第 5 章　多维随机向量

前面讨论了一维随机变量的概率分布, 然而很多在实际问题中, 随机现象可能需要两种或两种以上的随机因素来描述, 仅仅用一个随机变量是不够的, 需要多个随机变量. 例如, 为了考察某地区儿童的身体素质时, 可以同时考虑他们的身高、体重、肺活量、视力等, 此时至少需要四个随机变量来描述. 这些随机变量之间可能存在某些关联, 因此分别对每个随机变量单独进行研究是不够的, 需要将其看作一个整体来研究, 即多维随机向量.

定义 5.1　设 $X_1 = X_1(\omega), X_2 = X_2(\omega), \cdots, X_n = X_n(\omega)$ 是定义在同一样本空间 Ω 上的 n 个随机变量, 由它们构成的向量 $(X_1, X_2, \cdots, X_n)$ 称为 n **维随机向量**, 或称 n **维随机变量**.

一维随机变量可以看作多维随机变量的一种特殊情况, 本章主要讨论二维随机向量及其分布, 同理可讨论二维以上的随机向量.

5.1 二维联合分布函数

类似于一维随机变量, 我们用分布函数来研究二维随机向量的概率特性.

定义 5.2　设 (X, Y) 为二维随机向量, 对任意实数 x 和 y,

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

称为二维随机向量 (X,Y) 的 **分布函数**, 或称随机变量 X 和 Y 的 **联合分布函数** (joint cumulative probability distribution function).

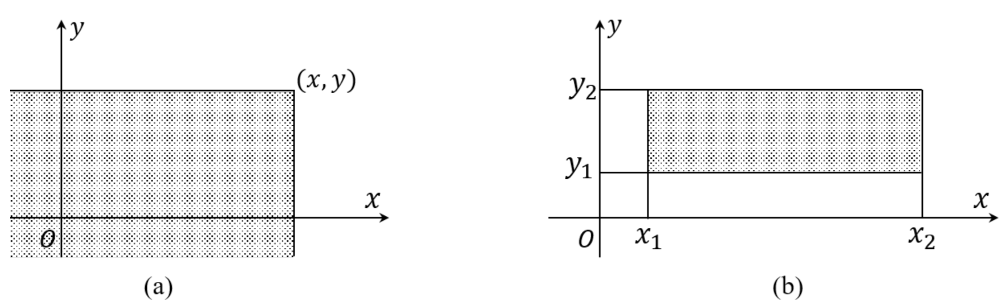


图 5.1　随机向量 (X,Y) 的分布函数 $F(x,y)$ 和概率 $P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$

若将 (X,Y) 看作平面上随机点的坐标, 则分布函数 $F(x,y)$ 的值表示随机向量 (X,Y) 落入以 (x,y) 为顶点的左下方无穷区域的概率, 如图 5.1(a) 所示. 再根据图 5.1(b) 可知, 随机向量 (X,Y) 落

入矩形区域 $\{(x,y)\colon x_1 < x \leqslant x_2, y_1 < y \leqslant y_2\}$ 的概率

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)\ .$$

二维随机向量 (X,Y) 的分布函数 $F(x,y)$ 具有以下性质:

1) 分布函数 $F(x,y)$ 对每个变量都是单调不减的, 即对任意固定的实数 y, 当 $x_1 > x_2$ 时有 $F(x_1,y) \geqslant F(x_2,y)$; 对任意固定的实数 x, 当 $y_1 > y_2$ 时有 $F(x,y_1) \geqslant F(x,y_2)$.

2) 对任意实数 x 和 y, 分布函数 $F(x,y) \in [0,1]$, 而且

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

3) 分布函数 $F(x,y)$ 关于每个变量右连续, 即

$$F(x,y) = F(x+0,y) \quad \text{和} \quad F(x,y) = F(x,y+0)\ .$$

4) 对任意实数 $x_1 < x_2$ 和 $y_1 < y_2$ 有

$$F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1) \geqslant 0\ .$$

任何的分布函数 $F(x,y)$ 都满足上述四条性质, 前三条性质与一维随机变量类似, 第四条性质根据图 5.1(b) 直接可证. 反之, 任何满足上面四条性质的二元函数 $F(x,y)$ 都可看成某二维随机向量的分布函数.

值得说明的是, 当二元函数 $F(x,y)$ 仅仅满足前面的三条性质时, 并不一定能成为某二维随机向量的分布函数, 例如

$$F(x,y) = \begin{cases} 1 & x+y \geqslant 0, \\ 0 & x+y < 0. \end{cases}$$

很容易验证 $F(x,y)$ 仅仅满足前面的三条性质, 但因为

$$F(1,1) - F(1,-1) - F(-1,1) + F(-1,-1) = -1\ ,$$

如图 5.2(a) 所示, 不满足第四条性质因此不构成一个分布函数.

根据随机向量 (X,Y) 的联合分布函数 $F(x,y)$, 还可以研究每个随机变量的统计特征, 即将 X 和 Y 看做单独的随机变量, 通过联合分布函数 $F(x,y)$ 来研究随机变量 X 和 Y 的分布函数 $F_X(x)$ 和 $F_Y(y)$, 即边缘分布函数.

定义 5.3 设二维随机向量 (X,Y) 的联合分布函数为 $F(x,y)$, 称

$$F_X(x) = P(X \leqslant x) = P(X \leqslant x, y < +\infty) = F(x,+\infty) = \lim_{y\to+\infty} F(x,y)\ ,$$

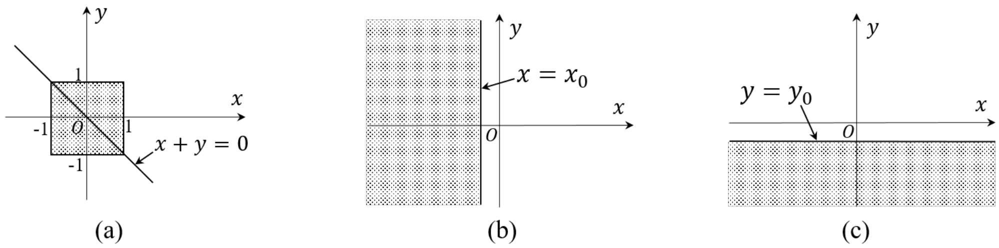


图 5.2 分布函数第四条性质的反例和边缘分布

为 (X,Y) 关于随机变量 X 的 **边缘分布函数** (marginal distribution function). 类似地定义 (X,Y) 关于随机变量 Y 的边缘分布函数

$$F_Y(y) = P(Y \leqslant y) = P(Y \leqslant y, x < +\infty) = F(+\infty, y) = \lim_{x\to+\infty} F(x,y) \ .$$

边缘分布函数 $F_X(x_0)$ 和 $F_Y(y_0)$ 的值分别表示随机向量 (X,Y) 落入图 5.2(b) 和 5.2(c) 中阴影部分的概率. 下面来看一个例子.

例 5.1 设二维随机向量 (X,Y) 的分布函数为

$$F(x,y) = A(B + \arctan\frac{x}{2})(C + \arctan\frac{y}{3}) \qquad (x, y \in \mathbb{R}) \ .$$

求随机变量 X 与 Y 的边缘分布函数, 以及概率 $P(Y > 3)$.

解 对任意 $x \in (-\infty, +\infty)$ 和 $y \in (-\infty, +\infty)$, 根据分布函数的性质有

$$\begin{aligned}
1 &= F(+\infty,+\infty) = A(B+\frac{\pi}{2})(C+\frac{\pi}{2}) \ , \\
0 &= F(x,-\infty) = A(B+\arctan\frac{x}{2})(C-\frac{\pi}{2}) \ , \\
0 &= F(-\infty,y) = A(B-\frac{\pi}{2})(C+\arctan\frac{y}{3}) \ .
\end{aligned}$$

求解上述方程可得

$$C = \frac{\pi}{2}, \quad B = \frac{\pi}{2}, \quad A = \frac{1}{\pi^2} \ .$$

从而得到 $F(x,y) = (\pi/2 + \arctan x/2)(\pi/2 + \arctan y/3)/\pi^2$, 进一步得到

$$F_X(x) = \lim_{y\to\infty} \frac{1}{\pi^2}(\frac{\pi}{2} + \arctan\frac{x}{2})(\frac{\pi}{2} + \arctan\frac{y}{3}) = \frac{1}{\pi}(\frac{\pi}{2} + \arctan\frac{x}{2}) \ ,$$

同理可得

$$F_Y(y) = \frac{1}{\pi}(\frac{\pi}{2} + \arctan\frac{y}{3}) \ .$$

最后得到

$$P(Y>3)=1-P(Y\leqslant 3)=1-F_Y(3)=1-\left(\frac{1}{2}+\frac{1}{\pi}\arctan 1\right)=\frac{1}{4}.$$

5.2 二维离散型随机向量

定义 5.4 若二维随机向量 (X,Y) 的取值是有限个或无限可列的, 则称 (X,Y) 为 **二维离散型随机向量**. 设离散型随机向量 (X,Y) 所有可能的取值为 $(x_i,y_j)\quad(i,j=1,2,\cdots)$, 则称

$$p_{ij}=P(X=x_i,Y=y_j)$$

为二维随机向量 (X,Y) 的 **联合分布列**, 简称 **分布列**.

二维随机向量分布列具有下列性质:

$$p_{ij}\geqslant 0 \qquad 和 \qquad \sum_{i,j}p_{ij}=1.$$

通过随机向量 (X,Y) 的联合分布列 p_{ij}, 还可以研究每个随机变量的统计特征, 例如随机变量 X 的 **边缘分布列** 为

$$P(X=x_i)=\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{j=1}^{\infty}p_{ij}=p_{i\cdot}\ ,$$

以及随机变量 Y 的 **边缘分布列** 为

$$P(Y=y_j)=\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{\infty}p_{ij}=p_{\cdot j}\ .$$

二维随机向量的联合分布列和边缘分布列可通过表 5.1 来进行表示.

表 5.1 二维随机向量的概率分布表

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}=\sum_j p_{ij}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
$p_{\cdot j}=\sum_i p_{ij}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

根据二维随机变量 (X,Y) 的联合分布列 p_{ij}, 可以得到它们的联合分布函数

$$F(x,y)=\sum_{x_i\leqslant x,y_j\leqslant y}p_{ij}\ ,$$

和边缘分布函数

$$F_X(x) = \sum_{x_i \leqslant x} p_{i\cdot} = \sum_{x_i \leqslant x} \sum_{j=1}^{+\infty} p_{ij} \quad \text{和} \quad F_Y(y) = \sum_{y_j \leqslant y} p_{\cdot j} = \sum_{y_j \leqslant y} \sum_{i=1}^{+\infty} p_{ij} \ .$$

例 5.2 假设某地区有 15% 的家庭没小孩, 20% 的家庭有一个小孩, 35% 的家庭有两个小孩, 30% 的家庭有三个小孩, 且假设每个小孩为男孩或女孩是相互独立且等可能的. 随机选择一个家庭, 用随机变量 X, Y 分别表示该家庭中男孩和女孩的个数, 求 $P(X \geqslant 1)$, $P(Y \leqslant 2)$ 和 $P(X \leqslant Y)$.

