工程概率统计

Probability and Statistics for Engineering

第三章联合分布

Chapter 3 Joint Distributions

3.1 随机向量与联合分布

我们之前每次只考虑一个随机变量，然而，通常我们需要处理多个随机变量。例如：

- 一个随机选择的人的身高和体重。
- 新加坡在随机选择的一天的温度、湿度、风力、降水量。

由于同一对象的不同度量通常相互关联，因此应同时考虑它们。

这里我们将讨论如何同时研究两个随机变量 X 和 Y，所有概念都可以推广到 n 个随机变量 $X_1, X_2, \ldots, X_n$。

随机向量 如果 $\omega \in \Omega \mapsto (X(\omega), Y(\omega))$ 是一个值在 $\mathbb{R}^2$ 上的函数，则称 (X, Y) 是一个**随机向量**。(X, Y) 也可以称为二维随机变量。

联合累积分布函数

对于随机向量 (X, Y)，无论是离散还是连续，其累积分布函数定义为

$$F(x, y) = P(X \le x, Y \le y), \quad \forall x, y \in \mathbb{R}.$$

$F(x, y)$ 也称为 X 和 Y 的**联合累积分布函数**。

你可以将 $\{X \le x\}$ 视为事件 A，将 $\{Y \le y\}$ 视为事件 B，那么

$$F(x, y) = P(A \cap B).$$

由此可知，对于 $\forall x, y \in \mathbb{R}$，我们有

$$\begin{aligned}
&0 \le F(x, y) \le 1 \\
&F(+\infty, +\infty) = 1, \quad F(-\infty, -\infty) = 0, \\
&F(-\infty, y) = 0, \quad F(x, -\infty) = 0.
\end{aligned}$$

联合 CDF 可用于计算对于任何 $-\infty < x_1 < x_2 < \infty, -\infty < y_1 < y_2 < \infty$ 的 $P(x_1 < X \le x_2, y_1 < Y \le y_2)$。

$$P\{x_1 < X \le x_2, \quad y_1 < Y \le y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \ge 0.$$

有了联合 CDF，就可以直接定义边缘 CDF。

边缘 CDF 设 $F(x, y)$ 是 (X, Y) 的联合 CDF，$F_X(x)$ 是不考虑 Y 时 X 的 CDF，$F_Y(y)$ 是不考虑 X 时 Y 的 CDF，则对于 $\forall x, y \in \mathbb{R}$：

$$F_X(x) = F(x, \infty), \quad F_Y(y) = F(\infty, y).$$

$F_X(F_Y)$ 也称为 $X(Y)$ 的**边缘累积分布函数**。

注意：虽然联合 CDF 唯一地确定了边缘 CDF，但反之则不成立。

例 3.1

假设随机向量 (X, Y) 的联合 CDF 为 (a, b, c 是常数)

$$F(x, y) = a\left(b + \arctan\frac{x}{2}\right)\left(c + \arctan\frac{y}{2}\right), \quad -\infty < x, y < \infty.$$

1. 确定 a, b, c 的值。
2. 计算 $P(-2 < X \leq 2, -2 < Y \leq 2)$。
3. 求 X 和 Y 的边缘 CDF。

解答

1. 根据联合 CDF 的基本性质，我们有：

$$\begin{aligned}
F(+\infty, +\infty) &= a\left(b + \frac{\pi}{2}\right)\left(c + \frac{\pi}{2}\right) = 1, \\
F(-\infty, +\infty) &= a\left(b - \frac{\pi}{2}\right)\left(c + \frac{\pi}{2}\right) = 0, \\
F(+\infty, -\infty) &= a\left(b + \frac{\pi}{2}\right)\left(c - \frac{\pi}{2}\right) = 0.
\end{aligned}$$

   解得 $b = \frac{\pi}{2}, c = \frac{\pi}{2}, a = \frac{1}{\pi^2}$。

   因此，联合 CDF 为

$$F(x, y) = \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)\left(\frac{\pi}{2} + \arctan\frac{y}{2}\right).$$

