第三章　随机变量的数字特征

在前章中,我们较仔细地讨论了随机变量的概率分布,这种分布是随机变量的概率性质最完整的刻画.而随机变量的数字特征,则是某些由随机变量的分布所决定的常数,它刻画了随机变量(或者说,刻画了其分布)的某一方面的性质.

例如,考虑某种大批生产的元件的寿命.如果知道了它的概率分布,就可以知道寿命在任一指定界限内的元件百分率有多少,这对该种元件寿命状况提供了一幅完整的图景.如下文将指出的,根据这一分布就可以算出元件的平均寿命 m,m 这个数虽则不能对寿命状况提供一个完整的刻画,但却在一个重要方面,且往往是人们最为关心的一个方面,刻画了元件寿命的状况,因而在应用上有极重要的意义.类似的情况很多,比如我们在了解某一行业工人的经济状况时,首先关心的恐怕会是其平均收入,这给了我们一个总的印象.至于收入的分布状况,除非为了特殊的研究目的,倒反而不一定是最重要的.

另一类重要的数字特征,是衡量一个随机变量(或其分布)取值的散布程度.例如,两个行业工人的平均收入大体相近,但一个行业中收入分配较平均:大多数人的收入都在平均值上下不远处,其"散布"小;另一个行业则相反:其收入远离平均值者甚多,散布较大,这二者的实际意义当然很不同.又如生产同一产品的两个工厂,各自的产品平均说来都能达到规格要求,但一个厂波动小,较为稳定,另一个厂则波动大,有时质量超标准,有时则低于标准不少,这二者的实际后果当然也不同.

上面论及的平均值和散布度,是刻画随机变量性质的两类最重要的数字特征.对多维变量而言,则还有一类刻画各分量之间的关系的数字特征.在本章中,我们将就以上各类数字特征中,举其

最重要者进行讨论.

3.1 数学期望(均值)与中位数

要说明这个名称的来由,让我们回到第一章的例 1.1. 甲乙二人赌技相同,各出赌金 100 元,约定先胜三局者为胜,取得全部 200 元. 现在甲胜 2 局乙胜 1 局的情况下中止,问赌本该如何分? 在那里我们已算出,如果继续赌下去而不中止,则甲有 3/4 的机会(概率)取胜,而乙胜的机会为 1/4. 所以,在甲胜 2 局乙胜 1 局这个情况下,甲能"期望"得到的数目,应当确定为

$$200 \times \frac{3}{4} + 0 \times \frac{1}{4} = 150(\text{元})$$

而乙能"期望"得到的数目,则为

$$200 \times \frac{1}{4} + 0 \times \frac{3}{4} = 50(\text{元})$$

如果引进一个随机变量 X, X 等于在上述局面(甲 2 胜乙 1 胜)之下,继续赌下去甲的最终所得,则 X 有两个可能值:200 和 0,其概率分别为 3/4 和 1/4. 而甲的期望所得,即 X 的"期望"值,即等于

X 的可能值与其概率之积的累加

这就是"数学期望"(简称期望)这个名词的由来. 这个名词源出赌博,听起来不大通俗化或形象易懂,本不是一个很恰当的命名,但它在概率论中已源远流长获得大家公认,也就站住了脚根. 另一个名词"均值"形象易懂,也很常用,将在下文解释.

3.1.1 数学期望的定义

先考虑一个最简单的情况.

定义 1.1 设随机变量 X 只取有限个可能值 $a_1, \cdots, a_m$. 其概率分布为 $P(X = a_i) = p_i, i = 1, \cdots, m$. 则 X 的数学期望,记为 $E(X)$* 或 EX,定义为

---

* E 是期望 Expectation 的缩写.

$$E(X) = a_1p_1 + a_2p_2 + \cdots a_mp_m \quad (1.1)$$

名词的来由已如前述.数学期望也常称为“均值”,即“随机变量取值的平均值”之意,当然这个平均,是指以概率为权的加权平均.

利用概率的统计定义,容易给均值这个名词一个自然的解释.假定把试验重复 N 次,每次把 X 取的值记下来,设在这 N 次中,有 N_1 次取 a_1,N_2 次取 a_2,$\cdots$,N_m 次取 a_m.则这 N 次试验中 X 总共取值为 $a_1N_1 + a_2N_2 + \cdots + a_mN_m$,而平均每次试验中 X 取值,记为 $\bar{X}$,等于

$$\begin{aligned}\bar{X} &= (a_1N_1 + a_2N_2 + \cdots + a_mN_m)/N \\ &= a_1(N_1/N) + a_2(N_2/N) + \cdots + a_m(N_m/N)\end{aligned}$$

N_i/N 是事件$\{X = a_i\}$在这 n 次试验中的频率.按概率的统计定义(见第一章,1.1 节),当 N 很大时,N_i/N 应很接近 p_i.因此,$\bar{X}$ 应接近于(1.1)式右边的量,就是说,X 的数学期望 $E(X)$不是别的,正是在大量次数试验之下,X 在各次试验中取值的平均.

