随机变量的数字特征

数学期望

定义

对于离散型随机变量 X ，设其分布律为 $P(X=x_i)=p_i$ ，若级数 $\sum_{i=1}^{+\infty}|x_i|p_i$ 收敛，则称 $\sum_{i=1}^{+\infty}x_ip_i$ 为 X 的数学期望，记为 $E[X]$，即

$$E[X]=\sum_{i=1}^{+\infty}x_ip_i$$

对于连续型随机变量 X ，设其概率密度为 $p(x)$ ，若 $\int_{-\infty}^{+\infty}|x|p(x)dx<\infty$ ，则 X 的数学期望为

$$E[X]=\int_{-\infty}^{+\infty}xp(x)dx$$

> 绝对收敛：对于一个数列 a_k ，若 $\sum|a_k|$ 收敛，则 $\sum a_k$ 收敛，反之不真
>
> 绝对可积：对于一个函数 $f(x)$ ，若 $\int|f(x)|dx$ 有限，则 $\int f(x)dx$ 有限，反之不真

中位数：设 X 为随机变量，若 $m\in\mathbb{R}$ 使得 $P(X\leqslant m)\geqslant\frac{1}{2},P(X\geqslant m)\geqslant\frac{1}{2}$ ，则称 m 为中位数

随机变量函数的期望

若随机变量 X 的函数 $Y=g(X)$ 也是随机变量，且其期望存在，则有

若 X 为离散型随机变量

$$E[g(X)]=\sum_{i=1}^{+\infty}g(x_i)p_i$$

若 X 为连续型随机变量

$$E[g(X)]=\int_{-\infty}^{+\infty}g(x)p(x)dx$$

该结论可推广至随机向量的情况，若随机向量 (X,Y) 的函数 $Z=g(X,Y)$ 是随机变量，且期望存在，则

若 (X,Y) 为离散型随机向量

$$E[g(X,Y)]=\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}g(x_i,y_j)p_{ij}$$

若 (X,Y) 为连续型随机向量

$$E[g(X,Y)]=\iint g(x,y)p(x,y)dxdy$$

期望的性质

对于常数 a,b ，若有 $a\leqslant X\leqslant b$ ，则 $a\leqslant E[X]\leqslant b$

对于常数 a ，有 $E[a]=a$

对任意 n 个实数 $c_1,c_2,\dots,c_n$ ，以及 n 个随机变量 $X_1,X_2,\dots X_n$ ，设其期望均存在，则有

$$E\left[\sum_{i=1}^{n}c_iX_i\right]=\sum_{i=1}^{n}c_iE[X_i]$$

这个性质被称为期望的**线性性质**，在**任何情况下都成立**

若随机变量 $X_1,X_2,\dots X_n$ **相互独立**，则

$$E\left[\prod_{i=1}^{n}X_i\right]=\prod_{i=1}^{n}E[X_i]$$

在求解复杂随机变量的期望时，可以将其化为多个简单随机变量的和，然后利用期望的线性性质求解

> 示性函数（indicator function）：对于事件 A ，其示性函数
>
> $$I_A=\begin{cases}1 & \text{A 发生}\\0 & \text{A 不发生}\end{cases}$$
>
> 若 A 发生的概率为 p ，则有 $E[I_A]=p$

方差

定义

期望用于描述随机变量的均值，而方差则用于描述随机变量相较于期望的分散程度

若 $E[X^2]<+\infty$ ，则称 $E[(X-E[X])^2]$ 为随机变量 X 的**方差**，记为 $D[X]$ ，即

$$D[X]=E[(X-E[X])^2]$$

而称

$$\sigma[X]=\sqrt{D[X]}$$

为随机变量 X 的标准差

一般常用下述公式计算方差

$$D[X]=E[X^2]-(E[X])^2$$

该公式可以利用期望的线性特性得到

方差的性质

对于常数 a ，有 $D[a]=0$

对于随机变量 X ，有 $D[X]=0 \iff P(X=E[X])=1$

对于常数 a,b ，若随机变量 X 的方差存在，则

$$D[aX+b]=a^2D[X]$$

对于任意方差存在的随机变量 X,Y ，其和或差的方差仍存在，且

$$D[X\pm Y]=D[X]+D[Y]\pm 2E[(X-E[X])(Y-E[Y])]$$

特别的，当 X,Y 独立时，有

$$D[X\pm Y]=D[X]+D[Y]$$

该结论可推广至 n 个随机变量，即

$$D\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}D[X_i]+2\sum_{1\leqslant i<j\leqslant n}E[(X_i-E[X_i])(X_j-E[X_j])]$$

切比雪夫不等式：设随机变量 X 的期望与方差均存在，对任意 $\varepsilon>0$ 有

$$P(|X-E[X]|\geqslant\varepsilon)\leqslant\frac{D[X]}{\varepsilon^2}$$

> Proof.
>
> $$P(|X-E[X]|\geqslant\varepsilon)=E[I_{\{|X-E[X]|\geqslant\varepsilon\}}]$$
>
> 放缩不等式