2. 根据联合 CDF 的定义和性质，我们有：

$$\begin{aligned}
&P(-2 < X \leq 2, -2 < Y \leq 2) = F(2,2) - F(2,-2) - F(-2,2) + F(-2,-2) \\
&= \frac{1}{\pi^2}\left[\left(\frac{\pi}{2} + \arctan(1)\right)^2 - 2\left(\frac{\pi}{2} + \arctan(1)\right)\left(\frac{\pi}{2} + \arctan(-1)\right) + \left(\frac{\pi}{2} + \arctan(-1)\right)^2\right] \\
&= \frac{1}{\pi^2}\left[\left(\frac{\pi}{2} + \arctan(1)\right) - \left(\frac{\pi}{2} + \arctan(-1)\right)\right]^2 = \frac{1}{\pi^2}\left[\frac{\pi}{4} - \left(-\frac{\pi}{4}\right)\right]^2 = \frac{1}{4}.
\end{aligned}$$

3. 根据边缘分布的定义，我们有：

$$\begin{aligned}
F_X(x) &= F(x, +\infty) = \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)\left(\frac{\pi}{2} + \frac{\pi}{2}\right) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2}, \\
F_Y(y) &= F(+\infty, y) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{2}.
\end{aligned}$$

离散随机向量的联合与边缘 PMF

对于随机向量 (X, Y)，设 $S_X = \{x_1, x_2, \ldots\}$ 和 $S_Y = \{y_1, y_2, \ldots\}$ 分别是 X 和 Y 的支撑集。则 (X, Y) 的**联合概率质量函数**定义为

$$p(x_i, y_j) = P(X = x_i, Y = y_j) \triangleq p_{ij}, \quad i, j = 1, 2, \ldots$$

联合 PMF 满足：

- 非负性：$p_{ij} \geq 0, i, j = 1, 2, \ldots$
- 归一性：$\sum_i \sum_j p_{ij} = 1$

X 的**边缘概率质量函数**是不考虑 Y 时 X 的 PMF：

$$p_X(x_i) = P(X = x_i) = \sum_{j=1}^{\infty} p_{ij} \triangleq p_i, \quad i = 1, 2, \ldots$$

类似地，Y 的边缘 PMF 为

$$p_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{\infty} p_{ij} \triangleq p_{,j}, \quad j = 1, 2, \ldots$$

联合 PMF 通常以表格形式显示：

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

例 3.2

独立地投掷两个骰子。设 X 为较小的点数，Y 为较大的点数。如果两个骰子显示相同的点数，例如 z 点，则 $X = Y = z$。

1. 求 (X, Y) 的联合 PMF；
2. 求 X 的边缘 PMF。

解答

1. 由于两个骰子是独立投掷的，不难得到 (X, Y) 的联合 PMF：

$X\backslash Y$	1	2	3	4	5	6
1	1/36	1/18	1/18	1/18	1/18	1/18
2	0	1/36	1/18	1/18	1/18	1/18
3	0	0	1/36	1/18	1/18	1/18
4	0	0	0	1/36	1/18	1/18
5	0	0	0	0	1/36	1/18
6	0	0	0	0	0	1/36

2. 有了 (X, Y) 的联合 PMF，可以直接得到 X 的边缘 PMF 为：

取值	1	2	3	4	5	6
概率	11/36	9/36	7/36	5/36	3/36	1/36

连续随机向量的联合 PDF

如果存在一个非负函数 $f(x,y)$，定义在所有 $x,y\in\mathbb{R}^2$ 上，满足对于任何 $D\subset\mathbb{R}^2$，

$$P((X,Y)\in D)=\iint_{(x,y)\in D} f(x,y)dxdy,$$

则称 (X,Y) 是一个连续随机向量。 $f(x,y)$ 称为 (X,Y) 的**联合概率密度函数**。

特别地，(X,Y) 的联合 CDF 为

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

由此可得

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

联合 PDF 满足：

- 非负性：$f(x,y)\geq 0,\forall x,y\in\mathbb{R}$
- 归一性：$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$

类似于单随机变量的 PDF，联合 PDF $f(x,y)\neq P(X=x,Y=y)$。相反，它反映了概率在 (x,y) 附近的集中程度。

例 3.3

随机选择的机器的两个电子元件的寿命（以年为单位）用随机变量 X 和 Y 表示，其联合 PDF 为

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

计算：

1. $P(X<1,Y<1)$
2. $P(X<Y)$

解答

1. 根据 (X,Y) 的联合 PDF 的定义，我们有：

$$\begin{aligned}P(X<1,Y<1)&=\int_{-\infty}^{1}\left(\int_{-\infty}^{1} f(x,y)dx\right)dy=\int_{0}^{1}\left(\int_{0}^{1} 2e^{-x-2y}dx\right)dy\\&=\int_{0}^{1}\left(2(1-e^{-1})e^{-2y}\right)dy=(1-e^{-1})\int_{0}^{1} 2e^{-2y}dy\\&=(1-e^{-1})(1-e^{-2})\approx 0.5466.\end{aligned}$$

