■ 数学期望与方差

定义 1 设 X 为一离散型随机变量，其分布列为 $P(X=x_i)=p_i,(i=1,2,\dots)$，若级数 $\sum_{i=1}^{\infty}x_ip_i$ 绝对收敛，则称这级数为 X 的数学期望，简称期望或均值，记为 $E(X)$ ，即

$$\mathrm{E}(X)=\sum_{i=1}^{\infty}x_ip_i$$

否则，称 X 的数学期望不存在。

$$D(X)=\sum_{i=1}^{\infty}[x_i-E(X)]^2p_i=\sigma^2$$

为 X 的方差，就是 X 的函数 $[X-E(X)]^2$ 的数学期望。

在定义 1 中，要求 $\sum_{i=1}^{\infty}x_ip_i$ 绝对收敛是必须的，因为 X 的数学期望是一确定的量，不受 x_ip_i 在级数中排列次序的影响，这在数学上就要求级数绝对收敛。X 的数学期望也称为数 x_i 以概率 p_i 为权的加权平均。

定义 2 设 X 为一连续型随机变量，其密度函数为 $f(x)$ ，若 $\int_{-\infty}^{+\infty}|x|f(x)<+\infty$ 时，则称

$$E(X)=\int_{-\infty}^{+\infty}f(x)dx=\mu$$

为 X 的数学期望，否则称 X 的数学期望不存在。

$$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2f(x)dx=\sigma^2$$

计算方差时常用简化公式

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

■ 常见分布的数学期望

1．两点分布

设 $X\sim B(1,p)$ ，即 $P(X=0)=q,P(X=1)=p,\quad 0<p<1,p+q=1$，则

$$E(X)=0\cdot q+1\cdot p=p$$

$$D(X)=pq$$

2．二项分布

设 $X\sim B(k;n,p)$ ，即 $P(X=k)=C_n^kp^kq^{n-k},k=0,1,2,\dots,n,\quad 0<p<1,p+q=1$，则

$$E(X)=\sum_{k=0}^{n}k\cdot P(X=k)=\sum_{k=0}^{n}k\cdot C_n^k p^k q^{n-k}=\sum_{k=1}^{n}\frac{np\cdot(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!}p^i q^{(n-1)-(k-1)}$$

$$=np(p+q)^{n-1}=np$$

$D(X)=npq$ ，证明较复杂，略。

在二项分布中，当$n\to\infty,p\to 0$ 时，概率是服从泊松分布的：

$$\lim_{n\to\infty,p\to 0}C_n^k p^k q^{n-k}=\lim_{n\to\infty,p\to 0}\frac{n(n-1)\dots(n-k+1)}{k!}p^k q^{\frac{np}{p}-k}=\lim_{n\to\infty,p\to 0}\frac{n^k}{k!}p^k q^{\frac{np}{p}-k}$$

$$=\lim_{n\to\infty,p\to 0}\frac{(np)^k}{k!}\left((1-p)^{\frac{1}{-p}}\right)^{-np}(1-p)^{-k}=\lim_{n\to\infty,p\to 0}\frac{\lambda^k}{k!}\left(\left(1+\frac{-p}{n}\right)^{\frac{n}{-p}}\right)^{-\lambda}$$

$$=\lim_{n\to\infty,p\to 0}\frac{\lambda^k}{k!}e^{-\lambda}$$

其中， λ 是泊松分布的期望。

3．泊松分布

设 $X\sim P(k;\lambda)$ ，即 $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\dots$ ，$\lambda>0$，则

$$E(X)=\sum_{k=0}^{n}k\cdot P(X=k)=\sum_{k=0}^{n}k\cdot\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum_{k=1}^{n}\frac{\lambda^k}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=1}^{n}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

其中用到泰勒级数展开式 $e^{\lambda}=\sum_{k=1}^{n}\frac{\lambda^{k-1}}{(k-1)!}$

这说明，泊松分布的参数 λ 就是服从泊松分布的随机变量的均值。

$$D(X)=\lambda$$

4．均匀分布

设 $X\sim U[a,b]$ ，即 X 的密度函数为 $f(x)=\begin{cases}\frac{1}{b-a},a<x<b\\ 0,其他\end{cases}$，则

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{\mathrm{a}}^{b}x\cdot\frac{1}{b-a}dx=\frac{a+b}{2}$$

恰好是 $[a,b]$ 的中点，这与均值意义相符。

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

5．指数分布

设 $X\sim E(\lambda)$ ，即 X 的密度函数为 $f(x)=\begin{cases}\lambda e^{-\lambda x},x>0\\ 0,x\le 0\end{cases}$ ，则

