中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

第五章 多元随机变量及其概率分布

洪永淼
中国科学院数学与系统科学研究院
中国科学院大学经济与管理学院

目　录

定义 5.1 [随机向量 (Random Vector)]

一个 n 维随机向量，记作 $(Z_1,\cdots,Z_n)'$，是从样本空间 (sample space) S 到 n 维欧式空间 (Euclidean space) $\mathbb{R}^n$ 的一个映射。对于样本空间内的任意结果 $s\in S$，$Z(s)$ 是一个 n 维实值向量，称为随机向量 Z 的一个实现。

- 本章主要关注：
  - ✓ 二元概率分布，它可反映多元概率分布的绝大多数（但非全部）本质特征。
  - ✓ 两个随机变量 (X,Y)，其中 X 和 Y 定义在同一概率空间 $(S,\mathbb{B},P)$ 上。
  - ✓ (X,Y) 的一组实现值记作 $(x,y)\in\mathbb{R}^2$。

定义 5.2 [联合 CDF]

X 和 Y 的联合 CDF 定义如下：

$$\begin{aligned} F_{XY}(x,y) &= P(X \le x, Y \le y) \\ &= P(X \le x \cap Y \le y) \end{aligned}$$

其中$(x,y) \in \mathbb{R}^2$ 是任意实数组。

引理 5.1 [$F_{XY}(x,y)$ 的性质]

对任意实数组 (x,y),

(1) $F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$; $F_{XY}(\infty,\infty) = 1$;

(2) $F_{XY}(x,y)$ 是关于 x 和 y 的**非递减**函数;

(3) $F_{XY}(x,y)$ 是关于 x 和 y 的**右连续**函数。

定理 5.1

$$F_X(x) = F_{XY}(x, \infty), \text{且 } F_Y(y) = F_{XY}(\infty, y)$$

证明:

- 定义两个事件：$A = \{X \le x\}$，$B = \{Y \le \infty\}$。
- 由于 B 总是成立，故有 $A \cap B = A$。
- 因此，$P(A) = P(A \cap B)$，即 $F_X(x) = F_{XY}(x, \infty)$。

证毕。

定理 5.1 含义

- 定理 5.1 表明可从 X 和 Y 的联合 CDF 求得其各自的 CDF 。X 和 Y 各自的 CDF 分别称为 X 和 Y 的边际 CDF 。
- 联合 CDF 和边际 CDF 在统计学和计量经济学有广泛应用。
  - ✓ 以下以关联（copula）函数为例，说明其重要性。关联函数可用于对随机变量之间的相依性的建模。

例 5.1： [二元关联 (Copula) 函数]

- 令 $U = F_X(X)$，$V = F_Y(Y)$ 分别为 X 和 Y 的概率积分变换。
- 这两个随机变量均为 $U[0,1]$ 随机变量，且 (U,V) 的联合 CDF
$$C(u,v) = P(U \le u, V \le v), -\infty \le u, v \le \infty$$
称为 (X,Y) 联合概率分布的关联函数。
- 关联函数 $C(u,v)$ 和联合 CDF $F_{XY}(x,y)$ 紧密相关。
- 假设 $F_X(\cdot)$ 和 $F_Y(\cdot)$ 为严格递增函数，则有
$$\begin{aligned} F_{XY}(x,y) &= P(X \le x, Y \le y) \\ &= P[F_X(X) \le F_X(x), F_Y(Y) \le F_Y(y)] \\ &= P[U \le F_X(x), V \le F_Y(y)] \\ &= C[F_X(x), F_Y(y)] \end{aligned}$$

例 5.1 (Cont.):

- 这表明联合分布 $F_{XY}(x,y)$ 可分解为两个成分：
  - ✓ 一部分是边际分布 $F_X(\cdot)$ 和 $F_Y(\cdot)$，
  - ✓ 另一部分是 X 和 Y 之间“纯粹”的关联函数 $C(\cdot,\cdot)$。
- 换言之，关联函数 $C(\cdot,\cdot)$ 仅刻画了 X 和 Y 之间的关联性而与其边际分布无关。关联函数可将边际行为与 X 和 Y 之间的关联性区分开来。给定函数形式 $C(\cdot,\cdot)$，不同的边际分布会导致 (X,Y) 的不同联合分布。
- 由于关联函数可用于刻画不同市场之间或不同资产之间的联动性 (Cherubini et al., 2004)，其被广泛应用于计量经济学和金融业界中。

5.1.1 离散情形

定义 5.3 [联合 PMF]

若 X 和 Y 为两个离散随机变量，则对任意 $(x,y)\in\mathbb{R}^2$，二者的联合 PMF 定义为

$$f_{XY}(x,y)=P(X=x\cap Y=y)=P(X=x,Y=y),-\infty<x,y<\infty$$

引理 5.2 [$f_{XY}(x,y)$ 的性质]

(1) 对任意 $(x,y)\in\mathbb{R}^2$，$f_{XY}(x,y)\geq\ 0$;

(2) $\sum_{z\in\Omega_X}\sum_{y\in\Omega_Y}f_{XY}(x,y)=1$，其中 Ω_X 和 Ω_Y 分别为 X 和 Y 的支撑集。

定义 5.4 [支撑 (Support)]

二维随机向量 (X,Y) 的支撑定义为所有取严格正概率的可能实现值 (x,y) 构成的集合，即

$$\mathrm{Support}(X,Y)=\Omega_{XY}=\{(x,y)\in\mathbb{R}^2: f_{XY}(x,y)>0\}$$

◆ 问题 5.1

假设 Ω_X 和 Ω_Y 分别为 X 和 Y 的支撑。是否有

$$\begin{aligned}\Omega_{XY} &= \Omega_X \times \Omega_Y \\ &= \{(x,y) \in \mathbb{R}^2: f_X(x) > 0, f_Y(y) > 0\}\end{aligned}$$

其中符号“×”代表笛卡尔乘积 (Cartesian product)。

- 答案是**否定**的。
- 考虑一个二维分布的例子，其中 X 和 Y 均为非负整数但满足约束条件 $X \leq Y$。则

$$\Omega_X = \Omega_Y = \{0, 1, 2, \cdots\}$$

而 $\Omega_{XY} = \{(x,y) : 0 \leq x \leq y < \infty,$其中$x,y$均为整数$\}$

- 显然，Ω_{XY} 是 $\Omega_X \times \Omega_Y$ 的一个子集。

◆ 问题 5.2

$$\sum_{(x,y)\in\Omega_{XY}} f_{XY}(x,y) = 1$$

是否成立？其中，Ω_{XY} 通常是 $\Omega_X \times \Omega_Y$ 的一个子集。

- 联合 PMF $f_{XY}(x,y)$ 可用于计算由 (X,Y) 定义的任意事件的概率。
- 对任意子集 $A \in \mathbb{R}^2$，有

$$P[(X,Y) \in \mathrm{A}] = \sum_{(x,y)\in A} f_{XY}(x,y)$$

例 5.2:

假设 X 和 Y 的联合 PMF 为

$$f_{XY}(x,y) = c|x+y|,$$

$$x = -1, 0, 1;\ y = 0, 1$$

其中 c 为未知常数。求

(1) X, Y 以及 (X, Y) 的支撑;

(2) c 的值;

(3) $P(X = 0$ 且 $Y = 1)$;

(4) $P(X = 1)$;

(5) $P(|X - Y| \leq 1)$。

解:

(1) $\Omega_X = \{-1,0,1\}$，$\Omega_Y = \{0,1\}$ 以及

$$\Omega_{XY} = \{(-1,0),(0,1),(1,0),(1,1)\}。$$

由于 $(0,0) \in \Omega_X \times \Omega_Y$ 但 $(0,0) \notin \Omega_{XY}$，Ω_{XY} 是 $\Omega_X \times \Omega_Y$ 的一个子集。

(2) 由等式 $\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} f_{XY}(x,y) = 1$，有

$$c[|-1+0|+|-1+1|+|0+0|+|0+1|+|1+0|+|1+1|] = 1$$

因此得 $c = 1/5$。

(3) 由联合 PMF 定义可知

$$P(X=0,Y=1) = f_{XY}(0,1) = \frac{1}{5}|0+1| = \frac{1}{5}$$

解 (Cont.)：

(4) 因事件 $\{X=1\}=\{X=1\}\cap\{Y\in\Omega_Y\}$，有

$$P(X=1)=\sum_{y\in\Omega_y} f_{XY}(1,y)=f_{XY}(1,0)+f_{XY}(1,1)=\frac{3}{5}$$

(5)

$$P(|X-Y|\leq 1)=\sum_{(x,y)\in\Omega_{XY}:|x-y|\leq 1} f_{XY}(x,y)=f_{XY}(-1,0)+f_{XY}(0,1)+f_{XY}(1,0)+f_{XY}(1,1)=1$$

$f_{XY}(x,y)$ 和 $F_{XY}(x,y)$ 之间的关系

- 这与一元变量情形下 $f_X(x)$ 和 $F_X(x)$ 之间的关系十分类似。

（1）可从 $f_{XY}(x,y)$ 求得 $F_{XY}(x,y)$

- 对于离散随机变量 X 和 Y，其联合 CDF 为

$$F_{XY}(x,y) = P(X \le x, Y \le y)$$

$$= \sum_{(u,v)\in\Omega_{XY}(z,y)} f_{XY}(u,v)$$

其中 $\Omega_{XY}(x,y)$ 是在 (X,Y) 的支撑 Ω_{XY} 上满足 $u \le x, v \le y$ 的所有可能 (u,v) 取值所构成的集合，即

$$\Omega_{XY}(x,y) = \{(u,v) \in \Omega_{XY}: u \le x, v \le y\}$$

$f_{XY}(x,y)$ 和 $F_{XY}(x,y)$ 之间的关系 (Cont.)

（2）通过对 $F_{XY}(x,y)$ 求差分，可从 $F_{XY}(x,y)$ 求得 $f_{XY}(x,y)$

- 不失一般性，假设
  - ✓ X 的可能取值按照升序排列：$x_1 < x_2 < x_3 < \cdots$，
  - ✓ Y 的可能取值也按升序排列：$y_1 < y_2 < y_3 < \cdots$。
- 则对 $i > 1$，$j > 1$，有

$$
\begin{aligned}
f_{XY}(x_i, y_j) &= \Delta_Y \Delta_X F_{XY}(x_i, y_j) \\
&= \Delta_Y \left[F_{XY}(x_i, y_j) - F_{XY}(x_{i-1}, y_j)\right] \\
&= \left[F_{XY}(x_i, y_j) - F_{XY}(x_i, y_{j-1})\right] - \left[F_{XY}(x_{i-1}, y_j) - F_{XY}(x_{i-1}, y_{j-1})\right] \\
&= F_{XY}(x_i, y_j) - F_{XY}(x_i, y_{j-1}) - F_{XY}(x_{i-1}, y_j) + F_{XY}(x_{i-1}, y_{j-1})
\end{aligned}
$$

- 其中，Δ_X 和 Δ_Y 分别是对 x 和 y 的差分算子 (difference operators)。例如，对任意给定的 y，有 $\Delta_X F_{XY}(x_i, y) = F_{XY}(x_i, y) - F_{XY}(x_{i-1}, y)$。

$f_{XY}(x,y)$ 和 $F_{XY}(x,y)$ 之间的关系 (Cont.)

- 上述公式不包含 $i=1$ 或 $j=1$ 的情况。
- 对这类情形，有

$$f_{XY}(x_i, y_j) = \begin{cases} F_{XY}(x_i, y_j) - F_{XY}(x_i, y_{j-1}), & i=1, j>1 \\ F_{XY}(x_i, y_i) - F_{XY}(x_{i-1}, y_j), & i>1, j=1 \\ F_{XY}(x_i, y_i), & i=1, j=1 \end{cases}$$

n 维随机向量联合 PMF 与联合 CDF 之间的关系

- 上述二元情形均可推广到 n 维随机向量的多元情形。
- 例如，n 个离散随机变量 $X_1, \cdots, X_n$ 的联合 PMF 为

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n) = P(X_1 = x_1, \cdots, X_n = x_n)$$

其中 $\boldsymbol{x}^n = (x_1, \cdots, x_n)' \in \mathbb{R}^n$ 是 n 维实数空间上的任意 n 维数组。

- 而 $\boldsymbol{X}^n = (X_1, \cdots, X_n)'$ 的联合 CDF 则为

$$F_{\boldsymbol{X}^n}(\boldsymbol{x}^n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$$

其中 $\boldsymbol{x}^n \in \mathbb{R}^n$。

5.1.2 连续情形

定义 5.5 [联合 PDF]

- 若两个随机变量 X 和 Y 的联合 CDF $F_{XY}(x,y)$ 对 x 和 y 均是绝对连续 (absolutely continuous) 的，则称其具有连续联合分布 (joint continuous distribution)。
- 此时，存在函数 $f_{XY}(x,y)$，对任意给定的 $(x,y)\in\mathbb{R}^2$，有

$$F_{XY}(x,y)=P(X\le x,Y\le y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f_{XY}(u,v)dudv$$

其中函数 $f_{XY}(x,y)$ 称为 (X,Y) 的联合 PDF。

引理 5.3 [联合 PDF $f_{XY}(x,y)$ 的性质]

(1) 对 xy 平面上的所有实数组 (x,y)，有 $f_{XY}(x,y) \geq 0$;

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$。

证明:

(1)

- 对任意给定的二元数组 $(x,y)\in\mathbb{R}^2$，定义 $A(x,y)=\{(u,v):u\le x,v\le y\}$，有

$$\begin{aligned}P[(X,Y)\in A(x,y)]&=P(X\le x,Y\le y)\\&=F_{XY}(x,y)\\&=\int_{-\infty}^{z}\int_{-\infty}^{y}f_{XY}(u,v)dudv\end{aligned}$$

- 该公式与在离散情形下对联合 PDF 二重求和类似，表明可通过对联合 PDF $f_{XY}(x,y)$ 二重积分而得 $F_{XY}(x,y)$。

证明 (Cont.):

- 对函数 $F_{XY}(x,y)$ 的任意可微点 (x,y)，有

$$f_{XY}(x,y)=\frac{\partial^2 F_{XY}(x,y)}{\partial x\partial y}\geq 0$$

- 其中，等式由微积分基本定理可得，不等式乃根据 $F_{XY}(x,y)$ 为 (x,y) 的非递减函数而得。该公式与离散情形下对联合 CDF $F_{XY}(x,y)$ 进行二重差分类似，表明在连续情形下可通过对联合 CDF $F_{XY}(x,y)$ 求导而得联合 PDF $f_{XY}(x,y)$。

(2) 根据 $F_{XY}(\infty,\infty)=1$ 即得

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

证毕。

◆ 问题 5.3

如何解释联合 PDF $f_{XY}(x,y)$?

- 给定 xy 平面上的任意数组 (x,y)，考虑事件

$$A(x,y)=\left\{x-\frac{\epsilon}{2}<X\leq x+\frac{\epsilon}{2} \text{ 且 } y-\frac{\epsilon}{2}<Y\leq y+\frac{\epsilon}{2}\right\}$$

其中 $\epsilon>0$ 为很小的常数。

- 几何上，$A(x,y)$ 是 (X,Y) 在以点 (x,y) 为中心、四边长度均为 ϵ 的一个小矩形区域内取值的事件。

CONTINUE

- 假设 $f_{XY}(x,y)$ 在点 (x,y) 处连续，则有

$$
\begin{aligned}
P[A(x,y)] &= P\left(x-\frac{\epsilon}{2}<X\le x+\frac{\epsilon}{2}, y-\frac{\epsilon}{2}<Y\le y+\frac{\epsilon}{2}\right)\\
&= \int_{y-\epsilon/2}^{y+\epsilon/2}\int_{x-\epsilon/2}^{x+\epsilon/2} f_{XY}(u,v)dudv\\
&= f_{XY}(\bar{x},\bar{y})\epsilon\epsilon \quad [\text{对某一对 } (x,y)]\\
&\approx f_{XY}(x,y)\epsilon^2 \quad (\text{当 } \epsilon \text{ 很小时})
\end{aligned}
$$

- 因此，$f_{XY}(x,y)$ 虽不是概率测度，但其与 (X,Y) 在以点 (x,y) 为中心的一个小矩形区域内取值的概率成正比。
- 换言之，$f_{XY}(x,y)$ 和 (X,Y) 在 xy 平面上以点 (x,y) 为中心的区域内取值的概率成正比。

CONTINUE

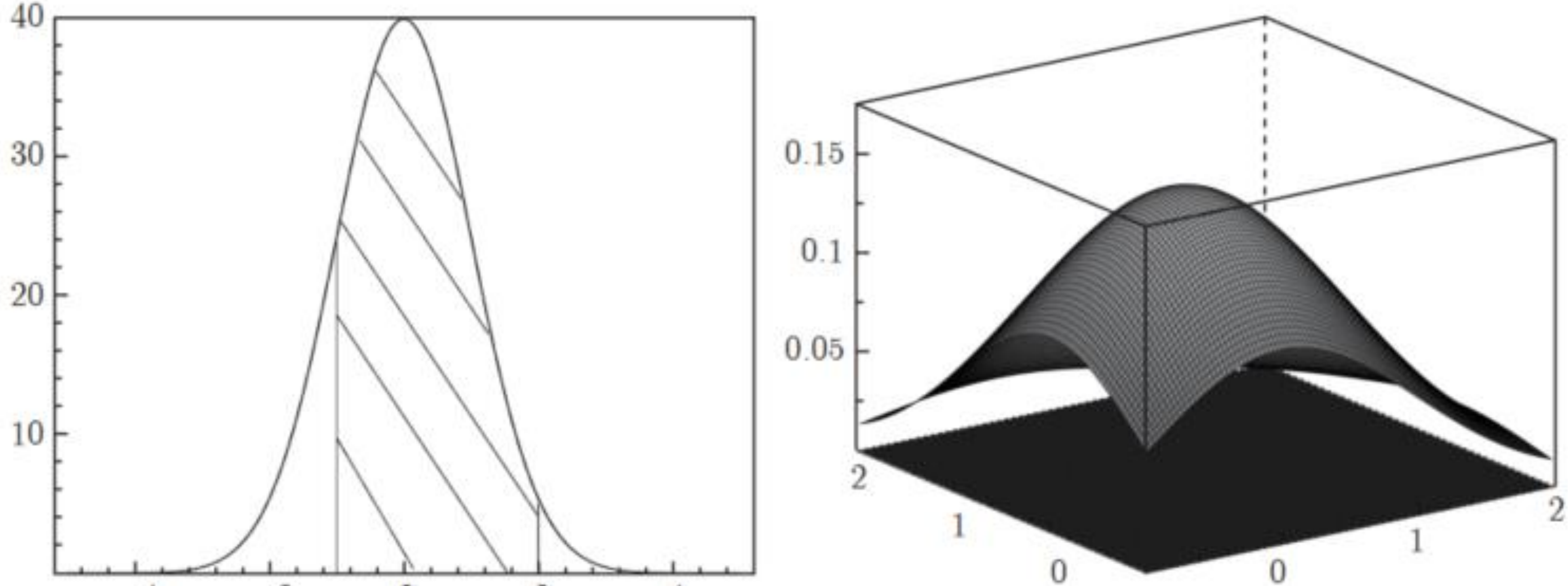


图 5.1：(a) $P[X \in [-1,2]]$ 的几何解释　(b) $P[(X,Y) \in [-1,2]^2]$ 的几何解释

定义 5.6 [(X,Y) 的支撑]

二元连续随机向量 (X,Y) 的支撑定义为

$$\mathrm{Support}(X,Y)=\Omega_{XY}=\{(x,y)\in\mathbb{R}^2: f_{XY}(x,y)>0\}$$

- 因此，(X,Y) 的支撑 $\mathrm{Support}(X,Y)$ 是 xy 平面中满足 (X,Y) 在该点的小邻域内取值的概率严格为正的所有点 (x,y) 构成的集合。
- 由于对 (X,Y) 支撑集外的所有点 (x,y)，有 $f_{XY}(x,y)=0$，故在计算与 (X,Y) 相关的任意事件的概率时，仅需在支撑 Ω_{XY} 上对 $f_{XY}(x,y)$ 求积分。

例 5.3：

假设 X 和 Y 的联合 PDF 为

$$f_{XY}(x,y) = cy^2, \qquad 0 \le x \le 2, 0 \le y \le 1$$

其中 c 为未知常数。

求：

(1) c 的值； (2) $P(X+Y>2)$；

(3) $P(X<0.5)$； (4) $P(X=3Y)$。

解：

- 支撑 $\Omega_{XY} = \{(x,y): 0 \le x \le 2, 0 \le y \le 1\}$ 为 xy 平面上的矩形区域。

CONTINUE

解 (Cont.):

(1) 因等式 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$ 则

$$\begin{aligned}
\int_0^1 \left(\int_0^2 cy^2dx\right)dy &= \int_0^1 cy^2\left(\int_0^2 dx\right)dy \\
&= c\int_0^1 y^2x\Big|_0^2 dy \\
&= 2c\int_0^1 y^2dy \\
&= \frac{2c}{3}y^3\Big|_0^1 \\
&= \frac{2c}{3} \\
&= 1
\end{aligned}$$

- 因此，得 $c = 3/2$。

CONTINUE

解 (Cont.):

(2) $P(X+Y>2)=\int_0^1\left(\int_{2-y}^2 \frac{3}{2}y^2dx\right)dy=\frac{3}{8}$

(3) $P(X<0.5)=\int_0^1\left(\int_0^{\frac{1}{2}} \frac{3}{2}y^2dx\right)dy=\frac{1}{4}$

(4) 因为 $x=3y$ 是一条直线，直线上方 $f_{xy}(x,y)$ 覆盖的体积为 0，故 $P(X=3Y)=0$。

例 5.4：

假设X和Y的联合PDF 为

$$f_{XY}(x,y)=cx^2y, x^2\le y\le 1$$

其中 c 为未知常数。求 (1) c 的值；(2) $P(X\ge Y)$。

解： 支撑 $\Omega_{XY}=\{(x,y):x^2\le y\le 1\}$,

(1) $\int_0^1\left(\int_{-\sqrt{y}}^{\sqrt{y}}cx^2y\,dx\right)dy=1$，可得 $c=21/4$。

(2) $P(X\ge Y)=\int_0^1\left(\int_y^{\sqrt{y}}cx^2y\,dx\right)dy=\frac{3}{20}$。

例 5.5：

假设 X 和 Y 的联合 CDF 为

$$F_{XY}(x,y)=\frac{1}{16}xy(x+y),0\le x\le 2,0\le y\le 2$$

求 (1) CDF $F_{XY}(x,y)$ 在整个 xy 平面上的完整表达式；

(2) 联合 PDF $f_{XY}(x,y)$；(3) $P(1\le X\le 2,1\le Y\le 2)$。

解：(1) 根据引理 5.1 和定理 5.1 关于联合 CDF $F_{XY}(x,y)$ 的性质，有

$$F_{XY}(x,y)=\begin{cases}0, & x<0\text{ 或 }y<0\\ \frac{1}{16}xy(x+y), & 0\le x\le 2,0\le y\le 2\\ \frac{1}{8}x(x+2), & 0\le x\le 2,y>2\\ \frac{1}{8}y(y+2), & x>2,0\le y\le 2\\ 1, & x>2,y>2\end{cases}$$

CONTINUE

解 (Cont.):

(2) 对联合 CDF $F_{XY}(x,y)$ 求偏导，可得

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$
$$= \frac{1}{8}(x+y), 0 \leq x \leq 2, 0 \leq y \leq 2$$

对所有其他的点 (x,y)，有 $f_{XY}(x,y) = 0$。

(3)

$$P(1 \leq X \leq 2, 1 \leq Y \leq 2) = \int_1^2 \int_1^2 \frac{1}{8}(x+y) dx dy$$
$$= \frac{3}{8}$$

多元连续随机变量

- 二元连续随机变量的相关概念可扩展到更一般的多元情形。
- n 维连续随机向量 $\boldsymbol{X}^n = (X_1,\cdots,X_n)'$ 的联合 CDF 为
$$F_{\boldsymbol{X}^n}(\boldsymbol{x}^n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{\boldsymbol{X}^n}(\mathbf{u}^n)d\mathbf{u}^n$$
- 其中 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n)$ 是 $\boldsymbol{X}^n$ 的联合 PDF，$\boldsymbol{x}^n = (x_1,\cdots,x_n)'$, 且 $\boldsymbol{u}^n = (u_1,\cdots,u_n)'$。
- 若对空间 $\mathbb{R}^n$ 上的点 $\boldsymbol{x}^n$，$F_{\boldsymbol{X}^n}(\boldsymbol{x}^n)$ 的偏导数存在，则有
$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n) = \frac{\partial^n F_{\boldsymbol{X}^n}(\boldsymbol{x}^n)}{\partial x_1\cdots\partial x_n}$$

目 录

◆ 问题 5.4

从联合 PMF/PDF $f_{XY}(x,y)$ 可获取哪些信息呢?

