2.2 多维随机变量(随机向量)

2.2.1 离散型随机向量的分布

随机向量的概念在2.1节2.1.1段中已提及过了.一般,设 $X=(X_1,X_2,\cdots,X_n)$为一 n 维向量,其每个分量,即 $X_1,\cdots,X_n$,都是一维随机变量,则称 X 是一个 n 维随机向量或 n 维随机变量.

与随机变量一样,随机向量也有离散型和连续型之分.本段先考虑前者,一个随机向量 $X=(X_1,\cdots,X_n)$,如果其每一个分量 X_i 都是一维离散型随机变量,则称 X 为离散型的.

定义 2.1 以$\{a_{i1},a_{i2},\cdots\}$记 X_i 的全部可能值,$i=1,2,\cdots$,则事件$\{X_1=a_{1_{j_1}},X_2=a_{2_{j_2}}\cdots,X_n=a_{n_{j_n}}\}$的概率

$$p(j_1,j_2,\cdots,j_n)=P(X_1=a_{1j_1},X_2=a_{2j_2},\cdots,X_n=a_{nj_n})$$

$$j_1=1,2,\cdots,j_2=1,2,\cdots,\cdots,j_n=1,2,\cdots \tag{2.1}$$

称为随机向量 $X=(X_1,\cdots,X_n)$的概率函数或概率分布,概率函数应满足条件:

$$p(j_1,j_2,\cdots,j_n)\geqslant 0,\sum_{j_n}\cdots\sum_{j_2}\sum_{j_1}p(j_1,j_2,\cdots,j_n)=1 \tag{2.2}$$

例 2.1 图2.5所示的二维离散型随机向量 $X=(X_1,X_2)$的概率分布为

$$P(X_1=2,X_2=1)=1/3$$
$$P(X_1=2,X_2=2.5)=1/4$$
$$P(X_1=5,X_2=3)=5/12$$

从图上看出,X_1 的可能值为2和5,X_2 的可能值为1,2.5和3.故形式上看,$X=(X_1,X_2)$应有6组可能值,即(2,1),(2,2.5),(2,3),(5,1),(5,2.5),(5,3).X 的概率分布告诉我们,实际上只

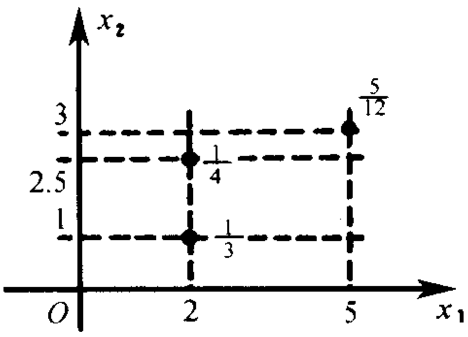

图 2.5

有第 1,2,6 组是真正的可能值,但这并无关系:对一组不可能的值,只要把它的概率定为 0 就行了.这一做法使我们可以把离散型分布统一写成(2.1)的格式,在理论上有其方便之处.自然,在具体例子中,如例 2.1 并无必要硬凑成那个形式,只要指出概率大于 0 的那部分就行了.

例 2.2 多项分布.

多项分布是最重要的离散型多维分布.设 $A_1, A_2, \cdots, A_n$ 是某一试验之下的完备事件群,即事件 $A_1, \cdots, A_n$ 两两互斥,其和为必然事件(每次试验时,事件 $A_1, \cdots, A_n$ 必发生一个且只发生一个).分别以 $p_1, p_2, \cdots, p_n$ 记事件 $A_1, A_2, \cdots, A_n$ 的概率,则 $p_i \geqslant 0, p_1 + \cdots + p_n = 1$.