很自然地,如果 X 为离散型变量,取无穷个值 $a_1, a_2, \cdots$,而概率分布为 $P(X = a_i) = p_i$, $i = 1, 2, \cdots$,则我们仿照(1.1),而把 X 的数学期望 $E(X)$定义为级数之和:

$$E(X) = \sum_{i=1}^{\infty} a_ip_i \quad (1.2)$$

但当然,必须级数收敛才行,实际上我们要求更多,要求这个级数绝对收敛:

定义 1.2 如果

$$\sum_{i=1}^{\infty} |a_i| p_i < \infty \quad (1.3)$$

则称(1.2)式右边的级数之和为 X 的数学期望.

为什么不就要求(1.2)右边收敛而必须要求(1.3)? 这就涉及

级数理论中的一个现象:如果某个级数,例如 $\sum_{i=1}^{\infty} a_i p_i$,只是收敛(称为条件收敛),而其绝对值构成的级数 $\sum_{i=1}^{\infty} |a_i| p_i$,并不收敛,则将这级数各项次序改排以后,可以使它变得不收敛,或者使它收敛而其和等于事先任意指定之值.这就意味着(1.2)右边的和存在与否,等于多少,与随机变量 X 所取之值的排列次序有关,而 $E(X)$作为刻画 X 的某种特性的数值,有其客观意义,不应与其值的人为排列次序有关.

在连续型随机变量的情况,以积分代替求和,而得到数学期望的定义:

定义 1.3 设 X 有概率密度函数 $f(x)$.如果

$$\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x < \infty \tag{1.4}$$

则称

$$E(X) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \tag{1.5}$$

为 X 的数学期望.

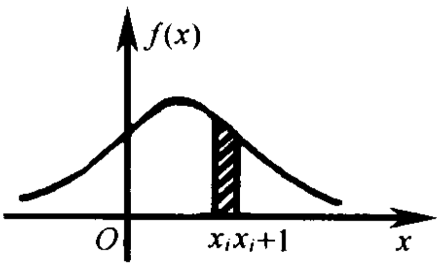

图 3.1

这个定义可以用离散化的方式来加以解释.如图 3.1,在 x 轴上用密集的点列 $\{x_i\}$ 把 x 轴分成很多小区间,长为 $x_{i+1} - x_i = \Delta x_i$.当 X 取值于区间 $[x_i, x_{i+1})$ 内时,可近似地认为其值就是 x_i.按密度函数的定义,X 取上述区间内之值的概率,即图中斜线标出部分的面积,近似地为 $f(x_i)\Delta x_i$.用这个方式,我们把原来的连续型随机变量 X 近似地离散化为一个取无穷个值 $\{x_i\}$ 的离散型变量 X',X' 的分布为 $P(X' = x_i) \approx f(x_i)\Delta x_i$.按定义 1.2,有

$$E(X') \approx \sum_i x_i f(x_i) \Delta x_i$$

随着区间 Δx_i 愈分愈小，X' 愈来愈接近 X，而上式右端之和也愈来愈接近于(1.5)式右边的积分，这样就得出定义 1.3. 至于要求积分绝对收敛即(1.4)式，其原因与定义 1.2 的情况有所不同，在此不能细论了.

例 1.1 设 X 服从波哇松分布 $P(\lambda)$(见第二章例 1.2)，则

$$E(X)=\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda}=\lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}$$

$$=\lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}=\lambda e^{-\lambda}e^{\lambda}=\lambda \tag{1.6}$$

这解释了波哇松分布 $P(\lambda)$中参数 λ 的意义，拿第二章例 1.2 的情况来说，λ 就是在所指定的时间段中发生事故的平均次数.

例 1.2 设 X 服从负二项分布(见第二章例 1.5 的(1.11)式)，则

$$E(X)=p^r\sum_{i=0}^{\infty} i\binom{i+r-1}{r-1}(1-p)^i \tag{1.7}$$

为求这个和，我们要用到在第二章例 1.5 中指出过的负指数二项展开式

$$(1-x)^{-r}=\sum_{i=0}^{\infty}\binom{i+r-1}{r-1}x^i$$

两边对 x 求导，得

$$r(1-x)^{-r-1}=\sum_{i=0}^{\infty} i\binom{i+r-1}{r-1}x^{i-1}$$

在上式中令 $x=1-p$，然后两边同乘 $1-p$ 得到

$$\sum_{i=0}^{\infty} i\binom{i+r-1}{r-1}(1-p)^i=rp^{-(r+1)}(1-p)$$

而

$$E(X)=p^r\cdot rp^{-(r+1)}(1-p)=r(1-p)/p \tag{1.8}$$

p 愈小，则此值愈大，这是自然的：若事件 A 的概率 p 很小，则等待它出现 r 次的平均时间也就愈长，当 $r=1$ 时，得到几何分布(第二章(2.12)式)的期望为$(1-p)/p$.