- 直观上，可得如下信息：

(1) X 的边际分布信息，由 X 的 PMF/PDF $f_X(x)$ 刻画；

(2) Y 的边际分布信息，由 Y 的 PMF/PDF $f_Y(y)$ 刻画；

(3) X 和 Y 之间的预测关系，由适当的条件分布刻画。

5.2.1 离散情形

定义 5.7 [边际概率质量函数 (Marginal PMF’s)]

- 假设 X 和 Y 具有联合离散分布，其联合 PMF 为 $f_{XY}(x,y)$。则 X 和 Y 的边际 PMF 分别定义为

$$f_X(x) = P(X = x) = \sum_{y \in \Omega_Y} f_{XY}(x,y), \quad -\infty < x < \infty$$

$$f_Y(y) = P(Y = y) = \sum_{x \in \Omega_X} f_{XY}(x,y), \quad -\infty < y < \infty$$

定义 5.7 含义

- 在二元情形下定义事件 $\{X=x\}$。注意到

$$\begin{aligned}\{X=x\} &= \{X=x\}\cap\{Y\in\Omega_Y\}\\ &= \{X=x\}\cap\left[\bigcup_{y\in\Omega_Y}\{Y=y\}\right]\\ &= \bigcup_{y\in\Omega_Y}\left[\{X=x\}\cap\{Y=y\}\right]\ (\text{定理 2.1 的分配律})\end{aligned}$$

- 全概率公式指出，对任意事件 A 以及一系列互斥且完全穷尽事件序列 $\{A_1,A_2,\cdots\}$，有

$$P(A)=\sum_{i=1}^{\infty}P(A\cap A_i)$$

- 与之类似，可得

$$\begin{aligned}F_X(x) &= P(X=x)\\ &= P(\{X=x\}\cap\{Y\in\Omega_Y\})\\ &= \sum_{y\in\Omega_Y}f_{XY}(x,y)\end{aligned}$$

定义 5.7含义 (Cont.)

- 直观上，X 的边际 PMF $f_X(x)$ 是不论 Y 取何值，X 取给定值 x 的概率。通过考虑 Y 的所有可能取值，消去了关于 Y 的全部信息而只保留 X 的信息。
- 之所以称 $f_X(x)$ 为 X 的边际 PMF，旨在强调其是在给定二元随机向量 (X,Y) 的联合分布下，随机变量 X 的 PMF。就技术层面来说，“边际”这一修饰词是多余的。

引理 5.4 [$f_X(x)$ 和 $f_Y(y)$ 的性质]

(1) 对所有 $x \in (-\infty, \infty)$，有 $f_X(x) \geq 0$；

(2) $\sum_{x\in\Omega_X} f_X(x) = 1$, 其中 Ω_X 为 X 的支撑。

类似结果对 $f_Y(y)$ 亦成立。

例 5.6:

假设 X 和 Y 的联合 PMF 为

$$f_{XY}(x, y) = \frac{1}{5}|x + y|, x = -1, 0, 1; y = 0, 1$$

求 (1) $f_X(x)$；(2) $f_Y(x)$。

解：(1)

- 对于 $x=-1$，事件 $\{X=-1\}$ 包含两个基本结果：$\{X=-1, Y=0\}$ 和 $\{X=-1, Y=1\}$。
- 这两个基本结果相互排斥。因此，

$$\begin{aligned} f_X(-1) &= P(X=-1) \\ &= f_{XY}(-1,0)+f_{XY}(-1,1) \\ &= \frac{1}{5} \end{aligned}$$

- 类似有

$$f_X(0)=f_{XY}(0,0)+f_{XY}(0,1)=\frac{1}{5}$$

$$f_X(1)=f_{XY}(1,0)+f_{XY}(1,1)=\frac{3}{5}$$

CONTINUE

解 (Cont.):

- 则 X 的边际 PMF 为

$$f_X(x) = \begin{cases} \frac{1}{5}, & x = -1 \\ \frac{1}{5}, & x = 0 \\ \frac{3}{5}, & x = 1 \end{cases}$$

(2)

- 类似可得 Y 的 PMF 为

$$f_Y(y) = \begin{cases} \frac{2}{5}, & y = 0 \\ \frac{3}{5}, & y = 1 \end{cases}$$

CONTINUE

- 可用矩阵形式完整地描述 (X,Y) 的联合 PMF $f_{XY}(x,y)$，并且对行或者列求和，即可分别得二者的边际 PMF：

Y/X	-1	0	1
0	$\frac{1}{5}$	0	$\frac{1}{5}$
1	0	$\frac{1}{5}$	$\frac{2}{5}$

- 可能出现联合 PMF 不同，但边际 PMF 相同的情形。
- 许多联合分布有相同的边际分布。联合 PMF 不仅提供了边际信息，而且包括无法从 X 和 Y 的边际分布中获取的二者之间关系的信息。尽管 X 和 Y 的边际分布可能相同，但若二者之间的关系有所差异时，则会产生不同的联合分布。见例 5.7。

例 5.7：

- 假设 X 和 Y 均为二值随机变量，其取值为 1 或 0。考虑以下 (X,Y) 的两个联合 PMF，分别是
  - ✓ 情形(1)：
$$f_{XY}(x,y)=\begin{cases}p, & x=y=1\\ 1-p, & x=y=0\\ 0, & \text{其他}\end{cases}$$
  - ✓ 情形(2)：
$$f_{XY}(x,y)=\begin{cases}p^{x+y}(1-p)^{2-(x+y)}, & x=1,0;y=1,0\\ 0, & \text{其他}\end{cases}$$
- 显然，情形 (1) 和 (2) 是 (X,Y) 的两个不同联合分布。但这在两种情况下，均有 $X\sim\mathrm{Bernoulli}(p)$，$Y\sim\mathrm{Bernoulli}(p)$。

5.2.2 连续情形

- 假设二维连续随机向量 (X,Y) 的联合 PDF 为 $f_{XY}(x,y)$。首先考虑如何求得 X 的边际 CDF。
- 由于事件 $\{X \le x\} = \{X \le x\} \cap \{-\infty < Y < \infty\}$，故

$$
\begin{aligned}
F_X(x) &\equiv P(X \le x) \\
&= P(X \le x, -\infty < Y < \infty) \\
&= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{XY}(u,y)dudy \\
&= \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{XY}(u,y)dy\right] du \\
&= \int_{-\infty}^{x} f_X(u)du
\end{aligned}
$$

- 对上述等式两边求微分，可得

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

- 因此，通过对联合 PDF $f_{XY}(x, y)$ 关于 y 求积分，可得边际 PDF $f_X(x)$。

定义 5.8 [边际PDF]

- 假设 X 和 Y 具有联合连续分布，并且其联合 PDF 为 $f_{XY}(x,y)$。则 X 和 Y 的边际 PDF 分别定义为：

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \qquad -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx, \qquad -\infty < y < \infty$$

- 与经济学中其他边际概念比如边际效用和边际生产率不同，一个变量的边际 PDF 是通过对另一个变量求积分而得，而并非对该变量求偏导数。
- 直观上，对联合 PDF $f_{XY}(x,y)$ 关于 y 求积分，余下即为关于 X 的信息。

引理 5.5 [$f_X(x)$ 和 $f_Y(y)$ 的性质]

(1) 对任意 $x \in (-\infty, \infty)$，有 $f_X(x) \geq 0$;

(2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$ 。

类似结果对 $f_Y(y)$ 也成立。

例 5.9：

假设对 $x^2 < y < 1$，X 和 Y 的联合 PDF 为 $f_{XY}(x,y) = cy^2$。
求 (1) $f_X(x)$；(2) $f_Y(y)$。

解：

- (X,Y) 的支撑与例 5.4 中 (X,Y) 的支撑相同。从中可求得：
  - ✓ X 的支撑为 $\Omega_X = \{x \in \mathbb{R} : -1 < x < 1\}$，
  - ✓ Y 的支撑为 $\Omega_Y = \{y \in \mathbb{R}:\ 0 < y < 1\}$。
- 首先用如下性质确定 c 的值：

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

解 (Cont.):

- 给定 (X,Y) 的支撑，有

$$
\begin{aligned}
c\int_{-1}^{1}\left(\int_{x^2}^{1} y^2dy\right)dx &= c\int_{-1}^{1}\frac{1}{3}\left(1-x^6\right)dx \\
&= \frac{2c}{3}\int_{0}^{1}\left(1-x^6\right)dx \\
&= \frac{4c}{7} \\
&= 1
\end{aligned}
$$

- 因此 $c = 7/4$.

解 (Cont.):

(1) 对 $-1 < x < 1$，有

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy \\
&= \int_{x^2}^{1} cy^2 dy \\
&= c\frac{1}{3}y^3\Big|_{x^2}^{1} \\
&= \frac{7}{12}\left(1-x^6\right)
\end{aligned}
$$

- 因此 X 的边际 PDF 为

$$
f_X(x) = \begin{cases} \frac{7}{12}\left(1-x^6\right), & -1 < x < 1 \\ 0, & \text{其他} \end{cases}
$$

解 (Cont.):

(2) 对 $0 < y < 1$，有

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} cy^2 dx \\
&= cy^2 \cdot 2\sqrt{y} \\
&= \frac{7}{2} y^{\frac{5}{2}}
\end{aligned}
$$

- 因此，Y 的边际 PDF 为

$$
f_Y(y) = \begin{cases} \frac{7}{2} y^{\frac{5}{2}}, & 0 < y < 1 \\ 0, & \text{其他} \end{cases}
$$

多个变量的边际分布

- 当有两个以上的随机变量时，不仅可定义每个随机变量的边际分布，还可对随机变量的任意子集定义联合分布。
- 例如，若有 n 个离散随机变量 $X_1,\cdots,X_n$，则 X_1 的边际 PMF 为

$$f_{X_1}(x_1)=\sum_{x_2\in\Omega_2}\cdots\sum_{x_n\in\Omega_n} f_{\boldsymbol{X}^n}(\boldsymbol{x}^n),-\infty<x_1<\infty$$

- 子集 (X_1,X_2,X_3) 的联合 PMF 为

$$f_{X_1X_2X_3}(x_1,x_2,x_3)=\sum_{x_4\in\Omega_4}\cdots\sum_{x_n\in\Omega_n} f_{\boldsymbol{X}^n}(\boldsymbol{x}^n),-\infty<x_1,x_2,x_3<\infty$$

- 其中 $\boldsymbol{X}^n=(X_1,\cdots,X_n)'$， $\boldsymbol{x}^n=(x_1,\cdots,x_n)'$，且 Ω_i，$i=1,\cdots,n$，是 X_i 的支撑。

目　录

- 通常情况下，两个随机变量 (X, Y) 的观测值是互相关联的。
- 因此，即使无法根据 X 的相关信息准确预测 Y 的实现值，也可据此获得 Y 的相关信息。

◆ **问题 5.5**

如何刻画 X 和 Y 之间的预测关系?

5.3.1 离散情形

定义 5.9 [条件 PMF]

- 假设 X 和 Y 具有联合离散分布，其联合 PMF 为 $f_{XY}(x,y)$，边际 PMF 分别为 $f_X(x)$ 和 $f_Y(y)$。
- 随机变量 Y 基于 $X=x$ 的条件 PMF 定义为
$$f_{Y|X}(y \mid x) = P(Y=y \mid X=x) = \frac{f_{XY}(x,y)}{f_X(x)}$$
其中 $f_X(x) > 0$。
- 类似地，随机变量 X 基于 $Y=y$ 的条件 PMF 定义为
$$f_{X|Y}(x \mid y) = P(X=x \mid Y=y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$
其中 $f_Y(y) > 0$。

定义 5.9 含义

- 直观上，条件 PMF $f_{Y|X}(y \mid x)$ 是在观测到随机变量 X 取值 x 的条件下，随机变量 Y 取任意值 y 的概率。
- 第二章曾指出，任意两个事件的条件概率公式为 $P(A|B) = P(A \cap B)/P(B)$。
- 若定义事件 $A = \{X = x\}$ 和 $B = \{Y = y\}$，则

$$f_{Y|X}(y \mid x) = P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

- 需要强调，给定任意 x 满足 $f_X(x)>0$，条件 PMF $f_{Y|X}(y \mid x)$ 是 Y 的 PMF。
- 即：给定任意满足 $f_X(x)>0$ 的 x，有
  - ✓ (1) 对任意 $y\in(-\infty,\infty)$，$f_{Y|X}(y \mid x)\geq 0$；
  - ✓ (2) $\sum_{y\in\Omega_{Y|X}(x)} f_{Y|X}(y \mid x)=1$，

    其中 $\Omega_{Y|X}(x)=\{y\in\Omega_Y: f_{Y|X}(y \mid x)>0\}$ 为满足 $f_{Y|X}(y \mid x)>0$ 的所有 y 的可能取值构成的集合

条件分布的应用

- 随机变量 X 的不同实现值 x 对应 Y 的不同条件分布。
  - ✓ 例如，假定 X 为取 0 和 1 两个可能实现值的状态变量。
    - ➢ 当 $X=0$ 时 (如代表股市为熊市)，股票收益率 Y 的波动较大；
    - ➢ 当 $X=1$ 时 (如代表股市为牛市)，股票收益率 Y 的波动较小。
  - ✓ 再如，假设 Y 表示雇员的工资收入，X 为表示性别的虚拟变量，对女性雇员其取值为 1，对男性雇员其取值为 0。
    - ➢ $f_{Y|X}(y \mid 1)$ 表示女性雇员工资收入的分布，
    - ➢ $f_{Y|X}(y \mid 0)$ 表示男性雇员工资收入的分布。

条件 PMF 的乘法法则

- 根据条件 PMF 的定义，有如下乘法法则：

$$f_{XY}(x,y) = f_{Y|X}(y \mid x)f_X(x) = f_{X|Y}(x \mid y)f_Y(y)$$

  - ✓ 其中，只要 $f_X(x) > 0$，第一个等式恒成立；
  - ✓ 只要 $f_Y(y) > 0$，第二个等式恒成立。
- 上述乘法法则说明联合 PMF 可等价地由条件 PMF $f_{Y|X}(y \mid x)$ 与边际 PMF $f_X(x)$ 联合描述。
  - ✓ 对 X 的不同取值 x，Y 基于 $X = x$ 条件下的概率分布可能并不相同，由此构成了一个关于 Y 的概率分布族，其中每个 x 对应 Y 的一个概率分布。当描述整个分布族时，可用"$Y \mid X$ 的分布"表示。

离散情形条件分布总结

- 随机变量 Y 基于 $X = x$ 的条件 PMF 定义为
$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x)$$
$$= \frac{f_{XY}(x, y)}{f_X(x)}$$
其中 $f_X(x) > 0$。
- 若定义事件 $A = \{X = x\}$ 和 $B = \{Y = y\}$，则
$$f_{Y|X}(y \mid x) = P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

5.3.2 连续情形

- 连续随机变量：需要扩展离散情形下 PMF 的概念
- 与离散情形不同，若 X 和 Y 均为连续随机变量，则对任意 x，有 $P(X = x) = 0$。
- 因此，在计算诸如 $P(Y > 5 \mid X = 10)$ 的条件概率时，由于分母 $P(X = 10) = 0$，无法使用定义 $P(A|B) = P(A \cap B)/P(B)$，其中 $B = \{X = 10\}$。
- 可证明当 X 和 Y 连续时，可将 PMF 替换为 PDF，采用与离散情形类似的方式定义给定 $X = x$ 下 Y 的条件概率分布。

定义 5.10 [条件 PDF]

- 假定 X 和 Y 具有联合连续分布，其联合 PDF 为 $f_{XY}(x,y)$，边际 PDF 分别为 $f_X(x)$ 和 $f_Y(y)$。则给定 $X=x$ 下 Y 的条件 PDF 定义为

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

其中 $f_X(x) > 0$。

- 类似地，X 基于 $Y=y$ 的条件 PDF 为

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

其中 $f_Y(y) > 0$。

定义 5.10 含义

- 注意
  - ✓ 对离散随机变量 (X,Y)，条件 PMF $f_{Y|X}(y \mid x)$ 是在给定事件 $\{X=x\}$ 发生时事件 $\{Y=y\}$ 发生的概率，
  - ✓ 而对于连续随机变量 (X,Y)，条件 PDF $f_{Y|X}(y \mid x)$ 则定义为联合 PDF $f_{XY}(x,y)$ 和边际 PDF $f_X(x)$ 的比值。

引理 5.6 [条件 PDF 的性质]

- 对任意满足 $f_X(x) > 0$ 的 $x \in \mathbb{R}$， $f_{Y|X}(y \mid x)$ 为 Y 的 PDF。即

(1) 对任意 $y \in (-\infty, \infty)$，有 $f_{Y|X}(y \mid x) \geq 0$；

(2) $\int_{-\infty}^{\infty} f_{Y|X}(y \mid x) = 1$。

- 上述性质对 $f_{X|Y}(x \mid y)$ 同样成立。

证明:

- 设 $f_X(x) > 0$，则对所有 $y \in (-\infty, \infty)$，$f_{Y|X}(y \mid x) = f_{XY}(x,y)/f_X(x)$ 有定义且非负。
- 此外，

$$\int_{-\infty}^{\infty} f_{Y|X}(y \mid x)dy = \int_{-\infty}^{\infty} \frac{f_{XY}(x,y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \frac{1}{f_X(x)} f_X(x) = 1$$

证毕。

- 给定任意满足 $f_X(x) > 0$ 的 x 值，$f_{Y|X}(y \mid x)$ 为 Y 的 PDF。不同的 x 值对应 Y 的不同分布。这意味着可用 X 的相关信息预测 Y 的分布。

例 5.11：

假设二维随机变量 (X, Y) 服从 $\{(x, y) : x^2 + y^2 \leq 1\}$ 上的均匀分布。求 (1) Y 基于 $X = x$ 的条件概率；(2) $P(Y > 0 \mid X = 0)$。

CONTINUE

解：

(1) (X, Y) 的联合 PDF 为

$$f_{XY}(x,y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1 \\ 0, & \text{其他} \end{cases}$$

且支撑 $\Omega_{XY} = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ 为 xy 平面上以原点 $(0,0)$ 为中心的单位圆。

- 对 $-1 \le x \le 1$，有

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

$$= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy$$

$$= \frac{2}{\pi}\sqrt{1-x^2}$$

CONTINUE

解（Cont.）:

- 因此，对于 $-1 < x < 1$，Y 基于 $X = x$ 的条件 PDF 为

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \begin{cases} \frac{1}{2\sqrt{1-x^2}}, & -\sqrt{1-x^2} \le y \le \sqrt{1-x^2} \\ 0, & \text{其他} \end{cases}$$

CONTINUE

解（Cont.）:

(2) Y 基于 $X = x$ 的条件分布依赖于 x 的大小。当 $x = 0$ 时，有

$$f_{Y|X}(y \mid 0) = \begin{cases} \frac{1}{2}, & 1 < y < 1 \\ 0, & \text{其他} \end{cases}$$

- 因此

$$P(Y > 0 \mid X = 0) = \int_0^{\infty} f_{Y|X}(y \mid 0) dx$$
$$= \int_0^1 \frac{1}{2} dy$$
$$= \frac{1}{2}$$

◆ 问题 5.6

如何解释条件 PDF $f_{Y|X}(y \mid x)$?

- 对满足 $f_{XY}(x,y) > 0$ 的任意给定 (x,y)，考虑两个事件

$A(x) = \left\{x - \frac{\epsilon}{2} < X \le x + \frac{\epsilon}{2}\right\}$ 和 $B(y) = \left\{y - \frac{\epsilon}{2} < Y \le y + \frac{\epsilon}{2}\right\}$,

其中 ϵ 为很小的正常数。

- 由中值定理 (the mean value theorem) 可得

$$
\begin{aligned}
P[A(x) \cap B(y)] &= \int_{y-\epsilon/2}^{y+\epsilon/2} \int_{x-\epsilon/2}^{x+\epsilon/2} f_{XY}(u,v)dudv \\
&\approx f_{XY}(x,y)\epsilon^2 \\
P[A(x)] &= \int_{x-\epsilon/2}^{x+\epsilon/2} f_X(u)du \\
&\approx f_X(x)\epsilon
\end{aligned}
$$

CONTINUE

- 则有

$$
\begin{aligned}
P[B(y) \mid A(x)] &= \frac{P[A(x) \cap B(y)]}{P[A(x)]} \\
&\approx \frac{f_{XY}(x,y)\epsilon^2}{f_X(x)\epsilon} \\
&= f_{Y|X}(y \mid x)\epsilon
\end{aligned}
$$

- 这意味着条件 PDF $f_{Y|X}(y \mid x)$ 和下式的条件概率成正比

$$P[B(y) \mid A(x)] = P\left(y - \frac{\epsilon}{2} < Y \le y + \frac{\epsilon}{2} \,\middle|\, x - \frac{\epsilon}{2} < X \le x + \frac{\epsilon}{2}\right)$$

- 即 $f_{Y|X}(y \mid x)$ 与给定 X 在 x 附近取值时，Y 在 y 附近取值的概率成正比。

连续情形的乘法法则（与离散情形类似）

- 乘法法则

$$f_{XY}(x,y) = f_{Y|X}(y \mid x)f_X(x) = f_{X|Y}(x \mid y)f_Y(y)$$

- 其中，
  - ✓ 当 $f_X(x) > 0$ 时，第一个等式成立；
  - ✓ 当 $f_Y(y) > 0$ 时，第二个等式成立。

当存在两个以上随机变量时，可考虑多种条件分布

- 例如，可定义

$$f_{X_i|\boldsymbol{X}^{i-1}}\left(x_i \mid \boldsymbol{x}^{i-1}\right) = \frac{f_{\boldsymbol{X}^i}\left(\boldsymbol{x}^i\right)}{f_{\boldsymbol{X}^{i-1}}\left(\boldsymbol{x}^{i-1}\right)}$$

其中 $f_{\boldsymbol{X}^{i-1}}\left(\boldsymbol{x}^{i-1}\right) > 0, \boldsymbol{X}^{i-1} = (X_1, \cdots, X_{i-1})', \boldsymbol{x}^{i-1} = (x_1, \cdots, x_{i-1})'$。

- 同样可定义

$$f_{(X_1,X_2)|(X_3,X_4)}(x_1, x_2 \mid x_3, x_4) = \frac{f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)}{f_{X_3X_4}(x_3, x_4)}$$

其中 $f_{X_3X_4}(x_3, x_4) > 0$。

目　录

定义 5.11 [独立性 (Independence)]

对于两个随机变量 X 和 Y，若

$$F_{XY}(x,y) = F_X(x)F_Y(y), \text{对所有} -\infty < x,y < \infty$$

则称二者相互独立，其中 $F_{XY}(\cdot), F_X(\cdot), F_Y(\cdot)$ 分别为联合和边际 CDF。

- 上述独立性的定义等价于以下定义，即对任意实数子集 $A \in \Omega_X$ 和 $B \in \Omega_Y$，有

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- 无论 X 和 Y 是离散还是连续随机变量，上述独立性的定义均成立。对于离散情形，可用联合和边际 PMF 刻画独立性。

定理 5.2

对于两个离散随机变量 (X,Y)， X 和 Y 相互独立，当且仅当

$$f_{XY}(x,y) = f_X(x)f_Y(y), \text{对所有 } (x,y) \in \mathbb{R}^2$$

其中 $f_{XY}(x,y)$， $f_X(x)$， $f_Y(y)$ 分别为联合和边际 PMF。

证明:

(1) [必要性]

- 根据独立性的定义，有

$$F_{XY}(x,y) = F_X(x)F_Y(y),\ \text{对所有}\ -\infty < x, y < \infty$$

- 不失一般性，假设:
  - ✓ X 的可能取值按照升序排列: $x_1 < x_2 < x_3 < \cdots$,
  - ✓ Y 的可能取值也按升序排列: $y_1 < y_2 < y_3 < \cdots$。
- 当 $i > 1$ 时，对上式关于 x 求差分，即从 x_{i-1} 到 x_i 的差分，得

$$\begin{aligned}\Delta_X F_{XY}(x_i, y) &= F_{XY}(x_i, y) - F_{XY}(x_{i-1}, y) \\ &= [F_X(x_i) - F_X(x_{i-1})]F_Y(y)\end{aligned}$$