现在将试验独立地重复 N 次,而以 X_i 记在这 N 次试验中事件 A_i 出现的次数,$i = 1, \cdots, n$,则 $X = (X_1, \cdots, X_n)$ 为一 n 维随机向量.它取值的范围是:$X_1, \cdots, X_n$ 都是非负整数且其和为 N. X 的概率分布就叫做多项分布,有时记为 $M(N; p_1, \cdots, p_n)$.为定出这分布,要计算事件

$$B = \{X_1 = k_1, \cdots, X_i = k_i, \cdots, X_n = k_n\}$$

的概率,只须考虑 k_i 都是非负整数且 $k_1 + \cdots + k_n = N$ 的情况,否则 $P(B) = 0$.为计算 $P(B)$,从 N 次试验的原始结果 $j_1 j_2 \cdots j_N$ 出发,它表示第一次实验事件 A_{j_1} 发生,第二次试验 A_{j_2} 发生,等等.为使事件 B 发生,在 $j_1, j_2, \cdots, j_N$ 中应有 k_1 个 1,k_2 个 2,…等等.这种序列的数目,等于把 N 个相异物体分成 n 堆,各堆依次有 $k_1, k_2, \cdots, k_n$ 件的不同分法.据第一章(2.6)式,不同分法有 $N! / (k_1! \cdots k_n!)$ 个.其次,由于独立性,利用概率乘法定理知,每个适合上述条件的原始结果序列 $j_1 j_2 \cdots j_n$ 出现的概率,应为 $p_1^{k_1} p_2^{k_2} \cdots$

$p_n^{k_n}$. 于是得到

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) = \frac{N!}{k_1! k_2! \cdots k_n!} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \tag{2.3}$$

$$(k \text{ 为非负整数}, k_1 + \cdots + k_n = N)$$

(2.3)就是多项分布,名称的来由是因多项展开式

$$(x_1 + \cdots + x_n)^N = \sum{}^{*} \frac{N!}{k_1! \cdots k_n!} x_1^{k_1} \cdots x_n^{k_n} \tag{2.4}$$

$\sum^{*}$ 表示求和的范围为: k_i 为非负整数, $k_1 + \cdots + k_n = N$. 在 (2.4) 中令 $x_i = p_i$ 并利用 $p_1 + \cdots + p_n = 1$, 得

$$\sum{}^{*} \frac{N!}{k_1! \cdots k_n!} p_1^{k_1} \cdots p_n^{k_n} = 1$$

这说明分布(2.3)适合条件(2.2).

多项分布在实用上颇常见:当一个体按某种属性分成几类时,就会涉及这个分布.例如,一种产品分成一等品(A_1)、二等品(A_2)、三等品(A_3)和不合格品(A_4)四类.若生产该产品的某厂,其一、二、三等品和不合格品的比率分别为 0.15,0.70,0.10 和 0.05,从该厂产品中抽出 N 个.若这 N 个只占其产品的极少一部分,则可以把这 N 个看成一个一个地独立抽出,且在抽取过程中,各等品的概率(即比率)不变.在这种情况下,若分别以 $X_1, \cdots, X_4$ 记这 N 个产品中一、二、三等和不合格品的个数,则 $X = (X_1, \cdots, X_4)$将有多项分布 $M(N; 0.15, 0.70, 0.10, 0.05)$. 又如在医学上,一种疾病的患者可按严重的程度分期等等,都属于这种情况.

如果 $n = 2$,即只有 A_1, A_2 两种可能,这时 A_2 就是 A_1 的对立事件.由于这时有 $X_1 + X_2 = N$, X_1 唯一地决定了 X_2,我们不必同时考虑 X_1 和 X_2,而只须考虑 X_1 就够了,这就回到二项分布的情形.

2.2.2 连续型随机向量的分布

设 $X = (X_1, \cdots, X_n)$是一个 n 维随机向量.其取值可视为 n

维欧氏空间R^n中的一个点.如果X的全部取值能充满R^n中某一区域,则称它是连续型的.

与一维连续型变量一样,描述多维随机向量的概率分布,最方便的是用概率密度函数.为此我们引进一个记号:$X\in A$,读作"X属于A"或"X落在A内",其中A是R^n中的集合.$\{X\in A\}$是一个随机事件,因为作了试验以后,X之值就知道了,因而也就能知道它是否落在A内.

