3.2 方差与矩

3.2.1 方差和标准差

现在我们转到本章开始时提到的另一类数字特征,即刻画随机变量在其中心位置附近散布程度的数字特征,其中最重要的是方差.

设随机变量 X 有均值 $a=E(X)$. 试验中,X 取的值当然不一定恰好是 a,而会有所偏离. 偏离的量 $X-a$ 本身也是随机的(因为 X 是随机的). 我们要取这个偏离 $X-a$ 的某种有代表性的数字,来刻画这偏离即散布的程度大小如何. 我们不能就取 $X-a$ 的均值,因为 $E(X-a)=E(X)-a=0$——正负偏离彼此抵消了. 一种解决办法是取 $X-a$ 的绝对值 $|X-a|$ 以消除符号,再取其

均值 $E|X-a|$，作为变量 X 取值的散布程度的数字特征. 这个量 $E|X-a|$ 叫做X(或其分布)的“平均绝对差”，是常用于刻画散布度的数字特征之一. 但是，由于绝对值在数学上处理甚不方便，人们就考虑了另一种作法：先把$(X-a)$平方以消去符号，然后取其均值得 $E(X-a)^2$，把它作为 X 取值散布度的衡量. 这个量就叫做 X 的“方差”(方差：“差”的“方”).

定义 2.1 设 X 为随机变量，分布为 F，则

$$\mathrm{Var}(X) = E(X - EX)^2 \tag{2.1}$$

称为 X(或分布 F)的方差*，其平方根$\sqrt{\mathrm{Var}(X)}$(取正值)称为 X(或分布 F)的标准差.

暂记 $EX=a$. 由于$(X-a)^2=X^2-2aX+a^2$，按定理 1.1 得

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - 2aE(X) + a^2 \\ &= E(X^2) - (EX)^2 \end{aligned}\tag{2.2}$$

方差的这个形式在计算上往往较为方便.

方差之所以成为刻画散布度的最重要的数字特征，原因之一是它具有一些优良的数字性质，反映在以下的几个定理中.

定理 2.1 1° 常数的方差为 0. 2° 若 C 为常数，则 $\mathrm{Var}(X+C)=\mathrm{Var}(X)$. 3° 若 C 为常数，则 $\mathrm{Var}(CX)=C^2\mathrm{Var}(X)$.

证 1° 若 $X=$常数 a，则 $E(X)=a$，故 $X-E(X)=0$，因而 $\mathrm{Var}(X)=0$.

2° 因为 $E(X+C)=E(X)+C$，故

$$\begin{aligned}\mathrm{Var}(X+C) &= E[(X+C)-(EX+C)]^2 = E[X-EX]^2 \\ &= \mathrm{Var}(X)\end{aligned}$$

3° 因 C 为常数，有 $E(CX)=CE(X)$. 故

$$\begin{aligned}\mathrm{Var}(CX) &= E[CX - CE(X)]^2 = C^2E(X-EX)^2 \\ &= C^2\mathrm{Var}(X)\end{aligned}$$

定理 2.2 独立随机变量之和的方差，等于各变量的方差之和：

---

* Var 是方差 Variance 的缩写.

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \quad (2.3)$$

证 记 $E(X_i)=a_i, i=1,\cdots,n$，则因 $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} a_i$，有

$$\mathrm{Var}(X_1 + \cdots + X_n) = E\left[\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} a_i\right]^2 = E\left[\sum_{i=1}^{n} (X_i - a_i)\right]^2$$

$$= \sum_{i,j=1}^{n} E[(X_i - a_i)(X_j - a_j)] \quad (2.4)$$

有两类项：一类是 i, j 相同，这类项，按方差的定义，即为 $\mathrm{Var}(X_i)$. 另一类项是 i, j 不同. 这时，因 X_i, X_j 独立，按定理 1.2 有 $E(X_iX_j) = E(X_i)E(X_j) = a_ia_j$. 所以

$$E[(X_i - a_i)(X_j - a_j)] = E(X_iX_j) - E(a_iX_j) - E(a_jX_i) + a_ia_j$$
$$= a_ia_j - a_ia_j - a_ia_j + a_ia_j = 0$$

这样，在(2.4)式最后一个和中，只剩下 $i=j$ 的那些项. 这些项之和即(2.3)式右边. 因而证明了本定理.

这个定理是方差的一个极重要的性质，它与均值的定理 1.1 相似. 但要注意的是：方差的定理要求各变量独立，而均值的定理则不要求.