解 根据题意有 X, Y 的所有可能取值为 $\{0, 1, 2, 3\}$, 进一步有联合分布列

$$\begin{aligned} P(X = i, Y = j) &= P(\text{选择的家庭有 } i + j \text{ 个小孩, 其中 } i \text{ 个男孩和 } j \text{ 个女孩}) \\ &= P(\text{选择的家庭有 } i + j \text{ 个小孩})P(i \text{ 个男孩和 } j \text{ 个女孩}|\text{选择的家庭有 } i + j \text{ 个小孩}) \\ &= \binom{i+j}{i} \frac{1}{2^{i+j}} P(\text{选择的家庭有 } i + j \text{ 个小孩}) \ , \end{aligned}$$

由此可得联合分布列和边缘分布列为

X \ Y	0	1	2	3	$p_{i\cdot}$
0	0.1500	0.1000	0.0875	0.0375	0.3750
1	0.1000	0.175	0.1125	0	0.3875
2	0.0875	0.1125	0	0	0.2000
3	0.0375	0	0	0	0.0375
$p_{\cdot j}$	0.3750	0.3875	0.2000	0.0375	1

最后得到

$$P(X \geqslant 1) = 0.625, \quad P(Y \leqslant 2) = 0.9625, \quad P(X \leqslant Y) = 0.6625 \ .$$

最后介绍一种常用的多维离散分布: 多项分布, 它本质上是二项分布的推广, 可用于机器学习中多分类问题. 假设试验 E 有 n 种可能的结果 $A_1, A_2, \cdots, A_n$, 每种结果发生的概率 $p_i = P(A_i)$, 则有 $p_1 + p_2 + \cdots + p_n = 1$.

将试验 E 独立重复地进行 m 次, 用 $X_1, X_2, \cdots, X_n$ 分别表示事件 $A_1, A_2, \cdots, A_n$ 发生的次数, 则每个随机变量 X_i 的取值为 $\{0, 1, 2, \cdots, m\}$ 且满足 $X_1 + X_2 + \cdots + X_n = m$, 则随机向量 $(X_1, X_2, \cdots, X_n)$ 服从多项分布, 其严格的定义如下:

定义 5.5 若 n 维随机向量 $(X_1, X_2, \cdots, X_n)$ 的分布列为

$$P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) = \binom{m}{k_1, k_2, \cdots, k_n} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} \ ,$$

其中 $k_1, k_2, \cdots, k_n$ 是非负的整数且满足 $k_1 + k_2 + \cdots + k_n = m$, 则称随机向量 $(X_1, X_2, \cdots, X_n)$ 服从参数为 $m, p_1, p_2, \cdots, p_n$ 的 **多项分布** (multinomial distribution), 记为 $(X_1, X_2, \cdots, X_n) \sim M(m, p_1, p_2, \cdots, p_n)$.

很容易验证 $P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) \geqslant 0$ 以及

$$\begin{aligned}&\sum_{k_i\geqslant 0, k_1+k_2+\cdots+k_n=m} P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n)\\ &= \sum_{k_i\geqslant 0, k_1+k_2+\cdots+k_n=m} \binom{m}{k_1, k_2, \cdots, k_n} p_1^{k_1} p_2^{k_2} \cdots p_n^{k_n} = (p_1 + p_2 + \cdots + p_n)^m = 1 \ .\end{aligned}$$

当 $n = 2$ 时多项分布简化为二项分布.

引理 5.1 若多维随机向量 $(X_1, X_2, \cdots, X_n) \sim M(m, p_1, p_2, \cdots, p_n)$, 则每个随机变量 X_i 的边缘分布是二项分布 $B(m, p_i)$.

根据 X_i 的实际含义, 考虑事件 A_i 发生或不发生的伯努利试验, 则有 $X_i \sim (m, p_i)$. 另一种方法是通过多项分布的定义直接计算, 我们将其作为一个作业题.

5.3 二维连续型随机向量

定义 5.6 设二维随机向量 (X, Y) 的分布函数为 $F(x, y)$, 如果存在二元非负可积函数 $f(x, y)$, 使得对任意实数 x 和 y 有

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv,$$

则称 (X, Y) 为 **二维连续型随机向量**, 称 $f(x, y)$ 为二维随机向量 (X, Y) 的 **密度函数**, 或称随机变量 X 和 Y 的 **联合密度函数**.

联合密度函数 $f(x, y)$ 满足如下性质:

1) 非负性: 对任意实数 x 和 y 有 $f(x, y) \geqslant 0$.

2) 规范性: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$.
   任何满足上面两条性质的二元函数 $f(x, y)$ 可以成为某随机向量 (X, Y) 的联合密度函数.

3) 若 G 为平面上的一个区域, 则点 (X, Y) 落入 G 的概率为

$$P((X, Y) \in G) = \iint_{(x,y)\in G} f(x, y) dx dy \ ,$$

在几何上可以看作是以 G 为底面, $z = f(x, y)$ 为顶面的柱体体积, 如图 5.3(a) 所示.

4) 若密度函数 $f(x,y)$ 在 (x,y) 连续, 则联合分布函数 $F(x,y)$ 和密度函数 $f(x,y)$ 满足

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}\ .$$

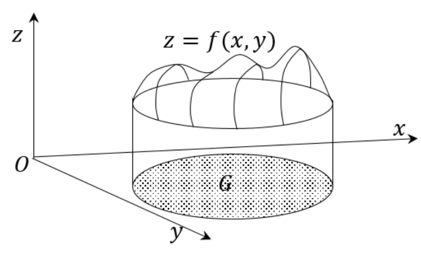
根据此性质、并利用多元泰勒展开式有

$$\begin{aligned}&\lim_{\substack{\Delta x\to 0^+\\ \Delta y\to 0^+}}\frac{P(x<X\leqslant x+\Delta x,y<Y\leqslant y+\Delta y)}{\Delta x\Delta y}\\ =&\lim_{\substack{\Delta x\to 0^+\\ \Delta y\to 0^+}}\frac{F(x+\Delta x,y+\Delta y)-F(x+\Delta x,y)-F(x,y+\Delta y)+F(x,y)}{\Delta x\Delta y}\\ =&\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)\ ,\end{aligned}$$

由此可知

$$P(x<X\leqslant x+\Delta x,y<Y\leqslant y+\Delta y)\approx f(x,y)\Delta x\Delta y\ ,$$

概率 $f(x,y)$ 的值反映了二维随机向量 (X,Y) 落入 (x,y) 邻域内概率的大小.



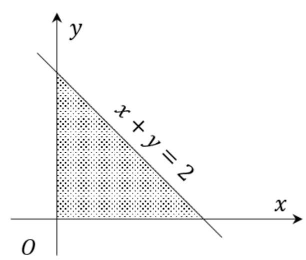


图 5.3 二维密度函数的几何意义和例 5.3 的积分区域

根据 (X,Y) 的联合密度函数 $f(x,y)$, 还可以研究每个随机变量 X 和 Y 的密度函数 $f_X(x)$ 和 $f_Y(y)$. 首先考虑随机变量 X 的边缘分布

$$\begin{aligned}F_X(x)&=P(X\leqslant x)=P(X\leqslant x,Y<\infty)=F(x,+\infty)\\ &=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(t,y)dtdy=\int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty}f(t,y)dy\right)dt\ ,\end{aligned}$$

对上式两边求导得到 X 的边缘概率密度

$$f_X(x)=F_X'(x)=\int_{-\infty}^{+\infty}f(x,y)dy\ .$$

同理分析随机变量 Y 的边缘分布, 于是得到边缘概率密度的严格定义.

定义 5.7 设二维随机向量 (X,Y) 的联合密度函数为 $f(x,y)$, 则随机变量 X 和 Y 的 **边缘密度函数** 分别为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy \qquad \text{和} \qquad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx\ .$$

例 5.3 设二维随机变量 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}ce^{-(3x+4y)} & x>0,y>0\\ 0 & \text{其它}\ .\end{cases}$$

求: 1) 常数 c; 2) 联合分布函数 $F(x,y)$; 3) X 和 Y 的边缘概率密度; 4) 概率 $P(X+Y\leqslant 2)$.

解 根据密度函数的性质可知

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}ce^{-(3x+4y)}dxdy=\frac{c}{12},$$

求解出 $c=12$. 当 $x>0$ 和 $y>0$ 时有

$$F(x,y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=\int_0^x\int_0^y 12e^{-(3x+4y)}dxdy=(1-e^{-3x})(1-e^{-4y})\ ,$$

进一步根据边缘概率密度的定义有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_0^{+\infty}12e^{-(3x+4y)}dy=3e^{-3x}\ ,$$

同理可得 $f_Y(y)=4e^{-4y}$. 最后计算概率 $P(X+Y\leqslant 2)$, 其积分区域如图 5.3(b) 所示, 有

$$P(X+Y\leqslant 2)=12\int_0^2 dx\int_0^{2-x}e^{-(3x+4y)}dy=3\int_0^2 e^{-3x}(1-e^{-8+4x})dx=1-4e^{-6}+3e^{-8}\ .$$

5.3.1 常用二维连续分布

下面介绍两种常用的二维连续分布: 均匀分布和正太分布.

定义 5.8 设 G 为平面上一个有界的区域, 其面积为 A_G, 若二维随机向量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}1/A_G & (x,y)\in G\\ 0 & (x,y)\notin G,\end{cases}$$

则称 (X,Y) 服从区域 G 上的 **二维均匀分布**.

二维均匀分布在区域 G 上每一点等可能发生, 本质上就是 (平面) 几何概型的随机向量描述. 这里以圆的均匀分布为例, 可类似考虑三角形、椭圆等平面上一个有界区域的均匀分布.

例 5.4 在一个以坐标原点为中心、半径为 R 的圆内等可能随机投点. 用随机向量 (X,Y) 分别表示落点的横坐标和纵坐标, 求: 随机向量 (X,Y) 的联合密度函数, 边缘密度函数, 以及 (X,Y) 落入 $X^2+Y^2 \leqslant r^2$ $(0<r\leqslant R)$ 的概率.