定义 2.2 若$f(x_1,\cdots,x_n)$是定义在R^n上的非负函数,使对R^n中的任何集合A,有

$$P(X\in A)=\int_A\cdots\int f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \qquad (2.5)$$

则称f是X的(概率)密度函数.

如果把A取成全空间R^n,则$\{X\in A\}$为必然事件,其概率为1.因此应有

$$\int_{-\infty}^{\infty}\cdots\int f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1 \qquad (2.6)$$

这是一个概率密度函数必须满足的条件.

例 2.3 考虑二维随机向量$X=(X_1,X_2)$,其概率密度函数为

$$f(x_1,x_2)=\begin{cases}1/[(b-a)(d-c)], & \text{当 } a\leqslant x_1\leqslant b,c\leqslant x_2\leqslant d\\ 0, & \text{其他处}\end{cases}$$

则f非负且条件(2.6)满足.从f的形状看出,它在图2.6中那个矩形之外为0,说明(X_1,X_2)只能取该矩形内的点为值.在这矩形内,密度各处一样,因而全部概率均匀地分布在这矩形内.从公式(2.5)看出:若集A在矩形内,则"X落在A内"的概率$P(X\in A)$,与A的面积成正比而与其位置及形状无关,这是均匀性的另一种说法.以此之故,人们把本例中X的分布称为上述矩形上的均匀分布.

例 2.4 向一个无限平面靶射击,设命中点$X=(X_1,X_2)$有

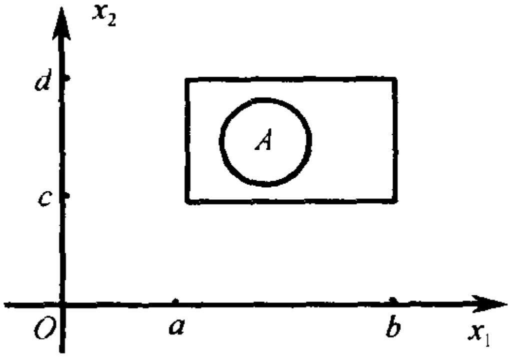

图 2.6

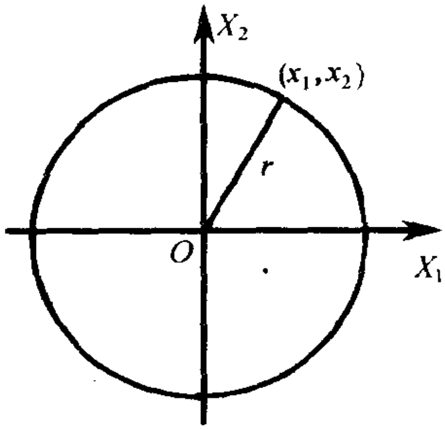

图 2.7

概率密度

$$f(x_1,x_2) = \pi^{-1}(1 + x_1^2 + x_2^2)^{-2}$$

从这个函数看出:命中点的密度只与该点与靶心的距离 r 有关.这可以解释为:在图 2.7 中以靶心 O 为中心的圆周上各点有同等被命中的机会.另外,$x_1^2 + x_2^2$ 愈小则 f 愈大,说明与靶心接近之点,较之远离靶心之点,有更大的命中机会.

为验证(2.6)式只须转到极坐标,得

$$\iint_{-\infty}^{\infty} f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$= \int_0^{2\pi}\mathrm{d}\theta\int_0^{\infty}\pi^{-1}(1 + r^2)^{-2}r\mathrm{d}r$$

$$= 2\pi\cdot\pi^{-1}\int_0^{\infty}(1 + t)^{-2}\mathrm{d}t/2 = 1$$

而"命中点与靶心之距离不超过 r_0"这个事件 A 的概率为

$$\iint_{x_1^2+x_2^2\leqslant r_0^2} f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$= \int_0^{2\pi}\mathrm{d}\theta\int_0^{r_0}\pi^{-1}(1 + r^2)^{-2}r\mathrm{d}r = r_0^2/(1 + r_0^2)$$

例 2.5 二维正态分布.