例 2.1 设 X 为一随机变量，$E(X)=a$ 而 $\mathrm{Var}(X)=\sigma^2$. 记 $Y=(X-a)/\sigma$，则 $E(Y)=0$，且按定理 2.1 易知 $\mathrm{Var}(Y)=1$. 这样，对 X 作一线性变换后，得到一个具均值 0、方差 1 的变量 Y. 常称 Y 是 X 的“标准化”.

例 2.2 设 X 服从波哇松分布 $P(\lambda)$. 求其方差. 前已求出 $E(X)=\lambda$. 又据定理 1.3，知

$$E(X^2) = \sum_{i=0}^{\infty} i^2 \mathrm{e}^{-\lambda}\lambda^i / i!$$

把 i^2 写为 $i(i-1)+i$，注意到 $\sum_{i=0}^{\infty} i\mathrm{e}^{-\lambda}\lambda^i/i!$ 就是 X 的均值，即 λ，而 $i(i-1)/i! = 1/(i-2)!$，有

$$E(X^2)=\sum_{i=2}^{\infty}e^{-\lambda}\lambda^i/(i-2)!+\lambda$$

$$=\lambda^2\sum_{i=2}^{\infty}e^{-\lambda}\lambda^{i-2}/(i-2)!+\lambda$$

$$=\lambda^2e^{-\lambda}\sum_{j=0}^{\infty}\lambda^j/j!+\lambda=\lambda^2e^{-\lambda}e^{\lambda}+\lambda=\lambda^2+\lambda$$

于是按公式(2.2)得到 $\mathrm{Var}(X)=\lambda^2+\lambda-\lambda^2=\lambda$. 即波哇松分布 $P(\lambda)$的均值方差相同,都等于其参数 λ.

例 2.3 设 X 服从二项分布$B(n,p)$,求 $\mathrm{Var}(X)$.

把 X 表为(1.21)的形式,其中 X_i 由(1.20)定义,因为 $X_1,\cdots,X_n$ 独立,有 $\mathrm{Var}(X)=\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)$. 现计算 $\mathrm{Var}(X_i)$. 因 X_i 只取 1,0 两个值,概率分别为 p 和 $1-p$,故

$$E(X_i)=p,E(X_i^2)=p,i=1,\cdots,n$$

因而得到 $\mathrm{Var}(X_i)=p-p^2=p(1-p)$,而

$$\mathrm{Var}(X)=np(1-p) \tag{2.5}$$

本题也可由定义直接计算,但比这麻烦些.

例 2.4 再考察例 1.7,求该例中变量 X 的方差.

仍如该例把 X 表为 $X_1+\cdots+X_n$. 麻烦的是,这里 $X_1,\cdots X_n$ 并非独立,因而不能用定理 2.2. 但这种表示仍可简化计算,有

$$E(X^2)=E\Big(\sum_{i=1}^{n}X_i\Big)^2=\sum_{i,j=1}^{n}E(X_iX_j) \tag{2.6}$$

分两类项:一类是 $i=j$. 这类项之和为 $\sum_{i=1}^{n}E(X_i^2)$. 由于 X_i 只取 1,0 两值,故 $X_i^2=X_i$,因而

$$\sum_{i=1}^{n}E(X_i^2)=\sum_{i=1}^{n}E(X_i)=E(X)=n/(2n-1)$$

(见例 1.7)

对 $i\neq j$,取 $i=1,j=2$ 为例,其他 i,j 一样,因为 X_i,X_j 都只取 1,0 为值,有

$$E(X_1X_2)=P(X_1=1,X_2=1)$$

即“第1,2堆都恰成一双”的概率.这概率计算的思想,与例1.7中阐明过的完全一样,结果为

$$P(X_1=1,X_2=1)=2n\cdot 1\cdot(2n-2)\cdot 1\cdot(2n-4)!/(2n)!$$
$$=1/[(2n-1)(2n-3)]$$

又在和(2.6)中,$i\neq j$ 的项的个数为 $n(n-1)$,故第二类项($i\neq j$ 的项)之和为 $n(n-1)/[(2n-1)(2n-3)]$.由此,用公式(2.2),得

$$\begin{aligned}\mathrm{Var}(X)&=E(X^2)-(EX)^2\\&=n/(2n-1)+n(n-1)/[(2n-1)(2n-3)]\\&\quad-[n/(2n-1)]^2=4n(n-1)^2/[(2n-1)^2(2n-3)]\end{aligned}$$

例2.5 设 X 服从正态分布 $N(\mu,\sigma^2)$.注意到 $E(X)=\mu$,有

$$\mathrm{Var}(X)=E(X-\mu)^2=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2\mathrm{e}^{-(x-\mu)^2/2\sigma^2}\mathrm{d}x$$