CONTINUE

证明 (Cont.):

- 进一步当 $j>1$ 时，对上式关于 y 求差分，可得

$$\Delta_Y\Delta_X F_{XY}\left(x_i,y_j\right)=[F_X(x_i)-F_X(x_{i-1})]\left[F_Y\left(y_j\right)-F_Y\left(y_{j-1}\right)\right]$$

- 由此可得联合 PMF 和边际 PMF 之间的关系为

$$f_{XY}(x_i,y_j)=f_X(x_i)f_Y(y_j),i,j>1$$

- 当 $i=1$ 或 $j=1$ 时，可得相同结果 (请自行证明)。

(2) [充分性]

- 假设有

$$f_{XY}\left(x_i,y_j\right)=f_X(x_i)f_Y(y_j),\text{对所有 } i,j=1,2,\cdots$$

- 进一步假设 $x_i\le x<x_{i+1},y_i\le y<y_{i+1}$。

CONTINUE

证明 (Cont.)：

- 则有

$$
\begin{aligned}
F_{XY}(x,y) &= P(X \le x, Y \le y) \\
&= \sum_{i'=1}^{i} \sum_{j'=1}^{j} f_{XY}\left(x_{i'}, y_{j'}\right) \\
&= \sum_{i'=1}^{i} f_X\left(x_{i'}\right) \sum_{j'=1}^{j} f_Y\left(y_{j'}\right) \\
&= F_X(x) F_Y(y)
\end{aligned}
$$

- 由于 i, j 为任意值，故 x 和 y 也为任意值。因此，对 xy 平面上的所有 (x,y)，有 $F_{XY}(x,y) = F_X(x)F_Y(y)$。

证毕。

定理 5.3

假设 X 和 Y 为两个连续随机变量，则 X 和 Y 相互独立，当且仅当

$$f_{XY}(x,y)=f_X(x)f_Y(y),\text{对所有 }(x,y)\in\mathbb{R}^2$$

其中 $f_{XY}(x,y)$，$f_X(x)$ 和 $f_Y(y)$ 为联合和边际 PDF。

证明：

(1) [必要性] 首先证明若 (X,Y) 相互独立，则对所有 $(x,y)\in\mathbb{R}^2$，有 $f_{XY}(x,y)=f_X(x)f_Y(y)$。

- 假设 (X,Y) 相互独立，根据定义可知

$$F_{XY}(x,y)=F_X(x)F_Y(y), \text{对所有 } x,y$$

- 对上式两边分别关于 x 和 y 求偏导数，可得

$$\frac{\partial^2 F_{XY}(x,y)}{\partial x\partial y}=\frac{\partial^2 F_X(x)F_Y(y)}{\partial x\partial y}$$
$$=\frac{\partial F_X(x)}{\partial x}\frac{\partial F_Y(y)}{\partial y}$$

- 这意味着

$$f_{XY}(x,y)=f_X(x)f_Y(y), \text{对所有 } (x,y)\in\mathbb{R}^2$$

CONTINUE

证明 (Cont.):

(2)[充分性] 需要证明当 $f_{XY}(x,y)=f_X(x)f_Y(y)$ 时，X 和 Y 相互独立。

- 假设对任意 $(u,v)\in\mathbb{R}^2$，有 $f_{XY}(u,v)=f_X(u)f_Y(v)$，则对其求积分可得

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

- 即 $F_{XY}(x,y)=F_X(x)F_Y(y)$，对所有 $-\infty<x,y<\infty$
- 根据定义可知 X 和 Y 相互独立。

证毕。

例 5.12：

假设 $f_{XY}(x,y)=4xy$，其中 $0\le x\le 1$ 和 $0\le y\le 1$。那么 X 和 Y 是否相互独立?

解： 在例 5.8 中，已求得

- ✓ 对 $0\le x\le 1$，有 $f_X(x)=2x$；
- ✓ 对 $0\le y\le 1$，有 $f_Y(y)=2y$。

- 因此 $f_X(x)f_Y(y)=4xy=f_{XY}(x,y), 0\le x\le 1, 0\le y\le 1$
- 同时，对所有由 $0\le x\le 1$，$0\le y\le 1$ 定义的矩形区域外的 (x,y)，有 $f_{XY}(xy)=f_X(x)f_Y(y)=0$。
- 由此可知，对 xy 平面上所有的 (x,y)，$f_{XY}(xy)=f_X(x)f_Y(y)$，即 X 和 Y 相互独立。

例 5.13：

假设 $f_{XY}(x,y)=8xy$，$0\le x\le y\le 1$。X 和 Y 是否相互独立?

解：

- 支撑 $\Omega_{XY}=\{(x,y):0\le x\le y\le 1\}$ 为上三角区域。
- 分别对 y 和 x 求积分，可得边际 PDF
  - ✓ $f_X(x)=4x(1-x^2),0\le x\le 1$
  - ✓ $f_Y(y)=4y^3,0\le y\le 1$
- 由于当 $0\le x\le y\le 1$ 时，$f_X(x)f_Y(y)\ne f_{XY}(x,y)$，因此，$X$ 和 Y 并不独立。

定理 5.4 [因子分解定理 (Factorization Theorem)]

两个随机变量 X 和 Y 相互独立，当且仅当联合 PMF/PDF 可写成

$$f_{XY}(x,y) = g(x)h(y), \text{对所有 } -\infty < x,y < \infty。$$

证明： 此处只证明连续情形。离散情形的证明类似可得。

(1)

- 若 X 和 Y 相互独立，则

$$f_{XY}(x,y) = f_X(x)f_Y(y) = g(x)h(y), \text{对所有} -\infty < x,\ y < \infty。$$

其中，令 $g(x) = f_X(x)$ 和 $h(y) = f_Y(y)$，即可得证。

CONTINUE

证明 (Cont.):

(2)

- 现假设存在某些函数 $g(\cdot)$ 和 $h(\cdot)$，使得

$$f_{XY}(x,y) = g(x)h(y), \text{对所有 } -\infty < x,\ y < \infty$$

- 则可得边际 PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$
$$= \int_{-\infty}^{\infty} g(x)h(y)dy$$
$$= g(x)\int_{-\infty}^{\infty} h(y)dy$$
$$f_Y(y) = h(y)\int_{-\infty}^{\infty} g(x)dx$$

CONTINUE

证明 (Cont.):

- 因此,

$$f_X(x)f_Y(y) = \left[g(x)\int_{-\infty}^{\infty} h(v)dv\right]\left[h(y)\int_{-\infty}^{\infty} g(u)du\right]$$
$$= g(x)h(y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)h(v)dudv$$
$$= f_{XY}(x,y),\ \text{对所有}\ -\infty < x,y < \infty$$

- 上述推导中使用了如下等式

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(u,v)dudv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)h(v)dudv = 1$$

证毕。

定理 5.5

两个随机变量 X 和 Y 相互独立，当且仅当联合 PMF/PDF 可写成

$$f_{Y|X}(y \mid x) = f_Y(y), \text{对所有}(x, y) \in \mathbb{R}^2$$

其中 $f_X(x) > 0$；并且

$$f_{X|Y}(x \mid y) = f_X(x), \text{对所有}(x, y) \in \mathbb{R}^2$$

其中 $f_Y(y) > 0$。

证明： 若 X 和 Y 相互独立，对 xy 平面上的所有 (x, y)，有

$f_{XY}(x, y) = f_X(x) f_Y(y)$。

CONTINUE

证明 (Cont.):

- 因此，

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y),\text{ 对所有 } (x,y) \in \mathbb{R}^2$$

其中 $f_X(x) > 0$。

- 类似有

$$f_{X|Y}(x \mid y) = f_X(x),\text{ 对所有 } (x,y) \in \mathbb{R}^2$$

其中 $f_Y(y) > 0$。

证毕。

例 5.14：

- 假设两个随机变量 X 和 Y 的联合 PDF 为

$$f_{XY}(x,y)=\begin{cases} e^{-y}, & 0<x<y<\infty \\ 0, & \text{其他} \end{cases}$$

(1) 求 $f_X(x)$ 和 $f_Y(y)$；

(2) 求 $f_{Y|X}(y \mid x)$ 和 $f_{X|Y}(x \mid y)$；

(3) 检验 X 和 Y 是否相互独立。

CONTINUE

解:

- 支撑 $\Omega_{XY} = \{(x, y): 0 < x < y < \infty\}$ 为一个上三角区域,

(1) X 的支撑为 $0 < x < \infty$。由定义

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy$$

- 对于 $x \le 0$，有 $f_{XY}(x, y) = 0$。故而对 $x \le 0$，有 $f_X(x) = 0$。
- 对 $x > 0$,

$$f_X(x) = \int_x^{\infty} e^{-y} dy$$
$$= e^{-x}$$

- 因此

$$f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

- 需要注意，不论 X 和 Y 是否相互独立，边际 PDF $f_X(x)$ 都与 y 无关。

CONTINUE

解 (Cont.):

- 下面计算 $f_Y(y)$。Y 的支撑为 $0 < y < \infty$。
- 由于对 $y \le 0$，有 $f_{XY}(x,y) = 0$，所以对 $y \le 0$，$f_Y(y) = 0$。
- 对 $y > 0$，

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx \\ &= \int_0^y e^{-y}dx \\ &= ye^{-y} \end{aligned}$$

- 因此

$$f_Y(y) = \begin{cases} ye^{-y}, & y > 0 \\ 0, & \text{其他} \end{cases}$$

CONTINUE

解 (Cont.):

(2) 首先计算 $f_{Y|X}(y \mid x)$。因为对 $x \le 0$，有 $f_X(x) = 0$，条件 PDF $f_{Y|X}(y \mid x)$ 仅对任意给定 $x > 0$ 有定义。

- 给定任意的 $x > 0$，有

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$
$$= \frac{e^{-y}}{e^{-x}}, 0 < x < y < \infty$$

- 因此

$$f_{Y|X}(y \mid x) = \begin{cases} e^{-(y-x)}, & y \in (x, \infty) \\ 0, & y \in (-\infty, x] \end{cases}$$

- 上式意味着，Y 基于 $X = x$ 的条件分布为支撑为 $y \in (x, \infty)$ 的指数分布。

CONTINUE

解 (Cont.):

- 以下计算 $f_{X|Y}(x|\,y)$。因为 $y \le 0$ 时，$f_Y(y) = 0$，条件 PDF $f_{X|Y}(x|\,y)$ 仅对任意给定 $y > 0$ 有定义。根据定义可得

$$
\begin{aligned}
f_{X|Y}(x \mid y) &= \frac{f_{XY}(x, y)}{f_Y(y)} \\
&= \frac{e^{-y}}{ye^{-y}} \\
&= \frac{1}{y}, x \in (0, y)
\end{aligned}
$$

CONTINUE

解 (Cont.):

- 因此，对任意给定 $y > 0$，

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{1}{y}, & x \in (0, y) \\ 0, & \text{其他} \end{cases}$$

- 这意味着 X 基于 $Y = y$ 服从区间 $(0, y)$ 上的均匀分布。

(3) 由于对 $0 < x < y < \infty$，$f_{X|Y}(x \mid y) \neq f_X(x)$，故 X 和 Y 并非相互独立。

定义 5.12

对于 n 个随机变量 $X_1,\cdots,X_n$，若联合 CDF 等于边际 CDF 的乘积

$$F_{\boldsymbol{X}^n}(\boldsymbol{x}^n)=\prod_{i=1}^{n}F_{X_i}(x_i),\ \text{对所有}\ -\infty<x_1,\cdots,x_n<\infty$$

则其相互独立，其中 $\boldsymbol{X}^n=(X_1,\cdots,X_n)'$ 并且 $\boldsymbol{x}^n=(x_1,\cdots,x_n)'$。

例 5.15：

- 假设随机变量 X_1, X_2, X_3 的联合 PDF 为

$$f_{X_1X_2X_3}(x_1,x_2,x_3)=\begin{cases}(x_1+x_2)e^{-x_3}, & 0<x_1<1, 0<x_2<1, x_3>0\\ 0, & \text{其他}\end{cases}$$

- 可以证明 X_1，X_2 和 X_3 两两独立 (pairwise independent)，但三者并非联合独立 (jointly independent)。
- 这与第二章例 2.58 中三个事件两两独立，但并非联合独立的情形类似。
- **何时两两独立意味着联合独立呢?**
  - ✓ 若存在此类情形，将大大简化对多维情形下相互独立性的判断。
  - ✓ 一个例子是多维正态分布，其两两独立意味着联合独立。参见本章第六节和第八节的讨论。

目 录

问题

现在假设有二元变换 (bivariate transformation)

$$\begin{cases} U = g_1(X,Y) \\ V = g_2(X,Y) \end{cases}$$

其中 (X,Y) 的联合 PMF/PDF $f_{XY}(x,y)$ 已给定。那么如何求 (U,V) 的联合 PMF/PDF $f_{UV}(u,v)$?

二元变换在经济学和金融学的重要性

例 5.16：

假设 X 和 Y 是引起欧元汇率 U 和日元汇率 V 波动的两个共同因素。那么，U 和 V 均为 X 和 Y 的函数。

例 5.17：

某商品的需求 U 和供给 V 均是商品价格 X 和消费者收入 Y 的函数。

n 元变换

- 对两个以上随机变量的情形，可推导类似的多元变换定理。

$$U_1 = g_1(X_1, X_2, \dots, X_n)$$
$$U_2 = g_2(X_1, X_2, \dots, X_n)$$
$$\dots$$
$$U_n = g_n(X_1, X_2, \dots, X_n)$$

- 给定联合分布 $f_{X_1 \cdots X_n}(x_1, x_2, \cdots, x_n)$，我们可以求得联合分布 $f_{U_1 \dots U_n}(u_1, u_2, \dots, u_n)$。

n 元变换 (Cont.)

- 通常我们不关心联合分布 $f_{U_1\ldots U_n}(u_1,u_2,\ldots,u_n)$，我们只关注一个变量的分布。例如，一个估计量或检验统计量通常表示为随机样本的一个函数 $U_1=g_1(X_1,X_2,\ldots,X_n)$，知道该估计量或检验统计量的分布有助于我们建立统计推断理论。
- 因此，得到联合分布后可以求得 U_1 的边际分布。

$$f_{U_1}(u_1)=\int\cdots\int f_{U_1\ldots U_n}(u_1,\ldots,u_n)du_2\cdots du_n$$

定义 5.13 [雅可比矩阵 (Jacobian Matrix) 和雅可比 (Jacobian)]

- 对于二元变换

$$\begin{cases} U = g_1(X,Y) \\ V = g_2(X,Y) \end{cases}$$

其中函数 $g_1(\cdot,\cdot)$ 和 $g_2(\cdot,\cdot)$ 关于 (x,y) 连续可导。

- 则 2×2 维矩阵

$$J_{UV}(x,y) = \begin{bmatrix} \dfrac{\partial g_1(x,y)}{\partial x} & \dfrac{\partial g_1(x,y)}{\partial y} \\ \dfrac{\partial g_2(x,y)}{\partial x} & \dfrac{\partial g_2(x,y)}{\partial y} \end{bmatrix}$$

称为 (U,V) 的雅可比矩阵，其行列式称为 (U,V) 的雅可比。

定义 5.14 [反函数 (Inverse Function)]

- 假设 A 和 B 是 $\mathbb{R}^2$ 的子集，$g_1\colon A\to B$ 和 $g_2\colon A\to B$ 是连续可导函数，且 $J_{UV}(x,y)$ 的行列式对任意 $(x,y)\in A$ 非零。则存在如下函数

$$\begin{cases} X = h_1(U,V) \\ Y = h_2(U,V) \end{cases}$$

- 其中 $h_1\colon B\to A$ 和 $h_2\colon B\to A$ 是 B 上的连续可导函数，且满足条件

$$h_1[g_1(x,y),g_2(x,y)] = x$$
$$h_2[g_1(x,y),g_2(x,y)] = y$$

- 则向量函数 $\{h_1(\cdot),\ h_2(\cdot)\}$ 称为向量函数 $\{g_1(\cdot),\ g_2(\cdot)\}$ 的反函数。

定理 5.6：

(X,Y) 的雅可比矩阵为

$$J_{XY}(u,v)=\begin{bmatrix}\dfrac{\partial h_1(u,v)}{\partial u} & \dfrac{\partial h_1(u,v)}{\partial v}\\ \dfrac{\partial h_2(u,v)}{\partial u} & \dfrac{\partial h_2(u,v)}{\partial v}\end{bmatrix}$$
$$=J_{UV}^{-1}(x,y)$$

其中 $x=h_1(u,v)$，$y=h_2(u,v)$。

CONTINUE

证明:

- 由反函数定义，已知如下恒等式成立

$$
\begin{gathered}
h_1[g_1(x,y),g_2(x,y)]=x \\
h_2[g_1(x,y),g_2(x,y)]=y
\end{gathered}
$$

- 令 $u=g_1(x,y)$，$v=g_2(x,y)$，并且对第一个恒等式分别关于 x 和 y 求微分可得

$$
\begin{gathered}
\frac{\partial h_1(u,v)}{\partial u}\frac{\partial g_1(x,y)}{\partial x}+\frac{\partial h_1(u,v)}{\partial v}\frac{\partial g_2(x,y)}{\partial x}=1 \\
\frac{\partial h_1(u,v)}{\partial u}\frac{\partial g_1(x,y)}{\partial y}+\frac{\partial h_1(u,v)}{\partial v}\frac{\partial g_2(x,y)}{\partial y}=0
\end{gathered}
$$

CONTINUE

证明 (Cont.):

- 类似地，对第二个恒等式分别关于 x 和 y 求微分可得

$$\frac{\partial h_2(u,v)}{\partial u}\frac{\partial g_1(x,y)}{\partial x}+\frac{\partial h_2(u,v)}{\partial v}\frac{\partial g_2(x,y)}{\partial x}=0$$

$$\frac{\partial h_2(u,v)}{\partial u}\frac{\partial g_1(x,y)}{\partial y}+\frac{\partial h_2(u,v)}{\partial v}\frac{\partial g_2(x,y)}{\partial y}=1$$

- 将上述四个微分方程表示成矩阵形式，可得

$$\begin{bmatrix}\frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v}\\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v}\end{bmatrix}\begin{bmatrix}\frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_1(x,y)}{\partial y}\\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial y}\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

CONTINUE

证明 (Cont.):

- 或

$$J_{XY}(u,v)J_{UV}(x,y) = I$$

- 其中 I 为 2×2 的单位矩阵。由于 $J_{UV}(x,y)$ 对所有 $(x,y) \in A$ 均是非奇异的，对上式右乘 $J_{UV}(x,y)$ 的逆，可得

$$J_{XY}(u,v) = J_{UV}^{-1}(x,y)$$

证毕。

例 5.19：

两个随机变量 $(U,V):\mathbb{R}^2\to\mathbb{R}^2$ 分别定义为 $U\equiv g_1(x,y)=XY$ 和 $V\equiv g_2(x,y)=X$。则 $g:\mathbb{R}^2\to\mathbb{R}^2$ 为一一映射（one-to-one mapping)，且具有反函数 $h:\mathbb{R}^2\to\mathbb{R}^2$。

(1) 求反函数;

(2) 证明 $J_{XY}(u,v)=J_{UV}^{-1}(x,y)$，其中 $u=xy$ 与 $v=x$。

CONTINUE

解：

(1) 给定

$$U = XY = g_1(X,Y)$$
$$V = X = g_2(X,Y)$$

有反函数

$$X = h_1(U,V) = V$$
$$Y = h_2(U,V) = \frac{U}{V}$$

CONTINUE

解 (Cont.):

(2) 根据定义，反函数的雅可比矩阵为

$$J_{XY}(u,v)=\begin{bmatrix}\dfrac{\partial h_1(u,v)}{\partial u} & \dfrac{\partial h_1(u,v)}{\partial v}\\ \dfrac{\partial h_2(u,v)}{\partial u} & \dfrac{\partial h_2(u,v)}{\partial v}\end{bmatrix}$$

$$=\begin{bmatrix}0 & 1\\ \dfrac{1}{v} & -\dfrac{u}{v^2}\end{bmatrix}$$

CONTINUE

解 (Cont.):

- 进一步地，原函数 $U=g_1(X,Y)=XY$ ，$V=g_2(X,Y)=X$ 的雅可比矩阵为

$$J_{UV}(x,y)=\begin{bmatrix}\frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_1(x,y)}{\partial y}\\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial y}\end{bmatrix}$$

$$=\begin{bmatrix}y & x\\ 1 & 0\end{bmatrix}$$

CONTINUE

解 (Cont.):

- 其逆矩阵为

$$
\begin{aligned}
J_{UV}^{-1}(x,y) &= -\frac{1}{x}\begin{bmatrix} 0 & -x \\ -1 & y \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 \\ \dfrac{1}{x} & -\dfrac{y}{x} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 \\ \dfrac{1}{v} & -\dfrac{u}{v^2} \end{bmatrix} \\
&= J_{XY}(u,v)
\end{aligned}
$$

其中 $u = xy$ 和 $v = x$。

◆ 问题

现在考察更为一般的情况：假设 (X,Y) 的联合 PDF 为 $f_{XY}(x,y)$，且

$$\begin{cases} U = g_1(X,Y) \\ V = g_2(X,Y) \end{cases}$$

那么，如何求 (U,V) 的联合 PDF $f_{UV}(u,v)$?

定理 5.7 [二元变换 (Bivariate Transformation)]

- 假设二元连续型随机变量 (X,Y) 的联合 PDF 为 $f_{XY}(x,y)$，并记 (X,Y) 的支撑为 $\Omega_{XY}=\{(x,y)\in\mathbb{R}^2: f_{XY}(x,y)>0\}$。定义

$$\begin{cases} U=g_1(X,Y) \\ V=g_2(X,Y) \end{cases}$$

- 其中，$g\colon\Omega_{XY}\to\mathbb{R}^2$ 为一一映射、在 Ω_{XY} 上连续可导，且对所有 $(x,y)\in\Omega_{XY}$，有行列式 $\det[J_{UV}(x,y)]\neq 0$。
- 则 (U,V) 的联合 PDF 为

$$f_{UV}(u,v)=f_{XY}(x,y)\left|\det[J_{XY}(u,v)]\right|, \text{对所有 } (u,v)\in\Omega_{UV}$$

- 其中 $x=h_1(u,v)$，$y=h_2(u,v)$，且 $\Omega_{UV}=\left\{(u,v)\in\mathbb{R}^2: u=g_1(x,y), v=g_2(x,y), \text{对所有 } (x,y)\in\Omega_{XY}\right\}$ 为 (U,V) 的支撑。

证明:

- 对 (U,V) 支撑 Ω_{UV} 上的任意 (u,v)，有

$$
\begin{aligned}
F_{UV}(u,v) &= P(U \le u, V \le v) \\
&= P[g_1(X,Y) \le u, g_2(X,Y) \le v] \\
&= \iint_{\mathbb{A}(u,v)} f_{XY}(x',y')dx'dy'
\end{aligned}
$$

- 其中，二重积分在集合 $A(u,v) = \{(x,y) \in \mathbb{R}^2 : g_1(x,y) \le u$ 和 $g_2(x,y) \le v\}$ 上运算。
- 令变量变换 $s = g_1(x',y')$ 和 $t = g_2(x',y')$，并且运用二重积分的变量替换公式，得

$$
F_{UV}(u,v) = \int_{-\infty}^{v}\int_{-\infty}^{u} \frac{1}{|\det[J_{UV}(x',y')]|} f_{XY}(x',y')dsdt
$$

证明 (Cont.):

- 其中 (x', y') 满足约束 $g_1(x', y') = s$ 和 $g_2(x', y') = t$。对上式两边关于 (u, v) 求偏导，可得

$$
\begin{aligned}
f_{UV}(u, v) &= \frac{\partial^2 F_{UV}(u, v)}{\partial u \partial v} \\
&= f_{XY}(x, y) \frac{1}{|\det[J_{UV}(x, y)]|}
\end{aligned}
$$

- 其中 $x = h_1(u, v)$，$y = h_2(u, v)$。根据 $J_{UV}(x, y) = J_{XY}^{-1}(u, v)$，同样有

$$
f_{UV}(u, v) = f_{XY}(x, y) |\det[J_{XY}(u, v)]|
$$

证毕。

定理 5.7 含义

- 为更好地理解二元变换定理，考察事件 $u-\frac{\epsilon}{2}<U\leq u+\frac{\epsilon}{2}$ 和 $v-\frac{\epsilon}{2}<V\leq v+\frac{\epsilon}{2}$，其中常数 $\epsilon>0$。该事件指的是 (U,V) 在以 (u,v) 为中心的矩形内取值，边长均为 ϵ。当 ϵ 很小时，其概率约等于 $f_{UV}(u,v)\epsilon^2$。
- 接着，考察等价事件 $x-\frac{\delta}{2}<X\leq x+\frac{\delta}{2}$ 与 $y-\frac{\delta}{2}<Y\leq y+\frac{\delta}{2}$，其中 $x=h_1(u,v)$, $y=h_2(u,v)$, $\delta^2=|\det[J_{XY}(u,v)]|\epsilon^2$。

定理 5.7 含义 (Cont.)