最重要的多维连续型分布是多维正态分布. 对二维的情况，其概率密度函数有形式

$$f(x_1,x_2)=(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-a)^2}{\sigma_1^2}\right.\right.$$
$$\left.\left.-\frac{2\rho(x_1-a)(x_2-b)}{\sigma_1\sigma_2}+\frac{(x_2-b)^2}{\sigma_2^2}\right)\right)\qquad(2.7)$$

这里为书写方便，引进了一个记号 exp. 其意义是：$\exp(c)=e^c$. f 中包含了五个常数 $a,b,\sigma_1^2,\sigma_2^2$ 和 ρ，它们是这个分布的参数，其可取值的范围为：

$-\infty<a<\infty,\ -\infty<b<\infty,\sigma_1>0,\sigma_2>0,\ -1<\rho<1.$

常把这个分布记为 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$. 这函数（在三维空间中）的图形，好像一个椭圆切面的钟倒扣在 Ox_1x_2 平面上，其中心在 (a,b) 点.

为了证明(2.7)式确实是一个密度函数，还须证明(2.6)式成立. 为此，作变数代换

$$u=(1-\rho^2)^{-1/2}\frac{x_1-a}{\sigma_1},v=(1-\rho^2)^{-1/2}\frac{x_2-b}{\sigma_2}$$

得

$$\iint_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$
$$=\frac{1}{2\pi}\sqrt{1-\rho^2}\iint_{-\infty}^{\infty}\exp\cdot\left[-\frac{1}{2}(u^2-2\rho uv+v^2)\right]\mathrm{d}u\mathrm{d}v$$

再作变数代换 $t_1=u-\rho v,t_2=\sqrt{1-\rho^2}v$. 注意到 $u^2-2\rho uv+v^2=(u-\rho v)^2+(1-\rho^2)v^2=t_1^2+t_2^2$，且变换的贾可比行列式为

$$\begin{vmatrix}\partial t_1/\partial u & \partial t_1/\partial v\\ \partial t_2/\partial u & \partial t_2/\partial v\end{vmatrix}=\begin{vmatrix}1 & -\rho\\ 0 & \sqrt{1-\rho^2}\end{vmatrix}=\sqrt{1-\rho^2}$$

得

$$\iint_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$= \frac{1}{2\pi}\sqrt{1-\rho^2}(\sqrt{1-\rho^2})^{-1}\iint_{-\infty}^{\infty}\exp\cdot\left[-\frac{1}{2}(t_1^2+t_2^2)\right]dt_1dt_2$$

$$= (2\pi)^{-1}\int_0^{\infty}e^{-t_1^2/2}dt_1\int_0^{\infty}e^{-t_2^2/2}dt_2$$

$$= (2\pi)^{-1}\sqrt{2\pi}\sqrt{2\pi} = 1$$

这里用到(1.15).

类似地可定义 n 维正态分布的概率密度函数,这里不细讲了.

在结束这一段之前,让我们指出几点有关事项:

1. 不论在一维或多维,在定义连续型随机变量时,实质之点都在于它有概率密度函数存在,即存在有函数 f,满足(1.13)或(2.5)式.在概率论理论上,把这一点直接取为连续型随机变量的定义:它就是有密度函数的随机变量.至于它可以在一个区间或区域上连续取值倒不是本质的,甚至也是不确切的.

2. 与离散型随机向量的定义不同,连续型随机向量不能简单地定义为"其各分量都是一维连续型随机变量的那种随机向量".举一个例子:设 $X_1 \sim R(0,1)$, $X_2 = X_1$,则随机向量 (X_1,X_2) 的两个分量 X_1,X_2 都是连续型的.但 (X_1,X_2) 却只能在图 2.8 中所示的单位正方形的对角线(图中的虚线)上取值.因而不可能存在一个函数 $f(x_1,x_2)$ 满足(2.6)式(二元函数在平面上任一线段上的积分都是 0),即 (X_1,X_2) 的概率密度函数不存在.

