2.3 条件概率分布与随机变量的独立性

2.3.1 条件概率分布的概念

一个随机变量或向量 X 的条件概率分布，就是在某种给定的

条件之下，X 的概率分布．一如以前我们在讨论条件概率时所指出的，任何事件的概率都是“有条件的”，即与这事件联系着的试验的条件，如骰子是均匀的立方体且抛掷的高度是足够大之类．以此，任何随机变量或向量的分布，也无不是在一定条件下．但此处所谈的条件分布，是在试验中所规定的“基本”条件之外再附加的条件．它一般采取如下的形式：设有两个随机变量或向量 X, Y，在给定了 Y 取某个或某些值的条件下，去求 X 的条件分布．

例如，考虑一大群人，从其中随机抽取一个，分别以 X_1 和 X_2 记其体重和身高，则 X_1, X_2 都是随机变量，它们都有一定的概率分布．现在如限制 $1.7 \leqslant X_2 \leqslant 1.8$(米)，在这个条件下去求 X_1 的条件分布，这就意味着要从这一大群人中把其身高在 1.7 米和 1.8 米的那些人都挑出来，然后在挑出的人群中求其体重的分布．容易想像，这个分布与不设这个条件的分布(无条件分布)会很不一样．例如，在条件分布中体重取大值的概率会显著增加．

从这个例子也看出条件分布这个概念的重要性．在本例中，弄清了 X_1 的条件分布随 X_2 之值而变化的情况，就能了解身高对体重的影响在数量上的表述．由于在许多问题中有关的变量往往是彼此有影响的，这使条件分布成为研究变量之间的相依关系的一个有力工具．这一点以后在第六章中还要作更深入的发挥．

2.3.2 离散型随机变量的条件概率分布

这个情况比较简单，实际上无非是第一章讲过的条件概率概念在另一种形式下的重复．设 (X_1, X_2) 为一个二维离散型随机向量，X_1 的全部可能值为 $a_1, a_2, \cdots$；X_2 的全部可能值为 $b_1, b_2, \cdots$，而 (X_1, X_2) 的联合概率分布为

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

现考虑 X_1 在给定 $X_2 = b_j$ 的条件下的条件分布，那无非是要找条件概率 $P(X_1 = a \mid X_2 = b_j)$．依条件概率的定义，有

$$\begin{aligned} P(X_1 = a_i \mid X_2 = b_j) &= P(X_1 = a_i, X_2 = b_j) / P(X_2 = b_j) \\ &= p_{ij} / P(X_2 = b_j) \end{aligned}$$

再据公式(2.8)($n=2$ 的情况),有 $P(X_2=b_j)=\sum\limits_k p_{kj}$. 于是

$$P(X_1=a_i \mid X_2=b_j)=p_{ij}\Big/\sum_k p_{kj},\ i=1,2,\cdots \qquad (3.1)$$

类似地有

$$P(X_2=b_j \mid X_1=a_i)=p_{ij}\Big/\sum_k p_{ik},\ j=1,2,\cdots \qquad (3.2)$$

例 3.1 再考虑例 2.6. 据公式(3.1)和(3.2),不难算出在给定 X_2 时 X_1 的条件分布,与给定 X_1 时 X_2 的条件分布. 例如,在给定 $X_2=0$ 时有

$$P(X_1=1 \mid X_2=0)=0.05/0.33=5/33$$

$$P(X_1=3 \mid X_2=0)=0.28/0.33=28/33$$

例 3.2 设$(X_1,X_2,\cdots,X_n)$服从多项分布 $M(N;p_1,\cdots,p_n)$. 试求在给定 $X_2=k_2$ 的条件下,X_1 的条件分布.