- 等价指的是事件 (U,V) 在以 (u,v) 为中心、边长为 ϵ 的矩形内取值，对应于事件 (X,Y) 在以 (x,y) 为中心、边长为 δ 的矩形内取值。
- 由于 (U,V) 在以 (u,v) 为中心的第一个矩形内取值的概率等于 (X,Y) 在以 (x,y) 为中心的第二个矩形内取值的概率，可知当 ϵ 充分小时，

$$\begin{aligned} f_{UV}(u,v)\epsilon^2 &\sim f_{XY}(x,y)\delta^2 \\ &\sim f_{XY}(x,y)|\det[J_{XY}(u,v)]|\epsilon^2 \end{aligned}$$

则

$$f_{UV}(u,v) = f_{XY}(x,y)|\det[J_{XY}(u,v)]|$$

定理 5.7 含义 (Cont.)

- 第三章讨论过，对一元变换 $Y = g(X)$，其中 $g(X)$ 为连续可导单调函数，其 PDF 为 $f_Y(y) = f_X(x)|h'(y)| = f_X(x)|g'(x)|^{-1}$ 其中 $x = h(y)$ 为 $y = g(x)$ 的反函数。具体参见第三章的定理 3.12。
- 因此，可将二元变换定理视为一元变换定理的扩展。事实上，对两个以上随机变量的情形，可推导类似的多元变换定理。
- 二元变换定理十分重要，可用于推导一元分布和二元分布。
- 需要强调，一一映射是应用二元变换的关键条件，这与一元变换要求函数 $g(\cdot)$ 为单调变换是类似的。

◆ **问题 5.7**

若 $g(\cdot)$ 是 (X,Y) 的支撑 A 上的多对一函数，又会怎么样呢?

- 若 $g(\cdot)$ 不是一一映射，则上述二元变换定理不适用。
- 然而，若 (X,Y) 的支撑 A 可分解为 k 个互斥子集 $\{A_i\}_{i=1}^{k}$，其中 $i=1,\cdots,k$，则对于 A_i 上的 (x,y)，有 $g(x,y)=g_i(x,y)$，其中 $g_i(x,y)$ 是 A_i 上的一一映射。
- 与第三章的定理 3.13 的一元情形类似，可证明 $f_{UV}(u,v)$ 等于 $f_{XY}(x,y)$ 在 k 个子集 $\{A_i\}$ 上各自变换的总和。

定理 5.8

假设 $U=g_1(X)$ 和 $V=g_2(Y)$ 是连续可导一一映射的可测函数。则 X 和 Y 相互独立，当且仅当 U 和 V 相互独立。

证明：(1) [必要性] 首先证明 X 和 Y 相互独立意味着 U 和 V 相互独立。根据定义可知，雅可比矩阵为

$$J_{UV}(x,y)=\begin{bmatrix} g_1'(x) & 0 \\ 0 & g_2'(y) \end{bmatrix}$$

- 雅可比为

$$\det[J_{UV}(x,y)]=g_1'(x)g_2'(y)$$

证明 (Cont.)：

- 根据二元变换定理，有

$$\begin{aligned} f_{UV}(u,v) &= f_{XY}(x,y)|\det[J_{UV}(x,y)]|^{-1} \\ &= f_X(x)f_Y(y)|g_1'(x)g_2'(y)|^{-1} \\ &= [f_X(x)|g_1'(x)|^{-1}][f_Y(y)|g_2'(y)|^{-1}] \\ &= f_U(u)f_V(v) \end{aligned}$$

- 其中，$x = g_1^{-1}(u)$ 和 $y = g_2^{-1}(v)$ 分别为 $g_1(x)$ 和 $g_2(y)$ 的反函数。因为 (u,v) 为任意值，故有 U 和 V 相互独立。

(2) [充分性] 其次证明 U 和 V 相互独立意味着 X 和 Y 相互独立。

- 对二元变换 $X = g_1^{-1}(U)$ 和 $Y = g_2^{-1}(V)$ 应用二元变换定理，参照与 (1) 类似的推导过程即可证明。

证毕。

例 5.20：

- 假设随机变量 X ~贝塔分布 $BETA(\alpha,\beta)$，其 PDF 为
$$f_X(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},0<x<1$$
- 随机变量 Y ~贝塔分布 $BETA(\alpha+\beta,\gamma)$，其 PDF 为
$$f_Y(y)=\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}y^{\alpha+\beta-1}(1-y)^{\gamma-1},0<y<1$$
- 并且 X 和 Y 相互独立。
- 定义 $U=XY$，$V=X$，求二者的联合 PDF $f_{UV}(u,v)$。

CONTINUE

解:

(1) 首先，(X,Y) 的支撑 Ω_{XY} 为一个矩形区域:

$$\begin{aligned}\Omega_{XY} &= \{(x,y)\in\mathbb{R}^2: f_{XY}(x,y)=f_X(x)f_Y(y)>0\}\\ &= \{(x,y)\in\mathbb{R}^2: 0<x<1, 0<y<1\}\end{aligned}$$

- 给定 $U=XY$ 和 $V=X$，求得 (U,V) 的支撑 Ω_{UV} 为

$$\begin{aligned}\Omega_{UV} &= \{(u,v)\in\mathbb{R}^2: f_{UV}(u,v)>0\}\\ &= \{(u,v)\in\mathbb{R}^2: 0<u<v, 0<v<1\}\\ &= \{(u,v)\in\mathbb{R}^2: 0<u<v<1\}\end{aligned}$$

(2) 以下求解反函数 $x=h_1(u,v)$ 和 $y=h_2(u,v)$。给定 $u=xy$，$v=x$，有

$$x=h_1(u,v)=v$$
$$y=h_2(u,v)=\frac{u}{v}$$

CONTINUE

解 (Cont.)：

(3) (X, Y) 的雅可比矩阵为

$$J_{XY}(u,v) = \begin{bmatrix} \dfrac{\partial h_1(u,v)}{\partial u} & \dfrac{\partial h_1(u,v)}{\partial v} \\ \dfrac{\partial h_2(u,v)}{\partial u} & \dfrac{\partial h_2(u,v)}{\partial v} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 \\ \dfrac{1}{v} & -\dfrac{u}{v^2} \end{bmatrix}$$

- 因此，雅可比为

$$\det[J_{XY}(u,v)] = \det\left(\begin{bmatrix} 0 & 1 \\ v^{-1} & -u/v^2 \end{bmatrix}\right) = -\frac{1}{v}$$

CONTINUE

解 (Cont.):

- 由二元变换定理，得

$$
\begin{aligned}
f_{UV}(u,v) &= f_{XY}(x,y)|\det[J_{XY}(u,v)]| \\
&= f_X(x)f_Y(y)\left|-\frac{1}{v}\right| \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta)\Gamma(\gamma)}y^{\alpha+\beta-1}(1-y)^{\gamma-1}\frac{1}{v} \\
&= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}v^{\alpha-1}(1-v)^{\beta-1}\left(\frac{u}{v}\right)^{\alpha+\beta-1}\left(1-\frac{u}{v}\right)^{\gamma-1}\frac{1}{v}
\end{aligned}
$$

其中 $0 < u < v < 1$，$x = v$，$y = \frac{u}{v}$。

例 5.21：

假设 X 和 Y 为相互独立的 $N(0,\sigma^2)$ 随机变量。定义 $U=X^2+Y^2$，$V=X/\sqrt{X^2+Y^2}$。

(1) 求 $f_{UV}(u,v)$；(2) 证明 U 和 V 相互独立。

解：(1) 由 $U=X^2+Y^2$， $V=X/\sqrt{X^2+Y^2}$ 和 $X\sim N(0,\sigma^2)$ ，$Y\sim N(0,\sigma^2)$ 可知，(U,V) 的支撑为

$$\Omega_{UV}=\{(u,v)\in\mathbb{R}^2: 0<u<\infty, -1<v<1\}$$

- 由于 (x,y) 和 $(x,-y)$ 对应相同的 (u,v) 值，(U,V) 并非一一映射，故无法直接应用二元变换定理。
- 令 $Z=Y^2$，则从 (X,Z) 到 (U,V) 为一一映射变换，因此，可应用二元变换定理。

解 (Cont.):

- 对任意 $z \geq 0$，$Z = Y^2$ 的分布为：

$$\begin{aligned} F_Z(z) &= P\left(Y^2 \leq z\right) \\ &= P(-\sqrt{z} \leq Y \leq \sqrt{z}) \\ &= F_Y(\sqrt{z}) - F_Y(-\sqrt{z}) \\ f_Z(z) &= F_Z'(z) \\ &= f_Y(\sqrt{z})\frac{1}{2\sqrt{z}} + f_Y(-\sqrt{z})\frac{1}{2\sqrt{z}} \\ &= \frac{1}{\sqrt{2\pi z}\sigma} e^{-z/2\sigma^2}, z \geq 0 \end{aligned}$$

- 由于 X 和 Y 相互独立，X 和 $Z = Y^2$ 也相互独立，所以 (X, Z) 的联合 PDF 为

$$f_{XZ}(x, z) = \frac{1}{2\pi\sigma^2\sqrt{z}} e^{-x^2/2\sigma^2} e^{-z/2\sigma^2}, -\infty < x < \infty, 0 \leq z < \infty$$

解 (Cont.):

- 另外，(X,Z) 的支撑为

$$\Omega_{XZ}(x,z)=\{(x,z)\in\mathbb{R}^2:-\infty<x<\infty,0\le z<\infty\}$$

- $U=X^2+Z$ 和 $V=X/\sqrt{X^2+Z}$ 的联合支撑为

$$\Omega_{UV}=\{(u,v)\in\mathbb{R}^2:0<u<\infty,-1<v<1\}$$

- 现在，求反函数 $X=h_1(U,V)$ 和 $Z=h_2(U,V)$。给定

$$U=X^2+Z,V=\frac{X}{\sqrt{X^2+Z}}$$

有

$$X=h_1(U,V)=V\sqrt{U}$$
$$Z=h_2(U,V)=U\left(1-V^2\right)$$

解 (Cont.):

- 则 (X,Z) 的雅可比矩阵为

$$J_{XZ}(u,v)=\begin{bmatrix}\frac{\partial h_1(u,v)}{\partial u} & \frac{\partial h_1(u,v)}{\partial v}\\ \frac{\partial h_2(u,v)}{\partial u} & \frac{\partial h_2(u,v)}{\partial v}\end{bmatrix}$$

$$=\begin{bmatrix}\frac{v}{2\sqrt{u}} & \sqrt{u}\\ 1-v^2 & -2uv\end{bmatrix}$$

解 (Cont.):

- 根据二元变换定理可得

$$
\begin{aligned}
f_{UV}(u,v) &= f_{XZ}(x,z)|\det[J_{XZ}(u,v)]| \\
&= \frac{1}{2\pi\sigma^2\sqrt{z}} e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{z}{2\sigma^2}} |\det[J_{XZ}(u,v)]| \\
&= \frac{1}{2\pi\sigma^2\sqrt{u(1-v^2)}} e^{-\frac{u}{2\sigma^2}} \sqrt{u} \\
&= \frac{1}{2\pi\sigma^2\sqrt{1-v^2}} e^{-\frac{u}{2\sigma^2}}, u>0, -1<v<1
\end{aligned}
$$

- 则

$$
f_{UV}(u,v) = \begin{cases} \frac{1}{2\pi\sigma^2\sqrt{1-v^2}} e^{-\frac{u}{2\sigma^2}}, & 0<u<\infty, -1<v<1 \\ 0, & \text{其他} \end{cases}
$$

解 (Cont.):

(2) 尽管 U 和 V 均为 (X,Y) 的函数，但对所有 $(u,v) \in \mathbb{R}^2$，其联合 PDF $f_{UV}(u,v)$ 可分解成如下两个分别关于 u 和 v 的函数的乘积，即

$$g(u) = \begin{cases} e^{-\frac{u}{2\sigma^2}}, & u > 0 \\ 0, & u \le 0 \end{cases}$$

和

$$h(v) = \begin{cases} \dfrac{1}{2\pi\sigma^2\sqrt{1-v^2}}, & -1 < v < 1 \\ 0, & \text{其他} \end{cases}$$

- 因此，根据因子分解定理 (定理 5.4) 可得 U 和 V 相互独立。

例 5.22：

假设 $X \sim N(\mu,\sigma^2)$，$Y \sim N(\mu,\sigma^2)$，并且 X, Y 相互独立。令 $U = X + Y$，$V = X - Y$。

(1) 求 (U,V) 的联合 PDF；(2) 证明 U 和 V 相互独立。

解：(1) (U,V) 的支撑为整个 xy 平面。根据 $U = X + Y$，$V = X - Y$，有 $X = h_1(U,V) = \frac{1}{2}(U+V)$，$Y = h_2(U,V) = \frac{1}{2}(U-V)$。$(X,Y)$ 的雅可比矩阵为

$$J_{XY}(u,v) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

解 (Cont.)：

- 根据二元变换定理（定理 5.7）以及 $x=\frac{1}{2}(u+v)$ 和 $y=\frac{1}{2}(u-v)$，可得

$$
\begin{aligned}
f_{UV}(u,v) &= f_{XY}(x,y)|\det[J_{XY}(u,v)]| \\
&= \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}[(x-\mu)^2+(y-\mu)^2]} \cdot \frac{1}{2} \\
&= \frac{1}{4\pi\sigma^2} e^{-\frac{1}{8\sigma^2}(u+v-2\mu)^2} e^{-\frac{1}{8\sigma^2}(u-v-2\mu)^2} \\
&= \frac{1}{4\pi\sigma^2} e^{-\frac{1}{4\sigma^2}(u-2\mu)^2} e^{-\frac{1}{4\sigma^2}v^2} \\
&= \frac{1}{\sqrt{2\pi 2\sigma^2}} e^{-\frac{1}{4\sigma^2}(u-2\mu)^2} \frac{1}{\sqrt{2\pi 2\sigma^2}} e^{-\frac{1}{4\sigma^2}v^2}, -\infty < u,v < \infty
\end{aligned}
$$

解 (Cont.):

(2) 从上述 $f_{UV}(u,v)$ 的表达式可得，$U \sim N(2\mu, 2\sigma^2)$，$V \sim N(0, 2\sigma^2)$。而且，其联合 PDF 可写为

$$f_{UV}(u,v) = g(u)h(v), -\infty < u, v < \infty$$

其中

$$g(u) = \frac{1}{\sqrt{2\pi 2\sigma^2}} e^{-\frac{1}{4\sigma^2}(u-2\mu)^2}, -\infty < u < \infty$$

为 $N(2\mu, 2\sigma^2)$ 的 PDF，而

$$h(v) = \frac{1}{\sqrt{2\pi 2\sigma^2}} e^{-\frac{1}{4\sigma^2}v^2}, -\infty < v < \infty$$

为 $N(0, 2\sigma^2)$ 的 PDF。

- 根据因子分解定理 (定理 5.4) 可得 U 和 V 相互独立。

例 5.23：

假设 $X \sim U[0,1]$，$Y \sim U[0,1]$，并且 X 和 Y 相互独立，求 $X+Y$ 的 PDF。

解：

- **(1)** 根据 $U = X+Y$，$V = X$，可得 (U,V) 的支撑为

$$\Omega_{UV} = \{(u,v) \in \mathbb{R}^2 : v \le u \le 1+v, 0 \le v \le 1\}$$

支撑 Ω_{UV} 如图 5.6 所示。

解 (Cont.)：

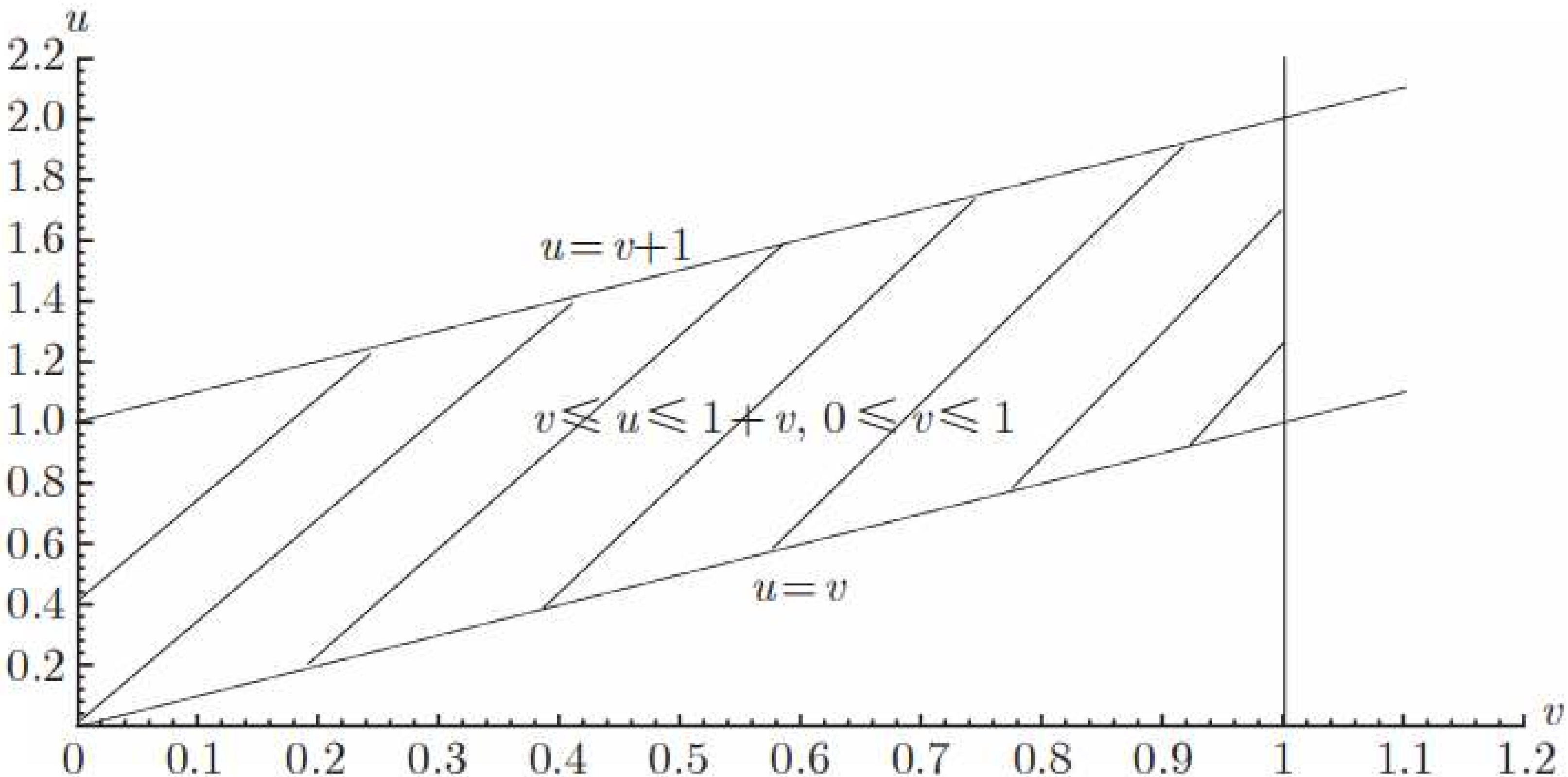


图 5.6：(U,V) 的支撑

解 (Cont.)：

- **(2)** (U,V) 的雅可比矩阵 $J_{UV}(x,y)$ 为

$$J_{UV}(x,y)=\begin{bmatrix}\frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_1(x,y)}{\partial y}\\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial y}\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}$$

- 因此，(U,V) 的雅可比为

$$\det[J_{UV}(x,y)]=-1$$

- **(3)** 根据二元变换定理，可得

$$\begin{aligned}f_{UV}(u,v)&=f_{XY}(x,y)|\det[J_{UV}(x,y)]|^{-1}\\&=f_X(x)f_Y(y)|-1|^{-1}\\&=1,\qquad 0\le v\le 1, v\le u\le 1+v\end{aligned}$$

解 (Cont.)：

- 对上式关于 v 求积分即可得 U 的 PDF $f_U(u)$。U 的支撑为 $0 \le u \le 2$。现将其分解为两个子区间：$0 \le u \le 1$ 和 $1 \le u \le 2$，并分别考虑。
  - ✓ 情形 (1)：对 $0 \le u \le 1$，有
$$f_U(u) = \int_0^u dv = u$$
  - ✓ 情形 (2)：对 $1 \le u \le 2$，有
$$f_U(u) = \int_{u-1}^1 dv$$
$$= 2 - u$$

解 (Cont.):

- 则

$$f_U(u) = \begin{cases} u, & 0 \le u \le 1 \\ 2-u, & 1 < u \le 2, \\ 0, & \text{其他} \end{cases}$$

这是一个在支撑 $u \in [0,2]$ 上，顶点位于 $u=1$ 处的三角形密度函数。

多元变换具有重要作用

- 如前文所述，可将**二元变换**进一步推广到涉及两个以上随机变量的**多元变换**。许多重要的统计量（例如参数估计量和检验统计量）一般都是两个以上随机变量的函数。因此，多元变换在求解和理解这些重要统计量的分布时具有十分重要的作用。
- 在实际应用中，多元变换往往十分繁琐，特别是涉及的随机变量个数较多时。幸运的是，在一些特殊情况下有其他更加简便的方法，例如可用**基于矩生成函数（moment generating function）的方法**求解所关注的统计量的概率分布，具体参见本章的后续示例。

目　录

定义 5.15 [二元正态分布 (Bivariate Normal Distribution)]：

- 若随机变量 (X,Y) 的联合 PDF 为

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]},$$

$$-\infty < x,y < \infty$$

  则称其服从联合正态分布，记作 $BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，其中 $|\rho| \le 1$。

- 当 $(\mu_1,\mu_2,\sigma_1,\sigma_2) = (0,0,1,1)$ 时，称 $BN(0,0,1,1,\rho)$ 为标准二元正态分布。

- 二元正态分布的联合 PDF $f_{XY}(x,y)$ 的另一种表述是

$$f_{XY}(x,y)=\frac{1}{\sqrt{(2\pi)^2\det(\Sigma)}}e^{-\frac{1}{2}(z-\mu)'\Sigma^{-1}(z-\mu)}$$

- 其中 $z=(x,y)'$，$\mu=(\mu_1,\mu_2)'$，且

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}=\begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}$$