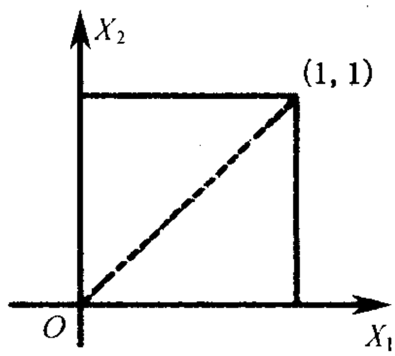

图 2.8

3. 与一维情况一样,也可以用概率分布函数去描述多维随机向量的概率分布,其定义为

$$F(x_1,x_2,\cdots,x_n) = P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n)$$

然而,在多维情况下,分布函数极少应用.

2.2.3 边缘分布

设 $X=(X_1,\cdots,X_n)$为一 n 维随机向量. X 有一定的分布F，这是一个 n 维分布. 因为 X 的每个分量X_i 都是一维随机变量，故它们都有各自的分布 F_i，$i=1,\cdots,n$，这些都是一维分布，称为随机向量 X 或其分布F 的“边缘分布”. 以下我们要指出：边缘分布完全由原分布 F 确定，且从这个事实的讲解中也就悟出“边缘”一词的含义.

表 2.1

X_1 \ X_2	-1	0	5	行合计
1	0.17	0.05	0.21	0.43
3	0.04	0.28	0.25	0.57
列合计	0.21	0.33	0.46	1.00

例 2.6 表 2.1 以列表的形式，显示了一个二维随机向量 $X=(X_1,X_2)$的概率分布. 比如

$$P(X_1=1,X_2=5)=0.21$$

等等. 现在如想求 X_1 的分布，则先注意到，X_1 只有两个可能值，即 1 和 3. 而$\{X_1=1\}$这个事件可以分解为三个互斥事件

$$\{X_1=1,X_2=-1\},\{X_1=1,X_2=0\},\{X_1=1,X_2=5\}$$

之和，故其概率应为上述三事件概率之和，即

$$P(X_1=1)=0.17+0.05+0.21=0.43$$

类似地得 $P(X_1=3)=0.04+0.28+0.25=0.57$. 用同样的方法确定 X_2 的概率分布为

$$P(X_2=-1)=0.21,P(X_2=0)=0.33,P(X_2=5)=0.46$$

注意这两个分布正好是表的中央部分的行和与列和. 它们都处在表的“边缘”位置上. 由此得出边缘分布这个名词. 也有称为边际分布的.

从这个例子就不难悟出，在一般的离散型情况下，怎样去求边缘分布. 回到定义 2.1 的记号，以 X_1 为例，它的全部可能值为 $a_{11},a_{12},a_{13},\cdots$. 例如，我们要求 $P(X_1=a_{1k})$. 它等于把(2.1)式那样的概率全加起来，但局限于 $j_1=k$（这相当于在上述简例 2.6 中求行和或列和）. 得

$$P(X_1=a_{1k})=\sum_{j_2,\cdots,j_n}p(k,j_2,\cdots,j_n),k=1,2,\cdots \quad (2.8)$$

例 2.7 设 $X=(X_1,\cdots,X_n)$ 服从多项分布 $M(N;p_1,\cdots,p_n)$，要求其边缘分布. 例如，考虑 X_1，我们把事件 A_1 作为一方，$A_2+\cdots+A_n$ 作为一方（它就是 $\overline{A}_1$），见例 2.2 的说明，那么，X_1 就是在 N 次独立试验中，事件 A_1 发生的次数，而在每次试验中 A_1 发生的概率保持为 p_1，经过这一分析，不待计算就可以明了：X_1 的分布就是二项分布 $B(N,p_1)$. 应用公式(2.8)也可以得出这个结果：按(2.8)，注意到多项分布的形式(2.3)，有

$$P(X_1=k)=\sum_{k_2,\cdots,k_n}{}' \frac{N!}{k_2!\cdots k_n!}p_2^{k_2}\cdots p_n^{k_n}\cdot p_1^k/k!$$