先计算概率 $P(X_1=k_2,X_2=k_2)$. 这里假定 k_1,k_2 都是非负整数,且 $k_1\leqslant N-k_2$. 按(2.3)式,有

$$P(X_1=k_1,X_2=k_2)=\sum_{k_3,\cdots,k_n}{}' \frac{N!}{k_1!k_2!k_3!\cdots k_n}p_1^{k_1}p_2^{k_2}p_3^{k_3}\cdots p_n^{k_n}$$

这里 $\sum\limits_{k_3,\cdots,k_n}{}'$ 表示求和的范围为$k_3,\cdots,k_n$ 都是非负整数,且 $k_3+\cdots+k_n=N-(k_1+k_2)$. 令 $p_i'=p_i/(1-p_1-p_2)$, $i\geqslant 3$,有

$$P(X_1=k_1,X_2=k_2)=\frac{N!}{k_1!k_2!(N-k_1-k_2)!} \cdot p_1^{k_1}p_2^{k_2}(1-p_1-p_2)^{N-k_1-k_2}C$$

其中

$$C=\sum_{k_3,\cdots,k_n}{}' \frac{(N-k_1-k_2)!}{k_3!\cdots k_n!}p_3'^{k_3}\cdots p_n'^{k_n}$$

由于 $p_3'+\cdots+p_n'=1$,考虑到上式求和的范围及多项展开式(2.4),即知 $C=1$,因此

$$P(X_1=k_1,X_2=k_2)$$

$$=\frac{N!}{k_1!k_2!(N-k_1-k_2)!}\cdot p_1^{k_1}p_2^{k_2}(1-p_1-p_2)^{N-k_1-k_2}$$

再根据例 2.7,X_2 的分布就是二项分布 $B(N,p_2)$.因此

$$
\begin{aligned}
& P(X_1 = k_1 \mid X_2 = k_2) \\
&= P(X_1 = k_1, X_2 = k_2)/P(X_2 = k_2) \\
&= \frac{N!}{k_1!k_2!(N-k_1-k_2)!} p_1^{k_1} p_2^{k_2} (1-p_1-p_2)^{N-k_1-k_2} \\
&\quad \cdot \Big/ \frac{N!}{k_2!(N-K_2)!} p_2^{k_2} (1-p_2)^{N-k_2} \\
&= \frac{(N-k_2)!}{k_1!(N-k_1-k_2)} \left(\frac{p_1}{1-p_2}\right)^{k_1} \left(1-\frac{p_1}{1-p_2}\right)^{N-k_1-k_2} \\
&= b(k_1; N-k_2, p_1/(1-p_2)), k = 0,1,\cdots,N-k_2
\end{aligned}
$$

由此可知:在给定 $X_2 = k_2$ 的条件下,X_1 的条件分布就是分布 $B(N-k_2, p_1/(1-p_2))$.

2.3.3 连续型随机变量的条件分布

设二维随机向量 $X=(X_1,X_2)$有概率密度函数 $f(x_1,x_2)$.我们先来考虑在限定 $a \leqslant x_2 \leqslant b$ 的条件下,X_1 的条件分布.有

$$
\begin{aligned}
& P(X_1 \leqslant x_1 \mid a \leqslant X_2 \leqslant b) \\
&= P(X_1 \leqslant x_1, a \leqslant X_2 \leqslant b)/P(a \leqslant X_2 \leqslant b)
\end{aligned}
$$

X_2 的边缘分布的密度函数 f_2 由(2.10)给出.有

$$
\begin{aligned}
& P(X_1 \leqslant x_1, a \leqslant X_2 \leqslant b) \\
&= \int_{-\infty}^{x_1} \mathrm{d}t_1 \int_a^b f(t_1,t_2)\mathrm{d}t_2
\end{aligned}
$$

$$
P(a \leqslant X_2 \leqslant b) = \int_a^b f_2(t_2)\mathrm{d}t_2
$$

由此得到

$$
\begin{aligned}
& P(X_1 \leqslant x_1 \mid a \leqslant X_2 \leqslant b) \\
&= \int_{-\infty}^{x_1} \mathrm{d}t_1 \int_a^b f(t_1,t_2)\mathrm{d}t_2 \Big/ \int_a^b f_2(t_2)\mathrm{d}t_2
\end{aligned}
$$