2. 再次根据 (X,Y) 的联合 PDF：

$$\begin{aligned}P(X<Y)&=\iint_{\{(x,y)|0<x<y<\infty\}} f(x,y)dxdy=\int_{0}^{\infty}\left(\int_{0}^{y} 2e^{-x-2y}dx\right)dy\\&=\int_{0}^{\infty} 2e^{-2y}(1-e^{-y})dy=\int_{0}^{\infty} 2e^{-2y}dy-\int_{0}^{\infty} 2e^{-3y}dy\\&=1-\frac{2}{3}=\frac{1}{3}.\end{aligned}$$

连续随机向量的边缘 PDF

(X,Y) 是一个具有联合 PDF $f(x,y)$ 的连续随机向量，则 X 的**边缘概率密度函数**，即不考虑 Y 时 X 的 PDF 为

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

类似地，Y 的边缘 PDF 为

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

注意：联合 PDF 唯一地定义了边缘 PDF，但反之则不成立。

例 3.4

假设随机向量 (X,Y) 的联合 PDF 为

$$f(x,y) = \begin{cases} c, & \text{若 } x^2+y^2 \le R^2 \\ 0, & \text{否则} \end{cases}$$

1. 确定常数 c
2. 求 X 的边缘 PDF

解答

1. 根据联合 PDF 的归一性：
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1,$$
由此可得
$$1 = \iint_{x^2+y^2\le R^2} cdxdy \implies c\cdot(\pi R^2) = 1 \implies c = \frac{1}{\pi R^2}.$$

2. 有了 (X,Y) 的联合 PDF，可以直接得到 X 的边缘 PDF 为：
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}, \quad \text{若 } |x| \le R$$
且 $f_X(x) = 0$，否则。

条件 PMF/PDF

对于具有联合 PMF $p(x,y)$ 的离散随机向量 (X,Y)，给定 $Y=y$ 时 X 的**条件概率质量函数**定义为

$$p_{X|Y}(x|y) \triangleq P(X=x|Y=y) = \frac{p(x,y)}{p_Y(y)},$$

对于所有满足 $p_Y(y)>0$ 的 y 值。

对于具有联合 PDF $f(x,y)$ 的连续随机向量 (X,Y)，给定 $Y=y$ 时 X 的**条件概率密度函数**定义为

$$f_{X|Y}(x|y) \triangleq \frac{f(x,y)}{f_Y(y)} \quad (-\infty < x < \infty),$$

对于所有满足 $f_Y(y)>0$ 的 y 值。

条件 PDF 反映了离散情况下条件 PMF 的定义，只是将联合/边缘 PMF 替换为联合/边缘 PDF。

条件 PMF/PDF 也满足非负性和归一性。

问题：对于连续情况，$P(Y=y)=0$，因此 $P(X=x|Y=y)$ 或 $P(X\leq x|Y=y)$ 没有定义。那么如何理解给定 $Y=y$ 时 X 的条件 PDF？

以 $Y=y$ 为条件可以理解为以 $\{y\leq Y\leq y+\epsilon\}$ 为条件，其中 $\epsilon\to 0$。

考虑条件 CDF：

$$\begin{aligned}
P\{X\leq x\mid y<Y\leq y+\epsilon\} &= \frac{P\{X\leq x, y<Y\leq y+\epsilon\}}{P\{y<Y\leq y+\epsilon\}} \\
&= \frac{\int_{-\infty}^{x}\int_{y}^{y+\epsilon} f(u,v)dvdu}{\int_{y}^{y+\epsilon} f_Y(v)dv} \\
&\to \int_{-\infty}^{x}\left(\frac{f(u,y)}{f_Y(y)}\right)du \quad (\epsilon\to 0).
\end{aligned}$$

根据条件 PDF 的定义，我们有

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

对两边关于 y 积分，X 的边缘分布可以表示为

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

这就是连续情况下的全概率定律。

例 3.5

对于随机选择的交通事故中的人，设 X 为他/她的受伤程度，Y 为事故时他/她佩戴的安全装置类型。X 和 Y 的联合 PMF 为

$X\backslash Y$	1 (无)	2 (仅安全带)	3 (安全带和安全带)	$p_X(x)$
1 (无)	0.065	0.075	0.060	0.20
2 (轻微)	0.165	0.160	0.125	0.45
3 (严重)	0.145	0.10	0.055	0.30
4 (死亡)	0.025	0.015	0.010	0.05
$p_Y(y)$	0.40	0.35	0.25	1.00