作变数代换 $x=\mu+\sigma t$,得

$$\mathrm{Var}(X)=\sigma^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2\mathrm{e}^{-t^2/2}\mathrm{d}t$$

式中的积分已在例1.8中计算过,为$\sqrt{2\pi}$.所以

$$\mathrm{Var}(X)=\sigma^2\tag{2.7}$$

由此得到正态分布 $N(\mu,\sigma^2)$中另一参数 σ^2 的解释:它就是分布的方差.正态分布完全由其均值 μ 和方差 σ^2 决定,故也常说“均值为 μ 方差为σ^2 的正态分布”.经过标准化 $Y=(X-\mu)/\sigma$,按例2.1得出均值为0方差为1的正态分布,即标准正态分布 .这一点早在第二章例1.6中,通过直接计算分布的方法证明过(第二章(1.17)式).

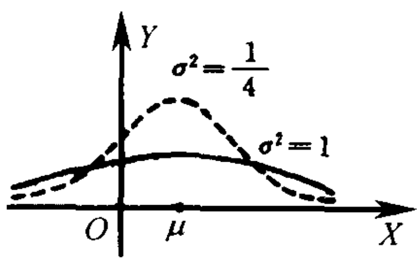

图3.3

方差 σ^2 愈小,则 X 的取值以更大的概率集中在其均值μ 附近,这一点也可从如下看出:正态分布 $N(\mu,\sigma^2)$ 的密度函数在 $x=\mu$ 点之值,等于 $(\sqrt{2\pi}\sigma)^{-1}$.它与 σ 成反比:σ 愈小,这个值愈大,而密度在 μ 点处有一个更

高的峰,显示概率更多地集中在 μ 点附近,见图 3.3.其中画出了正态 $N(\mu,\sigma^2)$ 当 $\sigma^2=1$ 和 $\sigma^2=1/4$ 时密度函数的图形.

例 2.6 指数分布(第二章例 1.7)的方差为 $1/\lambda^2$.区间 $[a,b]$ 上的均匀分布(第二章例 1.9)的方差为 $(b-a)^2/12$.这些都容易直接据公式(2.2)算出,留给读者.在均匀分布的情况,方差随区间 $[a,b]$ 之长 $b-a$ 的增大而增大,这当然,因区间长了,散布的程度也就大了.

例 2.7 求"统计三大分布"的方差.

先考虑卡方分布.设 $X\sim\chi_n^2$.把 X 表为 $X_1^2+\cdots+X_n^2$,X_1,$\cdots$,X_n 独立同分布且有公共分布 $N(0,1)$.有

$$\mathrm{Var}(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = E(X_i^4) - 1$$

而

$$E(X_i^4) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^4 \mathrm{e}^{-x^2/2}\mathrm{d}x = \frac{2}{\sqrt{2\pi}}\int_{-0}^{\infty} x^4 \mathrm{e}^{-x^2/2}\mathrm{d}x$$

作变数代换 $x=\sqrt{2t}$,有

$$E(X_i^4) = \frac{2}{\sqrt{2\pi}}2\sqrt{2}\int_0^{\infty} t^{3/2}\mathrm{e}^{-t}\mathrm{d}t = \frac{4}{\sqrt{\pi}}\Gamma\left(\frac{5}{2}\right)$$

$$= \frac{4}{\sqrt{\pi}}\,\frac{3}{2}\,\frac{1}{2}\sqrt{\pi} = 3$$

故 $\mathrm{Var}(X_i^4)=3-1=2$,而 $\mathrm{Var}(X)=2n$.

次考虑 t 分布.设 $X = X_1\Big/\sqrt{\frac{1}{n}X_2}$,X_1,X_2 独立而 $X_2\sim\chi_n^2$,$X_1\sim N(0,1)$.前已指出 $E(X)=0$.故由独立性有

$$\mathrm{Var}(X) = E(X^2) = E(X_1^2)E(n/X_2) = nE(1/X_2)$$

在例 1.8 中已算出 $E(1/X_2)=1/(n-2)$,故 $\mathrm{Var}(X)=n/(n-2)$,$(n>2)$.

自由度 n 的 t 分布 t_n 有期望 0,与标准正态 $N(0,1)$ 的期望同.其方差 $n/(n-2)$ 大于 1 但当 n 很大时接近 $N(0,1)$ 的方差 1.以后将指出:当 n 很大时,t_n 的分布确实接近 $N(0,1)$.