这是 X_1 的条件分布函数.对 x_1 求导数,得到条件密度函数为

$$
f_1(x_1 \mid a \leqslant X_2 \leqslant b) = \int_a^b f(x_1,t_2)\mathrm{d}t_2 \Big/ \int_a^b f_2(t_2)\mathrm{d}t_2 \qquad (3.3)
$$

更有兴趣的是 $a=b$ 的情况,即在 X_2 给定等于一个值之下,X_1 的条件密度函数.这不能通过直接在(3.3)中令 $a=b$ 得出,但可用极限步骤:

$$\begin{aligned}f_1(x_1|x_2) &= f_1(x_1|X_2=x_2)\\ &=\lim_{h\to 0}f_1(x_1|x_2\leqslant X_2\leqslant x_2+h)\\ &=\lim_{h\to 0}\frac{1}{h}\int_{t_2}^{x_2+h}f(x_1,t_2)\mathrm{d}t_2\Big/\lim_{h\to 0}\frac{1}{h}\int_{x_2}^{x_2+h}f_2(t_2)\mathrm{d}t_2\\ &=f(x_1,x_2)/f_2(x_2)\end{aligned}\tag{3.4}$$

这就是在给定 $X_2=x_2$ 的条件下,X_1 的条件密度函数.此式当然只有在 $f_2(x_2)>0$ 时才有意义.在上述取极限的过程中,还得假定函数 f_2 在 x_2 点连续,及 $f(x_1,t_2)$ 作为 t_2 的函数,在 $t_2=x_2$ 处连续.然而,用高等概率论的知识,可以在没有这种连续的假定下证明(3.4).

(3.4)式可改写为

$$f(x_1,x_2)=f_2(x_2)f_1(x_1|x_2)\tag{3.5}$$

就是说:两个随机变量 X_1 和 X_2 联合概率密度,等于其中之一的概率密度乘以在给定这一个之下另一个的条件概率密度.这个公式相应于条件概率的公式 $P(AB)=P(B)P(A|B)$.除(3.5)外,当然也有

$$f(x_1,x_2)=f_1(x_1)f_2(x_2|x_1)\tag{3.6}$$

其中 f_1 为 x_1 的边缘密度,而

$$f_2(x_2|x_1)=f(x_1,x_2)/f_1(x_1)\tag{3.7}$$

则是在给定 $X_1=x_1$ 的条件下,X_2 的条件密度.这些公式反映的实质可推广到任意多个变量的场合:设有 n 维随机向量$(X_1,\cdots,X_n)$,其概率密度函数为 $f(x_1,\cdots,x_n)$.则

$$f(x_1,\cdots,x_n)=g(x_1,\cdots,x_k)h(x_{k+1},\cdots,x_n|x_1,\cdots,x_k)\tag{3.8}$$

其中 g 是$(X_1,\cdots,X_k)$的概率密度,而 h 则是在给定 $X_1=x_1,\cdots,X_k=x_k$ 的条件下,$X_{k+1},\cdots,X_n$ 的条件概率密度.(3.8)可视为

(3.6)的直接推广，又可视为 $h(x_{k+1},\cdots,x_n|x_1,\cdots,x_k)$ 的定义.

例 3.3 设 (X_1,X_2) 服从二维正态分布 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$. 求在给定 $X_1=x_1$ 的条件下，X_2 的条件密度函数 $f_2(x_2|x_1)$.

利用公式(3.7)，(2.7)和(2.12)，经过简单的计算，得出

$$f_2(x_2|x_1)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left[-\frac{(x_2-(b+\rho\sigma_2\sigma_1^{-1}(x_1-a)))^2}{2(1-\rho^2)\sigma_2^2}\right] \tag{3.9}$$

这正是正态分布 $N(b+\rho\sigma_2\sigma_1^{-1}(x_1-a),\sigma_2^2(1-\rho^2))$ 的概率密度函数(注意在(3.9)式中，x_1 当常数看). 因此，正态变量的条件分布仍为正态，这是正态分布的一个重要性质.