这里，$\sum\limits_{k_2,\cdots,k_n}{}'$ 表示求和的范围为：$k_2,\cdots,k_n$ 都是非负整数，其和为 $N-k$. 令

$$p_2'=p_2/(1-p_1),\cdots,p_n'=p_n/(1-p_1)$$

则 $p_2'+\cdots+p_n'=(p_2+\cdots+p_n)/(1-p_1)=(1-p_1)/(1-p_1)=1$，且可把上式改写为

$$P(X_1=k)=\sum_{k_2,\cdots,k_n}{}' \frac{(N-k)!}{k_2!\cdots k_n!}p_2'^{k_2}\cdots p_n'^{k_n}$$

$$\cdot\frac{N!}{k!(N-k)!}p_1^k(1-p_1)^{N-k}$$

按多项展开式(2.4)，上式右边第一因子为

$$(p_2'+\cdots+p_n')^{N-k}=1^{N-k}=1$$

于是得到

$$P(X_1=k)=\frac{N!}{k!(N-k)!}p_1^k(1-p_1)^{N-k}$$

$$= b(k;N,p_1), k = 0,1,\cdots,N$$

正是二项分布 $B(N,p_1)$.

现在来考虑连续型随机向量的边缘分布. 为书写简单计,先考虑二维的情况,设 $X=(X_1,X_2)$有概率密度函数 $f(x_1,x_2)$. 我们来证明:这时 X_1 和 X_2 都具有概率密度函数.

为证明这一点,考虑 X_1 的分布函数 $F_1(x_1)=P(X_1\leqslant x_1)$. 它可以写为 $P(X_1\leqslant x_1,X_2<\infty)$. 注意到公式(2.5),得

$$F_1(x_1) = P(X_1 \leqslant x_1) = \int_{-\infty}^{x_1} \mathrm{d}t_1 \int_{-\infty}^{\infty} f(t_1,t_2)\mathrm{d}t_2$$

$\int_{-\infty}^{\infty} f(t_1,t_2)\mathrm{d}t_2$ 是 t_1 的函数,记之为 $f_1(t_1)$. 于是上式可写为

$$F_1(x_1) = \int_{-\infty}^{x_1} f_1(t_1)\mathrm{d}t_1$$

两边对 x_1 求导数,得到 X_1 的概率密度函数为

$$\mathrm{d}F_1(x_1)/\mathrm{d}x_1 = f_1(x_1) = \int_{-\infty}^{\infty} f(x_1,x_2)\mathrm{d}x_2 \qquad (2.9)$$

这不仅证明了 X_1 的密度函数的存在,而且还推出了其公式. 同理求出 X_2 的密度函数为

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1,x_2)\mathrm{d}x_1 \qquad (2.10)$$

这个结果很容易推广到 n 维的情形:设 $X=(X_1,\cdots,X_n)$有概率密度函数 $f(x_1,\cdots,x_n)$. 为求某分量 X_i 的概率密度函数,只须把 $f(x_1,\cdots,x_n)$中的 x_i 固定,然后对 $x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n$ 在 $-\infty$ 到 ∞ 之间作定积分. 例如,X_1 的密度函数为

$$f_1(x_1) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,x_2,\cdots,x_n)\mathrm{d}x_2\cdots\mathrm{d}x_n \quad (2.11)$$

例 2.8 再考虑例 2.3. 用公式(2.9),(2.10)很容易确定,X_1,X_2 的边缘分布分别是均匀分布 $R(a,b)$和 $R(c,d)$. 计算很容易,留给读者.

例 2.9 考虑例 2.4. 按(2.9),X_1 的边缘密度函数为

$$f_1(x_1) = \pi^{-1}\int_{-\infty}^{\infty}(1 + x_1^2 + x_2^2)^{-2}\mathrm{d}x_2$$

作变数代换 $t = x_2/\sqrt{1 + x_1^2}$,得

$$f_1(x_1) = \pi^{-1}(1 + x_1^2)^{-3/2}\int_{-\infty}^{\infty}(1 + t^2)^{-2}\mathrm{d}t = \frac{1}{2}(1 + x_1^2)^{-3/2}$$

积分 $\int_{-\infty}^{\infty}(1 + t^2)^{-2}\mathrm{d}t$ 通过变数代换 $t = \tan\theta$ 很易算出.