解 很容易得到圆的面积为 πR^2, 由此可知随机向量 (X,Y) 的联合密度函数

$$f(x,y)=\begin{cases}1/\pi R^2 & x^2+y^2\leqslant R^2\\ 0 & x^2+y^2>R^2\end{cases}.$$

对于随机变量 X 的边缘密度函数, 当 $x^2\leqslant R^2$ 时有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{x^2+y^2\leqslant R^2}\frac{1}{\pi R^2}dy=\int_{-\sqrt{R^2-x^2}}^{+\sqrt{R^2-x^2}}\frac{1}{\pi R^2}dy=\frac{2}{\pi R^2}\sqrt{R^2-x^2}\ ,$$

同理可得随机变量 Y 的边缘密度函数. 最后所求概率

$$P(X^2+Y^2\leqslant r^2)=\iint\limits_{x^2+y^2\leqslant r^2}\frac{1}{\pi R^2}dxdy=\frac{r^2}{R^2}\ .$$

二维连续分布中最重要的是二维正太分布, 其定义如下:

定义 5.9 对任意实数 x,y, 若随机向量 (X,Y) 的密度函数为

$$f(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)\ ,$$

其中常数 $\mu_x,\mu_y\in(-\infty,+\infty)$, $\sigma_x,\sigma_y\in(0,+\infty)$ 以及 $\rho\in(-1,1)$, 则称 (X,Y) 服从 **二维正太分布**, 记 $(X,Y)\sim\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$.

下面研究二维正态分布的性质:

定理 5.1 设二维随机向量 (X,Y) 服从正态分布 $\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$, 则有随机变量 X 和 Y 的边缘分布分别为 $X\sim\mathcal{N}(\mu_x,\sigma_x^2)$ 和 $Y\sim\mathcal{N}(\mu_y,\sigma_y^2)$.

证明 这里将证明随机变量 X 的边缘密度函数, 可同理证明 Y 的边缘密度函数. 首先将二维随机向量 (X,Y) 的联合密度函数 $f(x,y)$ 分解为

$$f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)\times\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\left(-\frac{(y-\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x)^2}{2\sigma_y^2(1-\rho^2)}\right)\ . \tag{5.1}$$

因此联合密度函数等于两个一维正太分布 $\mathcal{N}(\mu_x,\sigma_x)$ 和 $\mathcal{N}(\mu_y+\rho\sigma_y(x-\mu_x)/\sigma_x,\sigma_y^2(1-\rho^2))$ 的密度函数的乘积. 给定 $x,\mu_x\in(-\infty,+\infty),\sigma_x>0,\rho\in(-1,1)$ 则有

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\Big(-\frac{(y-\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x)^2}{2\sigma_y^2(1-\rho^2)}\Big)dy=1\ ,$$

于是得到

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\Big(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\Big)\ .$$

由此完成证明.

定理 5.1 说明正太分布的边缘分布还是正太分布, 并给出了二维正太分布前四个参数的意义, 即随机变量 X 和 Y 的期望和方差, 第五个参数反应了两个随机变量的密切程度, 我们将在后面介绍.

二维联合分布可以唯一确定它们的边缘分布, 但反之不成立, 即使知道两个随机变量的边缘分布, 也不足以决定联合分布. 例如, 两个边缘分布为 $\mathcal{N}(\mu_x,\sigma_x)$ 和 $\mathcal{N}(\mu_y,\sigma_y)$, 因为不能确定 ρ 的值而不能确定它们的联合分布. 基于 (5.1), 我们还可以验证二维正太分布的规范性

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=1\ ,$$

以及二维正太分布的密度函数本质是两个 (一维) 正太分布的密度函数的乘积.

5.4 随机变量的独立性

前面第二章介绍了随机事件的独立性, 即独立的的随机事件 A 和 B 满足 $P(AB)=P(A)P(B)$. 本节介绍概率统计中另一个重要的概念: 随机变量的独立性. 考虑两个随机变量, 若一个随机变量的取值对另一个随机变量没有什么影响, 则称两个随机变量相互独立. 下面给出严格的数学定义:

定义 5.10 设二维随机向量 (X,Y) 的联合分布函数为 $F(x,y)$, 以及 X 和 Y 的边缘分布函数分别为 $F_X(x)$ 和 $F_Y(y)$, 若对任意的实数 x 和 y 有

$$F(x,y)=F_X(x)F_Y(y)\ ,$$

则称随机变量 X 与 Y **相互独立**.

根据上面的定义可知, 随机变量 X 与 Y 相互独立等价于随机事件 $\{X\leqslant x\}$ 和 $\{Y\leqslant y\}$ 对任意实数 x 和 y 都相互独立; 容易发现常数 c 与任意随机变量相互独立.

对于离散型随机向量, 可以考虑通过分布列来刻画它的统计规律, 关于独立性有

定理 5.2 设二维离散型随机向量 (X,Y) 的分布列为 $p_{ij}=P(X=x_i,Y=y_j)$ $(i,j=1,2,\cdots)$, 以及 X 和 Y 的边缘分布列为 $p_{i\cdot}=P(X=x_i)$ 和 $p_{\cdot j}=P(Y=y_j)$, 则随机变量 X 和 Y 相互独立的充要条件是 $p_{ij}=p_{i\cdot}p_{\cdot j}$.

证明 首先证明必要性, 根据定义 5.10 分布函数的独立性有

$$\begin{aligned} p_{i,j} &= F(x_i, y_j) - F(x_{i-1}, y_j) - F(x_i, y_{j-1}) + F(x_{i-1}, y_{j-1}) \\ &= F_X(x_i)F_Y(y_j) - F_X(x_{i-1})F_Y(y_j) - F_X(x_i)F_Y(y_{j-1}) + F_X(x_{i-1})F_Y(y_{j-1}) \\ &= (F_X(x_i) - F_X(x_{i-1}))F_Y(y_j) - (F_X(x_i) - F_X(x_{i-1}))F_Y(y_{j-1}) \\ &= p_{i\cdot}F_Y(y_j) - p_{i\cdot}F_Y(y_{j-1}) = p_{i\cdot}p_{\cdot j} \ . \end{aligned}$$

其次证明充分性, 根据 $p_{ij} = p_{i\cdot}p_{\cdot j}$ $(i, j = 1, 2, \cdots)$ 有

$$F(x_m, y_n) = \sum_{i\leqslant m}\sum_{j\leqslant n} p_{ij} = \sum_{i\leqslant m}\sum_{j\leqslant m} p_{i\cdot}p_{\cdot j} = \sum_{i\leqslant m} p_{i\cdot} \times \sum_{j\leqslant n} p_{\cdot j} = F_X(x_m)F_Y(y_n) \ .$$

由此完成证明.

例 5.5 设离散型随机变量 X 和 Y 相互独立且它们的取值均为 $\{1, 2, 3\}$, 已知 $P(Y = 1) = 1/3$, $P(X = 1, Y = 1) = P(X = 2, Y = 1) = 1/8$ 和 $P(X = 1, Y = 3) = 1/16$, 求 X 和 Y 的联合分布列和边缘分布列.

解 根据边缘分布列的定义有

$$P(X = 3, Y = 1) = P(Y = 1) - P(X = 1, Y = 1) - P(X = 2, Y = 1) = 1/12 \ ,$$

再根据定理 5.2 有 $P(X = 1) = P(X = 2) = 3/8$ 和 $P(X = 3) = 1/4$, 同理计算其它概率, 最后得到的分布列为

X \ Y	1	2	3	$p_{i\cdot}$
1	1/8	3/16	1/16	3/8
2	1/8	3/16	1/16	3/8
3	1/12	1/8	1/24	1/4
$p_{\cdot j}$	1/3	1/2	1/6	

对于连续型随机向量, 一般可以通过密度函数来进行刻画, 关于独立性有

定理 5.3 设二维随机向量 (X, Y) 的联合密度函数为 $f(x, y)$, 及 X 和 Y 的边缘密度函数分别为 $f_X(x)$ 和 $f_Y(y)$, 则随机变量 X 和 Y 相互独立的充要条件是 $f(x, y) = f_X(x)f_Y(y)$.

证明 首先证明必要性: 若二维连续随机变量满足 $F(x, y) = F_X(x)F_Y(y)$, 则有

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)dudv = \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)dv \ ,$$

对上式两边同时求偏导有

$$f(x,y)=f_X(x)f_Y(y)\ .$$

其次证明充分性: 若 $f(x,y)=f_X(x)f_Y(y)$, 则有

$$\begin{aligned}F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv=\int_{-\infty}^{x}\int_{-\infty}^{y}f_X(u)f_Y(v)dudv\\ &= \int_{-\infty}^{x}f_X(u)du\int_{-\infty}^{y}f_Y(v)dv=F_X(x)F_Y(y)\ ,\end{aligned}$$

由此完成证明.

下面介绍关于随机变量独立性的一些性质:

性质 5.1 若随机变量 X 和 Y 相互独立, 则对任意给定的集合 $A,B\subseteq\mathbb{R}$, 事件 $\{X\in A\}$ 和事件 $\{Y\in B\}$ 相互独立.

证明 该引理对离散型和连续型随机变量均成立, 这里我们详细证明连续随机变量情形. 根据独立性有 $f(x,y)=f_X(x)f_Y(y)$, 由此可得

$$\begin{aligned}&P(X\in A,Y\in B)=\iint_{x\in A,y\in B}f(x,y)dxdy\\ &=\iint_{x\in A,y\in B}f_X(x)f_Y(y)dxdy=\int_{x\in A}f_X(x)dx\int_{y\in B}f_Y(y)dy=P(X\in A)P(Y\in B)\ ,\end{aligned}$$

引理得证.

性质 5.2 设随机变量 X 和 Y 相互独立, 以及 $f(x)$ 和 $g(y)$ 是连续或分段连续的函数, 则有 $f(X)$ 与 $g(Y)$ 相互独立.

该定理对离散型和连续型随机变量均成立, 这里没给出它的证明是因此其超出了本书的范围. 根据此引理, 若随机变量 X 与 Y 相互独立, 则 X^2 与 Y^3 相互独立, 以及 $\sin X$ 与 $\cos Y$ 也相互独立.

下面给出一种判断两个随机变量独立性的简单方法:

性质 5.3 如果存在两个函数 $h(x)$ 和 $g(y)$, 使得 X 和 Y 的联合密度函数 $f(x,y)$ 对任意实数 x 和 y 均有

$$f(x,y)=h(x)g(y)\ ,$$

则随机变量 X 和 Y 相互独立.

证明 不妨假设

$$C_1=\int_{-\infty}^{+\infty}h(x)dx\qquad 和\qquad C_2=\int_{-\infty}^{+\infty}g(y)dy\ .$$

根据密度函数的规范性有

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) = \int_{-\infty}^{+\infty} h(x) dx \int_{-\infty}^{+\infty} g(y) dy = C_1 C_2 \ .$$

随机变量 X 和 Y 的边缘分布分别为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = h(x) C_2 \quad \text{和} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = C_1 g(y) \ .$$

于是得到

$$f_X(x) f_Y(y) = C_1 C_2 h(x) g(y) = h(x) g(y) = f(x, y) \ ,$$

由此完成证明.

定理 5.4 设二维随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则 X 与 Y 独立的充要条件为 $\rho = 0$.