如我们在图 2.2b 中所显示的，正态分布 $N(\mu,\sigma^2)$ 关于 μ 点对称，μ 就是分布的中心位置，对正态分布(3.9)，这个中心位置在

$$m(x_1)=b+\rho\sigma_2\sigma_1^{-1}(x_1-a) \tag{3.10}$$

处，由这里可以看出 ρ 刻画了 X_1,X_2 之间的相依关系. 其解释如下：若 $\rho>0$，则随着 x_1 的增加，X_2(在 $X_1=x_1$ 之下)的条件分布的中心点 $m(x_1)$ 随 x_1 的增加而增加. 可以看出：这意味着当 x_1 增加时，X_2 取大值的可能性增加，即 X_2 有随着 X_1 的增长而增长的倾向(如体重与身高的关系那样). 反之，若 $\rho<0$，则 X_2 有随着 X_1 增长而下降的倾向. 由于这个原因，通常把 $\rho>0$ 的情况称为"正相关"，而 $\rho<0$ 的情况称为"负相关". 这一点在下一章中还要谈到.

把(3.5)两边对 x_2 积分，得

$$f_1(x_1)=\int_{-\infty}^{\infty}f(x_1,x_2)\mathrm{d}x_2=\int_{-\infty}^{\infty}f_1(x_1|x_2)f_2(x_2)\mathrm{d}x_2 \tag{3.11}$$

这个公式可解释为：X_1 的无条件密度 $f_1(x_1)$，是其条件密度 $f_1(x_1|x_2)$ 对"条件"x_2 的平均. 更确切地说，是按其概率大小为权的加权平均，因为，$f_2(x_2)\mathrm{d}x_2$ 正是 X_2 在 x_2 附近 $\mathrm{d}x_2$ 这么长的区

间内的概率. 从直观上看这应当是很自然的. 比如说,(X_1, X_2)代表一大群人中随机抽出的一个人的体重和身长,X_1(体重)有其(无条件)分布,这可以看作为各种不同的身高综合之后所呈现的分布,而不同于固定身长 $X_2=x_2$ 时的条件分布. 但把各种身长时体重的条件分布进行平均,也就实现了上述综合,即得到无条件分布. 公式(3.11)正好从数学上反映了这种综合(或平均)的过程.

还要注意:公式(3.11)也可以看作是全概率公式(第一章)(3.17)在概率密度这种情况下的表现形式. 在这里,$f_1(x_1)$相当于全概率公式中的 $P(A)$,$f_1(x_1|x_2)$相当于条件概率 $P(A|B_i)$,而(3.11)中的积分,正好相当于(3.17)式中的以 $P(B_i)$为权的加权和.

由此可见,在学习概率论时,不能光是形式地看待一些分析公式,更重要的是要分析其概率意义及直观意义,这样才能加深理解. 上述对公式(3.11)的分析是一个例子. 再如,在例 3.3 中我们用形式推导很容易得出了条件密度(3.9)式. 只看这形式推导,你可能地觉得这里没有什么特别值得注意的地方. 但经过分析(3.10)式中 ρ 的作用,再辅之以体重身高这个实例,我们就领悟到了 ρ 作为刻画二者的相依性的作用,理解就深一层了. 在下一章中我们还要进一步讨论(3.9)所反映出的其他概率含义.

2.3.4 随机变量的独立性

先考虑两个变量 X_1, X_2 的情况,并设(X_1, X_2)为连续型. 如前,分别以 $f(x_1,x_2)$,$f_1(x_1)$,$f_2(x_2)$,$f_1(x_1|x_2)$,$f_2(x_2|x_1)$记联合、边缘与条件概率密度.

一般,$f_1(x_1|x_2)$是随 x_2 的变化而变化的,这反映了 X_1 与 X_2 在概率上有相依关系的事实,即 X_1 的(条件)分布如何,取决于另一变量之值.

如果 $f_1(x_1|x_2)$不依赖于 x_2,因而只是 x_1 的函数,暂记为 $g(x_1)$,则表示 X_1 的分布情况与 X_2 取什么值完全无关,这时就称 X_1, X_2 这两个随机变量(在概率论意义上)独立. 这概念与事件

独立的概念完全相似.

