第 3 章　多元随机变量

由第 2 章可知，随机变量是对随机现象结果的记录．对一个样本点 ω 只用一个实数去描述往往是不够的．很多随机现象往往涉及多个随机变量，对随机现象的多个侧面进行描述，要把这些随机变量当作一个整体来看待，就形成了多元随机变量．

3.1　多元随机变量及其联合分布

例 3.1　打靶时，用 (X,Y) 表示/记录弹着点，其中 X 记录该点的横坐标，Y 记录该点的纵坐标．由于射击的随机性，X,Y 都各是一元随机变量．对弹着点的研究，仅研究 X 或仅研究 Y 都是片面的，应该把 X 和 Y 作为一个整体来考虑，讨论它们整体变化的概率特性，进一步可以讨论 X 与 Y 之间的关系．在有些随机现象中，甚至要同时考虑两个以上的随机变量．

直观来看，设 $X_1=X_1(\omega),X_2=X_2(\omega),\cdots,X_n=X_n(\omega)$ 是一个随机现象中的 n 个随机变量，则它们的整体，$(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))^{\mathrm{T}}$，称为 n 维/元随机向量（n 维/元随机变量），有时也简称为随机变量．更严谨的定义如下．

定义 3.1（多元随机变量）　设 $(\Omega,\mathscr{F},P)$ 是一个随机现象的概率空间，从样本空间 Ω 到 $\mathbb{R}^n$ 的函数 $X(\omega):\Omega\to\mathbb{R}^n$，称为 n 维随机变量（n 维随机向量）[①]，$\boldsymbol{X}(\omega)\overset{\text{def}}{=\!=}(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))^{\mathrm{T}}$．今后常省略符号 ω，把 $\boldsymbol{X}(\omega)$ 记为 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$，默认是列向量．另外，只要标识清楚，也可以把随机向量记成行向量，$\boldsymbol{X}(\omega)\overset{\text{def}}{=\!=}(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))$．

例 3.2　从某大学随机抽取一个同学 ω，其性别记为

$$X_1(\omega)=\begin{cases}0, & \omega\text{是男生},\\ 1, & \omega\text{是女生},\end{cases}$$

身高记为 $X_2(\omega)$，体重记为 $X_3(\omega)$，肺活量记为 $X_4(\omega)$，则 $X_1(\omega),X_2(\omega),X_3(\omega),X_4(\omega)$ 是对 ω（同学）的不同侧面的描述，应视为一个整体，这个整体构成一个 4 元随机变量 (X_1,X_2,X_3,X_4)．

注意，多维随机变量 $(X_1(\omega),X_2(\omega),\cdots,X_n(\omega))^{\mathrm{T}}$ 的关键是定义在同一样本空间上，各分量 $X_i(\omega)$（$i=1,\cdots,n$）依共同的自变量 ω 而变，各分量一般说来有某种联系，因

---

① 当然，为了 X 是数学上严格定义的 n 维随机变量，必须对函数 $X(\omega)$ 有所假定．类似一维随机变量，假定对任何 $\boldsymbol{x}\in\mathbb{R}^n$，集合 $\{\omega:\boldsymbol{X}(\omega)\leqslant\boldsymbol{x}\}$ 属于 $\mathscr{F}$．

而需要把它们作为一个整体（向量）来进行研究．从几何图像来看，二维随机向量可以看成是平面上的“随机点”，三维随机向量可以看成是空间（三维空间）中的“随机点”．当 $n \geqslant 4$ 时，n 维随机向量可以想象为 n 维空间中的“随机点”．

在引入多维随机变量后，可以用多维随机变量来表示随机事件．比如，我们在例 2.10 中，已经使用了一个三维随机变量 $(X_{\mathrm{A}}, X_{\mathrm{B}}, X_{\mathrm{C}})$ 来表示邮局中三个顾客 A, B, C 的办事时长，用 $\{\min(X_{\mathrm{A}}, X_{\mathrm{B}}) + X_{\mathrm{C}} > \max(X_{\mathrm{A}}, X_{\mathrm{B}})\}$ 表示 C 是最后一个办完事情．一般地，设有 n 维随机变量 $\boldsymbol{X} \in \mathbb{R}^n$，考虑数集 $B \subset \mathbb{R}^n$，则 $\boldsymbol{X}$ 取值于 B 表示随机事件

$$\{\boldsymbol{X} \in B\} = \{\omega | \boldsymbol{X}(\omega) \in B\}.$$

这与我们在一元随机变量中的做法是完全类似的．从认识一元随机变量上升到认识多元随机变量，是概率论学习的一个重要台阶．迈这个台阶，很关键的一点就是向量思维．有了向量观点，很多概念从一元到多元是一个很自然的推广，读者要多加体会．