类似地算出自由度 m,n 的 F 分布 $F_{m,n}$ 的方差为 $2n^2(m+$

$n-2)/[m(n-2)^2(n-4)]$(当 $n>4$).细节留给读者.

3.2.2 矩

定义 2.2 设 X 为随机变量,c 为常数,k 为正整数.则量 $E[(X-c)^k]$称为X 关于c 点的k 阶矩.

比较重要的有两个情况:

1. $c=0$.这时 $a_k=E(X^k)$称为 X 的k 阶原点矩.

2. $c=E(X)$.这时 $\mu_k=E[(X-EX)^k]$称为X 的k 阶中心矩.

一阶原点矩就是期望.一阶中心矩 $\mu_1=0$,二阶中心矩 μ_2 就是 X 的方差 $\mathrm{Var}(X)$.在统计学上,高于 4 阶的矩极少使用.三、四阶矩有些应用,但也不很多.

应用之一是用 μ_3 去衡量分布是否有偏.设 X 的概率密度函数为$f(x)$.若 f 关于某点a 对称,即

$$f(a+x)=f(a-x)$$

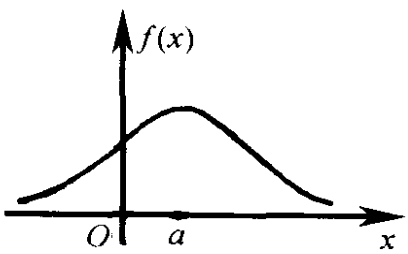

图 3.4

如图 3.4 所示,则 a 必等于$E(X)$,且 $\mu_3=E[X-E(X)]^3=0$.如果 $\mu_3>0$,则称分布为正偏或右偏.如果 $\mu_3<0$,则称分布为负偏或左偏.特别,对正态分布而言有 $\mu_3=0$,故如 μ_3 显著异于0,则是分布与正态有较大偏离的标志.由于 μ_3 的因次是 X 的因次的三次方,为抵消这一点,以 X 的标准差的三次方,即 $\mu_2^{3/2}$ 去除 μ_3.其商

$$\beta_1=\mu_3/\mu_2^{3/2} \tag{2.8}$$

称为 X 或其分布的"偏度系数".

应用之二是用 μ_4 去衡量分布(密度)在均值附近的陡峭程度如何.因为 $\mu_4=E[X-E(X)]^4$,容易看出,若 X 取值在概率上很集中在 $E(X)$附近,则 μ_4 将倾向于小,否则就倾向于大.为抵消尺度的影响,类似于 μ_3 的情况,以标准差四次方即 μ_2^2 去除,得

$$\beta_2 = \mu_4/\mu_2^2 \tag{2.9}$$

它称为 X 或其分布的“峰度系数”.

若 X 有正态分布 $N(\mu,\sigma^2)$,则 $\beta_2=3$,与 μ 和 σ^2 无关.为了迁就这一点,也常定义 μ_4/μ_2^2-3 为峰度系数,以使正态分布有峰度系数 0.

“峰度”这个名词,单从表面上看,易引起误解.例如,我们在例 2.4 中已指出,并由图 3.3 看出,就正态分布 $N(\mu,\sigma^2)$ 而言,σ^2 愈小,密度函数在 μ 点处“高峰”就愈高且愈陡峭.那么,为何所有的正态分布都又有同一峰度系数?这岂不与这个名词的直觉含义不符?原因在于:μ_4 在除以 μ_2^2 后已失去了因次,即与 X 的单位无关.或者换句话说,两个变量 X,Y,谁的峰度大,不能直接比其密度函数,而要调整到方差为 1 后再去比.就是说,找两个常数 c_1,c_2,使 c_1X 和 c_2Y 的方差都为 1,再比较其密度的“陡峭”程度如何.

在这个共同的标准下,“峰度”一词就好理解了.不信看图 3.5.为便于理解,我们在图中画了两条都以 μ 为对称中心的对称密度曲线,且峰的高度一样,但 f_1 在顶峰处很陡.而 f_2 则在顶峰处形成平台,较为平缓.这样,

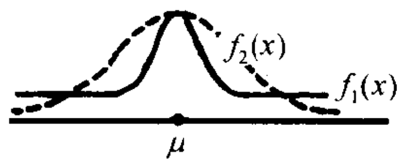

图 3.5

在 μ 附近,f_2 的概率多而 f_1 的概率少.而方差都为 1,故 f_1 的“尾巴”必比 f_2 的厚一些,这导致其 μ_4 较大,即有较大的峰度系数.