例 1.3 若 X 服从$[a,b]$区间的均匀分布(第二章例 1.9),则

$$E(X)=\frac{1}{b-a}\int_a^b x\mathrm{d}x=\frac{1}{2}(a+b) \tag{1.9}$$

即期望为区间中点,这在直观上很显然.

例 1.4 若 X 服从指数分布(第二章例 1.7,(1.20)式),则

$$E(X)=\lambda\int_0^{\infty}x\mathrm{e}^{-\lambda x}\mathrm{d}x=\lambda^{-1}\int_0^{\infty}x\mathrm{e}^{-x}\mathrm{d}x=\lambda^{-1}\Gamma(2)=\lambda^{-1} \tag{1.10}$$

这个结果的直观解释曾在第二章例 1.7 中指出过.

例 1.5 设 X 服从正态分布 $N(\mu,\sigma^2)$,则

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

作变数代换 $x=\mu+\sigma t$,化为

$$\begin{aligned}E(X)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\mu+\sigma t)\mathrm{e}^{-t^2/2}\mathrm{d}t\\&=\mu\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-t^2/2}\mathrm{d}t+\sigma\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t\mathrm{e}^{-t^2/2}\mathrm{d}t\end{aligned}$$

上式右边第一项为 μ,第二项为 0.因此

$$E(X)=\mu \tag{1.11}$$

这样,我们得到了正态分布 $N(\mu,\sigma^2)$中两个参数之一的 μ 的解释:μ 就是均值,这一点从直观上看很清楚,因为 $N(\mu,\sigma^2)$的密度函数关于 μ 点对称(见第二章图 2.2b),其均值自应在这个点.

因为数学期望是由随机变量的分布完全决定的,故我们可以而且常常说某分布 F 的期望是多少,某密度 f 的期望是多少等.期望是通过概率分布而决定这个事实,可能会被理解为:在任何应用的场合,当谈到某变量 X 的期望时,必须知道其分布,这话不完全确切.在有些应用问题中,人们难于决定有关变量的分布如何,甚至也难于对之提出某种合理的假定,但有相当的根据(经验的或理论的)对期望值提出一些假定甚至有不少了解.例如,我们可能比较确切地知道某行业工人的平均工资,而对工资的分布情况并不很清楚.另外,当需要通过观察或试验取得数据以进行估计时,

去估计一变量的期望，要比去估计其分布容易且更确切，因为期望只是一个数而分布（或密度）是一个函数. 以上所说对其他的数字特征也成立. 在本书后面讲到数理统计学时将更明白这一点.

3.1.2 数学期望的性质

数学期望之所以在理论和应用上都极为重要，除了它本身的含义（作为变量平均取值之刻画）外，还有一个原因，即它具有一些良好的性质，这些性质使得它在数学上很方便. 本段就是讨论这个问题.

定理 1.1 若干个随机变量之和的期望，等于各变量的期望之和，即

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n) \tag{1.12}$$

当然，这里要假定各变量 X_i 的期望都存在.

证 先就 $n=2$ 的情况来证，若 X_1,X_2 为离散型，分别以 $a_1,a_2,\cdots$ 和 $b_1,b_2,\cdots$ 记 X_1 和 X_2 的一切可能值，而记其分布为

$$P(X_1=a_i,X_2=b_j)=p_{ij},i,j=1,2,\cdots \tag{1.13}$$

当 $X_1=a_i,X_2=b_j$ 时，有 $X_1+X_2=a_i+b_j$. 故

$$E(X_1+X_2)=\sum_{i,j}(a_i+b_j)p_{ij}=\sum_{i,j}a_ip_{ij}+\sum_{i,j}b_jp_{ij}. \tag{1.14}$$

先看第一项，据第二章(2.8)式，有

$$P(X_1=a_i)=\sum_j p_{ij}$$

所以，按定义 1.2，有

$$\sum_{i,j}a_ip_{ij}=\sum_i a_i\sum_j p_{ij}=\sum_i a_iP(X_1=a_i)=E(X_1)$$

同理，(1.14)右边第二项为 $E(X_2)$. 这证明了所要结果.

若(X_1,X_2)为连续型，以 $f(x_1,x_2)$ 记其联合密度，按第二章(4.16)式，知 X_1+X_2 的密度函数为 $l(y)=\int_{-\infty}^{\infty}f(x,y-x)\mathrm{d}x$.