与一元随机变量中的结论类似，人们一般关心的多元随机变量 $\boldsymbol{X}$ 取值所表达的事件，均可以通过多元随机变量 $\boldsymbol{X}$ 取值于**基本博雷尔集**所表达的事件经集合运算（有限或可列个交、并运算）来表示．类比实轴（一维实数空间）上基本博雷尔集的定义 1.12，n 维实数空间的基本博雷尔集如下定义：对任意 $\boldsymbol{x} \stackrel{\text{def}}{=} (x_1, x_2, \cdots x_n)^{\mathrm{T}} \in \mathbb{R}^n$，

$$\{\boldsymbol{c} \stackrel{\text{def}}{=} (c_1, c_2, \cdots c_n)^{\mathrm{T}} | c_1 \in (-\infty, x_1], c_2 \in (-\infty, x_2], \cdots, c_n \in (-\infty, x_n)\}$$

称为以 $\boldsymbol{x}$ 为端点的（n 维）基本博雷尔集，即各维分别小于或等于 $x_1, x_2, \cdots, x_n$ 的 n 维实数向量构成的集合．$n = 2$ 的一个例子见图 3.1．如果将在 n 维实数空间的大小比较关系，“$x_1 \leqslant c_1, x_2 \leqslant c_2, \cdots, x_n \leqslant c_n$” 简写为“$\boldsymbol{x} \leqslant \boldsymbol{c}$”，则以 $\boldsymbol{x}$ 为端点的（n 维）基本博雷尔集可简写为

$$\{\boldsymbol{c} | \boldsymbol{c} \leqslant \boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^n\} \stackrel{\text{def}}{=} B_{\boldsymbol{x}}.$$

这显然是（一维）基本博雷尔集的自然推广．

定义 3.2（联合分布函数） 考虑 n 元随机变量 $\boldsymbol{X} \stackrel{\text{def}}{=} (X_1, X_2, \cdots, X_n)^{\mathrm{T}}$ 落在以 $\boldsymbol{x} \stackrel{\text{def}}{=} (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 为端点的基本博雷尔集的概率，让 $\boldsymbol{x} \in \mathbb{R}^n$ 变动起来得到一个 n 元函数，称为 $\boldsymbol{X}$ 的联合（累积）分布函数（joint CDF, JCDF），

$$F_{X_1 X_2 \cdots X_n}(x_1, x_2, \cdots, x_n) \stackrel{\text{def}}{=} P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n), \forall x_1, x_2, \cdots, x_n \in \mathbb{R}$$

$$\text{或简写为} F_{\boldsymbol{X}}(\boldsymbol{x}) \stackrel{\text{def}}{=} P(\boldsymbol{X} \leqslant \boldsymbol{x}),$$

记为 $\boldsymbol{X} \sim F_{\boldsymbol{X}}(\boldsymbol{x})$．这里用下标表示是 $\boldsymbol{X}$ 的分布函数，有时可省略下标 $\boldsymbol{X}$，记为 $F(\boldsymbol{x})$，指代函数本身，即为 $\boldsymbol{X}$ 的分布函数．

注 1 随机向量 $\boldsymbol{X}$ 取值于基本博雷尔集 $B_{\boldsymbol{x}}$ 是一个随机事件，

$$\{\boldsymbol{X} \in B_{\boldsymbol{x}}\} = \{\boldsymbol{X} \leqslant \boldsymbol{x}\} = \{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}$$

表示 n 个事件 $\{X_1 \leqslant x_1\}, \{X_2 \leqslant x_2\}, \cdots, \{X_n \leqslant x_n\}$ 同时发生（交）.

注 2 由以上定义可知，与一维情形完全类似，随机向量 $\boldsymbol{X}$ 取值于基本博雷尔集 $B_{\boldsymbol{x}}$ 的概率 $P(\boldsymbol{X} \leqslant \boldsymbol{x})$ 随博雷尔集的端点（$\boldsymbol{x}$）变动而形成的函数，称为分布函数.

注 3 简单起见，本章举例多为二维随机变量，二维以上的情况可类似进行．在二维随机变量 (X, Y) 场合，联合分布函数

$$F(x, y) = P(X \leqslant x, Y \leqslant y)$$

表示事件 $\{X \leqslant x\}$ 与 $\{Y \leqslant y\}$ 同时发生（交）的概率．如果将二维随机变量 (X, Y) 看成是平面上的随机点的坐标，那么联合分布函数 $F(x, y)$ 在 (x, y) 处的函数值就是随机点 (X, Y) 落在以 (x, y) 为端点的无穷直角区域（参见图 3.1）的概率.