$(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 在参数不同取值下的 $f_{XY}(x,y)$

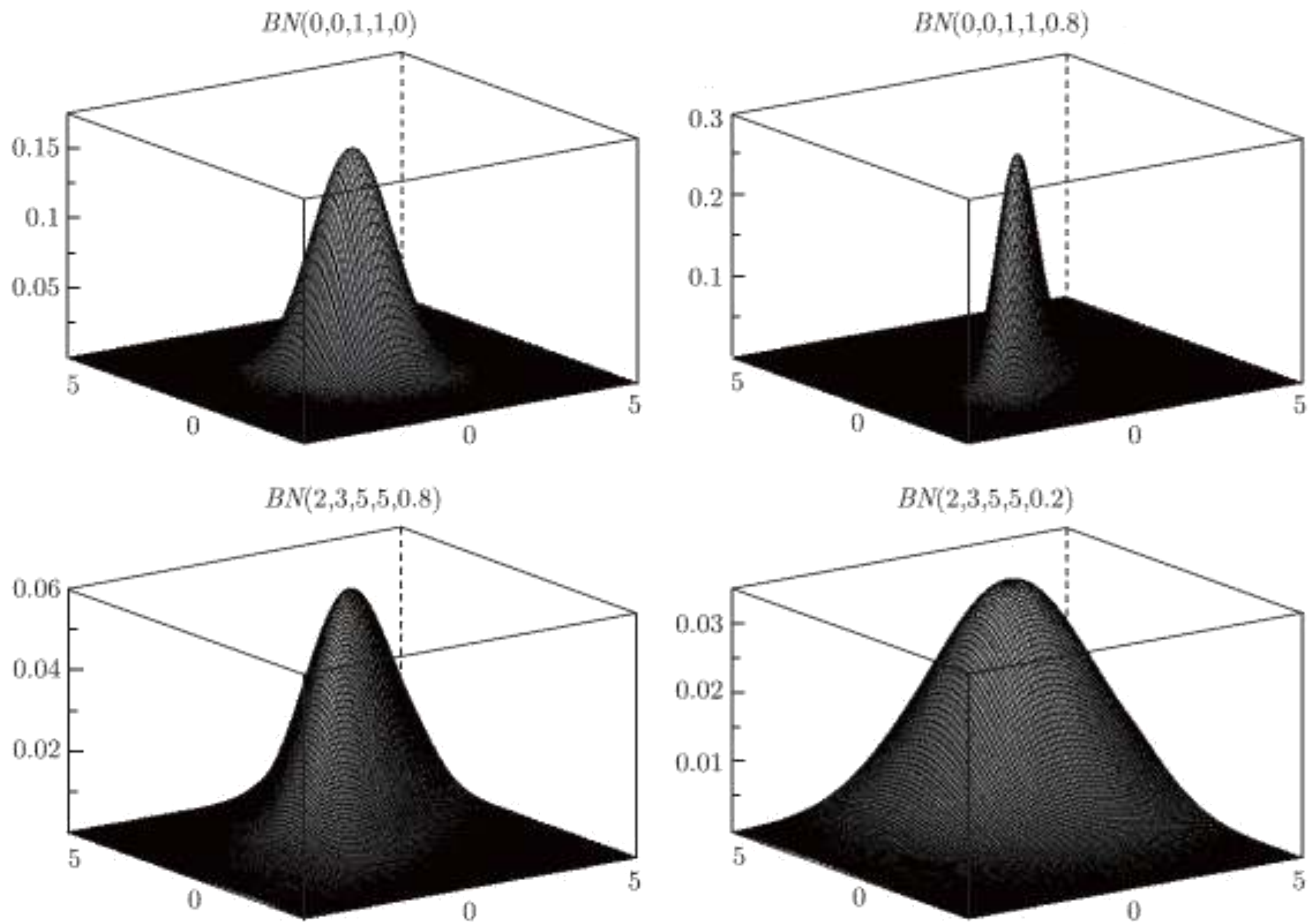


图 5.7：$(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的联合 PDF $f_{XY}(x,y)$

- 现在考察当 $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 时，如何求解 X 和 Y 的边际 PDF 和 Y 基于 $X=x$ 的条件 PDF 和 X 基于 $Y=y$ 的条件 PDF。
- 令

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

- 则二元正态分布的联合 PDF 可表述为

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}q(x,y)}$$

- 经简单代数运算，可得

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

$$=(1-\rho^2)\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu}{\sigma_2}\right)^2$$

其中 $\mu=\mu_2+\frac{\rho\sigma_2}{\sigma_1}(x-\mu_1)$

- 因此

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}e^{-\frac{(y-\mu)^2}{2\sigma_2^2(1-\rho^2)}}dy$$

$$=\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},-\infty<x<\infty$$

- 上式积分部分是对正态分布 $N[\mu,\sigma_2^2(1-\rho^2)]$ 的 PDF 之积分，故等于 1。
- 由对称性，同理可得

$$f_Y(y)=\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},-\infty<y<\infty$$

- 上述结果意味着当 X 和 Y 服从联合正态分布时，二者的边际分布均服从正态分布，即 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$。

- 当 $f_X(x) > 0$ 时，Y 基于 $X = x$ 的条件 PDF 为

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}} e^{-\frac{(y-\mu)^2}{2\sigma_2^2(1-\rho^2)}}, -\infty < y < \infty$$

其中 $\mu = \mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x - \mu_1)$。

- 因此，Y 基于 $X = x$ 的条件分布同样为正态分布

$$N[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)]$$

- 类似可得，X 基于 $Y = y$ 的条件分布为正态分布

$$N[\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)]$$

随机变量 $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的条件 PDF $f_{Y|X}(y|x)$

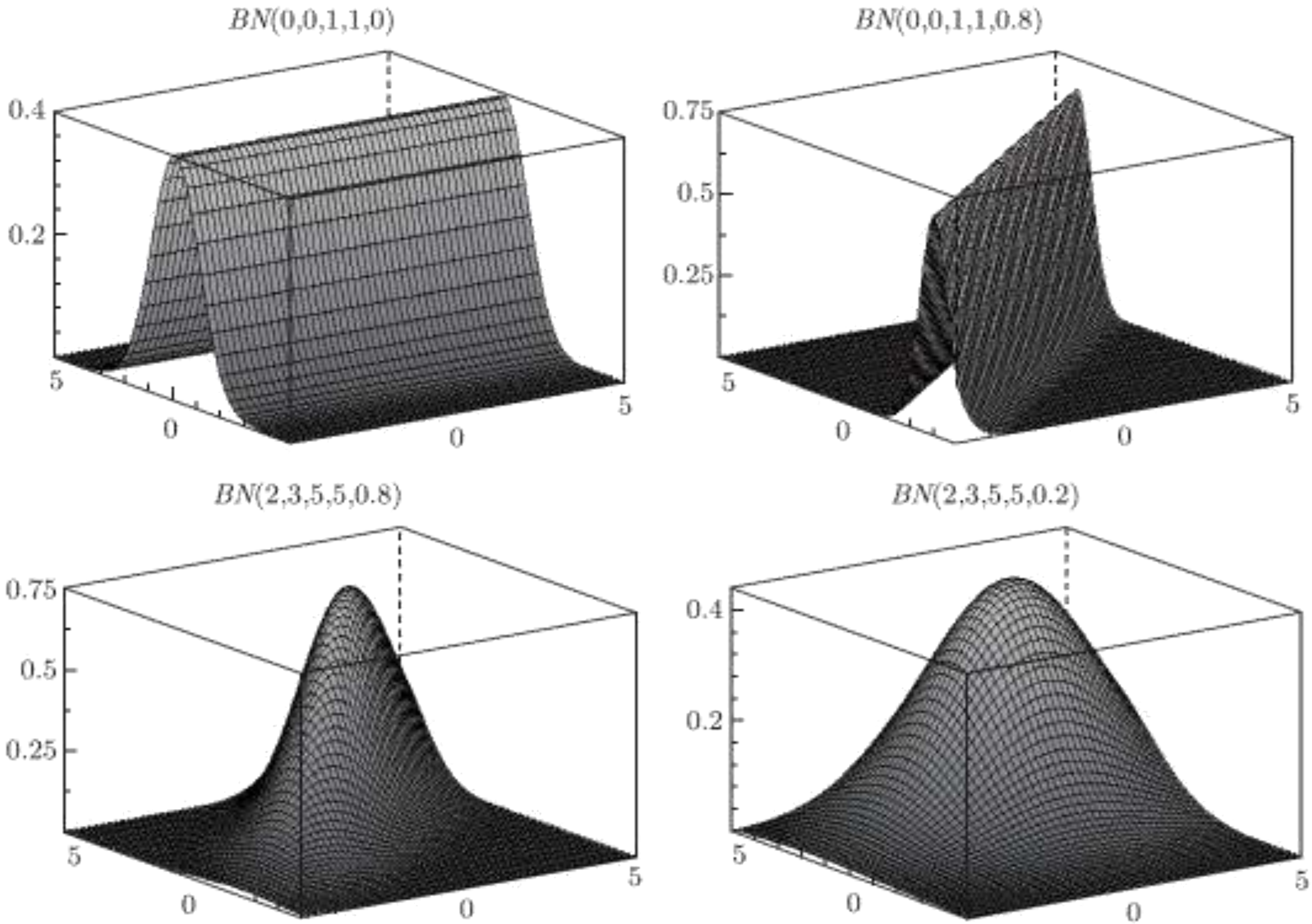


图 5.8： $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的条件 PDF $f_{Y|X}(y|x)$

◆ 问题 5.8

如何解释相依系数 (dependence parameter) ρ ?

- 常数 ρ 刻画了 X 和 Y 的**相依关系。**
- 当 $\rho = 0$ 时，(X, Y) 的联合 PDF 为

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2}$$

$$= f_X(x) f_Y(y), \text{ 对所有 } (x,y) \in \mathbb{R}^2$$

- 因此，当且仅当 $\rho = 0$ 时，两个服从联合正态分布的随机变量 X 和 Y 相互独立。

◆ **问题**

已经证明，两个联合正态随机变量的边际分布依然为正态分布。那么反之是否成立？即假设 X 和 Y 的边际分布均为正态分布，X 和 Y 是否服从联合正态分布？

- 答案是**否定**的。现在提供一个反例。

例 5.24：

假设两个随机变量 X 和 Y 的联合 PDF 为

$$f_{XY}(x,y)=\begin{cases}2f_X(x)f_Y(y), & xy>0\\ 0, & xy\le 0\end{cases}$$

其中

$$f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},f_Y(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$$

可以证明 X 和 Y 为正态随机变量，但二者显然不服从联合正态分布。

多元正态分布

- 若随机变量 $X_1,\cdots,X_n$ 的联合 PDF 为
$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n)=\frac{1}{\sqrt{(2\pi)^n\det(\Sigma)}}e^{-\frac{1}{2}(\boldsymbol{x}^n-\mu)'\Sigma^{-1}(\boldsymbol{x}^n-\mu)},\boldsymbol{x}^n\in\mathbb{R}^n$$
则称其服从联合正态分布，记作 $N(\mu,\Sigma)$，其中
  - ✓ $\boldsymbol{X}^n=(X_1,\cdots,X_n)'$，
  - ✓ $\boldsymbol{x}^n=(x_1,\cdots,x_n)'$，
  - ✓ $\mu=(\mu_1,\cdots,\mu_n)'$，
  - ✓ Σ 为 $n\times n$ 维对称正定矩阵。
- 矩阵 Σ 称为随机向量 $\boldsymbol{X}^n$ 的方差-协方差矩阵 (the variance-covariance matrix)，因为 Σ 的对角元素是 X_i 的方差，非对角元素是对于所有 $i\neq j$，X_i 与 X_j 之间的协方差。

二元变换的一个应用

- 令 $U = g_1(X_1), V = g_2(X_2)$。
- X 和 Y 相互独立当且仅当 U 和 V 相互独立。
- 一个想法是寻找一个特定函数 g_1 和 g_2 使得 U 和 V 服从二元正态分布。
- 进而，判断 U 和 V 的协方差是否为 0。若不为 0，则可以说明 X 和 Y 并非独立的。

目 录

◆ 问题 5.9

从二元分布中可获取什么信息呢?

定义 5.16 [二元联合分布下的期望]

- 假设 $g: \Omega_{XY} \to \mathbb{R}$ 为实值可测函数，其中 Ω_{XY} 是 (X,Y) 的支撑，则函数 $g(X,Y)$ 的期望定义为

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dF_{XY}(x,y)$$
$$= \begin{cases} \sum\sum_{(x,y)\in\Omega_{XY}} g(x,y)f_{XY}(x,y), & (X,Y) \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{XY}(x,y)dxdy, & (X,Y) \text{ 为连续随机变量} \end{cases}$$

其中上述二重求和或二重积分存在。与一元情形类似，若 $E|g(X,Y)| < \infty$，则称 $E[g(X,Y)]$ 存在。

定义 5.17 [乘积矩 (Product Moment)]

- (X,Y) 关于原点的第 r 阶和第 s 阶**乘积矩**定义如下

$$E(X^rY^s)=\begin{cases}\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} x^r y^s f_{XY}(x,y), & (X,Y)\text{ 为离散随机变量}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f_{XY}(x,y)dxdy, & (X,Y)\text{ 为连续随机变量}\end{cases}$$

- 类似地，第 r 阶和第 s 阶**中心化乘积矩** (central product moment) 定义为

$$E\{[X-E(X)]^r[Y-E(Y)]^s\}$$

$$=\begin{cases}\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}(x-\mu_X)^r(y-\mu_Y)^s f_{XY}(x,y), & (X,Y)\text{ 为离散随机变量}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)^r(y-\mu_Y)^s f_{XY}(x,y)dxdy, & (X,Y)\text{ 为连续随机变量}\end{cases}$$

若 X 和 Y 非相互独立，如何测度 X 和 Y 之间关系的强弱？

- 交叉乘积 $E(X^rY^s)$ 和 $E[(X-\mu X)^r(Y-\mu Y)^s]$ 提供了一种刻画 X 和 Y 之间关系或关联的方法。
- 不同的阶数 (r,s) 刻画了 X 和 Y 之间不同形式或不同方面的关联。
- 当 $(r,s)=(1,1)$ 时：这对应于函数 $g(X,Y)=(X-\mu_X)(Y-\mu_Y)$ 的期望。
  - ✓ 若 X 和 Y 倾向于在其均值上方或下方同方向移动，则该函数为正值；
  - ✓ 若 X 和 Y 倾向于在其均值上方或下方反方向移动，则该函数取负值。

定义 5.18 [协方差 (Covariance)]

假设 $E(X^2)<\infty$，$E(Y^2)<\infty$。随机变量 X 和 Y 的协方差定义为

$$\begin{aligned}\operatorname{cov}(X,Y)&=E[(X-\mu_X)(Y-\mu_Y)]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)dF_{XY}(x,y)\end{aligned}$$

协方差是对 X 和 Y 之间联动性的一种测度

- $\mathrm{cov}(X,Y)$ 的符号刻画了 X 和 Y **联动方向**。
  - ✓ $\boldsymbol{X}$ **和** $\boldsymbol{Y}$ **正相关**：若 X 和 Y 同时出现较大值或较小值的概率较大，则 $\mathrm{cov}(X,Y) > 0$。
  - ✓ $\boldsymbol{X}$ **和** $\boldsymbol{Y}$ **负相关**：若 X 的较大值和 Y 的较小值同时出现与 X 的较小值和 Y 的较大值同时出现的概率较大，则 $\mathrm{cov}(X,Y) < 0$。
  - ✓ $\boldsymbol{X}$ **和** $\boldsymbol{Y}$ **不相关**：若二者的变化不相关，则 $\mathrm{cov}(X,Y) = 0$ 。
- 注意 $\mathrm{cov}(X,X) = \mathrm{var}(X)$。

相关关系 ≠ 因果关系

- 非零协方差意味着 X 和 Y 存在相关性，但二者之间未必存在因果关系。
  - ✓ 例如，若发现石油价格上涨和经济增长放缓之间存在正相关关系，这无法说明石油价格上涨导致了经济放缓。
  - ✓ 又如，若发现抽烟和罹患癌症之间有正相关关系，这并不说明抽烟导致患癌。

定理 5.9

假设 (X,Y) 拥有有限二阶矩，则

$$\operatorname{cov}(X,Y)=E(XY)-\mu_X\mu_Y$$

证明：

- 因为期望算子 $E(\cdot)$ 为线性算子，则

$$\begin{aligned}\operatorname{cov}(X,Y)&=E[(X-\mu_X)(Y-\mu_Y)]\\&=E(XY-X\mu_Y-\mu_XY+\mu_X\mu_Y)\\&=E(XY)-\mu_X\mu_Y\end{aligned}$$

证毕。

定义 5.19 [相关系数 (Correlation Coefficient)]：

X 和 Y 的相关系数定义为

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

- 相关系数 ρ_{XY} 也称为**总体皮尔逊相关系数** (population Pearson correlation coefficient) 或**总体皮尔逊积矩相关系数** (population Pearson product-moment correlation coefficient)，是**标准化的协方差**，无度量单位，与偏度和峰度的定义类似。
- ρ_{XY} 的大小可以反映 ρ_{XY} 所刻画的关联程度。

定理 5.10

$|\rho_{XY}| \le 1$。

证明：

- 根据柯西-施瓦兹不等式 (Cauchy-Schwartz inequality)，对任意可测函数 $g(X)$ 和 $h(Y)$，有

$$E|g(X)h(Y)| \le \{E[g^2(X)]E[h^2(Y)]\}^{1/2}$$

- 令 $g(X) = X - \mu_X$，$h(Y) = Y - \mu_Y$，得

$$\begin{aligned}|\mathrm{cov}(X,Y)| &\le E|(X-\mu_X)(Y-\mu_Y)| \\ &\le \left(\sigma_X^2\sigma_Y^2\right)^{1/2} \\ &= \sigma_X\sigma_Y\end{aligned}$$

- 由此可知 $|\rho_{XY}| \le 1$。**证毕。**

定理 5.11

假设 $Y=a+bX$，$b\neq 0$，其中 $\sigma_X^2=\mathrm{var}(X)$ 存在。

则当 $b>0$ 时，$\rho_{XY}=1$；当 $b<0$ 时，$\rho_{XY}=-1$。

证明：

- 由于 $\mu_Y=a+b\mu_X$，并且 $\sigma_Y^2=b^2\sigma_X^2$，可得协方差

$$\begin{aligned}\mathrm{cov}(X,Y)&=E[(X-\mu_X)(Y-\mu_Y)]\\&=E[(X-\mu_X)(a+bX-a-b\mu_X)]\\&=bE(X-\mu_X)^2\\&=b\sigma_X^2\end{aligned}$$

证明 (Cont.)：

- 则

$$\rho_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{b\sigma_X^2}{|b|\sigma_X^2} = \frac{b}{|b|} = \begin{cases} 1, & b > 0 \\ -1, & b < 0 \end{cases}$$

证毕。

ρ_{XY} 所反映的关联性的本质

- ρ_{XY} 是 X 和 Y 之间线性关系的强弱测度：当 X 和 Y 存在完全线性关系时，总有 $\rho_{XY}=\pm 1$，即 ρ_{XY} 的绝对值取到最大值 1。
- 对 X 和 Y 的某些类型的联合分布而言，ρ_{XY} 是描述联合分布的一种十分有用的方式。然而，ρ_{XY} 的标准定义却没有揭示这一事实。
  - ✓ 已知对任意两个随机变量，有 $-1 \le \rho_{XY} \le 1$。若 $\rho_{XY}=1$，则存在一条由方程 $Y=a+bX$ 描述的直线，其中 $b>0$。该直线包含了 X 和 Y 概率分布的全部信息。在这种极端情况下，有 $P(Y=a+bX)=1$。若 $\rho_{XY}=-1$，情况类似，只是 $b<0$。

ρ_{XY} 所反映的关联性的本质 (Cont.)

- 一个值得关注的问题是：当 ρ_{XY} 并未达到极值时，xy 平面上是否存在一条直线，使得 X 和 Y 集中在该直线周围带状区域的概率较高?
  - ✓ 在某些条件下这种情况的确存在。此时，可将 ρ_{XY} 视作对 X 和 Y 在该直线附近取值的集中程度的一种测度。

例 5.25：

- 假设联合 PDF 为

$$f_{XY}(x,y)=\begin{cases}\dfrac{1}{4\varepsilon h}, & -\varepsilon+a+bx<y<\varepsilon+a+bx,-h<x<h\\ 0, & \text{其他}\end{cases}$$

  其中，常数 $h>0$，$\varepsilon>0$，而支撑 Ω_{XY} 如图 5.9 所示。
- 此处，$|Y-(a+bX)|$ 在带宽为 ε 的范围内。
- 当且仅当 $\varepsilon=0$ 时，存在精确的线性关系：$Y=a+bX$。
- 可以证明 $\rho_{XY}=\frac{bh}{\sqrt{\varepsilon^2+b^2h^2}}$。
- 显然，当 $b>0$ 时，ρ_{XY} 随着带宽 ε 收缩到零而收敛于 1。

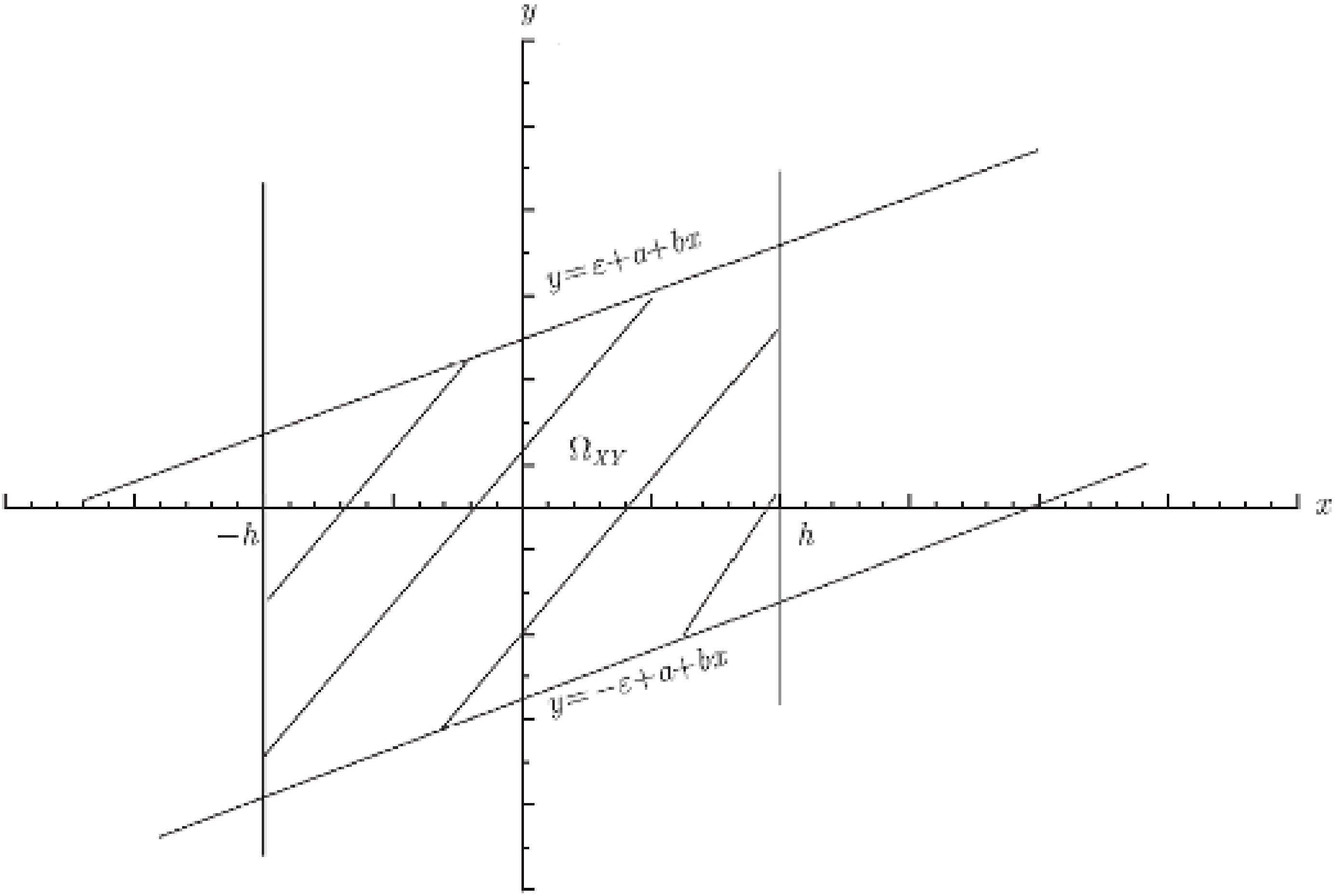


图 5.9：例 5.25 中的支撑 Ω_{XY}

例 5.26：[线性回归模型 (Linear Regression Model)]

- 假设

$$Y = a + bX + \varepsilon$$

其中随机变量 ε 满足 $E(\varepsilon) = 0$，$\mathrm{var}(\varepsilon) = \sigma_\varepsilon^2 > 0$，并且 ε 和 X 正交，即 $E(X\varepsilon) = 0$。

- 统计学和计量经济学常称此模型为线性回归模型。
- 随机变量 ε 可视为对完全线性关系 $Y = a + bX$ 的随机扰动。不同于例 5.25，$\varepsilon = Y - (a + bX)$ 可能具有无界支撑。

例 5.26 (Cont.):

- 当 $\sigma_\varepsilon^2 > 0$ 时，$|\rho_{XY}| < 1$，并且 $|\rho_{XY}|$ 将随 σ_ε^2 增加而减小。给定 $E(X\varepsilon) = 0$，可证

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{b}{\sqrt{b^2 + \sigma_\varepsilon^2/\sigma_X^2}}$$

- 因此，$|\rho_{XY}|$ 对 1 的偏离程度取决于比例 $\sigma_\varepsilon^2/\sigma_X^2$ 的大小。该比例常称为噪声信号比 (noise-to-signal ratio)。