故按定义 1.3，有

$$E(X_1+X_2)=\int_{-\infty}^{\infty} yl(y)\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y-x)\mathrm{d}x\mathrm{d}y$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} yf(x,y-x)\mathrm{d}y\right)\mathrm{d}x$$

在里面那个积分作变数代换 $y=x+t$，得

$$E(X_1+X_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+t)f(x,t)\mathrm{d}x\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,t)\mathrm{d}x\mathrm{d}t+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} tf(x,t)\mathrm{d}x\mathrm{d}t \tag{1.15}$$

按第二章(2.9)式，知 $\int_{-\infty}^{\infty} f(x,t)\mathrm{d}t$ 就是 X_1 的密度函数. 所以，(1.15)右边第一个积分等于

$$\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f(x,t)\mathrm{d}t\right)\mathrm{d}x=\int_{-\infty}^{\infty} xf_1(x)\mathrm{d}x=E(X_1)$$

同理证明第二个积分为 $E(X_2)$，于是证得了所要的结果.

一般情况可用归纳的方式得到. 例如，记 $Y=X_1+X_2$，有

$$E(X_1+X_2+X_3)=E(Y+X_3)=E(Y)+E(X_3)$$

$$=E(X_1)+E(X_2)+E(X_3)$$

等等. 定理 1.1 证毕.

定理 1.2 若干个独立随机变量之积的期望，等于各变量的期望之积：

$$E(X_1X_2\cdots X_n)=E(X_1)E(X_2)\cdots E(X_n)$$

当然，这里也要假定各变量 X_i 的期望都存在.

证 与定理 1.1 相似，只须对 $n=2$ 的情况证明即可. 先设 X_1,X_2 都为离散型，其分布为(1.13). 由独立性假定知 $p_{ij}=P(X_1=a_i)P(X_2=b_j)$.

因为当 $X_1=a_i,X_2=b_j$ 时有 $X_1X_2=a_ib_j$，故

$$E(X_1X_2)=\sum_{i,j}a_ib_jp_{ij}=\sum_{i,j}a_ib_jP(X_1=a_i)P(X_2=b_j)$$
$$=\sum_i a_iP(X_1=a_i)\sum_j b_jP(X_2=b_j)$$
$$=E(X_1)E(X_2)$$

如所欲证.若(X_1,X_2)为连续型,则因独立性,其联合密度$f(x_1,x_2)$等于各分量密度$f_1(x_1)$与$f_2(x_2)$之积,故

$$E(X_1X_2)=\iint\limits_{-\infty}^{\infty}x_1x_2f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$
$$=\int_{-\infty}^{\infty}x_1f_1(x_1)\mathrm{d}x_1\int_{-\infty}^{\infty}x_2f_2(x_2)\mathrm{d}x_2$$
$$=E(X_1)E(X_2)$$

细心的读者可能会注意到,在后一段证明中我们是从公式

$$E(X_1X_2)=\iint\limits_{-\infty}^{\infty}x_1x_2f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{1.16}$$

出发,而这公式并非直接从期望的定义而来,它也需要证明.因此,更严格的证法应如定理1.1那样,先推导出X_1X_2的密度g,计算$\int_{-\infty}^{\infty}xg(x)\mathrm{d}x$再通过积分变数代换.这不难做到,我们把它放在习题里留给读者去完成(习题21).

读者也许还会问:在以上两个定理中,如果一部分变量为离散型,一部分为连续型,结果如何? 答案是结论仍成立.对乘积的情况,由于有独立假定,证明不难.对和的情况则要用到高等概率论,这些都不在此细讲了.

要注意到定理1.2和1.1之间的区别:后者不要求变量有独立性.读者也可以思考一下这个问题:如果说,事件积的概率的定理(第一章定理3.3)与此处定理1.2完全对应,那么,为什么事件和概率的定理(第一章定理3.1)与此处的定理1.1并不完全对应(概率加法定理中有互斥要求而定理1.1无任何要求),道理何在?

定理1.3(随机变量函数的期望)　设随机变量X为离散型,

有分布 $P(X=a_i)=p_i, i=1,2,\cdots$,或者为连续型,有概率密度函数 $f(x)$.则

$$E(g(X)) = \sum_i g(a_i)p_i\left(当\sum_i |g(a_i)|p_i < \infty 时\right) \tag{1.17}$$

或

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x\left(当\int_{-\infty}^{\infty} |g(x)|f(x)\mathrm{d}x < \infty 时\right) \tag{1.18}$$

这个定理的实质在于:为了计算 X 的某一函数 $g(X)$ 的期望,并不需要先算出 $g(X)$ 的密度函数,而可以就从 X 的分布出发,这当然大大方便了计算,因为在 g 较为复杂时,$g(X)$ 的密度很难求.