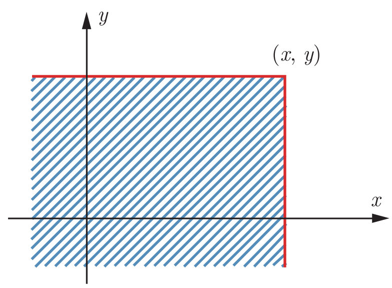


图 3.1 一个二维基本博雷尔集的示意图

定理 3.1 任一二维联合分布函数 $F(x, y)$ 必具有以下 4 点性质：

(1) 单调性 $F(x, y)$ 分别对 x 或 y 是单调非减的，即当 $x_1 < x_2$ 时，有 $F(x_1, y) \leqslant F(x_2, y)$，当 $y_1 < y_2$ 时，有 $F(x, y_1) \leqslant F(x, y_2)$.

(2) 有界性 对任意的 x 和 y，有 $0 \leqslant F(x, y) \leqslant 1$，且

$$F(-\infty, y) = \lim_{x \to -\infty} F(x, y) = 0,$$

$$F(x, -\infty) = \lim_{y \to -\infty} F(x, y) = 0,$$

$$F(+\infty, +\infty) = \lim_{x, y \to +\infty} F(x, y) = 1.$$

(3) 右连续性 对每个变量都是右连续的，即

$$F(x + 0, y) = F(x, y), \qquad F(x, y + 0) = F(x, y).$$

(4) 非负性 对任意的 $a < b$，$c < d$ 有

$$P(a < X \leqslant b, c < Y \leqslant d)$$

$$= F(b, d) - F(a, d) - F(b, c) + F(a, c) \geqslant 0.$$

证明 (1) 因为当 $x_1 < x_2$ 时，有 $\{X \leqslant x_1\} \subset \{X \leqslant x_2\}$，所以对于任意给定的 y 有

$$\{X \leqslant x_1, Y \leqslant y\} \subset \{X \leqslant x_2, Y \leqslant y\},$$

由此可得

$$F(x_1,y)=P(X\leqslant x_1,Y\leqslant y)\leqslant P(X\leqslant x_2,Y\leqslant y)=F(x_2,y),$$

即 $F(x,y)$ 关于 x 是单调非减的. 同理可证 $F(x,y)$ 关于 y 是单调非减的.

(2) 由概率的性质可知 $0\leqslant F(x,y)\leqslant 1$. 又因为对任意的正整数 n 有

$$\lim_{x\to-\infty}\{X\leqslant x\}=\lim_{n\to\infty}\bigcap_{m=1}^{n}\{X\leqslant -m\}=\varnothing,$$

$$\lim_{x\to+\infty}\{X\leqslant x\}=\lim_{n\to\infty}\bigcup_{m=1}^{n}\{X\leqslant m\}=\Omega,$$

对 $\{Y\leqslant y\}$ 也类似可得, 再由概率的连续性, 就可得

$$F(-\infty,y)=F(x,-\infty)=0,F(+\infty,+\infty)=1.$$

(3) 固定 y, 仿一维分布函数右连续的证明, 就可得知 $F(x,y)$ 关于 x 是右连续的. 同样固定 x, 可证得 $F(x,y)$ 关于 y 是右连续的.

(4) 只需证

$$P(a<X\leqslant b,c<Y\leqslant d)=F(b,d)-F(a,d)-F(b,c)+F(a,c).$$

为此记

$$A=\{X\leqslant a\},B=\{X\leqslant b\},C=\{Y\leqslant c\},D=\{Y\leqslant d\}.$$

考虑到

$$\{a<X\leqslant b\}=B-A=B\cap\overline{A},\{c<Y\leqslant d\}=D-C=D\cap\overline{C},$$

且 $A\subset B,C\subset D$, 由此可得

$$\begin{aligned}
0&\leqslant P(a<X\leqslant b,c<Y\leqslant d)\\
&=P(B\cap\overline{A}\cap D\cap\overline{C})\\
&=P(BD-(A\cup C))\\
&=P(BD)-P(ABD\cup BCD)\\
&=P(BD)-P(AD\cup BC)\\
&=P(BD)-P(AD)-P(BC)+P(ABCD)\\
&=P(BD)-P(AD)-P(BC)+P(AC)\\
&=F(b,d)-F(a,d)-F(b,c)+F(a,c).
\end{aligned}$$

■

还可证明, 具有上述 4 条性质的二元函数 $F(x,y)$ 一定是某个二维随机变量的分布函数.