证明 首先证明必要性. 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则 X 和 Y 的边缘分布分别为 $\mathcal{N}(\mu_x, \sigma_x^2)$ 和 $\mathcal{N}(\mu_y, \sigma_y^2)$. 当 $\rho = 0$ 时, 根据二维正太分布的定义有

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} \right) = f_X(x) f_Y(y) \ .$$

其次证明充分性. 若 X 与 Y 相互独立, 则对任意实数 x 和 y 均有 $f(x, y) = f_X(x) f_Y(y)$ 成立, 即

$$\begin{aligned} &\frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} \right) \\ &\quad = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y) \right] \right) \ , \end{aligned}$$

取 $x = \mu_x$ 和 $y = \mu_y$ 求解出 $\rho = 0$, 由此完成证明。

例 5.6 设二维随机向量的密度函数

$$f(x, y) = \begin{cases} cxe^{-y} & 0 < x < y < +\infty \\ 0 & \text{其它}, \end{cases}$$

问随机变量 X 与 Y 是否相互独立?

解 如图 5.4(a) 所示, 利用密度函数的规范性和分部积分有

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = c \int_0^{+\infty} dy \int_0^y x e^{-y} dx = c.$$

当 $x>0$ 时随机变量 X 的边缘概率密度为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_x^{+\infty}xe^{-y}dy=xe^{-x}.$$

同理当 $y>0$ 时随机变量 Y 的边缘概率密度为

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\int_0^{y}xe^{-y}dx=\frac{1}{2}y^2e^{-y}.$$

因为 $f(x,y)\neq f_X(x)f_Y(y)$ 可得随机变量 X 与 Y 不独立.

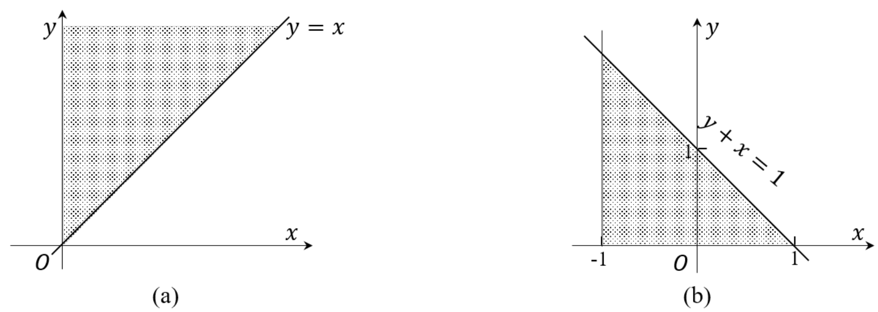


图 5.4 例 5.6 和 5.7 的积分区域

例 5.7 设随机变量 X 与 Y 相互独立, 且 X 服从 $[-1,1]$ 均匀分布, Y 服从参数为 $\lambda=2$ 的指数分布, 求 $P(X+Y\leqslant 1)$.

解 首先有随机变量 X 与 Y 的边缘概率密度分别为

$$f_X(x)=\begin{cases}\frac{1}{2} & x\in[-1,1]\\ 0 & \text{其它}\end{cases}\qquad \text{和}\qquad f_Y(y)=\begin{cases}2e^{-2y} & y\geqslant 0\\ 0 & \text{其它}\end{cases}$$

根据独立性可得随机变量 X 与 Y 的联合概率密度

$$f(x,y)=\begin{cases}e^{-2y} & -1\leqslant x\leqslant 1, y\geqslant 0\\ 0 & \text{其它}\ .\end{cases}$$

所求积分区域如图 5.4(b) 所示, 最后得到

$$P(X+Y\leqslant 1)=\int_{-1}^{1}dx\int_0^{1-x}e^{-2y}dy=\frac{3}{4}+\frac{1}{4}e^{-4}.$$

5.5 条件分布

前面第二章讨论了随机事件的条件概率, 即在事件 B 发生的条件下事件 A 发生的条件概率 $P(A|B) = P(AB)/P(B)$. 可同理考虑随机变量的条件分布, 在给定随机变量 Y 具体取值的条件下, 考虑随机变量 X 的概率分布. 下面分离散和连续两种情形进行讨论.

5.5.1 离散型随机变量的条件概率

定义 5.11 设离散型随机变量 (X,Y) 的分布列为 $p_{ij} = P(X = x_i, Y = y_j)$ $(i, j = 1, 2, \cdots)$, 对于给定的边缘概率 $p_{\cdot j} = P(Y = y_j) = \sum_{i=1}^{+\infty} p_{ij} > 0$,

$$P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}} \qquad (i = 1, 2, \cdots)$$

称为在 $Y = y_j$ 条件下随机变量 X 的 **条件分布列** (conditional probability distribution). 可以类似定义在 $X = x_i$ 条件下随机变量 Y 的条件分布列.

条件分布本质上也是一种概率分布, 具有分布的性质. 例如,

- **非负性**: 对任意整数 $i \geqslant 1$ 有 $P(X = x_i|Y = y_j) \geqslant 0$;
- **规范性**:
$$\sum_{i=1}^{\infty} P(X = x_i|Y = y_j) = \sum_{i=1}^{+\infty} \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \sum_{i=1}^{+\infty} \frac{p_{ij}}{p_{\cdot j}} = 1 \ .$$
- 若离散随机变量 X 和 Y 相互独立, 则有

$$P(X = x_i|Y = y_j) = P(X = x_i) \quad 和 \quad P(Y = y_j|X = x_i) = P(Y = y_j) \ .$$

在后面的章节中, 若出现条件概率 $P(X = x_i|Y = y_j)$, 一般默认概率 $P(Y = y_j) > 0$. 条件分布列也可以通过下面的表格给出:

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$P(X = x_i\|Y = y_j)$	$p_{1j}/p_{\cdot j}$	$p_{2j}/p_{\cdot j}$	$\cdots$	$p_{nj}/p_{\cdot j}$	$\cdots$

例 5.8 一个选手击中目标的概率为 p, 射中两次目标为止, 用 X 表示首次击中目标所进行的射击次数, 用 Y 表示第二次射中目标所进行的射击次数, 求 X 和 Y 的联合分布和条件分布.

解 随机变量 $X = m$ 表示首次击中目标射击了 m 次, $Y = n$ 表示第二次次击中目标射击了 n 次, 则 X 和 Y 的联合分布列为

$$P\{X = m, Y = n\} = f(x, y) = \begin{cases} p^2(1-p)^{n-2} & 1 \leqslant m < n < \infty \\ 0 & \text{其它} \ . \end{cases}$$

于是得到随机变量 X 的边缘分布列

$$P\{X=m\}=\sum_{n=m+1}^{\infty}P\{X=m,Y=n\}=\sum_{n=m+1}^{\infty}p^2(1-p)^{n-2}=p(1-p)^{m-1}\ ,$$

以及随机变量 Y 的边缘分布列

$$P\{Y=n\}=\sum_{m=1}^{n-1}P\{X=m,Y=n\}=\sum_{m=1}^{n-1}p^2(1-p)^{n-2}=(n-1)p^2(1-p)^{n-2}\qquad(n\geqslant 2)\ .$$

当 $n\geqslant 2$ 时, 随机变量 X 在 $Y=n$ 条件下的分布列

$$P\{X=m|Y=n\}=\frac{P\{X=m,Y=n\}}{P\{Y=n\}}=\frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}}=\frac{1}{n-1}\qquad(1\leqslant m\leqslant n-1)\ .$$

当 $m\geqslant 1$ 时, 随机变量 Y 在 $X=m$ 条件下的分布列

$$P\{Y=n|X=m\}=\frac{P\{X=m,Y=n\}}{P\{X=m\}}=\frac{p^2(1-p)^{n-2}}{p(1-p)^{m-1}}=p(1-p)^{n-m-1}\qquad(n>m)\ .$$

5.5.2 连续型随机向量的条件分布

连续型随机向量 (X,Y) 对任意实数 x,y 都有 $P(X=x)=0$ 和 $P(Y=y)=0$, 因此不能用条件概率的公式直接推导连续型随机向量的条件分布, 但可以通过下面的极限方式来考虑.

当 $P(y\leqslant Y\leqslant y+\epsilon)>0$ 时, 利用积分中值定理来求解条件分布函数

$$\begin{aligned}F_{X|Y}(x|y)&=\lim_{\epsilon\to 0^+}P\{X\leqslant x|y\leqslant Y\leqslant y+\epsilon\}=\lim_{\epsilon\to 0^+}\frac{P\{X\leqslant x,y\leqslant Y\leqslant y+\epsilon\}}{P\{y\leqslant Y\leqslant y+\epsilon\}}\\&=\lim_{\epsilon\to 0^+}\frac{\int_{-\infty}^{x}\int_{y}^{y+\epsilon}f(u,v)dudv}{\int_{y}^{y+\epsilon}f_Y(u)dv}=\lim_{\epsilon\to 0^+}\frac{\epsilon\int_{-\infty}^{x}f(u,y+\theta_1\epsilon)du}{\epsilon f_Y(y+\theta_2\epsilon)}\qquad\theta_1,\theta_2\in(0,1)\\&=\frac{\int_{-\infty}^{x}f(u,y)du}{f_Y(y)}=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du\ ,\end{aligned}$$

进一步得到条件密度函数 $f_{X|Y}(x|y)=f(x,y)/f_Y(y)$. 在后面的章节中, 若出现条件密度函数 $f_{X|Y}(x|y)$ (或 $f_{Y|X}(y|x)$), 一般都默认 $f_Y(y)>0$ (或 $f_X(x)>0$). 下面给详细的定义:

定义 5.12 设连续型随机变量 (X,Y) 的联合概率密度为 $f(x,y)$, 以及 X 和 Y 的边缘概率密度分别为 $f_X(x)$ 和 $f_Y(y)$. 对任意给定的 $f_Y(y)>0$, 称 $f(x,y)/f_Y(y)$ 为在 $Y=y$ 条件下随机变量 X 的 **条件密度函数** (conditional probability density function), 记为

$$f_{X|Y}(x|y)=f(x,y)/f_Y(y)\ ,$$

以及在 $Y=y$ 条件下 X 的 **条件分布函数** (conditional cumulative distribution function) 为

$$F_{X|Y}(x|y)=P\{X\leqslant x|Y=y\}=\int_{-\infty}^{x}f_{X|Y}(u|y)du\ .$$

对任意给定的 $f_X(x)>0$, 可类似定义在 $X=x$ 条件下 Y 的条件密度函数和条件分布函数分别为

$$f_{Y|X}(y|x)=f(x,y)/f_X(x)\quad \text{和}\quad F_{Y|X}(y|x)=\int_{-\infty}^{y}f_{Y|X}(v|x)dv\ .$$

条件密度函数本质上是密度函数, 具有以下性质:

- **非负性**: 对任意实数 x,y 有 $f_{Y|X}(y|x)\geqslant 0$.
- **规范性**: 对任意实数 y: $f_Y(y)>0$ 有

$$\int_{-\infty}^{+\infty}f_{X|Y}(x|y)dx=\int_{-\infty}^{+\infty}\frac{f(x,y)}{f_Y(y)}dx=\frac{1}{f_Y(y)}\int_{-\infty}^{+\infty}f(x,y)dx=1\ .$$

- **乘法公式**: $f(x,y)=f_X(x)f_{Y|X}(y|x)=f_Y(y)f_{X|Y}(x|y)$.
- 若随机变量 X 和 Y 相互独立, 则有

$$f_{Y|X}(y|x)=f_Y(y)\quad \text{和}\quad f_{X|Y}(x|y)=f_X(x)\ .$$

根据条件概率的乘法公式有

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}=\frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty}f(x,y)dx}=\frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty}f(y|x)f_X(x)dx}\ ,$$

可以将其看作 **密度函数的贝叶斯公式**. 目前可能有三种途径来构造二维随机向量的联合分布函数:

i) 根据实际问题或实际数据归纳为 $f(x,y)$;

ii) 根据随机变量的独立性有 $f(x,y)=f_X(x)f_Y(y)$;

iii) 根据乘法公式 $f(x,y)=f_X(x)f_{Y|X}(y|x)$.