证 离散型情况(1.17)好证,因为 $P(X=a_i)=p_i$,有 $P(g(X)=g(a_i))=p_i$ ($g(a_1), g(a_2), \cdots$ 中可以有相重的,但这并不影响下面的证明).由此立即得出(1.17).

连续型情况较复杂,我们只能就 g 为严格上升并可导的情况给出证明.按第二章(4.2)式,这时 $Y=g(X)$ 的密度函数为 $f(h(y))h'(y)$,其中 h 为 g 的反函数,即 $h(g(x))=x$.此式两边对 x 求导,得 $h'(y)|_{y=g(x)}g'(x)=1$,即 $h'(g(x))=1/g'(x)$.因此

$$E(g(X)) = \int_{-\infty}^{\infty} yf(h(y))h'(y)\mathrm{d}y$$

作积分变数代换 $y=g(x)$,注意到 $f(h(g(x)))=f(x)$, $h'(g(x))=1/g'(x)$ 及 $\mathrm{d}y=g'(x)\mathrm{d}x$,得

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$$

即(1.18).一般情况(g 非单调)的证明超出本书范围之外,但对有些简单情况,$g(X)$ 虽非单调,但 $g(X)$ 的密度不难求得,这时(1.18)也不难证.有几种这样的情况作为习题留给读者.

本定理的一个重要特例是

系 1.1 若 c 为常数,则

$$E(cX) = cE(X) \tag{1.19}$$

证明由取 $g(x)=cx$ 得出. 当然,直接证明也很容易.

这几个定理无论在理论上和实用上都有重大意义,这里我们举几个例子说明其应用.

例 1.6 设 X 服从二项分布 $B(n,p)$,求 $E(X)$.

此例不难由定义 1.1 直接计算,但如下考虑更简单:因 X 为 n 次独立试验中某事件 A 发生的次数,且在每次试验中 A 发生的概率为 p. 故如引进随机变量 $X_1,\cdots,X_n$,其中

$$X_t = \begin{cases} 1, & \text{若在第 } i \text{ 次试验时事件} A \text{ 发生} \\ 0, & \text{若在第 } i \text{ 次试验时事件} A \text{ 不发生} \end{cases} \tag{1.20}$$

则 $X_1,\cdots,X_n$ 独立,且

$$X = X_1 + \cdots + X_n \tag{1.21}$$

按定理 1.1 有,$E(X)=E(X_1)+\cdots+E(X_n)$. 为计算 $E(X_i)$,注意按定义(1.20),X_i 只取两个值 1 和 0,其取 1 的概率为 p,取 0 的概率为 $1-p$. 因而 $E(X_i)=1\times p+0\times(1-p)=p$. 由此得到

$$E(X) = np \tag{1.22}$$

这比直接计算要简单些,又注意:在上述论证中并未用到 $X_1,\cdots,X_n$ 独立这一事实.

例 1.7 再考虑第一章例2.2 那个"n 双鞋随机地分成 n 堆"的试验,以 X 记"恰好成一双"的那种堆的数目,求 $E(X)$.

此题若要直接用定义 1.1,就须计算 $P(X=i)$,即"恰好有 i 个堆各自成一双"的概率. 这个概率计算不易,但使用上例的方法不难求解:引进随机变量 $X_1,\cdots,X_n$,其中

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 堆的两只恰成一双} \\ 0, & \text{若第 } i \text{ 堆的两只不成一双} \end{cases}$$

则仍有 $X=X_1+\cdots+X_n$,且 $E(X_i)=P(X_i=1)=P$(第 i 堆恰成一双). 为算这个概率,我们取如下的分堆方法:先把 $2n$ 只鞋随机地自左至右排成一列,然后让排在 1,2 位置的成一堆,3,4 位置的为第二堆,等等. 总的排列方法有$(2n)!$ 种. 有利于事件{第 i 堆恰成一双}的排法可计算如下:第 i 堆占据排列中的第$(2i-1)$和

第 $2i$ 号位置.第 $(2i-1)$ 号位置可以从 $2n$ 只鞋中任取一只,有 $2n$ 种取法.这只定了以后,为使恰成一双,第 $2i$ 号位置就只有一种取法.取好后,剩下的 $2n-2$ 只则可任意排,有 $(2n-2)!$ 种排法.因此,有利于上述事件的总排列数为 $2n\cdot1\cdot(2n-2)!$,而所求的概率为

$$2n(2n-2)!/(2n)! = 1/(2n-1)$$

此即为 $E(X_i)$,而 $E(X)=E(X_1)+\cdots+E(X_n)=n/(2n-1)$

例 1.8 试计算"统计三大分布"的期望值.