任一二维分布函数 $F(x,y)$ 必具有上述 4 条性质, 其中性质 (4) 是二维场合特有的, 也是合理的. 但性质 (4) 不能由前三条性质推出, 必须单独列出, 因为存在这样的二元函数, 满足以上性质 (1)(2)(3), 但它不满足性质 (4).

怎么研究 n 维随机向量呢？在本节我们先介绍离散型随机向量、连续型随机向量，然后在 3.5.4 节讨论混合型随机向量. 有的地方为简单起见，常以二维随机向量为例，读者可类比将二维情形的结论推广到 n 维情形.

3.1.1　离散型多元随机变量

定义 3.3（离散型多元随机变量）　若 n 维随机变量 $\boldsymbol{X} \overset{\text{def}}{=\!=} (X_1, X_2, \cdots, X_n)^{\mathrm{T}} \in \mathbb{R}^n$ 的取值为有限或可列个（注意，每个值是一个 n 维实数向量），则称为 n 维/元离散随机变量；可能取值分别记为 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \in \mathbb{R}^n$，称一列数 $P(\boldsymbol{X} = \boldsymbol{x}_k) \overset{\text{def}}{=\!=} f_{\boldsymbol{X}}(\boldsymbol{x}_k)$ 为 $\boldsymbol{X}$ 的概率分布列或简称为分布列（PMF），记为 $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x}_k)$，也统称为概率分布. 这里用下标表示是 $\boldsymbol{X}$ 的分布函数，有时可省略下标 $\boldsymbol{X}$，记为 $f(\boldsymbol{x}_k)$，指代分布列本身.

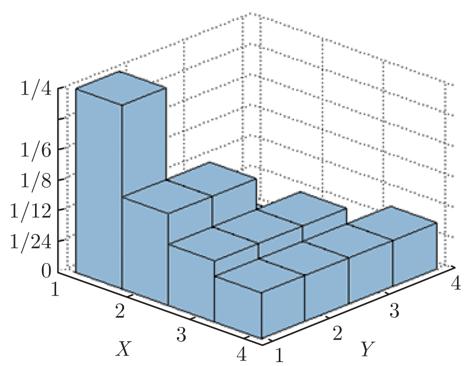
注　读者不难发现，离散型多元随机变量及其分布的定义，与前面离散型一元随机变量及其分布的定义完全类似，只是换成向量表达. 为强调是多元情形，$\boldsymbol{X}$ 的概率分布 $f_{\boldsymbol{X}}(\cdot)$ 也叫作 $(X_1, X_2, \cdots, X_n)$ 的联合分布列（Joint PMF，JPMF）.

定理 3.2（联合分布列的基本性质）　（1）非负性，$f(\boldsymbol{x}_k) \geqslant 0, k = 1, 2, \cdots$；（2）正则性，$\sum\limits_k f(\boldsymbol{x}_k) = 1$.

如果二维随机变量 (X, Y) 只取有限个或可列个数对 (x_i, y_j)，则 (X, Y) 为二维离散随机变量，其联合分布列 $f_{XY}(x_i, y_j), i, j = 1, 2, \cdots$，可以用一张二维表格来表示，或者更形象地，用二维平面上的柱状图来表示，参见图 3.2.

X \ Y	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

(a) 联合分布列（表格形式）



(b) 联合分布列(柱状图)

图 3.2　一个二维离散随机变量（例 3.3）的联合分布列（JPMF）

例 3.3　从 1,2,3,4 中任取一个数记为 X，再从 $1, 2, \cdots, X$ 中任取一数记为 Y，求 (X, Y) 的联合分布列.

解　不难看出 (X, Y) 构成一个二维随机变量，求出联合分布列如图 3.2 所示，用表格或者柱状图来表示. 我们还画出了 (X, Y) 的联合分布函数，见图 3.3，供读者形象理解.