关于二维正太分布有

定理 5.5 若随机向量 $(X,Y)\sim\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$, 则在 $Y=y$ 的条件下随机变量 X 服从正太分布 $\mathcal{N}(\mu_x-\rho\sigma_x(y-\mu_y)/\sigma_y,(1-\rho^2)\sigma_x^2)$, 以及在 $X=x$ 的条件下随机变量 Y 服从正太分布 $\mathcal{N}(\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x,(1-\rho^2)\sigma_y^2)$.

证明 若随机向量 $(X,Y)\sim\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$, 则随机变量 X 的边缘分布为 $\mathcal{N}(\mu_x,\sigma_x^2)$, 即

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\Big(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\Big)\ ,$$

此外可以将二维随机向量 (X,Y) 的联合密度函数 $f(x,y)$ 分解为

$$f(x,y)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\Big(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\Big)\times\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\Big(-\frac{(y-\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x)^2}{2\sigma_y^2(1-\rho^2)}\Big)\ .$$

根据乘法公式 $f(x,y)=f_X(x)f_{Y|X}(y|x)$ 可得

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}}\exp\Big(-\frac{(y-\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x)^2}{2\sigma_y^2(1-\rho^2)}\Big)\ ,$$

即正太分布 $\mathcal{N}(\mu_y-\rho\sigma_y(x-\mu_x)/\sigma_x,(1-\rho^2)\sigma_y^2)$. 同理可证在 $Y=y$ 的条件下随机变量 X 的条件分布为 $\mathcal{N}(\mu_x-\rho\sigma_x(y-\mu_y)/\sigma_y,(1-\rho^2)\sigma_x^2)$.

例 5.9 设二维随机向量 (X,Y) 的密度函数

$$f(x,y)=\begin{cases} e^{-x/y}e^{-y}/y & x>0,y>0\\ 0 & \text{其它,}\end{cases}$$

求 $P(X>1|Y=y)$.

解 积分区域如图 5.5(a) 所示, 求解随机变量 Y 的边缘分布为

$$f_Y(y)=\int_0^{+\infty}e^{-x/y}e^{-y}/ydx=e^{-y}\left[-e^{-x/y}\right]_0^{+\infty}=e^{-y}\qquad(y>0)\ .$$

进而得到在 $Y=y$ 条件下 X 的条件概率密度为

$$f_{X|Y}(x|y)=e^{-x/y}/y\ .$$

最后求解得到

$$P(X>1|Y=y)=\int_1^{\infty}e^{-x/y}/ydx=-\left[e^{-x/y}\right]_1^{\infty}=e^{-1/y}\ .$$

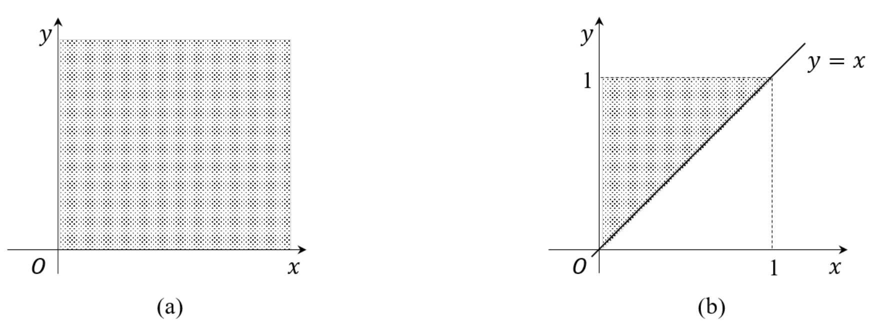


图 5.5 例 5.9 和 5.10 的积分区域

例 5.10 设随机变量 $X \sim U(0,1)$, 在观察到 $X=x$ 的条件下随机变量 $Y \sim U(x,1)$, 求随机变量 Y 的概率密度.

解 随机变量 $X \sim U(0,1)$, 在随机变量 $X=x$ 的条件下 $Y \sim U(x,1)$, 于是当 $x>0$ 时有

$$f_{Y|X}(y|x) = 1/(1-x)\ .$$

根据条件概率乘积公式有

$$f(x,y) = f_X(x)f_{Y|X}(y|x) = \begin{cases} 1/(1-x) & 0<x<y<1, \\ 0 & \text{其它}. \end{cases}$$

积分区域如图 5.5(b) 所示, 当 $y>0$ 时随机变量 Y 的边缘分布

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^y \frac{1}{1-x}dx = -\ln(1-y)\ .$$

5.6 多维随机变量函数的分布

已知二维随机向量 (X,Y) 的概率分布, 如何求解随机变量函数 $Z=g(X,Y)$ 的概率分布. 下面分离散型和连续型随机变量两种情况进行讨论.

5.6.1 二维离散型随机向量函数

已知二维离散型随机向量 (X,Y) 的联合分布列, 求函数 $Z=g(X,Y)$ 的分布列相对简单. 首先针对 X,Y 的各种取值, 计算随机变量 Z 的值, 然后对相同的 Z 值合并, 对应的概率相加. 下面研究两个相互独立的离散型随机变量之和, 即离散型随机变量的卷积公式:

定理 5.6 设离散型随机变量 X 与 Y 相互独立、且它们的分布列分别为 $a_i = P(X=i)$ 和 $b_j = P(Y=j)$ $(i,j=0,1,\cdots)$, 则随机变量 $Z=X+Y$ 的分布列为

$$P(Z=k) = \sum_{i=0}^{k} a_i b_{k-i}\ .$$

证明 对任意非负整数 i 和 j, 根据独立性可知

$$P(X=i,Y=j) = P(X=i)P(Y=j) = a_i b_j\ .$$

因此随机变量 Z 的分布列为

$$\begin{aligned} P(Z=k) &= P(X+Y=k) \\ &= \sum_{i=0}^{k} P(X=i,Y=k-i) = \sum_{i=0}^{k} P(X=i)P(Y=k-i) = \sum_{i=0}^{k} a_i b_{k-i}\ , \end{aligned}$$

定理得证.

基于定理 5.6, 可以得到一系列推论:

推论 5.1 若随机变量 $X \sim B(n_1, p)$ 和 $Y \sim B(n_2, p)$ 相互独立, 则随机变量

$$Z = X + Y \sim B(n_1 + n_2, p) \ .$$

证明 根据二项分布的定义, 当 $i = 0, 1, \cdots, n_1$ 和 $j = 0, 1, \cdots, n_2$ 有

$$P(X = i) = \binom{n_1}{i} p^i (1-p)^{n_1 - i} \quad \text{和} \quad P(X = j) = \binom{n_2}{j} p^j (1-p)^{n_2 - j} \ .$$

对 $k = 0, 1, \cdots, n_1 + n_2$, 根据定理 5.6 有

$$\begin{aligned} P[Z = k] &= \sum_{i=0}^{k} P[X = i] P[Y = k - i] = \sum_{i=0}^{k} \binom{n_1}{i} p^i (1-p)^{n_1 - i} \binom{n_2}{k-i} p^{k-i} (1-p)^{n_2 - (k-i)} \\ &= p^k (1-p)^{n_1 + n_2 - k} \sum_{i=0}^{k} \binom{n_1}{i} \binom{n_2}{k-i} = \binom{n_1 + n_2}{k} p^k (1-p)^{n_1 + n_2 - k} \ . \end{aligned}$$

利用归纳法和推论 5.1, 若相互独立的随机变量 $X_i \sim \mathrm{Ber}(p) = B(1, p)$ $(i \in [n])$, 则随机变量

$$X = X_1 + X_2 + \cdots + X_n \ \sim \ B(n, p) \ .$$

即随机变量 $X \sim B(n, p)$ 可以看作 n 个相互独立的服从参数为 p 的伯努利分布随机变量之和.

推论 5.2 若随机变量 $X \sim P(\lambda_1)$ 和 $Y \sim P(\lambda_2)$ 相互独立, 则随机变量

$$Z = X + Y \sim P(\lambda_1 + \lambda_2) \ .$$

证明 根据泊松分布的定义, 对任意非负整数 i 和 j 有

$$P(X = i) = \lambda_1^i e^{-\lambda_1} / i! \quad \text{和} \quad P(Y = j) = \lambda_2^j e^{-\lambda_2} / j! \ .$$

对任意非负整数 k, 根据定理 5.6 有

$$\begin{aligned} P(Z = k) &= \sum_{i=0}^{k} P(X = i, Y = k - i) = \sum_{i=0}^{k} P(X = i) P(Y = k - i) \\ &= \sum_{i=0}^{k} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-(\lambda_1 + \lambda_2)} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k \ . \end{aligned}$$

5.6.2 二维连续型随机向量函数

已知二维连续型随机向量 (X,Y) 的联合概率密度为 $f(x,y)$, 求随机变量 $Z=g(X,Y)$ 的概率密度, 一般先求解分布函数

$$F_Z(z)=P(Z\leqslant z)=P(g(x,y)\leqslant z)=\iint\limits_{g(x,y)\leqslant z} f(x,y)dxdy\ ,$$

再对分布函数 $F_Z(z)$ 求导得到密度函数

$$f_Z(z)=F_Z'(z)\ .$$

例 5.11　设服从标准正态分布的两个随机变量 X 和 Y 相互独立, 求随机变量 $Z_1=\sqrt{X^2+Y^2}$ 和 $Z_2=X^2+Y^2$ 的密度函数.