定理 5.12 [最优线性最小二乘预测 (Best Linear Least Squares Prediction)]

- 假设随机变量 X 和 Y 均具有有限二阶矩。若用线性函数 $\alpha + \beta X$ 预测 Y，则预测误差为 $Y-(\alpha+\beta X)$。评价预测优劣的一种常用准则为均方误 (mean squared error)，定义为

$$\mathrm{MSE}(\alpha,\beta)=E[Y-(\alpha+\beta X)]^2$$

- 则最小化 $\mathrm{MSE}(\alpha,\beta)$ 的最优参数 (α^*,β^*) 为

$$\alpha^*=\mu_Y-\frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}\mu_X$$

$$\beta^*=\frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}=\rho_{XY}\sqrt{\frac{\mathrm{var}(Y)}{\mathrm{var}(X)}}$$

证明：

- 用一阶条件 (FOC) 求解最优参数 (α^*, β^*)

$$\left.\frac{\partial \mathrm{MSE}(\alpha,\beta)}{\partial \alpha}\right|_{(\alpha^*,\beta^*)} = -2E[Y-(\alpha^*+\beta^* X)] = 0$$

$$\left.\frac{\partial \mathrm{MSE}(\alpha,\beta)}{\partial \beta}\right|_{(\alpha^*,\beta^*)} = -2E\{X[Y-(\alpha^*+\beta^* X)]\} = 0$$

- 求解上述联立方程，即可得 α^* 和 β^* 的表达式。

证毕。

- 定义最优线性最小二乘预测 $\alpha^* + \beta^* X$ 的预测误差为

$$\varepsilon = Y - (\alpha^* + \beta^* X)$$

- 则可将 Y 表述为

$$Y = \alpha^* + \beta^* X + \varepsilon$$

- 其中 ε 与 X 正交，即

$$E(X\varepsilon) = 0$$

- 最小化 $\mathrm{MSE}(\alpha, \beta)$ 的一阶条件确保 $E(X\varepsilon) = 0$。换言之，最优线性最小二乘预测从本质上确保了随机扰动项 ε 正交于 X，这意味着 ε 不包含任何可用于预测 Y 的 X 的线性成分。

- 注意最优斜率参数 β^* 与协方差 $\mathrm{cov}(X,Y)$ 成比例。
  - ✓ 若 $\mathrm{cov}(X,Y)=0$，则线性预测函数 $\alpha^*+\beta^*X$ 在 MSE 准则下对 Y 没有预测能力；
  - ✓ 若 $\mathrm{cov}(X,Y)\neq 0$，则可借助线性模型 $\alpha^*+\beta^*X$ 实现 X 对 Y 的预测。
- 需要再次强调，预测关系并不意味存在从 X 到 Y 的因果关系。
- 协方差在经济分析中有广泛应用。

例 5.27：[资本资产定价模型 (Capital Asset Pricing Model, CAPM)]

- 假设 R_{pt} 表示某投资组合在持有期 t 的收益率，r_{ft} 表示无风险利率，R_{mt} 为同一持有期内市场资产组合 (如用 S&P 500 价格指数代表) 的收益率。则 CAPM 假定

$$R_{pt} - r_{ft} = \beta_p\left(R_{mt} - r_{ft}\right) + \varepsilon_{pt}$$

  - ✓ $R_{pt} - r_{ft}$ 表示投资组合在第 t 期的超额收益率，
  - ✓ $R_{mt} - r_{ft}$ 表示同一时期内市场投资组合的超额收益率，代表无法避免的系统性风险。
  - ✓ ε_{pt} 表示该投资组合的特质风险 (idiosyncratic risk)，该风险可通过分散化投资而消除 (即用很多种资产形成资产组合，参见第六章例 6.6)。

例 5.27 (Cont.):

- 简便起见，这里假设 $R_{mt}-r_{ft}$ 和 ε_{pt} 相互独立。
- 可以证明**斜率系数 (slope coefficient)**

$$\beta_p=\frac{\operatorname{cov}\left(R_{pt}-r_{ft},R_{mt}-r_{ft}\right)}{\operatorname{var}\left(R_{mt}-r_{ft}\right)}$$

$$=\rho_{pm}\frac{\sqrt{\operatorname{var}\left(R_{pt}-r_{ft}\right)}}{\sqrt{\operatorname{var}\left(R_{mt}-r_{ft}\right)}}$$

其中，ρ_{pm} 为 $R_{pt}-r_{ft}$ 和 $R_{mt}-r_{ft}$ 的相关系数。

例 5.27 (Cont.):

- 根据上述关于 CAPM 的假设，有

$$\operatorname{var}\left(R_{pt}-r_{ft}\right)=\beta_p^2\operatorname{var}\left(R_{mt}-r_{ft}\right)+\operatorname{var}\left(\varepsilon_{pt}\right)$$

  - ✓ $\operatorname{var}(R_{pt}-r_{ft})$ 度量投资组合的**总风险** (total risk)，
  - ✓ $\operatorname{var}(\varepsilon_{pt})$ 度量投资组合的**特质风险** (idiosyncratic risk)，
  - ✓ $\beta_p^2\operatorname{var}\left(R_{mt}-r_{ft}\right)$ 度量投资组合无法避免的**市场系统风险** (market risk)。

例 5.27 (Cont.):

- 在金融学，参数 β_p 称为投资组合的“beta **系数**”。该系数具有重要的经济解释：其测量了投资组合的风险和与该风险相对应的回报。
- β_p 的大小取决于该投资组合和市场投资组合之间的相关程度以及该投资组合相对于市场投资组合的风险大小，反映了投资组合的风险程度。
  - ✓ 根据定义，市场投资组合的 β 等于 1，即 $\beta_m = 1$。
  - ✓ 若 $\beta_p > 1$，则投资组合较之市场投资组合风险更大，
  - ✓ 若 $\beta_p < 1$，则投资组合比市场投资组合风险更小。

定理 5.13

假设 $Z = a + bX + cY$，则有

(1) $E(Z) = a + b\mu_X + c\mu_Y$；

(2) $\mathrm{var}(Z) = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\,\mathrm{cov}(X, Y)$。

- 简便起见，假设 $a = 0$，$b = c = 1$，则定理 5.13 表明

$$\mathrm{var}(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2\mathrm{cov}(X, Y)$$

✓ 当 $\mathrm{cov}(X, Y) > 0$ 时，有 $\mathrm{var}(X + Y) > \sigma_X^2 + \sigma_Y^2$

➢ 风险资产之间正相关将增加投资组合的风险

✓ 当 $\mathrm{cov}(X, Y) < 0$ 时，有 $\mathrm{var}(X + Y) < \sigma_X^2 + \sigma_Y^2$

➢ 风险资产负相关则降低投资组合的风险

定理 5.14

假设 $X_1,\cdots,X_n$ 为 n 个随机变量组成的序列，且 $Y=a_0+\sum_{i=1}^{n} a_iX_i$，其中 a_i 为常数。则有

(1) $E(Y)=a_0+\sum_{i=1}^{n} a_iE(X_i)$;

(2) $\mathrm{var}(Y)=\sum_{i=1}^{n} a_i^2\mathrm{var}(X_i)+2\sum_{i=2}^{n}\sum_{j=1}^{i-1} a_ia_j\mathrm{cov}\left(X_i,X_j\right)$

定理 5.15

假设两个随机变量 (X,Y) 服从二元正态分布 $BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则相关系数 $\rho_{XY}=\rho$。

例 5.28:

给定两个相互独立的 $N(0,1)$ 随机变量 Z_1 和 Z_2，如何构造二元正态分布 $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$?

解:

- 定义 $X = \mu_1 + aZ_1 + bZ_2$，$Y = \mu_2 + cZ_1 + dZ_2$
- 其中，常数 a,b,c,d 满足约束条件

$$a^2 + b^2 = \sigma_1^2$$

$$c^2 + d^2 = \sigma_2^2$$

$$ac + bd = \rho\sigma_1\sigma_2$$

- 根据二元变换定理可证 $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$。

线性相关系数

- 由于 ρ_{XY} 是 X 和 Y 之间线性关系的测度，因此通常称为线性相关系数。同时，为避免混淆，“线性”一词通常用于描述 ρ_{XY} 测度的相关关系。比如，常用“正线性相关 (positively linearly correlated)”代替“正相关性 (positively correlated)”
- 由于相关系数 ρ_{XY} 仅刻画了线性相关性，相关系数为零的两个随机变量 X 和 Y 之间可能存在很强的非线性相依关系。换言之，ρ_{XY} 可能无法捕捉某些非线性关系，如下例所示。

例 5.29：

- 假设 $X \sim N(0,\sigma^2)$，并且 $Y = X^2$，则

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= E(XY) - \mu_X\mu_Y \\
&= E\left(X^3\right) \\
&= \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\
&= 0
\end{aligned}
$$

- 其中，积分等于零系根据被积函数为奇函数（即对所有 x，有 $g(-x) = -g(x)$）的性质而得。因此，尽管 Y 和 X 间存在确定的 (非线性) 函数关系，但二者并不相关。
- 这说明 $\operatorname{cov}(X,Y)$ 无法描述 X 和 Y 之间存在的某些重要的非线性关系。

目　录

定义 5.20 [联合MGF]

(X, Y) 的联合 MGF 定义为

$$M_{XY}(t_1, t_2) = E\left(e^{t_1 X + t_2 Y}\right), -\infty < t_1, t_2 < \infty$$

其中上述期望对在 $(0, 0)$ 的某邻域内所有 (t_1, t_2) 都存在。

- 联合 MGF $M_{XY}(t_1, t_2)$ 是第三章第九节 MGF $M_X(t)$ 的二元扩展，其包含了关于 (X, Y) 联合概率分布的信息。X 和 Y 的**边际 MGF** 可从联合 MGF $M_{XY}\ (t_1, t_2)$ 求得：

$$M_X(t_1) = M_{XY}(t_1, 0)$$
$$M_Y(t_2) = M_{XY}(0, t_2)$$

- 若联合 MGF $M_{XY}\ (t_1, t_2)$ 对 $(0, 0)$ 的某个邻域内所有 (t_1, t_2) 都存在，则乘积矩 (product moment) $E(X^r Y^s)$ 对所有阶数 (r, s) 均存在，且 $M_{XY}\ (t_1, t_2)$ 可用于生成各种阶数的乘积矩 $E(X^r Y^s)$。

定理 5.16

- 假设联合 MGF $M_{XY}(t_1,t_2)$ 对在 (0,0) 的某个邻域内所有 (t_1,t_2) 都存在，则对所有非负整数 $r,s \geq 0$，有

$$E(X^rY^s) = M_{XY}^{(r,s)}(0,0)$$

- 并且

$$\operatorname{cov}(X^r,Y^s) = M_{XY}^{(r,s)}(0,0) - M_X^{(r)}(0)M_Y^{(s)}(0)$$

- 特别地，

$$\operatorname{cov}(X,Y) = M_{XY}^{(1,1)}(0,0) - M_X^{(1)}(0)M_Y^{(1)}(0)$$

证明:

- 给定 $M_{XY}(t_1,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x+t_2y}dF_{XY}(x,y)$，有

$$M_{XY}^{(r,s)}(t_1,t_2)=\frac{\partial^{r+s}}{\partial t_1^r\partial t_2^s}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x+t_2y}dF_{XY}(x,y)$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial^{r+s}e^{t_1x+t_2y}}{\partial t_1^r\partial t_2^s}dF_{XY}(x,y)$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^r y^s e^{t_1x+t_2y}dF_{XY}(x,y)$$

- 则

$$M_{XY}^{(r,s)}(0,0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^r y^s dF_{XY}(x,y)=E(X^rY^s)$$

证明 (Cont.):

- 进一步地,

$$
\begin{aligned}
M_{XY}^{(r,s)}(0,0) - M_X^{(r)}(0)M_Y^{(s)}(0) &= E(X^rY^s) - E(X^r)E(Y^s) \\
&= \mathrm{cov}(X^r, Y^s)
\end{aligned}
$$

- 作为特例,当 $(r,s)=(1,1)$ 时,$\mathrm{cov}(X,Y)$ 可通过对联合 MGF $M_{XY}(t_1,t_2)$ 以及边际 MGF $M_X(t_1)$ 和 $M_Y(t_2)$ 求导得到。

证毕。

例 5.30:

- 假设 X, Y 服从二元正态分布 $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则其联合 MGF 为

$$M_{XY}(t_1, t_2) = e^{\mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1^2 t_1^2 + \sigma_2^2 t_2^2 + 2\rho\sigma_1\sigma_2 t_1 t_2}{2}}$$
$$= e^{\mu' \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}}$$

- 其中 $\boldsymbol{t} = (t_1, t_2)', \mu = (\mu_1, \mu_2)'$，且

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

例 5.30 (Cont.):

- 在多元框架下，可定义随机变量 $X_1,\cdots,X_n$ 的联合 MGF：

$$M_{\boldsymbol{X}^n}(\boldsymbol{t}) = E\left(e^{\boldsymbol{t}'\boldsymbol{X}^n}\right) = E\left(e^{\sum_{i=1}^n t_i X_i}\right)$$

- 其中 $\boldsymbol{X}^n = (X_1,\cdots,X_n)'$，$\boldsymbol{t} = (t_1,\cdots,t_n)'$，且上述期望对在原点 $(0,\cdots,0)'$ 的某个邻域内所有 $\boldsymbol{t} = (t_1,\cdots,t_n)'$ 都存在。

目　录

定理 5.17

- 假设 (X,Y) 相互独立，则对任意可测可积函数 $h(X)$ 和 $q(Y)$，有

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)],$$

- 或者等价地

$$\operatorname{cov}[h(X),q(Y)] = 0$$

证明：

$$E[h(X)q(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)q(y)dF_{XY}(x,y)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)q(y)dF_X(x)dF_Y(y) \quad \text{(由独立性)}$$
$$= \int_{-\infty}^{\infty} h(x)dF_X(x)\int_{-\infty}^{\infty} q(y)dF_Y(y)$$
$$= E[h(X)]E[q(Y)]$$

- 因此，若 X 和 Y 相互独立，则其任意线性或非线性可测变换均不相关。

证毕。

5.9.1 独立性和矩生成函数

推论 5.1

- 假设 X 和 Y 相互独立，并且二者的边际 MGF $M_X(t)$ 和 $M_Y(t)$ 对在 0 的某个邻域内所有 t 都存在。则 $M_{X+Y}(t)$ 对在 0 的某个邻域内所有 t 都存在，且

$$M_{X+Y}(t) = M_X(t)M_Y(t), \text{对 } 0 \text{ 的某个邻域内所有 } t$$

- 需要强调，$M_{X+Y}(t)$ 是随机变量 $X+Y$ 的 MGF，而非 (X,Y) 的联合 MGF $M_{XY}(t_1,t_2)$。

证明：

- 令

$$g(X,Y)=e^{t(X+Y)}=e^{tX}e^{tY}=h(X)q(Y)$$

- 根据定理 5.17 可知，若 X 和 Y 相互独立，则 $M_{X+Y}(t)$ 对在 0 的小邻域内所有 t 都存在且

$$E\left[e^{t(X+Y)}\right]=E\left(e^{tX}\right)E\left(e^{tY}\right)$$

- 即

$$M_{X+Y}(t)=M_X(t)M_Y(t)$$

证毕。

例 5.31 [正态分布的可加性 (Reproductivity of Normal Distribution)]：

- 假设 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，且 X 和 Y 相互独立。证明

$$X \pm Y \sim N\left(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

解：

- X 和 Y 的 MGF 分别为

$$M_X(t) = E\left(e^{tX}\right) = e^{\mu_1 t + \frac{\sigma_1^2}{2} t^2}$$

$$M_Y(t) = E\left(e^{tY}\right) = e^{\mu_2 t + \frac{\sigma_2^2}{2} t^2}$$

解 (Cont.):

- 根据独立性可得

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= e^{\mu_1 t+\frac{\sigma_1^2}{2}t^2} e^{\mu_2 t+\frac{\sigma_2^2}{2}t^2} \\ &= e^{\mu t+\frac{\sigma^2}{2}t^2} \end{aligned}$$

  其中 $\mu = \mu_1 + \mu_2$，$\sigma^2 = \sigma_1^2 + \sigma_2^2$。
- 因此 $X + Y \sim N\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$。
- 换言之，两个独立正态随机变量之和仍为正态变量。
- 事实上，该结果对多个相互独立正态随机变量的线性组合仍适用。该性质称为**正态分布的可加性**。

例 5.32 [泊松分布的可加性 (Reproductivity of Poisson Distribution)]:

- 假设 $X_1, \cdots, X_n$ 为 n 个独立随机变量，其中 X_i 服从泊松分布 $\text{Poisson}(\lambda_i)$，$i = 1, \cdots, n$。
- 证明 $\sum_{i=1}^{n} X_i$ 服从泊松分布 $\text{Poisson}(\lambda)$，其中 $\lambda = \sum_{i=1}^{n} \lambda_i$。

解:

- 若随机变量 X_i 服从泊松分布 $\text{Poisson}(\lambda_i)$，其 MGF

$$M_i(t) = e^{\lambda_i(e^t - 1)}, -\infty < t < \infty$$

解 (Cont.)：

- 因 $X = \sum_{i=1}^{n} X_i$，其 MGF

$$M_X(t) = E\left(e^{t\sum_{i=1}^{n} X_i}\right) = \prod_{i=1}^{n} M_i(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{\lambda(e^t-1)}$$

- 其中 $\lambda = \sum_{i=1}^{n} \lambda_i$。因此，$X \sim \text{Poisson}(\lambda)$。
- 换言之，n 个互相独立的 $\text{Poisson}(\lambda_i)$ 随机变量之和仍为 $\text{Poisson}(\lambda)$ 随机变量，其中 $\lambda = \sum_{i=1}^{n} \lambda_i$。
- 这称为**泊松分布的可加性**。

例 5.33 [χ^2 分布的可加性]:

假设 $X_1, X_2, \cdots, X_n$ 相互独立，且分别服从 $\chi^2_{\nu_i}$ 分布。证明 $\sum_{i=1}^{n} X_i$ 服从 χ^2_{ν} 分布，其中 $\nu = \sum_{i=1}^{n} \nu_i$。

解:

- 由第四章知 $\chi^2_{\nu_i}$ 分布的 MGF 为

$$M_i(t) = (1-2t)^{-\nu_i/2}, t < \frac{1}{2}$$

解 (Cont.):

- 令 $X=\sum_{i=1}^{n} X_i$, 则其 MGF 为

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \prod_{i=1}^{n} M_i(t) \\ &= \prod_{i=1}^{n} (1-2t)^{-\nu_i/2} \\ &= (1-2t)^{-\frac{1}{2}\sum_{i=1}^{n} \nu_i} \end{aligned}$$

- 由此可得，$X \sim \chi_\nu^2$，其中 $\nu=\sum_{i=1}^{n} \nu_i$。即，独立 χ^2 随机变量之和仍为 χ^2 随机变量，其自由度等于各随机变量自由度之和。
- 该性质称为 **χ^2 分布的可加性**。

例 5.34：

假设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，且均服从参数为 $\beta > 0$ 的指数分布，证明 $\sum_{i=1}^{n} X_i$ 服从伽玛分布 $G(n, \beta)$。

解:

- 由于 X_i 服从 $EXP(\beta)$ 分布，其 MGF 为

$$M_i(t) = (1-\beta t)^{-1}, t < \frac{1}{\beta}$$

- 记 $X = \sum_{i=1}^{n} X_i$, 则其 MGF 为

$$\begin{aligned} M_X(t) &= \prod_{i=1}^{n} M_i(t) \\ &= (1-\beta t)^{-n} \end{aligned}$$

- 这说明 $X \sim G(n,\beta)$。即，具有相同参数 β 的 n 个独立指数随机变量之和服从伽玛分布 $G(n,\beta)$。

5.9.2 独立性和不相关性

推论 5.2

假设 X 和 Y 相互独立，则 $\text{cov}(X,Y)=0$。

证明：

令 $g(X)=X-\mu_X$，$h(Y)=Y-\mu_Y$，应用定理 5.17 即可得证。

证毕。

- 定理 5.17 表明，独立性排除了 X 和 Y 之间的所有可能关联性，而不相关仅意味着不存在线性关系。显然，独立性比不相关更强。
- 因此，**相互独立 $\Rightarrow$ 不相关，但不相关 $\nRightarrow$ 相互独立。**

例 5.35：

令 $Y = X^2$，其中 X 服从关于 0 对称的连续概率分布（即 $f_X(x) = f_X(-x)$，$-\infty < x < \infty$）。则

$$\mathrm{cov}(X, Y) = 0$$

解：

- 由于 X 的分布关于 0 对称，故有 $E(X)=0$，且 $E(X^3)=0$。因此

$$
\begin{aligned}
\operatorname{cov}(X,Y) &= E(XY)-E(X)E(Y)\\
&= E\left(X^3\right)-0\cdot E\left(X^2\right)\\
&= E\left(X^3\right)\\
&= \int_{-\infty}^{\infty} x^3 f_X(x)dx\\
&= 0
\end{aligned}
$$

- 当 X 服从关于 0 对称的分布时，$\operatorname{cov}(X,Y)$ 无法捕获 X 和 $Y=X^2$ 之间的二次型关系。

例 5.36 [不相关的二元学生 t-分布]:

若随机变量 X 和 Y 的联合 PDF 为

$$f_{XY}(x,y)=\frac{1}{2\pi}\left[1+\frac{1}{\nu}\left(x^2+y^2\right)\right]^{-\frac{\nu}{2}},\nu\geq 3$$

则称其服从标准二元学生 t-分布，其中形状参数 ν 称为**自由度**

(degree of freedom)。

例 5.36 (Cont.):

- 可证明 $\mathrm{cov}(X, Y) = 0$ (问题：如何证明？)。
- 然而，由于 X 和 Y 的联合 PDF $f_{XY}(x, y)$ 无法分解为两个分别关于 x 和 y 的函数之积，故对任意给定 $\nu < \infty$，X 和 Y 并不相互独立。
- 当 $\nu \to \infty$ 时，有

$$f_{XY}(x, y) \to \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

这里用到了当 $\nu \to \infty$ 时，$\left(1 + \frac{a}{\nu}\right)^{\nu} \to e^{a}$ 这一极限性质。在这种情况下，X 和 Y 为相互独立的 $N(0, 1)$ 随机变量。

定理 5.18

假设 (X, Y) 服从联合正态分布，则当且仅当 $\mathrm{cov}(X, Y) = 0$ 时 (X, Y) 相互独立。

证明:

(1) [必要性] 由推论 5.2，若 X 和 Y 相互独立，则有 $\mathrm{cov}(X,Y)=0$。

(2) [充分性] 证明 $\mathrm{cov}(X,Y)=0$ 意味着 X 和 Y 相互独立。

- 对二维正态分布 $BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，定理 5.15 已证相关系数 $\rho_{XY}=\rho$。
- 因此，当 $\mathrm{cov}(X,Y)=0$ 时，有 $\rho=0$。于是联合 PDF 为

$$f_{XY}(x,y)=\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2}\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{1}{2\sigma_2^2}(y-\mu_2)^2}$$
$$=f_X(x)f_Y(y),\text{ 对所有 }-\infty<x,y<\infty$$

- 因此，X 和 Y 相互独立。

证毕。

定理 5.19

假设 $X \sim \text{Bernoulli}(p_1)$，$Y \sim \text{Bernoulli}(p_2)$。

则当且仅当 $\text{cov}(X, Y) = 0$ 时，X 和 Y 相互独立。

证明：

- 只需证明若 $\text{cov}(X, Y) = 0$，则 X 和 Y 相互独立。现考察对所有 $x = 1, 0$ 和 $y = 1, 0$，$f_{XY}(x, y) = f_X(x) f_Y(y)$ 是否成立。
- 首先，注意到
  - ✓ $f_X(1) = p_1$，$f_X(0) = 1 - p_1$，
  - ✓ $f_Y(1) = p_2$，$f_Y(0) = 1 - p_2$，
  - ✓ $\mu_X = p_1$，且 $\mu_Y = p_2$。

证明（Cont.）：

- 假设 $\mathrm{cov}(X, Y) = 0$，则

$$E(XY) = E(X)E(Y)$$

或等价地

$$\sum_{x=0}^{1} \sum_{y=0}^{1} xyf_{XY}(x, y) = p_1 p_2$$

- 由于 $\sum_{x=0}^{1} \sum_{y=0}^{1} xyf_{XY}(x, y) = f_{XY}(1, 1)$，故有

$$f_{XY}(1,1) = p_1 p_2 = f_X(1) f_Y(1)$$

证明 (Cont.)：

- 另外，给定 $f_X(1)=\sum_{y=0}^{1} f_{XY}(1,y)$,

$$f_{XY}(1,0)=f_X(1)-f_{XY}(1,1)$$
$$=p_1(1-p_2)$$
$$=f_X(1)f_Y(0)$$

- 类似地，可证

$$f_{XY}(0,1)=(1-p_1)p_2=f_X(0)f_Y(1)$$

及

$$f_{XY}(0,0)=(1-p_1)(1-p_2)=f_X(0)f_Y(0)$$

- 这表明，对所有 $x=1,0$ 和 $y=1,0$，有

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$

- 因此，X 和 Y 相互独立。**证毕。**

◆ 问题 5.10

- 定理 5.17 说明若 X 和 Y 相互独立，则对任意可测函数 $h(X)$ 和 $q(Y)$，有

$$\operatorname{cov}[h(X), q(Y)] = 0$$

- 现在，假设对任意可测函数 $h(X)$ 和 $q(Y)$，有 $\operatorname{cov}[h(X), q(Y)] = 0$。那么 X 和 Y 是否相互独立?