例 2.10 二维正态分布 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$的边缘分布密度.若$(X_1,X_2)$有二维正态分布 $N(a,b,\sigma_2^1,\sigma_2^2,\rho)$,我们来证明:$X_1$,$X_2$ 的边缘分布分别是一维正态分布 $N(a,\sigma_1^2)$和 $N(b,\sigma_2^2)$.为证此,要计算 $\int_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_2$,其中 f 由(2.7)式定义.注意到

$$\frac{(x_1 - a)^2}{\sigma_1^2} - \frac{2\rho(x_1 - a)(x_2 - b)}{\sigma_1\sigma_2} + \frac{(x_2 - b)^2}{\sigma_2^2}$$

$$= (1 - p^2)\frac{(x_1 - a)^2}{\sigma_1^2} + \left(\rho\frac{x_1 - a}{\sigma_1} - \frac{x_2 - b}{\sigma_2}\right)^2$$

得到

$$f_1(x_1) = \int_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_2 = (2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2})^{-1}$$

$$\cdot\exp\left(-\frac{(x_1 - a)^2}{2\sigma_1^2}\right)C$$

其中

$$C = \int_{-\infty}^{\infty}\exp\left[-\frac{1}{2(1 - \rho^2)}\left(\rho\frac{x_1 - a}{\sigma_1} - \frac{x_2 - b}{\sigma_2}\right)^2\right]\mathrm{d}x_2$$

作变数代换(注意 x_1 为常数,非积分变量)

$$t = \left(\frac{x_2 - b}{\sigma_2} - \rho\frac{x_1 - a}{\sigma_1}\right)\Big/\sqrt{1 - \rho^2}$$

得

$$C = \int_{-\infty}^{\infty}\exp(-t^2/2)\mathrm{d}t \cdot \sigma_2\sqrt{1 - \rho^2} = \sqrt{2\pi}\sigma_2\sqrt{1 - \rho^2}$$

以此代入前式,即得

$$f_1(x_1) = (\sqrt{2\pi}\sigma_1)^{-1}\exp\left(-\frac{(x_1-a)^2}{2\sigma_1^2}\right) \qquad (2.12)$$

这正是 $N(a,\sigma_1^2)$的概率密度函数.

从这个例子看出一个有趣的事实:虽则一个随机向量 $X=(X_1,\cdots,X_n)$的分布 F 足以决定其任一分量 X_i 的(边缘)分布 F_i,但反过来不对:即使知道了所有 X_i 的边缘分布 $F_i,i=1,\cdots,n$,也不足以决定 X 的分布 F. 例如,考虑两个二维正态分布

$$N(0,0,1,1,1/3) \text{ 和 } N(0,0,1,1,2/3)$$

它们的任一边缘分布都是标准正态分布 $N(0,1)$,但这两个二维分布是不同的分布,因为 ρ 的数值不相同. 这个现象的解释是:边缘分布只分别考虑了单个变量 X_i 的情况,而未涉及它们之间的关系,而这个信息却是包含在$(X_1,\cdots,X_n)$的分布之内的. 如就本例来说,在下一章(见第三章 3.3 节)将指出:ρ 这个参数正好刻画了两分量 X_1 和 X_2 之间的关系.

在结束这一节之前,我们再强调指出:"边缘"分布就是通常的分布,并无任何特殊的含义. 如果说有什么意思的话,它不过是强调了:这个分布是由于 X_i 作为随机向量$(X_1,\cdots,X_n)$之一分量,从后者的分布中派生出的分布而已,别无其他. 至于"边缘"一词的由来,已在例 2.6 中解释过了.

与此相应,为了强调$(X_1,\cdots,X_n)$的分布是把 $X_1,\cdots,X_n$ 作为一个有联系的整体来考虑,有时把它称为 $X_1,\cdots,X_n$ 的"联合分布".

另外,边缘分布也可以不只是单个的. 例如 $X=(X_1,X_2,X_3)$它的分布也决定了其任一部分,例如(X_1,X_3)的二维分布,这也称为边缘分布. 有关公式也不难导出,此处不细讲了.