解　根据独立性有随机变量 X 和 Y 的联合密度函数

$$f(x,y)=f_X(x)f_Y(y)=e^{-(x^2+y^2)/2}/2\pi\qquad(x,y\in\mathbb{R})\ .$$

当 $z_1\leqslant 0$ 时, 根据 $Z_1=\sqrt{X^2+Y^2}$ 很显然有分布函数 $F_{Z_1}(z_1)=0$; 当 $z_1>0$ 时有

$$F_{Z_1}(z_1)=P(Z_1\leqslant z_1)=P\left(\sqrt{X^2+Y^2}\leqslant z_1\right)=\iint\limits_{X^2+Y^2\leqslant z_1^2} e^{-(x^2+y^2)/2}/2\pi dxdy\ ,$$

利用极坐标积分变换 $x=r\cos\theta$ 和 $y=r\sin\theta$ 有

$$F_{Z_1}(z_1)=\int_0^{2\pi}\int_0^{z_1}\frac{r}{2\pi}e^{-r^2/2}d\theta dr=\int_0^{z_1}re^{-r^2/2}dr=1-e^{-z_1^2/2}\ .$$

由此得到随机变量 Z_1 的密度函数为

$$f_{Z_1}(z_1)=\begin{cases}z_1e^{-z_1^2/2} & z_1>0\\ 0 & z_1\leqslant 0\ .\end{cases}$$

上述分布称为 **瑞利分布** (Rayleigh distribution), 该分布常用于通信等领域. 同理可证随机变量 $Z_2\sim e(1/2)$, 即

$$f_{Z_2}(z_2)=\begin{cases}z_2e^{-z_2/2}/2 & z_2>0\\ 0 & z_2\leqslant 0\ .\end{cases}$$

5.6.2.1 和的分布 $Z=X+Y$

引理 5.2 设二维随机向量 (X,Y) 的联合密度函数为 $f(x,y)$, 则有 $Z=X+Y$ 的密度函数

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx \quad \text{或} \quad f_Z(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy\ .$$

解 首先求解 $Z=X+Y$ 的分布函数

$$F_Z(z)=P(Z\leqslant z)=P(X+Y\leqslant z)=\iint_{x+y\leqslant z}f(x,y)dxdy=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{z-x}f(x,y)dy\ ,$$

这里考虑的积分区域为 $\{(x,y)\colon x+y\leqslant z\}$, 如图 5.6(a) 所示. 利用变量替换 $u=y+x$ 并积分换序有

$$F_Z(z)=\int_{-\infty}^{z}\left(\int_{-\infty}^{+\infty}f(x,u-x)dx\right)du\ ,$$

两边同时对 z 求导数可得

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx\ .$$

同理可证 $f_Z(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy$.

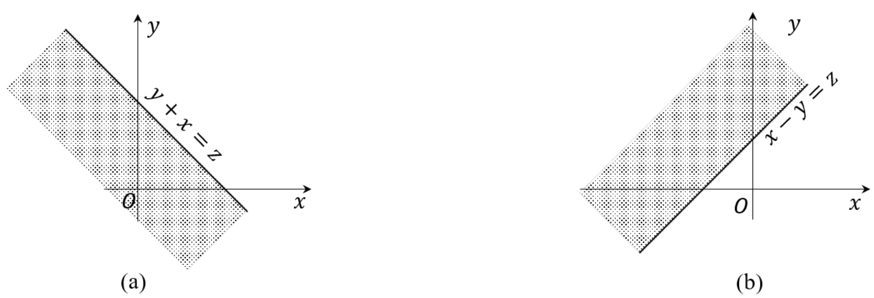


图 5.6 函数 $Z=X+Y$ 和 $Z=X-Y$ 的积分区域

类似考虑随机变量 $Z=X-Y$, 其积分区域 $\{(x,y)\colon x-y\leqslant z\}$ 如图 5.6(b) 所示, 得到 $Z=X-Y$ 的密度函数为

$$f_Z(z)=\int_{-\infty}^{+\infty}f(x,x-z)dx \quad \text{或} \quad f_Z(z)=\int_{-\infty}^{+\infty}f(z+y,y)dy\ .$$

若随机变量 X 和 Y 相互独立, 则有 $f(x,y)=f_X(x)f_Y(y)$. 结合引理 5.2 给出下面著名的定理:

定理 5.7 (卷积公式) 若连续型随机变量 X 与 Y 相互独立, 且它们的密度函数分别为 $f_X(x)$ 和 $f_Y(y)$, 则随机变量 $Z=X+Y$ 的密度函数为

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)dy\ .$$

推论 5.3 若随机变量 $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ 和 $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ 相互独立, 则随机变量

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)\ .$$

根据上面的推论很容易得到 $X - Y \sim \mathcal{N}(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$. 该结论可以推广到 n 个随机变量, 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立、且 $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, 则随机变量

$$Z = X_1 + X_2 + \cdots + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)\ .$$

证明 若随机变量 $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ 和 $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, 则根据正太分布的性质有

$$X' = X - \mu_x \sim \mathcal{N}(0, \sigma_x^2) \qquad \text{和} \qquad Y' = Y - \mu_y \sim \mathcal{N}(0, \sigma_y^2)\ .$$

因此只需证明 $Z = X' + Y' \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$. 根据卷积公式有

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma_1^2} - \frac{(z-x)^2}{2\sigma_2^2}\right) dx \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\big(x - \frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\big)^2 - \frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right) dx \\
&= \frac{\exp\left(-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \times \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\big(x - \frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\big)^2\right) dx \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}\right)\ ,
\end{aligned}$$

最后一个等式成立是因为正太分布的规范性.

例 5.12 设随机变量 $X \sim U(0,1)$ 和 $Y \sim U(0,1)$ 相互独立, 求 $Z = X + Y$ 的概率密度.

解 根据卷积公式有

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx.$$

根据区间 $(0,1)$ 上的均匀分布, 当 $x \in [0,1]$ 时有 $f_X(x) = 1$; 当 $z - x \in [0,1]$ 时有 $f_Y(z-x) = 1$. 由此可得非零的积分区域为 $\{x \in [0,1], z - x \in [0,1]\}$, 如图 5.7(a) 所示. 当 $z \leqslant 0$ 或 $z \geqslant 2$ 时有 $f_Z(z) = 0$; 当 $z \in (0,1)$ 时有

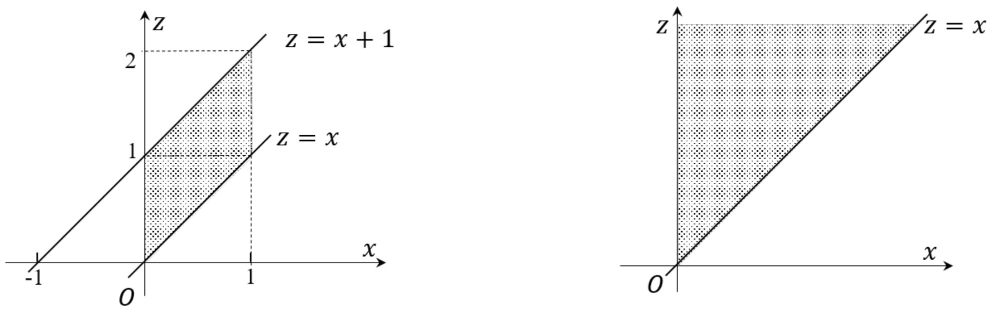
$$f_Z(z) = \int_0^z 1 dz = z\ ;$$

当 $z \in [1,2)$ 时有

$$f_Z(z) = \int_{z-1}^1 dx = 2 - z\ .$$

综上所述, 随机变量 $Z = X + Y$ 的概率密度

$$f_Z(z) = \begin{cases} z & z \in [0,1] \\ 2 - z & z \in [1,2] \\ 0 & \text{其它}\ . \end{cases}$$



(a) (b)

图 5.7 例 5.12 和 5.13 中积分区域示意图

例 5.13 设随机变量 $X \sim e(\lambda)$ 和 $Y \sim e(\lambda)$ 相互独立, 求 $Z = X + Y$ 的概率密度.

解 由卷积公式可得

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx\ .$$

根据指数分布的定义, 当 $x \geqslant 0$ 时有 $f_X(x) = \lambda \exp(-\lambda x)$; 当 $z - x \geqslant 0$ 时有 $f_Y(z - x) = \lambda \exp(-\lambda(z - x))$, 因此积分区域 $\{x \in [0, +\infty), z - x \in [0, +\infty)\}$ 如图 5.7(b) 所示. 当 $z \geqslant 0$ 时有

$$f_Z(z) = \lambda^2 \int_0^z \exp(-\lambda x) \exp(-\lambda(z - x)) dx = \lambda^2 z \exp(-\lambda z)\ .$$

5.6.3 随机变量的乘/除法分布

定理 5.8 设二维随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则随机变量 $Z = XY$ 的概率密度为

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx\ ;$$

随机变量 $Z = Y/X$ 的概率密度为

$$f_{Y/X}(z) = \int_{-\infty}^{+\infty} |x| f(x, xz) dx\ .$$

证明 这里给出随机变量 $Z=Y/X$ 的概率密度详细证明, 同理给出 $Z=XY$ 的概率密度. 首先考虑分布函数

$$\begin{aligned}
F_{Y/X}(z) &= P(Y/X\leqslant z)=\iint\limits_{y/x\leqslant z} f(x,y)dxdy\\
&= \iint\limits_{x<0,y\geqslant zx} f(x,y)dxdy+\iint\limits_{x>0,y\leqslant zx} f(x,y)dxdy\\
&= \int_{-\infty}^{0} dx\int_{zx}^{+\infty} f(x,y)dy+\int_{0}^{+\infty} dx\int_{-\infty}^{xz} f(x,y)dy\ ,
\end{aligned}$$

如图 5.8 所示, 这里考虑积分区域为 $\{(x,y)\colon x>0,y<xz\}\cup\{(x,y)\colon x<0,y>xz\}$. 变量替换 $t=y/x$ 有

$$\begin{aligned}
F_{Y/X}(z) &= \int_{-\infty}^{0} dx\int_{z}^{-\infty} xf(x,tx)dt+\int_{0}^{+\infty} dx\int_{-\infty}^{z} xf(x,tx)dt\\
&= \int_{-\infty}^{0}\int_{-\infty}^{z}(-x)f(x,tx)dtdx+\int_{0}^{+\infty}\int_{-\infty}^{z} xf(x,tx)dtdx\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{z}|x|f(x,tx)dtdx=\int_{-\infty}^{z} dt\int_{-\infty}^{+\infty}|x|f(x,tx)dx\ ,
\end{aligned}$$

对分布函数求导即可完成证明.

(a) $z>0$ (b) $z<0$

图 5.8 随机变量 $Z=Y/X$ 的积分区域

推论 5.4 若标准正太分布的随机变量 X 和 Y 相互独立, 则随机变量 $Z=Y/X$ 服从柯西分布.

证明 根据独立性和定理 5.8, 对任意实数 z 有

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{+\infty}|x|f(x,xz)dx=\int_{-\infty}^{+\infty}|x|f_X(x)f_Y(xz)dx\\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}|x|e^{-x^2(1+z^2)/2}dx=\frac{1}{\pi}\int_{0}^{+\infty}xe^{-x^2(1+z^2)/2}dx
\end{aligned}$$

$$= \frac{1}{\pi}\left[-\frac{e^{-x^2(1+z^2)/2}}{1+z^2}\right]_0^{+\infty} = \frac{1}{\pi(1+z^2)} \ ,$$

推论得证.