对自由度 n 的卡方分布,直接用其密度函数的形式(第二章(4.26)),Γ 函数的公式(第二章(4.23))及数学期望的定义 1.3,不难算出其期望为 n.略简单一些是用第二章例 4.9,把 X 表为 $X_1^2+\cdots+X_n^2$,$X_1,\cdots,X_n$ 独立且各服从标准正态分布 $N(0,1)$.按定理 1.3,有

$$E(X_i^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-x^2/2} x^2 dx$$

把 $e^{-x^2/2}x^2dx$ 写为 $-xd(e^{-x^2/2})$,用分部积分,得到

$$\int_0^{\infty} e^{-x^2/2} x^2 dx = \int_0^{\infty} e^{-x^2/2} dx = \frac{1}{2}\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}/2$$

后一式用第二章(1.15).于是得到 $E(X_i^2)=1$,而 $E(X)=E(X_1^2)+\cdots+E(X_n^2)=n$.

对自由度 n 的 t 分布,由于其密度函数(第二章(4.31)式)关于 0 对称,易见其期望为 0.但是有一个条件,就是自由度 n 必须大于 1.这是因为

$$\int_{-\infty}^{\infty} |x|(1+x^2/n)^{-\frac{n+1}{2}} dx = \infty \text{ 当 } n = 1$$

因而条件(1.4)不适合,当 $n>1$ 时上式的积分有限.

对自由度为 m,n 的 F 分布,写

$$X = \frac{1}{m}X_2\Big/\frac{1}{n}X_1 = m^{-1}nX_2/X_1$$

其中 X_1,X_2 独立,分别服从分布 χ_n^2 和 χ_m^2.由于 X_1,X_2 独立,按

第二章定理 3.3,知 X_1^{-1} 和 X_2 也独立,故按定理 1.2,有

$$E(X)=m^{-1}nE(X_2)E(X_1^{-1})=m^{-1}nmE(X_1^{-1})=nE(X_1^{-1}) \tag{1.23}$$

于是问题归结为计算 $E(X_1^{-1})$. 按定理 1.3,有

$$\begin{aligned}
E(X_1^{-1}) &= \left(2^{n/2}\Gamma\left(\frac{n}{2}\right)\right)^{-1}\int_0^\infty x^{-1}\mathrm{e}^{-x/2}x^{n/2-1}\mathrm{d}x \\
&= \left(2^{n/2}\Gamma\left(\frac{n}{2}\right)\right)^{-1}\int_0^\infty \mathrm{e}^{-x/2}x^{(n-2)/2-1}\mathrm{d}x \\
&= \left(2^{n/2}\Gamma\left(\frac{n}{2}\right)\right)^{-1}2^{(n-2)/2}\Gamma\left(\frac{n}{2}-1\right) \\
&= \frac{1}{2}\Gamma\left(\frac{n}{2}-1\right)\Big/\Gamma\left(\frac{n}{2}\right) \\
&= \frac{1}{2}\Gamma\left(\frac{n}{2}-1\right)\Big/\left[\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)\right] \\
&= 1/(n-2)
\end{aligned}$$

由此及(1.23),知

$$E(X)=n/(n-2) \quad (X\sim F_{m,n}) \tag{1.24}$$

此式只在 $n>2$ 时才有效. 当 $n=1,2$ 时,$F_{m,n}$的期望不存在.

3.1.3 条件数学期望(条件均值)

与条件分布的定义相似,随机变量 Y 的条件数学期望,就是它在给定的某种附加条件之下的数学期望. 对统计学来说,最重要的情况是:在给定了某些其他随机变量 $X,Z,\cdots$等的值 $x,z,\cdots$等的条件之下,Y 的条件期望,记为 $E(Y|X=x,Z=z,\cdots)$. 以只有一个变量 X 为例,就是 $E(Y|X=x)$. 在 X 已明确而不致引起误解的情况下,也可简记为 $E(Y|x)$.

如果知道了(X,Y)的联合密度,则 $E(Y|x)$的定义就可以具体化为:先定出在给定 $X=x$ 之下,Y 的条件密度函数 $f(y|x)$,然后按定义 1.3 算出

$$E(Y|x)=\int_{-\infty}^{\infty}yf(y|x)\mathrm{d}y \tag{1.25}$$

如果说,条件分布是变量 X 与 Y 的相依关系在概率上的完全刻画,那么,条件期望则在一个很重要的方面刻画了二者的关系,它反映了随着 X 取值 x 的变化,Y 的平均变化的情况如何,而这常常是研究者所关心的主要内容.例如,随着人的身高 x 的变化,具身高 x 的那些人的平均体重的变化情况如何;随着其受教育年数 x 的变化,其平均收入的变化如何等等.在统计学上,常把条件期望 $E(Y|x)$ 作为 x 的函数称为 Y 对 X 的"回归函数"(回归这个名词将在第六章中解释),而"回归分析",即关于回归函数的统计研究,构成统计学的一个重要分支.