问题 5.10 (Cont.)

- 为回答这一问题，对任意给定 $(x,y)\in R^2$，考察指示函数 $h_x(X)=\mathbf{1}(X\le x)$ 和 $q_y(Y)=\mathbf{1}(Y\le y)$。此时，

$$E[h_x(X)]=E[\mathbf{1}(X\le x)]=F_X(x)$$

$$E[q_y(Y)]=E[\mathbf{1}(Y\le y)]=F_Y(y)$$

$$E[h_x(X)q_y(Y)]=E[\mathbf{1}(X\le x)\mathbf{1}(Y\le y)]=F_{XY}(x,y)$$

- 假设 $\mathrm{cov}[h_x(X),q_y(Y)]=0$, 对所有 $-\infty<x,y<\infty$，或等价地 $E[h_x(X)q_y(Y)]=E[h_x(X)]E[q_y(Y)]$, 对所有 $-\infty<x,y<\infty$，
- 则 $F_{XY}(x,y)=F_X(x)F_Y(y)$, 对所有 $-\infty<x,y<\infty$。
- 因此，X 和 Y 相互独立。

定理 5.20

- 假设联合 MGF $M_X(t_1, t_2) = E(e^{t_1 X + t_2 Y})$ 对在原点 $(0,0)$ 的某个邻域内所有 (t_1, t_2) 都存在，则当且仅当对原点 $(0,0)$ 某个邻域内的所有点 (t_1, t_2)，有

$$M_{XY}(t_1, t_2) = M_X(t_1) M_Y(t_2)$$

成立时，X 和 Y 相互独立。

证明:

(1) **[必要性]** 假设 X 和 Y 相互独立。则由定理 5.17，有

$$\begin{aligned} M_{XY}(t_1,t_2) &= E\left(e^{t_1X+t_2Y}\right) \\ &= E\left(e^{t_1X}e^{t_2Y}\right) \\ &= E\left(e^{t_1X}\right)E\left(e^{t_2Y}\right) \\ &= M_X(t_1)M_Y(t_2) \end{aligned}$$

(2) **[充分性]** 现在证明若对所有在原点 $(0,0)$ 的某个邻域内所有 (t_1,t_2)， $M_{XY}(t_1,t_2)=M_X(t_1)M_Y(t_2)$，则 X 和 Y 相互独立。

- 根据 MGF 的唯一性定理，
  - ✓ 二元分布 $F_{XY}(x,y)$ 是对应联合 $M_{XY}(t_1,t_2)$ 的唯一分布，
  - ✓ 而二元分布 $F_X(x)F_Y(y)$ 是对应联合 MGF $M_X(t_1)M_Y(t_2)$ 的唯一分布。

证明 (Cont.):

- 若对在关于原点 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $M_{XY}(t_1,t_2)=M_X(t_1)M_Y(t_2)$，则 $F_X(x)F_Y(y)$ 也是对应于 $M_{XY}(t_1,t_2)$ 的唯一分布。
- 因此，对所有 $-\infty<x$，$y<\infty$，有 $F_{XY}(x,y)=F_X(x)F_Y(y)$。

证毕。

定理 5.21

假设对在关于原点 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $M_{XY}(t_1,t_2)$ 存在。则 X 和 Y 相互独立，当且仅当对 $(0,0)$ 的某个邻域内的所有 (t_1,t_2)，有

$$\operatorname{cov}\left(e^{t_1X},e^{t_2Y}\right)=0$$

证明：

- 根据公式 $\operatorname{cov}(U,V)=E(UV)-E(U)E(V)$，可直接证明
$$\begin{aligned}\operatorname{cov}\left(e^{t_1X},e^{t_2Y}\right)&=E\left(e^{t_1X}e^{t_2Y}\right)-E\left(e^{t_1X}\right)E\left(e^{t_2Y}\right)\\&=M_{XY}(t_1,t_2)-M_X(t_1)M_Y(t_2)\end{aligned}$$
- 由于当且仅当 X 和 Y 相互独立时，对在关于 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $M_{XY}(t_1,t_2)=M_X(t_1)M_Y(t_2)$。
- 因此，当且仅当 (X,Y) 相互独立时，对位于 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $\operatorname{cov}\left(e^{t_1X},e^{t_2Y}\right)=0$ 成立。

证毕。

定理 5.21 含义

- 由于 $\mathrm{cov}\left(e^{t_1X}, e^{t_2Y}\right)$ 是指数变换 e^{t_1X} 和 e^{t_2Y} 的协方差，其可视作**广义协方差** (generalized covariance)。
- 定理 5.21 表明:
  - ✓ 当且仅当对在关于 (0,0) 的某个邻域内的所有 (t_1, t_2)，该广义协方差 $\mathrm{cov}\left(e^{t_1X}, e^{t_2Y}\right) = 0$ 时，X 和 Y 相互独立。
- 以下定理（定理 5.22）说明:
  - ✓ 广义协方差 $\mathrm{cov}\left(e^{t_1X}, e^{t_2Y}\right)$ 实际上可视为**协方差生成函数** (covariance generating function)。即，其可用于生成各种协方差 $\mathrm{cov}(X^r, Y^s)$。

定理 5.22

假设对在 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $M_{XY}(t_1,t_2)$ 存在。则

$$\operatorname{cov}(X,Y)=\left.\frac{\partial^2\operatorname{cov}\left(e^{t_1X},e^{t_2Y}\right)}{\partial t_1\partial t_2}\right|_{(t_1,t_2)=(0,0)}$$

此外，对任意正整数 r,s，有

$$\operatorname{cov}(X^r,Y^s)=\left.\frac{\partial^{r+s}\operatorname{cov}\left(e^{t_1X},e^{t_2Y}\right)}{\partial t_1^r\partial t_2^s}\right|_{(t_1,t_2)=(0,0)}$$

定理 5.22 含义

- 若对在关于 $(0,0)$ 的某个邻域内所有 (t_1,t_2)，有 $M_{XY}(t_1,t_2)$ 存在，则 $\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)$ 在原点 $(0,0)$ 处的各阶导数均存在。
- 根据麦克劳伦级数展开 (MacLaurin's series expansion) $e^{tX}=\sum_{r=0}^{\infty}\frac{(tX)^r}{r!}$ 可得

$$\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)=\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{t_1^r t_2^s}{r!\,s!}\frac{\partial^{r+s}\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)}{\partial t_1^r\partial t_2^s}\Bigg|_{(t_1,t_2)=(0,0)}$$

$$=\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{t_1^r t_2^s}{r!\,s!}\mathrm{cov}(X^r,Y^s)$$

定理 5.22 含义 (Cont.)

- 因此，$\operatorname{cov}\left(e^{t_1X}, e^{t_2Y}\right)$ 包含了各阶协方差 $\{\operatorname{cov}(X^r, Y^s)\}$ 的所有信息。
  - ✓ 若 X 和 Y 相互独立，则对所有 $r, s > 0$，有 $\operatorname{cov}(X^r, Y^s) = 0$，当然也包括 $\operatorname{cov}(X, Y) = 0$。
  - ✓ 一般而言，$\operatorname{cov}(X, Y) = 0$ 仅是独立性所隐含的性质形成的无穷集中的一个元素。可能存在 $\operatorname{cov}(X, Y) = 0$ 但是对某些 (r, s)，$\operatorname{cov}(X^r, Y^s) \neq 0$ 的情形。

定理 5.22 含义 (Cont.)

- ✓ 在 这种情况下，X 和 Y 不相关，但并非相互独立。
- ✓ 一个经济学例子是高频资产收益率的时间序列 $\{X_t\}$。实际研究表明，高频资产收益率 $\{X_t\}$ 在不同时期互不相关，即对所有 $j > 0$，有 $\text{cov}(X_t, X_{t-j}) = 0$。然而，该序列存在持续的波动聚类 (volatility clustering) 现象，即序列 $\{X_t\}$ 的高波动和低波动往往会各自集聚在某一时间段，并且高低波动集聚的时期会交替出现。这表明至少对某些 $j > 0$，有 $\text{cov}(X_t^2, X_{t-j}^2) > 0$。

◆ 问题

假设 X 的各阶矩存在，且对所有整数 $r, s > 0$ 有 $\mathrm{cov}(X^r, Y^s) = 0$。$X$ 和 Y 是否相互独立?

- 这与一元情形下两个随机变量 X 和 Y 的各阶矩相等是否意味着相同分布的问题类似 (参见定理 3.25)。在随机变量 X 和 Y 均存在有界支撑的情况下，答案是肯定的。

定理 5.23

假设随机变量 X 和 Y 具有有界支撑，则当且仅当 X 和 Y 相互独立时，对所有整数 $r,s>0$，有 $\mathrm{cov}(X^r,Y^s)=0$。

证明： 由于 X 和 Y 均具有有界支撑，所以存在常数 $M<\infty$ 使得 $P(|X|<M)=1$，$P(|Y|<M)=1$。

- 因此，$E|X|^r \le M^r$，并且 $E|Y|^s \le M^s$。由于

$$
\begin{aligned}
\left|\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)\right| &= \left|\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{t_1^r t_2^s}{r!s!}\mathrm{cov}(X^r,Y^s)\right| \\
&\le \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{|t_1|^r|t_2|^s}{r!s!}|\mathrm{cov}(X^r,Y^s)| \\
&\le \sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{|t_1|^r|t_2|^s}{r!s!}(2M)^r(2M)^s \\
&= e^{2(|t_1|M+|t_2|M)}<\infty
\end{aligned}
$$

- 故对任意给定 $t_1,t_2\in(-\infty,\infty)$，有 $\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)$ 存在。

证明 (Cont.):

- 上式推导用到了 $|\mathrm{cov}(X^r,Y^s)| \le [\mathrm{var}(X^r)\mathrm{var}(Y^s)]^{\frac{1}{2}} \le (2M)^{r+s}$ 这一事实。
- **(1) [必要性]** 假设 X 和 Y 相互独立，则根据定理 5.17，且令 $h(X)=X^r$，$q(Y)=Y^s$，可得 $\mathrm{cov}(X^r,Y^s)=0$，对所有 $r,s>0$。
- **(2) [充分性]** 假设对所有 $r,s>0$，$\mathrm{cov}(X^r,Y^s)=0$。则对所有 $t_1,t_2\in(-\infty,\infty)$, 有

$$\mathrm{cov}\left(e^{t_1X},e^{t_2Y}\right)=\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\frac{t_1^r}{r!}\frac{t_2^s}{s!}\mathrm{cov}(X^r,Y^s)=0$$

- 由定理 5.21 即得 X 和 Y 相互独立。**证毕。**

- 定理 5.23 可视为定理 3.25 从一维到二维的推广。
- 这表明可通过检验 X^r 和 Y^s 所有可能的协方差是否为零，来判断 X 和 Y 是否相互独立。当然，一个前提条件是 X 和 Y 的支撑是有界的。

其他广义协方差 (generalized covariance) 的定义

- 例如，可定义**广义协方差**为：

$$\sigma_{XY}(x,y) \equiv \text{cov}[\mathbf{1}(X \le x), \mathbf{1}(Y \le y)], -\infty < x, y < \infty$$

其中 $\mathbf{1}(\cdot)$ 为指示函数，若括号内的条件成立，其取值为 1，否则取值为 0。

- 通过简单计算可证

$$\text{cov}[\mathbf{1}(X \le x), \mathbf{1}(Y \le y)] = F_{XY}(x,y) - F_X(x)F_Y(y)$$

- 显然，当且仅当对所有 (x,y)，广义协方差 $\sigma_{XY}(x,y) = 0$ 时，有 X 和 Y 相互独立。

协方差 $\mathrm{cov}(X,Y)$ 与广义协方差 $\sigma_{XY}(x,y)$ 的关系

- 虽然协方差 $\mathrm{cov}(X,Y)$ 不能刻画非高斯情况下 X 和 Y 所有的相依关系，但是广义协方差 $\sigma_{XY}(x,y)$ 可以刻画所有相依关系。
- 事实上，根据 Hoeffding & Korrelationtheorie (1940)，协方差 $\mathrm{cov}(X,Y)$ 与广义协方差 $\sigma_{XY}(x,y)$ 紧密相关，即

$$\begin{aligned}&\mathrm{cov}(X,Y)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[F_{XY}(x,y)-F_X(x)F_y(y)\right]dxdy\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sigma_{XY}(x,y)dxdy\end{aligned}$$

- 协方差 $\mathrm{cov}(X,Y)$ 不能刻画所有相依关系，因为对所有 (x,y)，有可能广义协方差 $\sigma_{XY}(x,y)$ 不为 0，但是二重积分为 0。

霍夫丁测度法可以刻画 X 和 Y 的所有相依关系

- 霍夫丁测度法 (Hoeffding measure)：

$$D^2(j)$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sigma_{XY}^2(x,y)dF_{XY}(x,y)$$

$$=E[F_{XY}(X,Y)-F_X(X)F_Y(Y)]^2$$

- 若 $F_{XY}(x,y)$ 为连续分布，则当且仅当 X 和 Y 相互独立，$D^2(j)$ 取值为 0。

其他类型的相关系数

- 另外，**肯德尔（Kendall）秩相关系数** τ (即样本秩相关统计量，用于测度两个测量量之间的序数关联) 的总体形式 (或概率极限) 为

$$\tau_{XY} = 4\mathrm{cov}[F_X(x), F_Y(y)] = 4E[\sigma_{XY}(X, Y)]$$
$$= 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma_{XY}(x, y) dF_{XY}(x, y)$$

- 在统计学，著名的**斯皮尔曼（Spearman）相关系数**或斯皮尔曼 ρ (即样本观测值的秩次相关系数) 的总体形式 (或概率极限) 为

$$S_{XY} = \frac{\mathrm{cov}[F_X(X), F_Y(Y)]}{\sqrt{\mathrm{var}[F_X(X)]\mathrm{var}[F_Y(Y)]}}$$

- 可以证明 $\mathrm{cov}[F_X(X), F_Y(Y)] = E[\sigma_{XY}(X, Y)]$

总结

- $\mathrm{cov}(X,Y)$：刻画线性相关关系，但它只能判断方向，即正相关或负相关。
- $\rho_{XY}=\frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}$：刻画相关关系，绝对值越接近一，相关性越高。
- 考虑一个线性回归 $Y=\alpha+X\beta+\varepsilon$，估计量表达式 $\beta=\frac{\mathrm{cov}(X,Y)}{\mathrm{Var}(X)}$，因此，线性回归分析即相关分析。
- X 和 Y 是独立的可以推出 $\mathrm{cov}(X,Y)=0$，但反过来不成立。
- X 和 Y 是独立的当且仅当 $\mathrm{cov}(g(X),h(Y))=0$，$\forall g(\cdot),h(\cdot)$。下一节将给出 $g(\cdot),h(\cdot)$ 的选择。

目　录

◆ 问题 5.12

从条件分布 $f_{Y|X}(y \mid x)$ 可获取什么信息?

定义 5.21 [条件期望 (Conditional Expectation)]

- 可测函数 $g(X,Y)$ 基于 $X=x$ 的条件期望定义为

$$E[g(X,Y) \mid X=x] = E[g(X,Y) \mid x] = \begin{cases} \sum_{y\in\Omega_Y(x)} g(x,y) f_{Y|X}(y \mid x), & (X,Y) \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty} g(x,y) f_{Y|X}(y \mid x) dy, & (X,Y) \text{ 为连续随机变量} \end{cases}$$

- 其中 $\Omega_Y(x) = \{y \in \Omega_Y\colon f_{Y|X}(y \mid x) > 0\}$ 为 Y 基于 $X=x$ 条件下的支撑。

$E[g(X,Y)|X]$ 为随机变量 X 的函数

- 在求条件期望 $E[g(X,Y)|X=x]$ 时，x 被视为固定值，且条件期望是对 Y 基于 $X=x$ 的条件分布而非 Y 的无条件分布求期望。
- 由于积分已消去 y，条件期望 $E[g(X,Y)|X=x]$ 仅为 x 的函数。

定理 5.24 [重复期望法则 (Law of Iterated Expectations)]

假设 $g(X,Y)$ 为可测函数且 $E[g(X,Y)]$ 存在，则

$$E[g(X,Y)] = E_X\{E[g(X,Y) \mid X]\} = E_Y\{E[g(X,Y) \mid Y]\}$$

证明:

- 只考虑连续随机变量的情形。
- 假设 (X,Y) 的联合 PDF 为 $f_{XY}(x,y)$。
- 由联合概率密度的乘法法则，当 $f_X(x) > 0$ 时，有

$$f_{XY}(x,y) = f_{Y|X}(y \mid x) f_X(x)$$

证明 (Cont.)：

- 因此

$$
\begin{aligned}
E[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{XY}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{Y|X}(y \mid x)f_X(x)dxdy \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y)f_{Y|X}(y \mid x)dy\right]f_X(x)dx \\
&= \int_{-\infty}^{\infty} E[g(X,Y) \mid X = x]f_X(x)dx \\
&= E_X\{E[g(X,Y) \mid X]\}
\end{aligned}
$$

- 对离散随机变量的情形同理可证。

证毕。

重复期望法则提供一种**计算无条件期望的两阶段方法**，因此也常称为**全期望法则 (the law of total expectation)**，类似于第二章的全概率公式。

定义 5.22 [条件均值 (Conditional Mean)]：

Y 基于 $X = x$ 的条件均值定义为

$$E(Y \mid x) = E(Y \mid X = x)$$

$$= \begin{cases} \displaystyle\sum_{y \in \Omega_Y(x)} y f_{Y|X}(y \mid x), & (X,Y) \text{ 为离散随机变量} \\ \displaystyle\int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy & (X,Y) \text{ 为连续随机变量} \end{cases}$$

例 5.38 [平均工资和重复期望法则]:

- 假设 Y 为员工的工资，X 为员工的性别虚拟变量，若为女性员工，其取值为 0；若为男性，其取值为 1。则
  - ✓ $E(Y|X=0)$ 为女性员工的平均工资，
  - ✓ $E(Y|X=1)$ 为男性员工的平均工资。
- 根据重复期望法则，可得全体员工的平均工资为

$$
\begin{aligned}
E(Y) &= E[E(Y \mid X)] \\
&= P(X=0)E(Y \mid X=0) + P(X=1)E(Y \mid X=1)
\end{aligned}
$$

其中 $P(X=0)$ 为女性员工在劳动力中的比例，$P(X=1)$ 为男性员工在劳动力中的比例。

- 重复期望法则的使用可清楚地了解劳动力市场的收入分配情况。

条件均值刻画 X 对 Y 的预测关系

- 条件均值 $E(Y|X)$ 仅为 X 的函数。
- 为了明确 X 对 Y 的预测关系，也被称为 Y 对 X 的回归函数。
- 该预测关系是统计学和计量经济学的一个主要研究对象。

定义 5.25 [均方误准则 (Mean Squared Error Criterion)]

- 假设 X 和 Y 是定义在同一样本空间上的随机变量，且 Y 具有有限方差，则条件均值 $E(Y|X)$ 是 $E[Y-g(X)]^2$ 最小化问题的最优解，即

$$E(Y \mid X) = \arg\min_{g(\cdot)} E[Y-g(X)]^2$$

其中最小化是对所有可测且平方可积函数求得。

证明：

- 令 $g_0(X)=E(Y|X)$，则有

$$
\begin{aligned}
\mathrm{MSE}(g) &= E[Y-g(X)]^2 \\
&= E\{[Y-g_0(X)]+[g_0(X)-g(X)]\}^2 \\
&= E\{[Y-g_0(X)]^2+[g_0(X)-g(X)]^2\} \\
&\quad +2E\{[Y-g_0(X)][g_0(X)-g(X)]\} \\
&= E\{[Y-g_0(X)]^2\}+E\{[g_0(X)-g(X)]^2\}
\end{aligned}
$$

- 根据重复期望法则以及 $E\{[Y-g_0(X)]\,|X\}=0$，可证倒数第二个等式中的交叉项为零：

$$
\begin{aligned}
E\{[Y-g_0(X)][g_0(X)-g(X)]\} &= E_X\{E[(Y-g_0(X))(g_0(X)-g(X))\mid X]\} \\
&= E_X\{(g_0(X)-g(X))E[(Y-g_0(X))\mid X]\} \\
&= E_X[(g_0(X)-g(X))\cdot 0] \\
&= 0
\end{aligned}
$$

证明 (Cont.)：

- 在 MSE(g) 的分解式中，第一项 $E[Y-g_0(X)]^2$ 与 $g(\cdot)$ 无关。
- 因此，求 $g(\cdot)$ 最小化 MSE(g) 等价于最小化第二项 $E[g_0(X)-g(X)]^2$。
- 当且仅当 $g(X)=g_0(X)$ 时，该项达到最小。

证毕。

定理 5.26 [回归恒等式 (Regression Identity)]

假设 $E(Y \mid X)$ 存在，则存在随机变量 ε 使得 $Y = E(Y \mid X) + \varepsilon$

其中 ε 满足 $E(\varepsilon \mid X) = 0$。

证明：

- 定义 $\varepsilon = Y - E(Y \mid X)$，则 $Y = E(Y|X) + \varepsilon$，其中

$$
\begin{aligned}
E(\varepsilon \mid X) &= E\{[Y - E(Y \mid X)] \mid X\} \\
&= E(Y \mid X) - E(Y \mid X) \\
&= 0
\end{aligned}
$$

证毕。

定理 5.26 的含义

- 随机变量 ε 通常称为回归函数 $E(Y \mid X)$ 的随机扰动项，反映了 X 和 Y 之间关系的不确定性程度。当 $\varepsilon = 0$ 时，X 和 Y 之间的关系完全确定。此时，$Y = g_0(X)$，其中 $g_0(\cdot)$ 为某个可测函数。
- 条件 $E(\varepsilon \mid X) = 0$ 的含义：该条件表明 ε 不包含可用于预测 Y 的条件期望值的任何有关 X 的有用信息。X 的系统信息中能够用于预测 Y 之条件均值的部分已完全反映在回归函数 $E(Y \mid X)$ 中。
- 回归函数 $E(Y \mid X)$ 是 X 的线性或非线性函数。当 $E(Y \mid X) = a + bX$ 时，称为线性回归函数。第十章将系统讨论线性回归模型的估计和检验理论。

例 5.40：

- 假设 (X,Y) 服从二元正态分布 $BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则条件均值

$$E(Y \mid X)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)$$

- 此为 X 的线性函数。
- 当 $E(Y|X)=a+bX$ 时，线性回归方程为

$$Y=a+bX+\varepsilon$$

- 其中 $E(\varepsilon|X)=0$。

引理 5.7：

假设 $Y=a+bX+\varepsilon$，其中 $E(\varepsilon|X)=0$，则 $E(X\varepsilon)=0$。

证明：

- 根据定理 5.24 的重复期望法则，有

$$\begin{aligned} E(X\varepsilon) &= E[E(X\varepsilon \mid X)] \\ &= E[XE(\varepsilon \mid X)] \\ &= E(X\cdot 0) \\ &= 0 \end{aligned}$$

- 这意味着 $cov(X,\varepsilon)=0$。(为什么？)