5.6.4 最大值和最小值的分布

定理 5.9 设随机变量 $X_1, \cdots, X_n$ 相互独立、且其分布函数分别为 $F_{X_1}(x_1), \cdots, F_{X_n}(x_n)$, 则随机变量 $Y = \max(X_1, X_2, \cdots, X_n)$ 的分布函数为

$$F_Y(y) = F_{X_1}(y)F_{X_2}(y)\cdots F_{X_n}(y) \ ,$$

以及随机变量 $Z = \min(X_1, X_2, \cdots, X_n)$ 的分布函数为

$$F_Z(z) = 1-(1-F_{X_1}(z))(1-F_{X_2}(z))\cdots(1-F_{X_n}(z)) \ .$$

证明 根据独立性, 随机变量 $Y = \max(X_1, X_2, \cdots, X_n)$ 的分布函数为

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) = P(\max(X_1, X_2, \cdots, X_n) \leqslant y) = P(X_1 \leqslant y, X_2 \leqslant y, \cdots, X_n \leqslant y) \\ &= P(X_1 \leqslant y)P(X_2 \leqslant y)\cdots P(X_n \leqslant y) = F_{X_1}(y)F_{X_2}(y)\cdots F_{X_n}(y) \ . \end{aligned}$$

随机变量 $Z = \min(X_1, X_2, \cdots, X_n)$ 的分布函数为

$$\begin{aligned} F_Z(z) &= P(Z \leqslant z) = P(\min(X_1, X_2, \cdots, X_n) \leqslant z) = 1 - P(\min(X_1, X_2, \cdots, X_n) > z) \\ &= 1 - P(X_1 > z, X_2 > z, \cdots X_n > z) = 1 - P(X_1 > z)P(X_2 > z)\cdots P(X_n > z) \\ &= 1-(1-F_{X_1}(z))(1-F_{X_2}(z))\cdots(1-F_{X_n}(z)) \ , \end{aligned}$$

定理得证.

根据定理 5.9 有

推论 5.5 设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量, 其分布函数和密度函数分别为 $F(x)$ 和 $f(x)$, 则随机变量 $Y = \max(X_1, X_2, \cdots, X_n)$ 的分布函数和密度函数分别为

$$F_Y(y) = (F(y))^n \qquad \text{和} \qquad f_Y(y) = n(F(y))^{n-1}f(y) \ ,$$

以及随机变量 $Z = \min(X_1, X_2, \cdots, X_n)$ 的分布函数和密度函数分别为

$$F_Z(z) = 1-(1-F(z))^n \qquad \text{和} \qquad f_Z(z) = n(1-F(z))^{n-1}f(z) \ .$$

例 5.14 假设随机变量 X 与 Y 相互独立, 且有 $X \sim e(\alpha)$ 和 $Y \sim e(\beta)$, 求随机变量 $Z_1 = \max(X, Y)$ 和 $Z_2 = \min(X, Y)$ 的概率密度.

解 根据指数随机变量的定义可知随机变量 X 和 Y 的概率密度为

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & x > 0 \\ 0 & x \leqslant 0 \end{cases} \quad \text{和} \quad f_Y(y) = \begin{cases} \beta e^{-\beta y} & y > 0 \\ 0 & y \leqslant 0 \ . \end{cases}$$

于是得到随机变量 Z_1 的分布函数为

$$F_{Z_1}(z_1) = F_X(z_1)F_Y(z_1) = \int_{-\infty}^{z_1} f_X(t)dt \int_{-\infty}^{z_1} f_Y(t)dt \ .$$

当 $z_1 \leqslant 0$ 时由 $F_{Z_1}(z_1) = 0$; 当 $z_1 > 0$ 时

$$F_{Z_1}(z_1) = \int_0^{z_1} f_X(t)dt \int_0^{z_1} f_Y(t)dt = \int_0^{z_1} \alpha e^{-\alpha t}dt \int_0^{z_1} \beta e^{-\beta y}dt = (1 - e^{-\alpha z_1})(1 - e^{-\beta z_1}) \ .$$

两边对 z_1 求导可得其概率密度为

$$f_{Z_1}(z_1) = \begin{cases} \alpha e^{-\alpha z_1} + \beta e^{-\beta z_1} - (\alpha+\beta)e^{-(\alpha+\beta)z_1} & z_1 > 0 \\ 0 & z_1 \leqslant 0 \ . \end{cases}$$

同理可得随机变量 Z_2 的分布函数和概率密度分别为

$$F_{Z_2}(z_2) = \begin{cases} 1 - e^{-(\alpha+\beta)z_2} & z_2 > 0 \\ 0 & z_2 \leqslant 0 \end{cases} \qquad f_{Z_2}(z_2) = \begin{cases} (\alpha+\beta)e^{-(\alpha+\beta)z_2} & z_2 > 0 \\ 0 & z_2 \leqslant 0 \ . \end{cases}$$

5.6.5 随机变量的联合分布函数

已知随机向量 (X, Y) 的联合概率密度为 $f(x, y)$, 随机变量 U 和 V 是 X 和 Y 的函数, 如何求解 (U, V) 的联合分布. 具体而言, 设

$$\begin{cases} U = u(X, Y) \\ V = v(X, Y) \ , \end{cases}$$

已知随机向量 (X, Y) 的联合概率密度为 $f(x, y)$, 如何求解二维随机向量 (U, V) 的联合分布. 这里二元函数 $u(\cdot, \cdot)$ 和 $v(\cdot, \cdot)$ 具有连续的偏导, 并满足

$$\begin{cases} u = u(x, y) \\ v = v(x, y) \end{cases} \quad \text{存在唯一的反函数} \quad \begin{cases} x = x(u, v) \\ y = y(u, v) \ . \end{cases}$$

定理 5.10 设随机变量 $U=u(X,Y)$ 和 $V=v(X,Y)$ 有连续偏导, 且存在反函数 $X=x(U,V)$ 和 $Y=y(U,V)$. 若 (X,Y) 的联合概率密度为 $f(x,y)$, 则 (U,V) 的联合密度为

$$f_{UV}(u,v)=f_{XY}(x(u,v),y(u,v))|J| \ ,$$

其中 J 为变换的雅可比行列式不为零, 即

$$J=\left|\frac{\partial(x,y)}{\partial(u,v)}\right|=\left|\frac{\partial(u,v)}{\partial(x,y)}\right|^{-1}=\begin{vmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{vmatrix}^{-1} .$$

上述结论可推广到一般的 n 维随机变量.

推论 5.6 设 X 和 Y 是相互独立的标准正太分布随机变量, 则有随机变量 $R=X^2+Y^2$ 与 $\theta=\arctan(Y/X)$ 相互独立, 且有 $R\sim e(1/2)$ 以及 $\theta\sim U(0,2\pi)$.

证明 令 $R=u(x,y)=x^2+y^2$ 和 $\Theta=v(x,y)=\arctan(y/x)$. 于是得到雅可比行列式

$$J=\begin{vmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{vmatrix}^{-1}=\begin{vmatrix}2x & 2y\\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2}\end{vmatrix}^{-1}=\frac{1}{2} .$$

由此可得 R 与 Θ 的联合分布为

$$f_{R\times\Theta}(r,\theta)=f_{X\times Y}(\sqrt{r}\cos\theta,\sqrt{r}\sin\theta)|J|=\frac{1}{4\pi}\exp(-r/2)=\frac{1}{2}\exp(-r/2)\times\frac{1}{2\pi} .$$

由此可以发现 $R\sim e(1/2)$ 和 $\Theta\sim U(0,2\pi)$, 且 R 和 Θ 相互独立. 推论得证.

5.7 多维正太分布

本节将二维随机向量及其分布推广到多维随机向量, 二维与多维随机变量没有本质性的区别, 只是相关的概念和结论的扩展.

定义 5.13 设 $(X_1,X_2,\cdots,X_n)$ 为 n 维随机向量, 对任意实数 $x_1,x_2,\cdots,x_n$, 称函数

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

为 n **维随机向量** $(X_1,X_2,\cdots,X_n)$ **的分布函数**, 或 **随机变量** $X_1,X_2,\cdots,X_n$ **的联合分布函数**. 若存在可积函数 $f(x_1,x_2,\cdots,x_n)$, 使得对任意实数 $x_1,x_2,\cdots,x_n$ 有

$$F(x_1,x_2,\cdots,x_n)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n}f(u_1,u_2,\cdots,u_n)du_1du_2\cdots du_n \ ,$$

则称 $(X_1,X_2,\cdots,X_n)$ 为 **连续型随机向量**, 以及 $f(x_1,x_2,\cdots,x_n)$ 为 n **维联合密度函数**.

类似于二维密度函数, n 维联合密度函数具有以下性质:

- 非负性: 对任意实数 $x_1, x_2, \cdots, x_n$ 有 $f(x_1, x_2, \cdots, x_n) \geqslant 0$.
- 规范性: $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(u_1, u_2, \cdots, u_n)du_1du_2\cdots du_n = 1$.
- 设 G 是 n 维空间的一片区域, 则有

$$P((X_1, X_2, \cdots, X_n) \in G) = \underset{G}{\int\cdots\int} f(u_1, u_2, \cdots, u_n)du_1du_2\cdots du_n \ .$$

- 若 $f(x_1, x_2, \cdots, x_n)$ 在点 $(x_1, x_2, \cdots, x_n)$ 处连续, 则有

$$\frac{\partial F(x_1, x_2, \cdots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n} = f(x_1, x_2, \cdots, x_n) \ .$$

随机向量 $(X_1, X_2, \cdots, X_n)$ 中任意 k 个向量所构成的随机向量 $(k \leqslant n)$, 它的分布函数和密度函数被称为 k **维边缘分布函数** 和 k **维边缘密度函数**. 例如随机向量 $(X_1, X_2, \cdots, X_n)$ 前 k 维随机向量的边缘分布函数和边缘密度函数分布为

$$\begin{aligned} F_{X_1,X_2,\cdots,X_k}(x_1, x_2, \cdots, x_k) &= P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_k \leqslant x_k) = \lim_{\substack{x_{k+1}\to+\infty \\ \cdots \\ x_n\to+\infty}} F(x_1, x_2, \cdots, x_n) \\ f_{X_1,X_2,\cdots,X_k}(x_1, x_2, \cdots, x_k) &= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(u_1, \cdots, u_k, u_{k+1}, \cdots, u_n)du_{k+1}\cdots du_n \ . \end{aligned}$$

还可以定义 n 个随机变量的独立性和两个随机向量的独立性.

定义 5.14 若随机向量 $(X_1, X_2, \cdots, X_n)$ 的联合分布函数 $F(x_1, x_2, \cdots, x_n)$ 满足

$$F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \ ,$$

则称 $X_1, X_2, \cdots, X_n$ **相互独立**. 若随机向量 $X = (X_1, X_2, \cdots, X_m)$ 和 $Y = (Y_1, Y_2, \cdots, Y_n)$ 的联合分布函数 $F(x_1, \cdots, x_m, y_1, \cdots, y_n)$ 满足

$$F(x_1, \cdots, x_m, y_1, \cdots, y_n) = F_X(x_1, \cdots, x_m)F_Y(y_1, \cdots, y_n) \ ,$$

则称 **随机向量 X 和 Y 相互独立**.