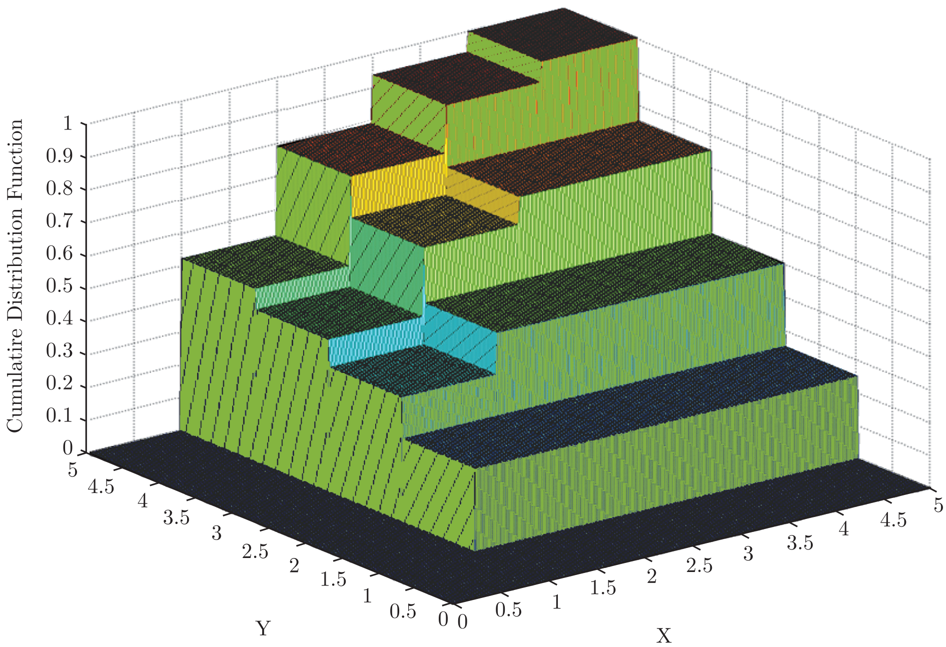


图 3.3　一个二维离散随机变量（例 3.3）的联合分布函数（JCDF）

例 3.4（多项分布）　多项分布是重要的多维离散分布，它是二项分布的推广，其定义如下：进行 n 次独立重复试验，如果每次试验有 r 个互不相容的结果：$A_1, A_2, \cdots, A_r$ 之一发生，每次试验中 A_i 发生的概率为 $p_i = P(A_i), i = 1, 2, \cdots, r$，且 $p_1 + p_2 + \cdots + p_r = 1$. 记 X_i 为 n 次独立重复试验中 A_i 出现的次数，$i = 1, 2, \cdots, r$，则 $(X_1, X_2, \cdots, X_r)$ 取值 $(n_1, n_2, \cdots, n_r)$ 的概率，即 A_1 出现 n_1 次，A_2 出现 n_2 次，$\cdots\cdots, A_r$ 出现 n_r 次的概率为

$$P(X_1 = n_1, X_2 = n_2, \cdots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r},$$

其中 $n = n_1 + n_2 + \cdots + n_r$. 这个概率是多项式 $(p_1 + p_2 + \cdots + p_r)^n$ 展开式中的一项，故其和为 1.

这个联合分布列称为 r 项分布，又称多项分布，记为 $M(n, p_1, p_2, \cdots, p_r)$. 当 $r = 2$ 时，即为二项分布.

注意，一个 r 项分布也可用 $r - 1$ 维随机向量的分布来表示. 以三项分布为例，设 $\boldsymbol{X} = (X_1, X_2)$ 取值于集合 $\mathcal{X} = \{(i, j); i$ 和 j 都是自然数且 $i + j \leqslant n\}$, $\boldsymbol{X}$ 的如下概率分布表示了一个三项分布 $M(n, p_1, p_2, 1 - p_1 - p_2)$

$$P(X_1 = i, X_2 = j) = \frac{n!}{i! j! (n - i - j)!} p_1^i p_2^j (1 - p_1 - p_2)^{n-i-j},$$

其中 $n \geqslant 1, 0 < p_1, p_2 < 1, p_1 + p_2 < 1, (i, j) \in \mathcal{X}$.

例 3.5（三项分布的例子） 今有一大批量粉笔，其中 60% 是白的，25% 是黄的，15% 是红的．先从中随机地、依次取出 6 支．问：这 6 支中恰有 3 支白、1 支黄、2 支红的概率是多少？