证毕。

- 由于 ε 不包含可用于预测 Y 的条件期望的 X 的信息，ε 应当与 X 正交，即 $E(X\varepsilon)=0$。根据正交性可得，ε 和 X 不相关，即 $\operatorname{cov}(X,\varepsilon)=0$。
- 事实上，根据重复期望法则，$E(\varepsilon|X)=0$ 意味着 ε 与 X 的任意可测函数 $h(X)$ 均正交，即 $E[\varepsilon h(X)]=0$。

例 5.42 [消费函数 (Consumption Function) 和边际消费倾向 (Marginal Propensity to Consume)]：

- 假设 Y 为消费，X 为收入，且 $Y = a + bX + \varepsilon$
- 其中 ε 代表了其他影响消费的随机因素，满足 $E(\varepsilon \mid X) = 0$。条件期望

$$E(Y \mid X) = a + bX$$

称为消费函数，其导数

$$\frac{dE(Y \mid X)}{dX} = b$$

为预期边际消费倾向，即当收入增加 1 个单位时消费的预期增加量。这是凯恩斯有效需求不足理论中最重要的一个概念。

例 5.43 [条件均值和有效市场假说 (Efficient Market Hypothesis, EMH)]：

- 假设 Y_t 表示第 t 期的资产收益率，I_{t-1} 为第 $t-1$ 期可获得的信息。
- 假设投资者试图用第 $t-1$ 期获得的信息预测第 t 期资产收益率 Y_t。若基于 $t-1$ 期信息的预期收益率 $E(Y_t \mid I_{t-1})$ 和长期市场平均收益率 $E(Y_t)$ 相等，则 $t-1$ 期信息对未来预期收益率 Y_t 无预测能力。
- 此时，称资本市场对 $t-1$ 期信息是有效的。该假说的正式表述如下

$$E(Y_t \mid I_{t-1}) = E(Y_t)$$

存在三种形式的有效市场假说：

(1) 弱式有效市场，其中 I_{t-1} 仅包含第 $t-1$ 期**可获取的资产收益率所有历史信息。**

(2) 半强式有效市场，其中 I_{t-1} 包含第 $t-1$ 期**可获取的所有公开信息。**

(3) 强式有效市场，其中 I_{t-1} 不仅包含第 $t-1$ 期**可获取的公开信息，还包括一部分未公开的内部信息。**

例 5.44 [期望损失 (Expected Shortfall) 与金融风险管理 (Financial Risk Management)]：

- 例 3.33 介绍了投资组合在某个时期的风险价值的概念。
- 在 α 水平上的风险价值 $V_t(\alpha)$ 定义如下

$$P[X_t < -V_t(\alpha) \mid I_{t-1}] = \alpha$$

- 其中
  - ✓ X_t 为投资组合在第 t 时期的收益，
  - ✓ I_{t-1} 为第 $t-1$ 期可获得的信息。
  - ✓ 风险价值 $V_t(\alpha)$ 是实际损失以概率 α 超出的临界值，通常用于确定资产充足水平以防止极端损失事件的发生。

例 5.44 (Cont.):

- 在实际应用中，仅基于风险价值确定资产充足水平可能还不够谨慎。
- 一些金融监管者建议使用期望损失即条件均值 $ES_t(\alpha)$ 设定资产充足水平。

$$ES_t(\alpha) = -E[X_t \mid X_t < -V_t(\alpha)]$$

  - ✓ 在 α 水平的期望损失是假定危机发生情况下的预期损失(当实际损失超出风险价值时即视为危机发生)。
  - ✓ 当 α 很小时，期望损失总是大于风险价值，从而提供了更为谨慎的风险管理方式。

例 5.45 [动态资产定价模型 (Dynamic CAPM) 与欧拉方程 (Euler Equation)]：

- 在基于消费的标准资产定价模型中，代表性投资者在跨期预算约束下通过最大化如下预期终身效用以选择最优消费路径 $\{C_t, t=1,2,\cdots\}$：

$$\max_{\{C_t\}} E\sum_{j=0}^{\infty}\beta^j u(C_{t+j})$$

- 其中
  - ✓ β 为时间折现因子 (time discount factor)，
  - ✓ $u(\cdot)$ 为投资者在每个时期的效用函数。

例 5.45 (Cont.)：

- 该跨期效用最大化问题的**一阶条件**为：

$$E(M_{t+1}R_{t+1} \mid I_t) = 1$$

- 其中
  - ✓ $M_{t+1} = \beta\frac{u'(C_{t+1})}{u'(C_t)}$ 称为随机折现因子（stochastic discount factor），反映了代表性投资者的风险态度，
  - ✓ 而 R_{t+1} 则为从第 t 期到第 $t+1$ 期的总资产收益率。
- 上述以条件均值刻画的**一阶条件** (first order conditions, FOC) 通常称为**欧拉方程** (Euler equation)。

例 5.45 (Cont.):

- 定义**随机定价误差**

$$\varepsilon_{t+1} = M_{t+1}R_{t+1} - 1$$

- 则欧拉方程可等价表示为

$$E(\varepsilon_{t+1} \mid I_t) = 0$$

- 直观上，欧拉方程揭示了第 $t+1$ 期的预期风险补偿收益应等于 1 (即第 t 期的投资成本)，也就是在每个时期均不存在系统定价偏差。
- 它刻画了不确定条件下最优资产投资路径，从而也等价地刻画了不确定条件下最优消费路径。

定义 5.23 [条件方差 (Conditional Variance)]

- 给定 $X = x$ 时 Y 的条件方差定义为

$$\operatorname{var}(Y \mid X = x) = \int_{-\infty}^{\infty} [y - E(Y \mid X = x)]^2 dF_{Y|X}(y \mid x)$$

- 直观上，$\varepsilon = Y - E(Y \mid X)$ 为用 $E(Y \mid X)$ 预测 Y 时，Y 中不可预测的成分。
- 条件方差 $\operatorname{var}(Y \mid X) = \operatorname{var}(\varepsilon \mid X)$ 度量了在给定信息 X 条件下，Y 的波动性或变异性如何随 X 的变化而变化。

条件方差函数 $\mathrm{var}(Y|X)$

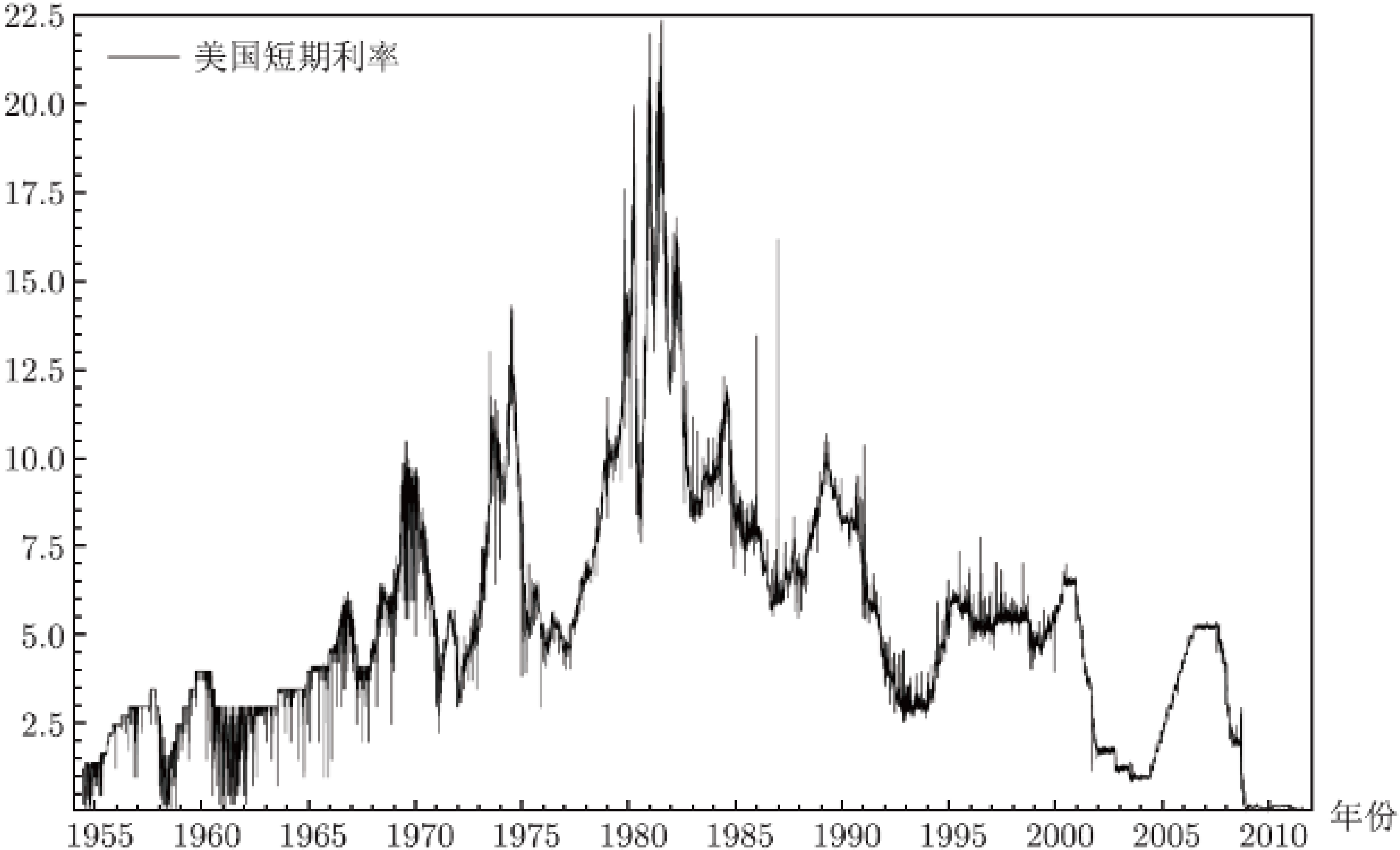
- **条件同方差（conditional homoskedasticity)**： $\mathrm{var}(Y|X)=\sigma^2$ 为常数，即 $\mathrm{var}(Y|X)$ 不依赖于 X。
  - ✓ 条件同方差是经典回归模型的一个重要假设（参考第十章)。
- **条件异方差 (conditional heteroskedasticity)**： $\mathrm{var}(Y\mid X)\neq\sigma^2$，$Y$ 基于 X 的条件方差是 X 的函数，即条件方差依赖于 X。
  - ✓ 因为不同经济主体之间、不同时间段之间存在异质性，所以条件异方差现象并不罕见。比如，规模较大的企业，其产出变化常常较大。

例 5.46 [利率波动的水平效应 (Level Effect)]:

- 利率波动的一个重要经验典型特征事实是：利率波动取决于利率水平，即利率水平越高，其波动越大，如图 5.12 所示的美国短期利率情形。
- 这一现象称为**利率波动的水平效应**，通常建模如下

$$\text{var}(r_t \mid I_{t-1}) = \alpha r_{t-1}^{\rho}$$

- 其中
  - ✓ r_t 为第 t 期的短期利率，
  - ✓ I_{t-1} 为第 $t-1$ 期可获得的信息。
- 更多讨论参见 Chan et al. (1992)。



美国短期利率
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
年份

图 5.12：美国短期利率

例 5.47 [ARCH 模型和波动聚类 (Volatility Clustering)]：

- 在金融市场上，资产价格的当期较大波动倾向于伴随未来较大波动，当期较小的波动倾向于伴随未来较小波动。该现象称为**波动聚类** (Mandelbrot, 1963)。
- 为解释这一经验典型特征事实，Engle (1982) 提出了一类**自回归条件异方差 (ARCH) 模型**对波动进行建模。
- 假设 Y_t 为第 t 时期的资产收益率，那么 **ARCH(1) 模型**为

$$\mathrm{var}(Y_t \mid I_{t-1}) = \alpha + \beta Y_{t-1}^2$$

其中参数 $\alpha, \beta > 0$，且 I_{t-1} 包含了所有过去的资产收益信息。

引理 5.8：

$$\operatorname{var}(Y \mid X) = E\left(Y^2 \mid X\right) - [E(Y \mid X)]^2$$

例 5.48：

与例 5.39 相同，令 $f_{XY}(x,y) = e^{-y}$，其中 $0 < x < y < \infty$。根据例 5.39 可知 $E(Y| x) = 1 + x$，则

CONTINUE

例 5.48 (Cont.):

$$
\begin{aligned}
\operatorname{var}(Y \mid x) &= E\left(Y^2 \mid x\right) - [E(Y \mid x)]^2 \\
&= \int_x^\infty y^2 e^{-(y-x)} dy - (1+x)^2 \\
&= e^x \int_x^\infty y^2 e^{-y} dy - (1+x)^2 \\
&= -e^x \int_x^\infty y^2 de^{-y} - (1+x)^2 \text{(这里 } de^{-y} = -e^{-y} dy\text{)} \\
&= -e^x \left(y^2 e^{-y} \Big|_x^\infty - \int_x^\infty e^{-y} dy^2 \right) - (1+x)^2 \\
&= -e^x \left(0 - x^2 e^{-x} - 2 \int_x^\infty y e^{-y} dy \right) - (1+x)^2 \\
&= x^2 + 2e^x \int_x^\infty y e^{-y} dy - (1+x)^2 \\
&= x^2 + 2 \int_x^\infty y e^{-(y-x)} dy - (1+x)^2 \\
&= x^2 + 2(1+x) - (1+x)^2 \\
&= 1
\end{aligned}
$$

例 5.49 [二维正态分布与条件同方差]:

假设 X 和 Y 服从二元正态分布 $BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，证明条件方差 $\mathrm{var}(Y|X)=\sigma_2^2(1-\rho^2)$。

解:

- 在第五章第六节中，已证对二元正态分布，给定 X 时，Y 的条件分布服从正态分布:
  - ✓ **均值**为 $\mu_2+\frac{\rho\sigma_2}{\sigma_1}(X-\mu_1)$,
  - ✓ **方差**为 $\sigma_2^2(1-\rho^2)$。
- 因此，**条件方差** $\mathrm{var}(Y|X)=\sigma_2^2(1-\rho^2)$，其不依赖于 X。

定理 5.27 [方差分解 (Variance Decomposition)]

对任意两个存在有限二阶矩的随机变量 X 和 Y，有

$$\mathrm{var}(Y) = \mathrm{var}[E(Y \mid X)] + E[\mathrm{var}(Y \mid X)]$$

证明：

- 令 $g_0(X) = E(Y \mid X)$，有 $Y = g_0(X) + \varepsilon$，其中 $E(\varepsilon \mid X) = 0$。则

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{var}[g_0(X) + \varepsilon] \\ &= \mathrm{var}[g_0(X)] + \mathrm{var}(\varepsilon) + 2\mathrm{cov}[g_0(X), \varepsilon] \\ &= \mathrm{var}[g_0(X)] + \mathrm{var}(\varepsilon)\end{aligned}$$

- 其中最后一个等式由定理 5.24 的重复期望法则以及 $E(\varepsilon \mid X) = 0$ 而得，因为

$$\begin{aligned}\mathrm{cov}[g_0(X), \varepsilon] &= E[g_0(X)\varepsilon] - E[g_0(X)]E(\varepsilon) \\ &= 0\end{aligned}$$

证明 (Cont.)：

- 由于第一项 $\mathrm{var}[g_0(X)] = \mathrm{var}[E(Y \mid X)]$，只需证明 $\mathrm{var}(\varepsilon) = E[\mathrm{var}(Y \mid X)]$。
- 由 $E(\varepsilon \mid X) = 0$ 以及重复期望法则，有 $E(\varepsilon) = 0$ 且

$$\begin{aligned}\mathrm{var}(\varepsilon) &= E\left(\varepsilon^2\right) \\ &= E\left[E\left(\varepsilon^2 \mid X\right)\right] \\ &= E[\mathrm{var}(Y \mid X)]\end{aligned}$$

- 其中，由定义有 $\mathrm{var}(Y \mid X) = E(\varepsilon^2 \mid X)$。

证毕。

方差分解定理的含义

方差分解定理也称为**总方差定律** **(the law of total variance)** 或**迭代方差定律** **(the law of iterated variances)**。

- 直观上，方差分解定理表示 Y 的总方差 $\text{var}(Y)$ 等于两部分之和：
  - ✓ $\text{var}[E(Y|X)]$：最优 MSE 预测量 $E(Y|X)$ 的波动，其测度了 $E(Y|X)$ 对 Y 的预测能力。$E(Y|X)$ 波动越大，其对 Y 的预测能力越强；
  - ✓ $E[\text{var}(Y|X)]$：预测误差 $\varepsilon = Y - E(Y|X)$ 之平方的均值，即 X 对 Y 的均方预测误差。

条件偏度与条件峰度

- 在计量经济学，前两阶条件矩非常重要。然而，经济学和金融学研究越来越关注高阶条件矩。以下是两个例子：
  - ✓ (1) **条件偏度** (conditional skewness)

$$S(Y \mid X) = \frac{E\left(\varepsilon^3 \mid X\right)}{[\mathrm{var}(\varepsilon \mid X)]^{3/2}}$$

  - ✓ (2) **条件峰度** (conditional kurtosis)

$$K(Y \mid X) = \frac{E(\varepsilon^4 \mid X)}{[\mathrm{var}(\varepsilon \mid X)]^2}$$

例 5.52：

设 (X,Y) 服从二维正态分布 $NB(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求 $E(Y \mid X)$，$\mathrm{var}(Y \mid X)$，$S(Y \mid X)$ 和 $K(Y \mid X)$。

解：

- 前面已证，Y 基于 X 的条件分布为正态分布

$$Y \Big| X \sim N\left[\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(X-\mu_1), \sigma_2^2\left(1-\rho^2\right)\right]$$

解 (Cont.)：

- 因此，**条件均值**

$$E(Y \mid X) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(X - \mu_1)$$

- **条件方差**

$$\mathrm{var}(Y \mid X) = \sigma_2^2(1 - \rho^2)$$

- **条件偏度**

$$S(Y \mid X) = 0$$

- **条件峰度**

$$K(Y \mid X) = 3$$

- 本例说明对二元正态变量，仅一阶条件矩 $E(Y \mid X)$ 依赖于 X，其他高阶条件矩均为常数，即不依赖于 X。

在经济实证研究中，应当采用哪个条件矩进行建模呢?

- 这取决于所研究的经济问题的本质。
  - ✓ 对一些具体应用，如市场有效性假说和动态资产定价，需要对经济变量或其函数的**条件均值**建模。
  - ✓ 对诸如波动溢出 (volatility spillover) 之类的研究，则需要对**条件方差**建模。
  - ✓ 在研究金融风险管理(如，金融危机管理)、对冲和衍生品定价时，则需要对**高阶条件矩乃至整个条件分布**建模。

在经济实证研究中，应当采用哪个条件矩进行建模呢？(Cont.)

- 更一般地，考察以下不确定条件下的决策问题。假设决策者的损失函数 (loss function) 为 $l(\cdot)$。该决策者在已知信息 $X=x$ 的基础上作出决策 a 以最小化预期损失函数。最小化问题为

$$\min_a E[l(Y-a)\mid X=x]=\min_a \int_{-\infty}^{\infty} l(y-a)f_{Y|X}(y\mid x)dy$$

- 其中 Y 为决策者制定决策时不可预知的随机结果，$f_{Y|X}(y\mid x)$ 为 Y 基于 $X=x$ 的条件 PDF。
- 在实际应用中，条件 PDF $f_{Y|X}(y\mid x)$ 通常未知，从而需要对其建模。

在经济实证研究中，应当采用哪个条件矩进行建模呢？(Cont.)

- 当损失函数为二次型时：

$$l(e) = e^2$$

✓ 最优决策为条件均值

$$a^*(X) = E(Y \mid X)$$

- 当损失函数为所谓的 Linexp 函数，且 Y 基于 X 的条件分布为正态分布时：

$$l(e) = \frac{1}{\alpha^2}\left[e^{\alpha e} - (1 + \alpha e)\right]$$

✓ 最优决策为条件均值和条件方差的线性组合

$$a^*(X) = E(Y \mid X) + \frac{\alpha}{2}\operatorname{var}(Y \mid X)$$

✓ 除了 $\alpha = 0$ 的情形，Linexp 损失函数为不对称函数。

✓ 当 $\alpha = 0$ 时，Linexp 函数变成平方损失函数。

不同 α 值对应的 *Linexp* 损失函数

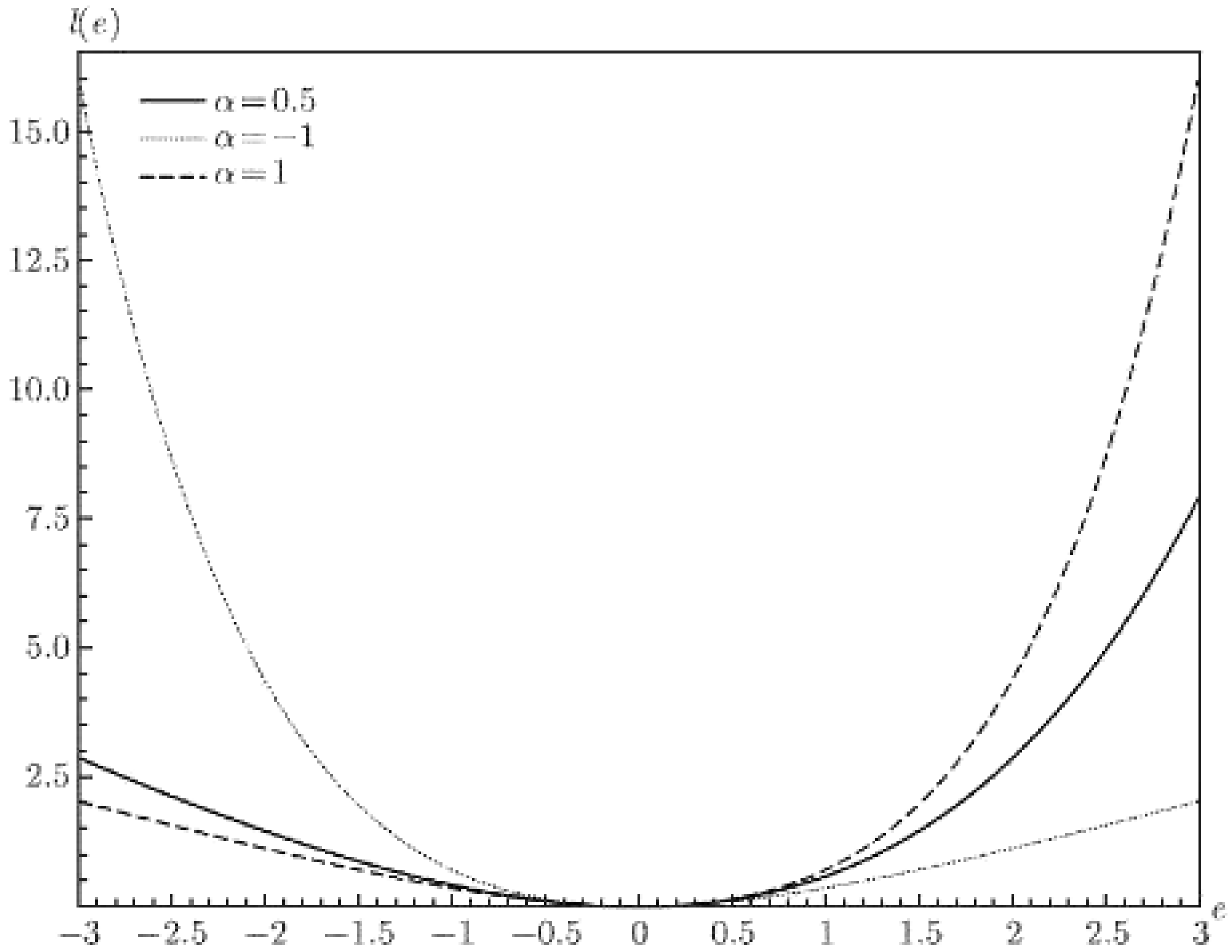


图 5.13：不同 α 值对应的 Linexp 损失函数

目 录

总结

- 计量经济学最重要目的之一是识别各种经济关系。
- 本章主要内容：
  - ✓ 引入**联合** CDF 刻画随机变量 X 和 Y 的联合概率分布，并且针对 (X,Y) 为离散随机变量或连续随机变量情形，分别给出了**联合** PMF、**联合** PDF 及其相应的**条件分布**形式
  - ✓ 介绍了**二元变换联合分布**的计算方法、借助条件分布考察了 X 和 Y 之间的预测关系
  - ✓ 详细探讨了**独立性**概念及其对联合分布、条件分布、相关性以及联合矩生成函数的含义
  - ✓ 介绍了一类**二元正态分布**，并讨论了其重要性质

中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Thank You !