1. 对于随机选择的未使用任何安全装置的人，其受伤程度的 PMF 是什么？
2. 对于随机选择的佩戴了安全带和安全带的人，其受伤程度的 PMF 是什么？

解答

1. 根据条件 PMF 的定义，给定 $Y=1$ 时 X 的条件分布为

取值	1 (无)	2 (轻微)	3 (严重)	4 (死亡)
概率	$\frac{0.065}{0.4}=0.1625$	$\frac{0.165}{0.4}=0.4125$	$\frac{0.145}{0.4}=0.3625$	$\frac{0.025}{0.4}=0.0625$

2. 类似地，给定 $Y=3$ 时 X 的条件分布为

   佩戴安全带和安全带时，严重受伤或死亡的概率比未使用任何安全装置的情况低 16.5%。

取值	1 (无)	2 (轻微)	3 (严重)	4 (死亡)
概率	$\frac{0.06}{0.25} = 0.24$	$\frac{0.125}{0.25} = 0.50$	$\frac{0.055}{0.25} = 0.22$	$\frac{0.01}{0.25} = 0.04$

例 3.4 (续)

确定给定 $Y = y$ (其中 $|y| \leq R$) 时 X 的条件 PDF。

解答　与例 3.4 类似，我们得到 Y 的边缘 PDF：

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \begin{cases} \frac{2\sqrt{R^2-y^2}}{\pi R^2}, & \text{若 } |y| \leq R, \\ 0, & \text{否则} \end{cases}$$

然后，根据定义，给定 $Y = y$ 时 X 的条件 PDF 为

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{2\sqrt{R^2-y^2}}, & \text{若 } |x| \leq \sqrt{R^2-y^2}, \\ 0, & \text{否则} \end{cases}$$

这表明给定 $Y = y$，X 服从均匀分布 $\text{Uniform}[-\sqrt{R^2-y^2}, \sqrt{R^2-y^2}]$。由于 $f_{X|Y}(x|y) \neq f_X(x)$，我们说 X 不独立于 Y。

3.2 两个随机变量之间的关系

在 3.1 节的末尾，我们提到了随机变量之间独立性的概念。回顾随机事件之间的独立性，随机变量之间的独立性可以类似定义：Y 的值不影响 X 的分布，反之亦然。例如，对于连续情况：

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

由 $f(x,y) = f_X(x)f_Y(y)$，我们有

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

随机变量的独立性

设 $F(x_1, x_2, \ldots, x_n)$ 是 $(X_1, X_2, \ldots, X_n)$ 的联合 CDF，$F_{X_i}(x_i)$ 是 X_i 的边缘 CDF，那么如果对于所有 $x_1, x_2, \ldots, x_n \in \mathbb{R}$ 我们有

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

则我们说随机变量 $X_1, X_2, \ldots, X_n$ 是（相互）独立的。

对于离散随机变量 $X_1, X_2, \ldots, X_n$，如果它们是独立的，则 PMF 满足

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n).$$

对于连续随机变量 $X_1, X_2, \ldots, X_n$，如果它们是独立的，则 PDF 满足

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n).$$

问题：三个或更多随机事件之间的独立性定义需要多个方程，为什么三个或更多随机变量之间的独立性只需要一个方程？

例 3.6

标准正态随机变量 Z 的 PDF 是 $f(z) = ce^{-z^2/2}$，$-\infty < z < \infty$。我们已经知道 $c = 1/\sqrt{2\pi}$，但是，这个值是如何得到的？令人惊讶的是，确定 c 的最简单方法是定义两个独立的标准正态随机变量，并利用它们的联合 PDF 必须积分为 1 这一事实。

解答 令随机变量 $X \sim N(0,1)$，$Y \sim N(0,1)$，且 X 和 Y 独立。则 X 和 Y 的联合 PDF 为

$$f(x,y) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}e^{\frac{-y^2}{2}} = \frac{1}{2\pi}e^{\frac{-(x^2+y^2)}{2}}$$

由于联合 PDF 必须积分为 1，我们有

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}e^{\frac{-(x^2+y^2)}{2}}dxdy$$

为了计算这个二重积分，我们转换到极坐标，使用代换 $x = r\cos\theta$，$y = r\sin\theta$ 以及 $dxdy = rdrd\theta$：