解 设 $X=$“6 支中白粉笔的支数”，$Y=$“6 支中黄粉笔的支数”，$Z=$“6 支中红粉笔的支数”，则 (X,Y,Z) 构成一个三维离散随机变量，不难看出其服从三项分布．则事件“6 支中恰有 3 支白、1 支黄、2 支红”就是事件 $\{X=3,Y=1,Z=2\}$，即$\{(X,Y,Z)=(3,1,2)\}$．该事件的概率为

$$P((X,Y,Z)=(3,1,2))=\frac{6!}{3!1!2!}(0.6)^3(0.25)(0.15)^2=0.0729.$$

更详细的推导如下．用“白白黄白红红”表示第一支是白的，第二支是白的，第三支是黄的，第四支是白的，第五支是红的，第六支是红的．由于是大批量，我们可以认为各次抽取是独立的且抽取到黄、红、白的概率不变，有

$$P(\text{“白白黄白红红”})=P(\text{白})P(\text{白})P(\text{黄})P(\text{白})P(\text{红})P(\text{红})=(0.6)^2(0.25)(0.6)(0.15)^2.$$

于是

$$P(\text{“6 支中恰有 3 支白、1 支黄、2 支红”})=m\cdot(0.6)^3(0.25)(0.15)^2,$$

其中 m 是由三白、一黄、二红组成的六维向量的个数．据排列组合知识得 $m=\dfrac{6!}{3!1!2!}=60$．因此所求得概率为

$$60\cdot(0.6)^3(0.25)(0.15)^2=0.0729. \qquad \blacksquare$$

3.1.2 连续型多元随机向量

定义 3.4（连续型多元随机向量） 如果存在二元非负函数 $f_{XY}(x,y)$，使得二维随机向量 (X,Y) 的分布函数可表示为

$$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(u,v)\mathrm{d}u\mathrm{d}v,$$

则称 (X,Y) 为二维连续随机向量，称 $f_{XY}(\cdot,\cdot)$ 为 (X,Y) 的联合概率密度函数（joint PDF，JPDF），简称密度函数或分布密度，记为 $(X,Y)\sim f_{XY}(x,y)$，也统称为概率分布．这里用下标表示是 XY 的分布函数，有时可省略下标 XY，记为 $f(x,y)$，指代联合概率密度函数本身．

推广到 n 维随机向量 $\boldsymbol{X}\in\mathbb{R}^n$，若存在 n 元非负函数 $f_{\boldsymbol{X}}(\boldsymbol{x}),\boldsymbol{x}\in\mathbb{R}^n$，使得 $\boldsymbol{X}$ 的分布函数可表示为

$$F_{\boldsymbol{X}}(\boldsymbol{x})=\int_{\boldsymbol{u}\leqslant\boldsymbol{x}}f_{\boldsymbol{X}}(\boldsymbol{u})\mathrm{d}\boldsymbol{u},$$

则称 $\boldsymbol{X}$ 为 n 维/元连续随机向量，记为 $\boldsymbol{X}\sim f_{\boldsymbol{X}}(\boldsymbol{x})$．上式中积分表示在以 $\boldsymbol{x}$ 为端点的基本博雷尔集 $B_{\boldsymbol{x}}$ 上的 n 重积分．

定理 3.3（联合概率密度函数的基本性质） （1）非负性，$f_{XY}(x,y)\geqslant 0$；

（2）正则性，$\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f_{XY}(u,v)\mathrm{d}u\mathrm{d}v=1$.

注 1 不难看出，与一维连续变量类似，对二维连续随机向量 (X,Y)，我们有：

- 其分布函数 $F_{XY}(x,y)$ 是 (x,y) 的连续函数；
- 在 $F(x,y)$ 偏导数存在的点上，$f_{XY}(x,y)=\dfrac{\partial^2}{\partial x\partial y}F_{XY}(x,y)$；
- 连续随机向量在单点上的取值概率等于 0，即 $P(X=x,Y=y)=0,\forall(x,y)\in\mathbb{R}^2$；
- (X,Y) 取值位于 (u,v) 附近一个微元区域的概率 $P(u<X\leqslant u+\mathrm{d}u,v<Y\leqslant v+\mathrm{d}v)\approx f_{XY}(u,v)\mathrm{d}u\mathrm{d}v$.

注 2 对连续型的随机向量 (X,Y)，可以证明对于平面上相当任意①的点集 $B\subset\mathbb{R}^2$ 均成立

$$P((X,Y)\in B)=\iint_{(x,y)\in B}f_{XY}(x,y)\mathrm{d}x\mathrm{d}y. \tag{3.1}$$

更一般地，对 n 维随机向量 $\boldsymbol{X}\in\mathbb{R}^n$ 以及 n 维实数空间中相当任意的点集 $B\subset\mathbb{R}^n$ 均成立

$$P(\boldsymbol{X}\in B)=\int_{\boldsymbol{x}\in B}f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{3.2}$$