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_{-\infty}^{+\infty} x\cdot\lambda e^{-\lambda x}dx=\int_{-\infty}^{+\infty} -xde^{-\lambda x}=\frac{1}{\lambda}$$

$$D(X)=\frac{1}{\lambda^2}$$

6．正态分布

设 $X\sim N(\mu,\sigma^2)$ ，即 X 的密度函数为 $f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty,\sigma>0$ ，则

$$E(X)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\int_{-\infty}^{+\infty} x\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,\text{ 令 } a=\frac{x-\mu}{\sigma},\text{则}$$

$$E(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\mu+a\cdot\sigma)\,e^{-\frac{a^2}{2}}da=\frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{a^2}{2}}da+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} ae^{-\frac{a^2}{2}}da=\mu$$

这说明正太分布的参数 μ 是正态随机变量的均值。

$$D(X)=\sigma^2$$

■ 数学期望的性质

定理： （1） $E(c)=c$ ；

（2） $E(cX)=cE(X)$；

（3） $E(X+Y)=E(X)+E(Y)$ ；

（4）设 $X\ Y$ 相互独立 $E(XY)=E(X)\cdot E(Y)$ ；

其中 c 为常数，且数学期望都存在。由此定理可推论：

推论 1 $E(c_1X_1+c_2X_2+\cdots+c_nX_n+b)=c_1E(X_1)+c_2E(X_2)+\cdots+c_nE(X_n)+b$

其中 c,b 均是常数，特别有 $E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$

推论 2 若 $X_1,X_2,\ldots,X_n$ 相互独立，则 $E(X_1\cdot X_2\cdot\ldots\cdot X_n)=E(X_1)\cdot E(X_2)\cdot\ldots\cdot E(X_n)$

■ 方差、协方差的性质

$$D(cX)=c^2D(X)$$

$$D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$$

$$D(X-Y)=D(X)+D(Y)-2Cov(X,Y)$$

若 $X\ Y$ 是相互独立的随机变量，即 $X\ Y$ 不相关，则协方差 $Cov(X,Y)=0$ ，因为

$$Cov(X,Y)=Cov(Y,X)=E\big[\big(X-E(X)\big)\big(Y-E(Y)\big)\big]=E(XY)-E(X)\cdot E(Y)$$

$$Cov(aX,bY)=abCov(X,Y)$$

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

引申：相关系数 $\boldsymbol{\rho_{XY}}$

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)\cdot D(Y)}}$$

若 $\rho_{XY}=0$ ，则称 X Y 不线性相关。充要性满足 $\rho_{XY}=0 \Leftrightarrow Cov(X,Y)=0$

■ **总体样本的无偏估计**

定义　设统计量 $\widehat{\Theta}=\widehat{\Theta}(X_1,X_2,\ldots,X_n)$ 是总体 Θ 中未知参数的估计量，若 $E(\widehat{\Theta})=\Theta$ ，则称 $\widehat{\Theta}$ 为 Θ 的无偏估计量；否则称为有偏估计量。

上面这个定义的意思就是说如果拿到了一堆样本观测值，然后想通过这一堆观测值去估计某个统计量 Θ （一般就是想估计总体的期望或方差），如果选择的方法所估计出来的统计量 $\widehat{\Theta}$ 的期望值与总体样本的统计量 Θ 相等，那么我们称这种方法下的估计量是无偏估计量，否则为有偏估计量。

有偏无偏是针对选择估计的方法所说的，它并不是针对具体某一次估计出来的估计量结果。如果方法不对，即使恰好在某一次计算出来一个值和总体样本统计量值相同，也并不代表所选的这个方法是无偏的。因为单次 $\widehat{\Theta}$ 值是和选取的样本相关的，每次样本（更加严格的意义是某次样本快照）的值变化了，那么每次 $\widehat{\Theta}$ 的值就有可能跟着变化，需对多个 $\widehat{\Theta}$ 求期望值来判断 $\widehat{\Theta}$ 的可信程度，如果一直重复这个试验，然后它的期望值与总体样本的统计量 Θ 一样，那么称按照这种方法估计出来的统计量是无偏的。

只有在无法通过整体直接“统计”获得你想要的“量”时，只能通过“部分样本”来做“整体样本”“量”的估计时，谈估计方法的“有偏”还是“无偏”才是有意义的。

重要假设　随机选取的样本 X_i 与总体样本 X 同分布，就是说他们的统计特性是完全一样的，即他们的期望值一样，方差值也一样，即

$$E(X_i)=E(X)=\mu$$

$$D(X_i)=D(X)=\sigma^2$$

定理 1　样本均值 $\bar{X}$ 是总体样本均值 μ 的无偏估计。

$$E(\bar{X})=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$$

$$D(\bar{X})=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\sigma^2}{n}$$

定理 2　样本方差 s^2 是总体样本方差 σ^2 的无偏估计。即

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$E(s^2)=E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)=\sigma^2$$

证明网查。之所以不是 $\frac{1}{n}$ ，是因为 $D(\bar{X})\neq 0$ 。或者如果已知总体均值 μ ，则

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

但实际上，总体均值 μ 都是未知的。