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{\frac{-(x^2+y^2)}{2}}dxdy = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty}e^{\frac{-r^2}{2}}rdrd\theta = \frac{1}{2\pi}\int_{0}^{2\pi}1d\theta = 1$$

这验证了 $c = 1/\sqrt{2\pi}$。

例 3.7

假设在随机选择的工作日进入购物中心的人数服从参数为 λ 的泊松分布。如果每个进入购物中心的人是男性的概率为 0.2，是女性的概率为 0.8。证明进入购物中心的男性和女性人数是独立的泊松随机变量，参数分别为 0.2λ 和 0.8λ。

解答 令随机变量 X_1 和 X_2 分别为进入购物中心的男性和女性人数。根据独立性的定义，我们需要证明对于 $\forall i_1, i_2 = 0, 1, \ldots$,

$$P(X_1 = i_1, X_2 = i_2) = P(X_1 = i_1)P(X_2 = i_2).$$

令 $Y = X_1 + X_2$ 为进入购物中心的总人数，则 $Y \sim \text{Poisson}(\lambda)$，即

$$P(Y = i_1 + i_2) = e^{-\lambda}\frac{\lambda^{i_1+i_2}}{(i_1+i_2)!}$$

给定 $Y = i_1 + i_2$，则 $X_1 \sim \text{Binomial}(i_1 + i_2, 0.2)$，所以

$$P(X_1 = i_1 \mid Y = i_1 + i_2) = \binom{i_1+i_2}{i_1}(0.2)^{i_1}(0.8)^{i_2} = \frac{(i_1+i_2)!}{i_1!i_2!}(0.2)^{i_1}(0.8)^{i_2}$$

因此，

$$\begin{aligned}
P(X_1 = i_1, X_2 = i_2) &= P(X_1 = i_1, X_2 = i_2 \mid Y = i_1 + i_2)P(Y = i_1 + i_2) \\
&= \frac{(i_1+i_2)!}{i_1!i_2!}(0.2)^{i_1}(0.8)^{i_2} \times e^{-\lambda}\frac{\lambda^{i_1+i_2}}{(i_1+i_2)!} \\
&= \left(e^{-0.2\lambda}\frac{(0.2\lambda)^{i_1}}{i_1!}\right)\left(e^{-0.8\lambda}\frac{(0.8\lambda)^{i_2}}{i_2!}\right)
\end{aligned}$$

有了联合 PMF，不难确定边缘 PMF，即 $X_1 \sim \text{Poisson}(0.2\lambda)$ 和 $X_2 \sim \text{Poisson}(0.8\lambda)$，并且 X_1 和 X_2 之间的独立性得证。

如果 X 和 Y 是独立的随机变量，那么，对于任何函数 g 和 h，我们有

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)],$$

$$\mathrm{Var}[g(X) \pm h(Y)] = \mathrm{Var}[g(X)] + \mathrm{Var}[h(Y)].$$

证明： 不失一般性，展示连续情况。

假设 X 和 Y 具有联合密度 $f(x,y)$，则

$$\begin{aligned}
\mathrm{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy = \mathrm{E}[g(X)]\mathrm{E}[h(Y)].
\end{aligned}$$

令

$$\mathrm{E}[g(X)] = a \quad 和 \quad \mathrm{E}[h(Y)] = b,$$

则

$$\begin{aligned}
\mathrm{Var}[g(X)+h(Y)] &= \mathrm{E}[(g(X)+h(Y)-a-b)^2] \\
&= \mathrm{E}[(g(X)-a)^2] + \mathrm{E}[(h(Y)-b)^2] + 2\mathrm{E}[(g(X)-a)(h(Y)-b)] \\
&= \mathrm{Var}[g(X)] + \mathrm{Var}[h(Y)] + 2\mathrm{Cov}(g(X),h(Y)).
\end{aligned}$$

由于 X 和 Y 独立，$g(X)$ 和 $h(Y)$ 也独立（对于适当的函数），所以 $\mathrm{Cov}(g(X),h(Y))=0$，因此

$$\mathrm{Var}[g(X)+h(Y)] = \mathrm{Var}[g(X)] + \mathrm{Var}[h(Y)].$$

对于减法的情况类似（方差性质）。

特殊情况：如果 X 和 Y 独立，则 $E(XY)=E(X)E(Y)$ 且 $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$。