式 (3.1), 式 (3.2) 是本章的基本公式之一．它的证明要用到较深的数学知识，超出了本书的范围，证明从略．读者要理解这个公式的意义和用法．

从式 (3.1) 知道，随机向量 (X,Y) 落入平面上任一区域 B 的概率等于联合密度 $f_{XY}(x,y)$ 在 B 上的二重积分．这就把概率的计算化为二重积分的计算．由此可知，事件 $\{(X,Y)\in B\}$ 的概率等于以曲面 $z=f_{XY}(x,y)$ 为顶，以平面区域 B 为底的曲顶柱体的体积．更一般地，从式 (3.1) 知道，n 维随机向量 $\boldsymbol{X}$ 落入 n 维实数空间中任一区域 B 的概率等于概率密度函数 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 在 B 上的 n 重积分．

以二维为例，在具体使用式 (3.1) 时，要注意积分范围是联合密度 $f_{XY}(x,y)$ 的非零区域（支集）与 B 的交集部分，然后设法化成累次积分，最后计算出结果．

例 3.6 设两个灯管的寿命为二元随机向量 (X,Y), 其联合概率密度函数为

$$f(x,y)=\begin{cases}2\mathrm{e}^{-x}\mathrm{e}^{-2y}, & 0<x<+\infty,0<y<+\infty,\\ 0, & \text{其他}.\end{cases}$$

求 $P(X<Y)$.

解
$$\begin{aligned}P(X<Y)&=\iint_{x<y}f(x,y)\mathrm{d}x\mathrm{d}y=\int_{0<x<+\infty}\int_{x<y<+\infty}2\mathrm{e}^{-x}\mathrm{e}^{-2y}\mathrm{d}x\mathrm{d}y\\&=\int_0^{+\infty}\mathrm{e}^{-x}\left(\int_x^{+\infty}2\mathrm{e}^{-2y}\mathrm{d}y\right)\mathrm{d}x=\int_0^{+\infty}\mathrm{e}^{-x}\mathrm{e}^{-2x}\mathrm{d}x=\frac{1}{3}.\end{aligned}$$

定义 3.5（二维正态分布） 若二元随机向量 (X,Y) 的联合概率密度函数为

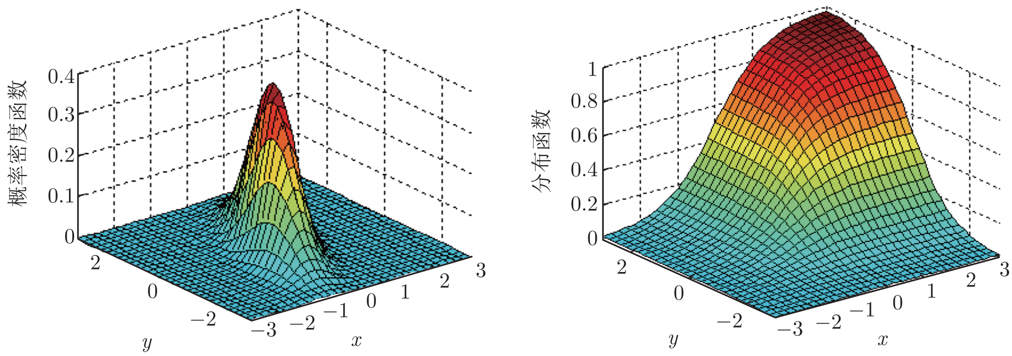
$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot$$

① “相当任意” 的集合 B 是指：B 为平面上的博雷尔集.

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}, \tag{3.3}$$

$-\infty < x, y < +\infty$，则称 (X, Y) 服从**二维正态分布**，称 (X, Y) 为（二维）正态变量，记为 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，其中 $-\infty < \mu_1, \mu_2 < +\infty, \sigma_1, \sigma_2 > 0, |\rho| \leqslant 1$ 为参数. 形如式 (3.3) 的概率分布称为二维正态分布，其概率密度函数和分布函数如图 3.4 所示. 式 (3.3) 所示正态分布的支集为二维平面 $\mathbb{R}^2$. 式 (3.3) 常简写为 $N(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. 同一维情形，正态分布可称为高斯分布，正态变量称为高斯变量.

以后将指出：μ_1, μ_2 分别是 X 与 Y 的均值，σ_1^2, σ_2^2 分别是 X 与 Y 的方差，ρ 是 X 与 Y 的相关系数.



