数字特征之方差

Guangyao Zhao

2022-08-11

Contents

随机变量 X 的期望是 $a = E(X)$，但是样本众多，不可能全部都一定恰好是 a，所以就会有所偏离，而描述这个『偏离』的程度便是方差。方差大偏离程度就大，反之小。

方差

方差的计算公式如下：

$$D(X) = Var(X) = E\{[X - E(X)]^2\}$$

标准差为：$\sigma(X) = \sqrt{D(X)}$。方差还有一种表达方式，在之后的计算的时候也经常使用这个公式：

$$\begin{aligned} D(X) &= E\{[X - E(X)]^2\} \\ &= E[X^2 + E^2(X) - 2XE(X)] \\ &= E(X^2) + E^2(X) - 2E(X)E(X) \\ &= E(X^2) - E^2(X) \end{aligned}$$

还记得初中数学老师很感兴地说过这对公式:『差的平方的期望』等于『期望的平方的差』。

例子

0-1 分布

X^2	0	1
P	p	1-p

$$\begin{aligned} D(X) &= E(X^2) - E^2(X) \\ &= p - p^2 \\ &= p(1-p) \end{aligned}$$

泊松分布

已知泊松分布的期望 $E(X) = \lambda$, 所以只需要求出 $E(X^2)$ 即可:

$$\begin{aligned} E\left(X^2\right) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} \\ &= \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k e^{\lambda}}{(k-1)!} \\ &= \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!} + \sum_{k=0}^{\infty} \frac{\lambda^{k+1} e^{-\lambda}}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\lambda^{k+2} e^{-\lambda}}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^{k+1} e^{-\lambda}}{k!} \\ &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \left(\lambda^2 e^{-\lambda} + e^{-\lambda}\right) \\ &= \lambda^2 + \lambda \end{aligned}$$

即: $D(X) = E(X^2) - E^2(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

均匀分布

均匀分布的期望 $E(X)=\frac{b-a}{2}$，所以只需要求出 $E(X^2)$ 即可，因为均匀分布是连续变量，所以就需要使用微积分的算法了：

$$
\begin{aligned}
E\left(X^2\right) &= \int_a^b x^2 \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \cdot \frac{1}{3} x^3 |_a^b \\
&= \frac{a^2+ab+ab}{3}
\end{aligned}
$$

即均匀分布的方差为 $D(X)=E(X^2)-E^2(X)=\frac{a^2+ab+ab}{3}-\left(\frac{b-a}{2}\right)^2=\frac{(b-a)^2}{12}$

指数分布

已知泊松分布的期望 $E(X)=\theta$，所以只需要求出 $E(X^2)$ 即可：

$$
\begin{aligned}
E\left(X^2\right) &= \int_0^{+\infty} x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} dx \\
&= 2\theta^2
\end{aligned}
$$

即均匀分布的方差为 $D(X)=E(X^2)-E^2(X)=2\theta^2-\theta^2=\theta^2$

方差的性质

- $D(C)=0$
- $D(C+X)=D(X)$
- $D(CX)=C^2D(X)$
- $D(X+Y)=D(X)+D(Y)-2E\{[X-E(X)]\cdot[Y-EY]\}$

特别的证明下最后一条性质：

$$
\begin{aligned}
D(X+Y) &= E\{[X+Y-E(X+Y)]^2\} \\
&= E\{[X+Y-E(X)-E(Y)]^2\} \\
&= E\{[X-E(X)+Y-E(Y)]^2\} \\
&= E\{[X-E(X)]^2+[Y-E(Y)]^2-2[X-E(X)]\cdot[Y-E(Y)]\} \\
&= D(X)+D(Y)-2E\{[X-E(X)]\cdot[Y-E(Y)]\}
\end{aligned}
$$

其中对于 $2E\{[X-E(X)]\cdot[Y-E(Y)]\}$:

$$\begin{aligned}E\{[X-E(X)]\cdot[Y-E(Y)]\} &= E[XY-E(X)Y-XE(Y)+E(X)E(Y)] \\ &= E(XY)-E(X)E(Y)-E(XE(Y)+E(X)E(Y) \\ &= E(XY)-E(X)(Y)\end{aligned}$$

即如果 X,Y 相互独立, 则 $D(X+Y)=D(X)+D(Y)$