(x_k))=P(X=x_k)=p_k$$

若 $y_1,y_2,\ldots,y_k,\ldots$ 中有取值相同的， e. g. $y_i=y_j$ ，即 $g(x_i)=g(x_j)$ ，则

$$P(Y=y_i)=P(Y=g(x_i)\cup Y=g(x_j))=P(X=x_i)+P(X=x_j)=x_i+x_j$$

即多个 y_k 取值相同时，应当把对应的概率加起来

连续型随机变量

分布函数法：先求 Y 的分布函数 $F_Y(y)$

$$F_Y(y)=P(Y\leqslant y)=P(g(x)\leqslant y)=\int\limits_{x:g(x)\leqslant y} p_X(x)dx$$

然后根据积分变限函数的求导法则求得

$$p_Y(y)=F_Y'(y)$$

> 公式法：
>
> 设随机变量 X 的可能取值范围为 (a,b) ，X 的 pdf 为 $p_X(x)$，$a<x<b$ ，其中 a 可为 $-\infty$ ，b 可为 $+\infty$ ，设 $y=g(x)$ 处处可导，且严格单调（恒有 $g'(x)<0$ 或 $g'(x)>0$） ，则 $Y=g(X)$ 为连续型随机变量，且其 pdf 为
>
> $$p_Y(y)=\begin{cases} p_X[g^{-1}(y)]\cdot|g^{-1}(y)|' & \alpha<y<\beta \\ 0 & \text{o. t.} \end{cases}$$
>
> 其中 $\alpha=\min\{g(a),g(b)\},\beta=\max\{g(a),g(b)\}$ ，$g^{-1}(y)$ 为 $g(x)$ 反函数
>
> 公式法不如分布函数法泛用