(a) 概率密度函数 $f(x, y)$　　(b) 分布函数 $F(x, y)$

图 3.4　二维正态分布的概率密度函数与分布函数

定义 3.6（n 维正态分布） 若 n 维随机向量 $(X_1, X_2, \cdots, X_n) \overset{\text{def}}{=\!=} \boldsymbol{X} \in \mathbb{R}^n$ 的联合概率密度函数为

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \boldsymbol{x} \in \mathbb{R}^n, \tag{3.4}$$

则称 $\boldsymbol{X}$ 服从 n 维正态分布，称 $\boldsymbol{X}$ 为（n 维）正态变量，记为 $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中 $\boldsymbol{\mu} \in \mathbb{R}^n$ 是 n 维实向量，$\boldsymbol{\Sigma}$ 为 $n \times n$ 正定矩阵，是 n 元正态分布的表征参数. 式 (3.4) 所示正态分布的支集为 n 维实数空间 $\mathbb{R}^n$. 式 (3.4) 常简写为 $N(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

以后将指出：$\boldsymbol{\mu}$ 是 $\boldsymbol{X}$ 的均值向量，$\boldsymbol{\Sigma}$ 是 $\boldsymbol{X}$ 的协方差矩阵.

不难看出，二维正态分布（式 (3.3)）是 n 元正态分布（式 (3.4)）在 $n = 2$ 的特例. 由二维正态分布的参数 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 可定义

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

那么式 (3.4) 变为

$$f_{XY}(x,y)=\frac{1}{2\pi\begin{vmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{vmatrix}^{\frac{1}{2}}}\cdot$$
$$\exp\left\{-\frac{1}{2}\begin{pmatrix}x-\mu_1\\ y-\mu_2\end{pmatrix}^{\mathrm{T}}\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}^{-1}\begin{pmatrix}x-\mu_1\\ y-\mu_2\end{pmatrix}\right\},$$

稍加整理可知，这与二维正态分布（式 (3.3)）代表相同的概率密度函数.

例 3.7（蒲丰投针问题） 平面上画有间隔为 d（$d>0$）的等距平行线，向平面任意投掷一枚长为 l（$l<d$）的针，求针与任一平行线相交的概率.

解 以 X 表示针的中点与最近一条平行线的距离，以 Θ 表示针与此直线间的交角，见图 3.5（a），易知

$$0\leqslant X\leqslant\frac{d}{2},\quad 0\leqslant\Theta\leqslant\pi.$$

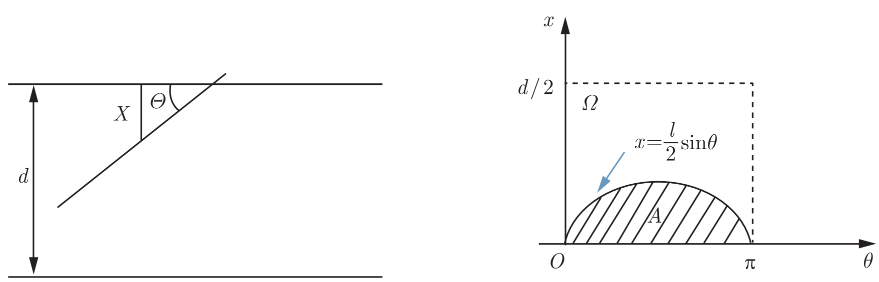
由这两式可以确定 x-θ 平面上的一个矩形 Ω 是随机变量 (X,Θ) 的支集，其面积为 $S_\Omega=\dfrac{d}{2}\pi$. 针与平行线相交（记为事件 A），可以等价表示为

$$X\leqslant\frac{l}{2}\sin\Theta.$$

由这个不等式表示的区域是图 3.5（b）中的阴影部分，记为区域 B.

由于针是向平面任意投掷的，可以认为 (X,Θ) 在矩形区域（支集）内服从二维均匀分布，密度为 $1/(\mathrm{d}\pi/2)$，由此得

$$\begin{aligned}P(A)&=P\left(X\leqslant\frac{l}{2}\sin\Theta\right)=\iint_{x\leqslant\frac{l}{2}\sin\theta}f(x,\theta)\mathrm{d}x\mathrm{d}\theta\\&=\int_0^{\pi}\int_0^{\frac{l}{2}\sin\theta}\frac{1}{\frac{d}{2}\pi}\mathrm{d}x\mathrm{d}\theta=\frac{1}{\frac{d}{2}\pi}\int_0^{\pi}\frac{l}{2}\sin\theta\mathrm{d}\theta=\frac{2l}{d\pi}.\end{aligned}\tag{3.5}$$



(a) 投针示意图 (b) 针与平行线相交时，(X,Θ) 的取值区域

图 3.5 蒲丰投针问题