例 1.9 条件期望的一个最重要的例子是 (X,Y) 服从二维正态分布 $N(a,b,\sigma_1^2,\sigma_1^2,\rho)$.根据第二章例 3.3,在给定 $X=x$ 时 Y 的条件分布为正态分布 $N(b+\rho\sigma_2\sigma_1^{-1}(x-a),\sigma_2^2(1-\rho^2))$.因为正态分布 $N(\mu,\sigma^2)$ 的期望就是 μ,故有

$$E(Y|x) = b + \rho\sigma_2\sigma_1^{-1}(x-a) \tag{1.26}$$

它是 x 的线性函数.如果 $\rho>0$,则 $E(Y|x)$ 随 x 增加而增加,即 Y"平均说来"有随 X 的增长而增长的趋势,这就是我们以前提到的"正相关"的解释.若 $\rho<0$,则为负相关,当 $\rho=0$ 时,X 与 Y 独立,$E(Y|x)$ 当然与 x 无关.

从条件数学期望的概念,可得出求通常的(无条件的)数学期望的一个重要公式.这个公式与计算概率的全概率公式相当.回想全概率公式 $P(A)=\sum_i P(B_i)P(A|B_i)$.它可以理解为通过事件 A 的条件概率 $P(A|B_i)$ 去计算其(无条件)概率 $P(A)$.更确定地说,$P(A)$ 就是条件概率 $P(A|B_i)$ 的某种加权平均,权即为事件 B_i 的概率.以此类推,变量 Y 的(无条件)期望,应等于其条件期望 $E(Y|x)$ 对 x 取加权平均,x 的权与变量 X 在 x 点的概率密度 $f_1(x)$ 成比例,即

$$E(Y) = \int_{-\infty}^{\infty} E(Y|x) f_1(x)\mathrm{d}x \tag{1.27}$$

此式很容易证明:以 $f(x,y)$ 记 (X,Y) 的联合密度函数,则 X,Y 的(边缘)密度函数分别为

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y \text{ 和 } f_2(y) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}x$$

按定义，$E(Y) = \int_{-\infty}^{\infty} y f_2(y)\mathrm{d}y$，可写为

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x, y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f(x, y)\mathrm{d}y\right]\mathrm{d}x \end{aligned}$$

由于 $E(Y|x) = \int_{-\infty}^{\infty} y f(x, y)\mathrm{d}y / f_1(x)$，有 $\int_{-\infty}^{\infty} y f(x, y)\mathrm{d}y = E(Y|x)f_1(x)$，而上式转化为(1.27).

公式(1.27)可给以另一种写法，记 $g(x) = E(Y|x)$，它是 x 的函数，则(1.27)成为

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_1(x)\mathrm{d}x \tag{1.28}$$

但据(1.18)，上式右边就是 $E(g(X))$. 从 $g(x)$ 的定义，$g(X)$ 是 $E(Y|x)|_{x=X}$，可简写为 $E(Y|X)$. 于是由(1.28)得

$$E(Y) = E[E(Y|X)] \tag{1.29}$$

这个公式可以形象地叙述为：一个变量 Y 的期望，等于其条件期望的期望. $E(Y|X)$ 这个符号的意义，从上面的叙述中已明确交代了，只须记住：在求 $E(Y|X)$ 时，先设定 X 等于一固定值 x，x 无随机性，这样可算出 $E(Y|x)$，其表达式含 x，再把 x 换成 X 即得.

公式(1.29)虽可算是概率论中一个比较高深的公式，它的实际含义其实很简单：它可理解为一个“分两步走”去计算期望的方法. 因为在不少情况下，迳直计算 $E(Y)$ 较难，而在限定某变量 X 之值后，计算条件期望 $E(Y|x)$ 则较容易. 因此我们分两步走：第一步算出 $E(Y|x)$，再借助 X 的概率分布，通过 $E(Y|x)$ 算出 $E(Y)$. 更直观一些，你可以把求 $E(Y)$ 看成为在一个很大的范围求平均. 限定 X 之值从这个很大的范围内界定了一个较小的部分. 先对这较小的部分求平均，然后再对后者求平均. 比如要求全

校学生的平均身高，你可以先求出每个班的学生的平均身高，然后再对各班的平均值求一次平均. 自然，在作后一平均时，要考虑到各班人数的不同，是以各班人数为权的加权平均. 这个权的作用相当于公式(1.27)中的 $f_1(x)$.