上面的独立性也可以通过联合密度函数和边缘密度函数来定义. 多维随机向量中最重要的常用分布是多维正太分布.

定义 5.15 给定一个向量 $\boldsymbol{\mu} \in \mathbb{R}^n$ 和正定矩阵 $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, 对任意实数向量 $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, 若随机向量 $X = (X_1, X_2, \cdots, X_n)$ 的密度函数为

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right) ,$$

则称随机向量 X 服从参数为 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 的多维正态分布 (multivariate normal distribution), 记

$$X = (X_1, X_2, \cdots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

在上面的定义中, $|\boldsymbol{\Sigma}|$ 表示矩阵 $\boldsymbol{\Sigma}$ 的行列式, 因为其正定性可以确保 $|\boldsymbol{\Sigma}|^{-1/2}$ 有意义. 特别地, 当 $n = 2$ 时, 设

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \qquad \text{和} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} ,$$

则定义 5.9 和定义 5.15 中关于二维正太分布的密度函数尽管表达形式不同, 但两者完全相等, 相关证明将作为一个练习题.

当 $\boldsymbol{\mu} = \mathbf{0}_n$ (全为零的 n 维向量), 以及 $\boldsymbol{\Sigma} = I_n$ ($n \times n$ 单位阵) 时, 正太分布 $\mathcal{N}(\mathbf{0}_n, I_n)$ 被称为 n **维标准正太分布**, 此时它的密度函数为

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{2}\right) .$$

不难发现, n 维标准正太分布可以看作是相互独立的 n 个标准正太分布随机变量的联合分布, 也容易验证 n 标准正太分布的密度函数满足

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(\boldsymbol{x}) dx_1 \cdots dx_n = \prod_{i=1}^{n} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right) dx_i = 1 .$$

对于正定矩阵 $\boldsymbol{\Sigma}$ 通过特征值分解有

$$\boldsymbol{\Sigma} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{U} ,$$

这里 $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ 是由特征值构成的对角阵, $\boldsymbol{U}$ 是特征向量所构成的正交矩阵. 基于特征值分解可以将任意 n 维正态分布转化为 n 维标准正态分布.

定理 5.11 设 n 维随机向量 $X = (X_1, X_2, \cdots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 以及正定矩阵 $\boldsymbol{\Sigma}$ 的特征值分解为 $\boldsymbol{\Sigma} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{U}$, 则随机向量

$$Y = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}_n, I_n) .$$

证明 根据 $Y=\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})$ 可得 $X=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}Y+\boldsymbol{\mu}$, 已知 X 的概率密度函数为

$$f_X(\boldsymbol{x})=(2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right)\ .$$

根据 n 维随机变量函数 (定理 5.10 的多维情况) 的概率密度公式有

$$f_Y(\boldsymbol{y})=f_X\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{y}+\boldsymbol{\mu}\right)\left|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\right|\ ,$$

其中 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$. 根据特征值分解 $\boldsymbol{\Sigma}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{U}$ 有

$$\left|\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\right|=|\boldsymbol{\Sigma}|^{1/2}\ ,$$

以及将 $\boldsymbol{x}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{y}+\boldsymbol{\mu}$ 代入有

$$(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\ .$$

由此可得随机向量 $Y=\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})$ 的密度函数为

$$f_Y(\boldsymbol{y})=(2\pi)^{-n/2}\exp\left(-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}/2\right)\ ,$$

定理得证.

多维正太分布有下面的性质, 其证明将作为一个练习题.

定理 5.12 设随机向量 $X=(X_1,X_2,\cdots,X_n)\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 则有

$$Y=\boldsymbol{A}X+\boldsymbol{b}\sim\mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b},\boldsymbol{A}\Sigma\boldsymbol{A}^{\mathrm{T}})$$

其中 $|\boldsymbol{A}|\neq 0, \boldsymbol{A}\in\mathbb{R}^{n\times n}$ 和 $\boldsymbol{b}\in\mathbb{R}^{n\times 1}$.

对于多维正太分布还有下面一些重要的性质:

定理 5.13 设随机向量 $X=(X_1,X_2,\cdots,X_n)^{\mathrm{T}}$ 和 $Y=(Y_1,Y_2,\cdots,Y_m)^{\mathrm{T}}$, 以及

$$\begin{pmatrix}X\\Y\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}\boldsymbol{\mu}_x\\\boldsymbol{\mu}_y\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{xx}&\boldsymbol{\Sigma}_{xy}\\\boldsymbol{\Sigma}_{yx}&\boldsymbol{\Sigma}_{yy}\end{pmatrix}\right)\ ,$$

其中 $\boldsymbol{\mu}_x=(\mu_{x_1},\mu_{x_2},\cdots,\mu_{x_n})^{\mathrm{T}}$, $\boldsymbol{\mu}_y=(\mu_{y_1},\mu_{y_2},\cdots,\mu_{y_m})^{\mathrm{T}}$, $\boldsymbol{\Sigma}_{xy}=\boldsymbol{\Sigma}_{yx}^{\mathrm{T}}\in\mathbb{R}^{m\times n}$, $\boldsymbol{\Sigma}_{xx}\in\mathbb{R}^{m\times m}$ 和 $\boldsymbol{\Sigma}_{yy}\in\mathbb{R}^{n\times n}$, 则有

- 随机向量 X 和 Y 的边缘分布分别为 $X\sim\mathcal{N}(\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_{xx})$ 和 $Y\sim\mathcal{N}(\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_{yy})$;
- 随机向量 X 与 Y 相互独立的充要条件是 $\boldsymbol{\Sigma}_{xy}=(\mathbf{0})_{m\times n}$ (元素全为零的 $m\times n$ 矩阵);

- 在 $X = \boldsymbol{x}$ 的条件下随机向量 $Y \sim \mathcal{N}(\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})$;
- 在 $Y = \boldsymbol{y}$ 的条件下随机向量 $X \sim \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx})$.

我们这里仅仅给出结论, 不给出具体的证明. 有兴趣的读者可以查询资料或自己动手, 证明的核心是矩阵的分块, 可以借鉴二维正太分布的证明.

习题

5.1 设二维随机变量 (X,Y) 的分布函数为 $F(x,y)$, 求概率 $P(X>x,Y>y)$.

5.2 设随机变量 (X,Y) 的分布函数

$$F(x,y)=\begin{cases}1-e^{-x}-e^{-y}+e^{-x-y} & x\geqslant 0,y\geqslant 0\\ 0 & \text{其它},\end{cases}$$

求 X 和 Y 的边缘分布函数和边缘密度函数.

5.3 设连续非负的随机变量 X 和 Y 相互独立, X 的边缘分布函数为 $F_X(x)$, Y 的边缘密度函数为 $f_Y(y)$, 证明

$$P(X<Y)=\int_0^{+\infty}F_X(x)f_Y(x)dx\ .$$

5.4 设随机变量 (X,Y) 的密度函数

$$f(x,y)=\begin{cases}ce^{-y} & 0\leqslant x<y<+\infty\\ 0 & \text{其它},\end{cases}$$

求 X 和 Y 的边缘密度函数.

5.5 设相互独立的随机变量 $X\sim e(\lambda_1)$ 和 $Y\sim e(\lambda_2)$, 其中 $\lambda_1>0,\lambda_2>0$. 求 $P(X<Y)$.

5.6 给定 $\alpha>0$, 设随机变量 (X,Y) 的分布函数

$$F(x,y)=\begin{cases}(1-e^{-\alpha x})y & x\geqslant 0,0\leqslant y\leqslant 1\\ 1-e^{-\alpha x} & x\geqslant 0,y>1\\ 0 & \text{其它},\end{cases}$$

求 X 和 Y 的独立性.

5.7 设二维随机变量 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}x^2+axy & 0\leqslant x\leqslant 1,0\leqslant y\leqslant 2\\ 0 & \text{其它},\end{cases}$$

求 $P(X+Y\geqslant 1)$.

5.8 设二维随机变量 (X, Y) 的概率密度

$$f(x,y)=\begin{cases} cxy & 0\leqslant x\leqslant y\leqslant 1\\ 0 & \text{其它}, \end{cases}$$

求 $P(X\leqslant 1/2)$.

5.9 若多维随机向量 $(X_1, X_2, \cdots, X_r) \sim M(n, p_1, p_2, \cdots, p_r)$, 则每个随机变量 X_i $(i\in[r])$ 的边缘分布为二项分布 $B(n, p_i)$. (利用联合分布函数的定义证明)

5.10 设随机变量 X, Y, Z 服从 $(0,1)$ 上的均匀分布且相互独立, 求概率 $P(X\geqslant YZ)$.

5.11 设随机向量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases} 1 & x\in(0,1), |y|<x\\ 0 & \text{其它 }, \end{cases}$$

求条件概率密度 $f_{Y|X}(y|x)$ 和 $f_{X|Y}(x|y)$.

5.12 设随机向量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases} e^{-y} & y>x>0\\ 0 & \text{其它} \end{cases}$$

求条件概率密度 $f_{X|Y}(x|y)$.

5.13 设随机向量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases} x+y & x,y\in(0,1)\\ 0 & \text{其它 }, \end{cases}$$

求 $Z_1 = X + Y$ 和 $Z_2 = XY$ 的密度函数.

5.14 设随机向量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases} A(x+y)e^{-x-y} & x>0, y>0\\ 0 & \text{其它 }, \end{cases}$$

求随机变量 X 与 Y 的独立性, 以及 $Z = X + Y$ 的密度函数.

5.15 设随机向量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases}Ae^{-x-y} & 0<x<1, y>0\\ 0 & \text{其它},\end{cases}$$

求 1) 常数 A; 2) X 与 Y 的边缘密度函数, 3) $Z=\max(X,Y)$ 的密度函数.

5.16 设随机变量 $X\sim U(0,1)$ 和 $Y\sim e(1)$ 相互独立, 求 $Z=X+Y$ 的概率密度.

5.17 若随机变量 $X\sim e(\lambda_1)$ 和 $Y\sim e(\lambda_2)$ 相互独立, 求 $Z=X+Y$ 的分布函数和概率密度.

5.18 若随机变量 $X\sim e(1)$ 和 $Y\sim e(1)$ 相互独立, 求 $Z=Y/X$ 的概率密度.

5.19 若随机变量 $X\sim G(p_1)$ 和 $Y\sim G(p_2)$ 相互独立, 求 $Z=X+Y$ 的分布列.

5.20 若相互独立的随机变量 X 和 Y 分别服从参数为 λ_1 和 λ_2 的泊松分布, 求在 $X+Y=n$ 的条件下 X 的条件分布.

5.21 证明: 定义 5.9 和定义 5.15 中关于二维正太分布的密度函数完全相等.

5.22 证明定理 5.12.

5.23 证明定理 5.13.