把 $f_1(x_1|x_2)=g(x_1)$代入(3.11),得

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} g(x_1) f_2(x_2) \mathrm{d}x_2 \\ &= g(x_1) \int_{-\infty}^{\infty} f_2(x_2) \mathrm{d}x_2 \\ &= g(x_1) \end{aligned}$$

因此,X_1 的无条件密度 $f_1(x_1)$,就等于其条件密度 $f_1(x_1|x_2)$,这也可取为独立性的定义.

再次,把 $f_1(x_1)=f_1(x_1|x_2)$代入(3.5),得

$$f(x_1,x_2) = f_1(x_1)f_2(x_2) \tag{3.12}$$

即(X_1,X_2)的联合密度,等于其各分量的密度之积.这也可取作为 X_1,X_2 独立的定义(此式相应于第一章(3.7)式),比之上述定义,它有其优越性:一是其形式关于两个变量对称,二是它总有意义,而在用条件密度去定义时,可能碰到条件密度在个别点无法定义(分母为 0)的情况.

这个形式的另一个好处是它可以直接推广到任意多个变量的情形.我们就把它取为一般情况下的正式定义:

定义 3.1 设 n 维随机向量$(X_1,\cdots,X_n)$的联合密度函数为 $f(x_1,\cdots,x_n)$,而 X_i 的(边缘)密度函数为 $f_i(x_i)$, $i=1,\cdots,n$.如果

$$f(x_1,\cdots,x_n) = f_1(x_1)\cdots f_n(x_n) \tag{3.13}$$

就称随机变量 $X_1,\cdots,X_n$ 相互独立或简称独立.

变量独立性的概念还可以从另外的角度去考察.按前面的分析,它含有这种意思:如果 $X_1,\cdots,X_n$ 独立,则各变量取值的概率如何,毫不受其他变量之影响,因此,若考察 n 个事件

$$A_1 = \{a_1 \leqslant X_1 \leqslant b_1\},\cdots,A_n = \{a_n \leqslant X_n \leqslant b_n\} \tag{3.14}$$

则因各事件只涉及一个变量,它们应当是相互独立的事件,我们可以把这个要求取为变量 $X_1,\cdots,X_n$ 独立的定义.下面的定理证明,

这与定义 3.1 是等价的,即同一件事的两种不同的说法.

定理 3.1 如果连续变量 $X_1,\cdots,X_n$ 独立时,则对任何 $a_i<b_i,i=1,\cdots,n$,由(3.14)定义的 n 个事件 $A_1,\cdots,A_n$ 也独立.

反之,若对任何 $a_i<b_i,i=1,\cdots,n$,事件 $A_1,\cdots,A_n$ 独立,则变量 $X_1,\cdots,X_n$ 也独立.

证 先设 $X_1,\cdots,X_n$ 独立,因而(3.13)成立.为证事件 $A_1,\cdots,A_n$ 独立,按第一章定义 3.3,必须对任何 $i_1,\cdots,i_m(1\leqslant i_1<i_2<\cdots<i_m\leqslant n)$ 去证明

$$P(A_{i_1}A_{i_2}\cdots A_{i_m}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_m})$$