公式(1.29)虽来自(1.27)，但因为其形式并不要求对 X,Y 有特殊的假设，故可适用于更为一般的情形. 例如，X 不必是一维的，如果 X 为 n 维随机向量$(X_1,\cdots,X_n)$，有概率密度 $f(x_1,\cdots,x_n)$，则公式(1.29)有形式

$$E(Y)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}E(Y|x_1,\cdots,x_n)f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \tag{1.30}$$

这里 $E(Y|x_1,\cdots,x_n)$就是在 $X_1=x_1,\cdots,X_n=x_n$ 的条件下，Y 的条件期望，又 X,Y 都可以是离散型的. 例如，设 X 为一维离散型变量，有分布

$$P(X=a_i)=p_i,\quad i=1,2,\cdots$$

则公式(1.29)有形式

$$E(Y)=\sum_{i=1}^{\infty}p_iE(Y|a_i) \tag{1.31}$$

3.1.4 中位数

刻画一个随机变量 X 的平均取值的数学特征，除了数学期望以外，最重要的是中位数.

定义 1.4 设连续型随机变量 X 的分布函数为 $F(x)$，则满足条件

$$P(X\leqslant m)=F(m)=1/2 \tag{1.32}$$

的数 m 称为 X 或分布 F 的中位数.

由于连续型变量取一个值的概率为 0，$P(X=m)=0$，由(1.32)知

$$P(X\leqslant m)=P(X<m)=P(X>m)=P(X\geqslant m)=1/2$$

就是说，m 这个点把 X 的分布从概率上一切两半：在 m 左边(包

括点 m 与否无所谓)占一半,m 右边也占一半,从概率上说,m 这个点正好居于中央,这就是“中位数”得名的由来.

在实用上,中位数用得很多,特别有不少社会统计资料,常拿中位数来刻画某种量的代表性数值,有时它比数学期望更说明问题.例如,某社区内人的收入的中位数告诉我们:有一半的人收入低于此值,另一半高于此值.我们直观上感觉到这个值对该社区的收入情况,的确很具代表性.它和期望值相比它的一个优点是:它受个别特大或特小值的影响很小,而期望则不然.举例而言,若该社区中有一人收入在百万元以上,则该社区的均值可能很高,而绝大多数人并不富裕,这个均值并不很有代表性.中位数则不然:它不受少量这种特大值的影响.

从理论上说,中位数与均值相比还有一个优点,即它总存在,而均值则不是对任何随机变量都存在.

虽则中位数有这些优点,但在概率统计中,无论在理论和应用上,数学期望的重要性都超过中位数,其原因有以下两方面:

一是均值具有很多优良的性质,反映在前面的定理 1.1—1.3.这些性质使得在数学上处理均值很方便.例如,$E(X_1+X_2)=E(X_1)+E(X_2)$,这公式既简单又毫无条件(除了均值存在以外).中位数则不然:X_1+X_2 的中位数,与 X_1,X_2 各自的中位数之间,不存在简单的联系,这使中位数在数学上的处理很复杂且不方便.

二是中位数本身所固有的某些缺点.首先,中位数可以不唯一.例如,考察图 3.2 的密度函数 f.它只在两个分开的区间 (a,b) 和 (c,d) 内不为 0,且在这两段区间上围成的面积都是 1/2.这时,按中位数的定义 1.4,区间 $[b,c]$ 中任何一点 m 都是中位数.它没有一个唯一的值.

次一个问题是:在 X 为离散型的情况,虽也可以定义中位数(其定义与定义 1.4 有所不同),但并不很理想,不完全符合“中位”这名词所应有的含义.考察一个简单例子,设 X 取三个值 1,2,3,概率分布为

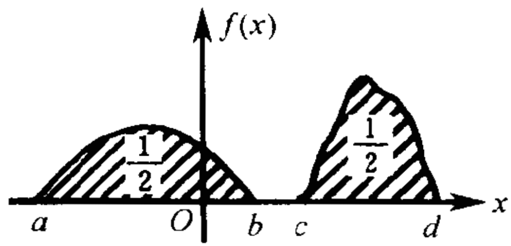

图 3.2

$$P(X=1)=2/7, P(X=2)=4/7, P(X=3)=1/7$$

这时就不存在一个点 m，使 m 两边的概率恰好一样，不得已只好退而求其次：找一个点 m，使其左右两边的概率差距最小，在本例中这个点是 2. 从 2 这个位置看，左边的概率（2/7）要比右边的概率（1/7）大. 故并不是理想的“中位”数.

例 1.10 正态分布 $N(\mu,\sigma^2)$ 的中位数就是 μ，这从 $N(\mu,\sigma^2)$ 的密度函数关于 μ 点对称可以看出. 指数分布函数已在第二章(1.21)式中列出，故其中位数 m 为方程 $1-e^{-\lambda m}=1/2$ 的解，即 $m=(\log 2)/\lambda$（本书中，log 都是以 e 为底）.