> $$
> \begin{aligned}
> E\left[I_{\{|X-E[X]|\geqslant\varepsilon\}}\right] &\leqslant E\left[\frac{(X-E[X])^2}{\varepsilon^2}I_{\{|X-E[X]|\geqslant\varepsilon\}}\right] \\
> &\leqslant E\left[\frac{(X-E[X])^2}{\varepsilon^2}\right] \\
> &= \frac{D[X]}{\varepsilon^2}
> \end{aligned}
> $$

矩

对于随机变量 X 和非负整数 k ，若 $E[|X^k|] < \infty$ ，则称

$$E[X^k]$$

为 X 的 k 阶原点矩，简称 k 阶矩

若 $E[|X-E[X]|^k] < \infty$ ，则称

$$E[(X-E[X])^k]$$

为 X 的 k 阶中心矩

常见分布的期望与方差

0-1 分布 $B(p)$

若 $X \sim B(p)$，则

$$
\begin{aligned}
E[X] &= p \\
D[X] &= p(1-p)
\end{aligned}
$$

二项分布 $B(n,p)$

若 $X \sim B(n,p)$ ，则

$$
\begin{aligned}
E[X] &= np \\
D[X] &= np(1-p)
\end{aligned}
$$

并非使用定义而是利用期望的线性性质计算

泊松分布 $P(\lambda)$

若 $X \sim P(\lambda)$ ，则

$$
\begin{aligned}
E[X] &= \lambda \\
D[X] &= \lambda
\end{aligned}
$$

均匀分布 $U(a,b)$

若 $X \sim U(a,b)$，则

$$E[X] = \frac{a+b}{2}$$
$$D[X] = \frac{(b-a)^2}{12}$$

指数分布 $e(\lambda)$

若 $X \sim e(\lambda)$，则

$$E[X] = \lambda^{-1}$$
$$D[X] = \lambda^{-2}$$

正态分布 $N(\mu,\sigma^2)$

若 $X \sim N(\mu,\sigma^2)$，则

$$E[X] = \mu$$
$$D[X] = \sigma^2$$

且任意独立的正态分布的线性组合仍是正态分布，设 $X_i \sim N(\mu_i,\sigma_i^2)$

$$\sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

利用期望和方差的性质即可证明

协方差与相关系数

协方差

对于随机变量 X,Y，若 $E[|X|],E[|Y|],E[(X-E[X])(Y-E[Y])]$ 均有限，则定义其协方差 $cov(X,Y)$ 为

$$cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

随机变量和与差的方差可表示为

$$D[X \pm Y] = D[X] + D[Y] \pm 2cov(X,Y)$$

一般常用下述公式计算协方差

$$cov(X,Y) = E[XY] - E[X]E[Y]$$

协方差有性质：

若 X,Y 独立，可以得出 $cov(X,Y)=0$ ，但是反之不一定成立

方差是特殊的协方差：$cov(X,X)=D[X]$

协方差对称：$cov(X,Y)=cov(Y,X)$

对于常数 $a,b,c,d\in\mathbb{R}$ ，有

$$cov(aX+c,bY+d)=abcov(X,Y)$$

且

$$cov(X_1+X_2,Y)=cov(X_1,Y)+cov(X_2,Y)$$

Cauchy-Schwarz 不等式：

$$(cov(X,Y))^2\leqslant D[X]D[Y]$$

等号成立的条件是存在不全为 0 的常数 a,b 使得

$$P(a(X-E[X])+b(Y-E[Y])=0)=1$$

相关系数

引入相关系数是为了消除协方差中量纲的影响。若 X,Y 的二阶矩有限，且 $D[X]>0,D[Y]>0$ ，则定义其相关系数为

$$\rho_{XY}=\frac{cov(X,Y)}{\sqrt{D[X]D[Y]}}$$

相关系数有性质：

$|\rho_{XY}|\leqslant 1$

上述等号成立的充要条件是存在 $a,b,c\in\mathbb{R}$ ，满足 $P(aX+cY=b)=1$ ，即 X,Y 以概率 1 具有线性关系。可看出相关性刻画变量间线性相关的程度，$|\rho_{XY}|$ 越接近 1 表示 X,Y 线性相关的程度越大，其为正则为正相关，为负则为负相关

若 $\rho_{XY}=0$ ，称为 X,Y 不相关，其等价于

- $\rho_{XY}=0$
- $cov(X,Y)=0$
- $E[XY]=E[X]E[Y]$
- $D[X\pm Y]=D[X]+D[Y]$

对于变量的相关性和独立性，X,Y 独立 $\Rightarrow$ X,Y 不相关，但反向不成立。但是例外是服从二维正态分布的 (X,Y) ，其独立性和不相关性等价。有

$$cov(X, Y) = \sigma_1 \sigma_2 \rho$$
$$\rho_{XY} = \rho$$