为书写简单计,我们对 $i_1=1,i_2=2,\cdots,i_m=m$ 来证此式,这不影响普遍性.按联合分布密度的定义(2.5)式,有

$$\begin{aligned}
&P(A_1A_2\cdots A_m)\\
&\quad= P(a_1\leqslant X_1\leqslant b_1,\cdots,a_m\leqslant X_m\leqslant b_m)\\
&\quad= P(a_1\leqslant X_1\leqslant b_1,\cdots,a_m\leqslant X_m\leqslant b_m,\\
&\qquad -\infty<X_{m+1}<\infty,\cdots,-\infty<X_n<\infty)\\
&\quad= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{a_m}^{b_m}\cdots\int_{a_1}^{b_1}f(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\\
&\quad= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{a_m}^{b_m}\cdots\int_{a_1}^{b_1}f_1(x_1)\cdots f_n(x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n\\
&\quad= \int_{-\infty}^{\infty}f_n(x_n)\mathrm{d}x_n\cdots\int_{-\infty}^{\infty}f_{m+1}(x)\mathrm{d}x_{m+1}\int_{a_1}^{b_1}f_1(x_1)\mathrm{d}x_1\cdots\\
&\qquad\times\int_{a_m}^{b_m}f_m(x_m)\mathrm{d}x_m\\
&\quad= \int_{a_1}^{b_1}f_1(x_1)\mathrm{d}x_1\cdots\int_{a_m}^{b_m}f_m(x_m)\mathrm{d}x_m\\
&\quad= P(a_1\leqslant X_1\leqslant b_1)\cdots P(a_m\leqslant X_m\leqslant b_m)
\end{aligned}$$

这证明了所要的结果.

另一方面,若对任何 $a_i<b_i,i=1,\cdots,n$,(3.14)中的 n 个事件独立,则取 $A_i=\{-\infty<X_i\leqslant x_i\},i=1,\cdots,n$,由

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)$$

即得

$$\int_{-\infty}^{x_n}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1}f(t_1,t_2,\cdots t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n$$

$$=\int_{-\infty}^{x_n}f_n(t_n)\mathrm{d}x_n\cdots\int_{-\infty}^{x_2}f_2(t_2)\mathrm{d}t_2\int_{-\infty}^{x_1}f_1(t_1)\mathrm{d}t_1$$

上式两边依次对 $x_1,x_2,\cdots,x_n$ 取偏导数(即作$\partial^n/\partial x_1\partial x_2\cdots\partial x_n$),即得(3.13)式,因而证明了 $X_1,\cdots,X_n$ 独立.

下面再提出两个有关独立性的有用的结果.

定理 3.2 若连续型随机向量$(X_1,\cdots,X_n)$的概率密度函数 $f(x_1,\cdots,x_n)$可表为 n 个函数$g_1,\cdots,g_n$ 之积,其中 g_i 只依赖于 x_i,即

$$f(x_1,\cdots,x_n)=g_1(x_1)\cdots g_n(x_n) \tag{3.15}$$

则 $X_1,\cdots,X_n$ 相互独立,且 X_i 的边缘密度函数$f_i(x_i)$与 $g_i(x_i)$只相差一个常数因子.

证 按(2.11)式,知 X_1 的密度函数为

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

$$=g_1(x_1)\int_{-\infty}^{\infty}g_2(x_2)\mathrm{d}x_2\cdots\int_{-\infty}^{\infty}g_n(x_n)\mathrm{d}x_n=C_1g_1(x_1)$$

其中 C_1 为常数.同法证明 X_i 的密度函数$C_ig_i(x_i)$.

因此 $g_i(x_i)=C_i^{-1}f_i(x_i)$,其中 f_i 是x_i 的密度函数.以此代入(3.15),知 $f(x_1,\cdots,x_n)=C_1^{-1}C_2^{-1}\cdots C_n^{-1}f_1(x_1)\cdots f_n(x_n)$.由此式两边对 $x_1,\cdots,x_n$ 的积分都为 1,知 $C_1^{-1}C_2^{-1}\cdots C_n^{-1}=1$,因而 $f(x_1,\cdots,x_n)=f_1(x_1)\cdots f_n(x_n)$.按定义(3.1),知 $X_1,\cdots,X_n$ 独立.

定理 3.3 若 $X_1,\cdots,X_n$ 相互独立,而

$$Y_1=g_1(X_1,\cdots,X_m),Y_2=g_2(X_{m+1},\cdots,X_n)$$

则 Y_1 和 Y_2 独立.

这个定理直观上的意义很明白:因为 $X_1,\cdots,X_n$ 相互独立,把

它分成两部分 $X_1,\cdots,X_m$ 及 $X_{m+1},\cdots,X_n$，二者没有关系. 因为 Y_1,Y_2 分别只与前者和后者有关，它们之间也不应有相依关系. 证明细节也不难写出，在此从略了.

以上讨论的是关于连续型变量的独立性，至于离散型则更为简单.

定义 3.2 设 $X_1,\cdots,X_n$ 都是离散型随机变量. 若对任何常数 $a_1,\cdots,a_n$，都有

$$P(X_1=a_1,\cdots,X_n=a_n)=P(X_1=a_1)\cdots P(X_n=a_n)$$

则称 $X_1,\cdots,X_n$ 相互独立.

所有关于独立性的定理，如定理 3.1—3.3，全都适用于离散型. 唯一的变动是：凡是在这些定理中提到"密度函数"的地方，现在要改为"概率函数".

例 3.4 设 (X_1,X_2) 服从二维正态分布 $N(a,b,\sigma_1^2,\sigma_2^2,\rho)$. 由其联合密度函数 $f(x_1,x_2)$ 的形式(2.7)看出：当且仅当 $\rho=0$ 时，$f(x_1,x_2)$ 才可以表为两个边缘密度 $f_1(x_1)$ 和 $f_2(x_2)$ 之积. 因此，当且仅当 $\rho=0$ 时，X_1 和 X_2 独立. 这进一步反映了我们以前提及的一点事实：ρ 这个参数与 X_1,X_2 的相依性有关.

例 3.5 考虑例 2.4 的随机向量 (X_1,X_2). 据例 2.9 的结果，不难知道：X_1,X_2 不为独立.

与事件的独立性一样，在实际问题中，变量的独立性往往不是从其数学定义去验证出来的. 相反，常是从变量产生的实际背景判断它们独立(或者其相依性很微弱因而可近似地认为是独立)，然后再使用独立性定义中所赋予的性质和独立性的有关定理. 例如，一城市中两个相距较远的路段在一定时间内各自发生的交通事故数，一个人的姓氏笔划与其智商. 在实际中，n 个变量 $X_1,\cdots,X_n$ 的独立性通常是这样产生的：有 n 个彼此无关联的试验 $E_1,\cdots,E_n$，而 X_i 只依赖于试验 E_i 的结果. 形式上我们可以构作一个复合试验 $E=(E_1,\cdots,E_n)$，以把这 n 个变量都包容在这个试验 E 之下. 这种观点在讲事件独立性时已提到过了.

然而，在主要是理论的情况下，需要直接借助于定义来验证变

量的独立性.举一个例子.

例 3.6 设 X_1, X_2 独立,都服从标准正态分布 $N(0,1)$. 把点 (X_1, X_2) 的极坐标记为 (R, Θ), $0 \leqslant R < \infty, 0 \leqslant \Theta < 2\pi$. 求证:$R$ 和 Θ 独立(图 2.9).

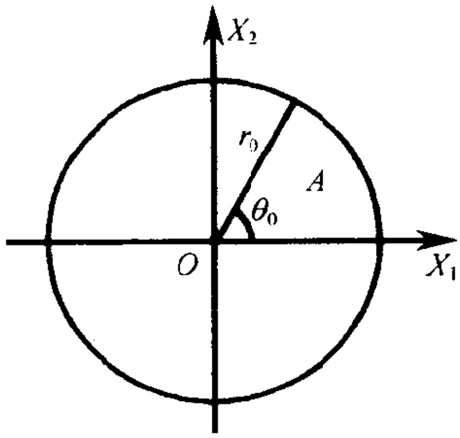

图 2.9

取定 $r_0 > 0, 0 < \theta_0 < 2\pi$. 考虑事件 $B = \{0 \leqslant R \leqslant r_0, 0 \leqslant \Theta \leqslant \theta_0\}$. 由于 X_1, X_2 独立且各自的密度函数分别为 $\sqrt{2\pi}^{-1} \cdot e^{-x_1^2/2}$ 和 $\sqrt{2\pi}^{-1} \cdot e^{-x_2^2/2}$, 由独立性定义 3.1 知 (X_1, X_2) 的联合密度为 $(2\pi)^{-1}\exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right)$. 因此,按密度函数的定义(2.5)式,有

$$P(B) = \iint_A (2\pi)^{-1}\exp\left[-\frac{1}{2}(x_1^2 + x_2^2)\right]dx_1dx_2$$

化为极坐标,得

$$P(0 \leqslant R \leqslant r_0, 0 \leqslant \Theta \leqslant \theta_0) = (2\pi)^{-1}\int_0^{\theta_0}\int_0^{r_0} e^{-r^2/2} r\,dr\,d\theta$$

由这个等式直接看出:(R, Θ) 的概率密度函数就是 $(2\pi)^{-1}e^{-r^2/2}r$ (当 $0 \leqslant r < \infty, 0 \leqslant \theta < 2\pi$, 其他处为 0). 它是下述两个函数的乘积:

$$f_1(r) = \begin{cases} e^{-r^2/2}r, & \text{当 } r \geqslant 0 \\ 0, & \text{当 } r < 0 \end{cases}$$

$$f_2(\theta) = \begin{cases} 1/2\pi, & 0 \leqslant \theta < 2\pi \\ 0, & \text{当他 } \theta \end{cases}$$

按定理 3.2,即得知 R 与 Θ 独立,且 R 与 Θ 的密度函数分别是 $f_1(r)$ 和 $f_2(\theta)$.

离散型变量独立性的一个重要例子涉及事件独立性与随机变

量独立性之间的关系.

例 3.7 设有 n 个事件 $A_1, A_2, \cdots, A_n$. 针对每个事件 A_i, 可定义一随机变量 X_i 如下:

$$X_i = 1, \text{当事件 } A \text{ 发生}; X_i = 0, \text{当 } A \text{ 不发生} \tag{3.16}$$

常把 X_i 称为事件 A 的指示变量或指示函数、示性函数(indicator), 意思是其值"指示"了 A 是否发生. 这个写法表明: 事件可视作随机变量的一种特例.

不难证明: 若事件 $A_1, \cdots, A_n$ 独立, 则其指示变量 $X_1, \cdots, X_n$ 独立. 反之亦成立. 证明是基于第一章的系 3.3, 我们把细节留给读者自己去完成.

利用指示变量的概念, 可以对第一章系 3.2 后面那段话作出统一而简洁的论证. 若事件 $A_1, \cdots, A_n$ 独立, 而事件 B_1 取决于 $A_1, \cdots, A_m$(这意思是说: 一旦知道了事件 $A_1, \cdots, A_m$ 中每一个发生与否, 就能定下 B_1 发生与否), 事件 B_2 取决于 $A_{m+1}, \cdots, A_n$, 则 B_1 与 B_2 独立. 转到指示变量: 分别以 $X_1, \cdots, X_n$ 记 $A_1, \cdots, A_n$ 的指示变量, 以 Y_1 和 Y_2 分别记 B_1 和 B_2 的指示变量. 按假定, 后者分别是 $X_1, \cdots, X_m$ 与 $X_{m+1}, \cdots, X_n$ 的函数:

$$Y_1 = g_1(X_1, \cdots, X_m), Y_2 = g_2(X_{m+1}, \cdots, X_n)$$

由 $A_1, \cdots, A_n$ 独立知随机变量 $X_1, \cdots, X_n$ 独立. 再据定理 3.3, 即知 Y_1 与 Y_2 独立, 因而事件 B_1 和 B_2 独立.

例 3.8 设$(X_1, \cdots, X_n)$服从多项分布 $M(N; p_1, \cdots, p_n)$, $p_i > 0, i = 1, \cdots, n$. 对任何 $u \neq v$, X_u 和 X_v 不独立.

这个结论从直观上看至为明显: 按多项分布的定义有 $X_1 + \cdots + X_n = N$. 若 $n > 2$, 则虽然 X_u 并不足以唯一决定 X_v, 但二者有关. 例如, 当 X_u 取很大的值时, X_v 取大值的可能性就降低. 这说明: X_v 在给定 X_u 之下的条件分布, 取决于 X_u 的给定值, 因而不符合独立性的要求. 形式的证明也不难作出, 留给读者去完成.