如果 X 和 Y 不独立呢？考虑 $\mathrm{Var}(X+Y)$ 和 $\mathrm{Var}(X)+\mathrm{Var}(Y)$ 之间的差异：

$$\mathrm{Var}(X+Y) - \mathrm{Var}(X) - \mathrm{Var}(Y) = 2E[(X-E(X))(Y-E(Y))] = 2[E(XY)-E(X)E(Y)].$$

所以，如果 $E[(X-E(X))(Y-E(Y))] \neq 0$，则 X 和 Y 不可能独立。

因此，$E[(X-E(X))(Y-E(Y))]$ 可以用来衡量 X 和 Y 之间的关系。

协方差

X 和 Y 之间的**协方差**，记为 $\mathrm{Cov}(X,Y)$，定义为

$$\mathrm{Cov}(X,Y) \triangleq E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y)$$

如果 X 和 Y 独立，则 $\mathrm{Cov}(X,Y)=0$。然而，如果 $\mathrm{Cov}(X,Y)=0$，X 和 Y 可能不独立，我们只能说 X 和 Y 是**不相关的**。

例 3.4 (续)

我们已经知道 X 和 Y 不独立，但是，证明 $\mathrm{Cov}(X,Y)=0$。

证明 由于 X 和 Y 的边缘 PDF 都是偶函数，直接可知

$$E(X) = E(Y) = 0.$$

然后，我们计算 $E(XY)$：

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \iint_{\{x^2+y^2\le R^2\}} \frac{xy}{\pi R^2}dxdy$$

$$= \frac{1}{\pi R^2} \int_{-R}^{R} \left(\int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} x dx \right) y dy = 0.$$

(因为内层关于 x 的积分是奇函数在对称区间上的积分)

因此，$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$。

注意，当 X 和 Y 倾向于同方向变化时，Cov(X, Y) 为正；当它们倾向于反方向变化时，Cov(X, Y) 为负。

协方差具有以下性质：（a, b, c 是常数）

- 协方差-方差关系：Cov(X, X) = Var(X)
- 对称性：Cov(X, Y) = Cov(Y, X)
- 常数不协变：Cov(X, c) = 0
- 提取常数：Cov(aX, bY) = ab Cov(X, Y)
- 分配律：Cov(X + X , Y) = Cov(X , Y) + Cov(X , Y)
- 双线性性质：

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j)$$

例 3.8

盖革计数器是一种用于检测和测量电离辐射的设备。每次检测到放射性粒子时，它会发出咔哒声。假设在一个城市，放射性粒子以 $\lambda = 0.8$ 粒子/秒的速率根据泊松过程到达盖革计数器。检测到第一个粒子的时间和检测到第二个粒子的时间分别记为 X 和 Y。计算 X 和 Y 之间的协方差。

解答 令 $Z = Y - X$，则 Z 表示第一个和第二个粒子到达之间的时间间隔。根据我们之前的知识，我们知道 $X \sim \mathrm{Exp}(\lambda)$，$Z \sim \mathrm{Exp}(\lambda)$，且 X 和 Z 独立。因此，我们有

$$E(X) = E(Z) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \mathrm{Var}(Z) = \frac{1}{\lambda^2}, \quad 且 \quad \mathrm{Cov}(X, Z) = 0.$$

然后，

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X, X+Z) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Z) = \mathrm{Var}(X) + 0 = \frac{1}{\lambda^2} = \frac{1}{0.8^2} = 1.5625.$$

$\mathrm{Cov}(X,Y) > 0$ 与我们的直觉一致：第一次到达等待的时间越长，第二次到达等待的时间也越长，因为第二次到达必须在第一次之后发生。

虽然协方差衡量了两个随机变量之间的关系，但其值取决于我们测量随机变量所使用的单位/尺度。例如，设 X (单位：米) 和 Y (单位：千克) 是一个随机选择的人的身高和体重，且 $\bar{X} = 100X$ (即 $\bar{X}$ 是以厘米为单位的身高)，则 $\mathrm{Cov}(\bar{X}, Y) = 100\mathrm{Cov}(X,Y)$。因此，较大的协方差不一定意味着较强的关系。

为了使度量具有可比性，我们需要消除单位/尺度的影响。

相关系数

X 和 Y 之间的**相关系数**，记为 $\mathrm{Cor}(X,Y)$ 或 ρ_{XY}，定义为

$$\rho_{XY}=\mathrm{Cor}(X,Y)\triangleq E\left[\frac{(X-E(X))}{\mathrm{SD}(X)}\cdot\frac{(Y-E(Y))}{\mathrm{SD}(Y)}\right]=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

ρ_{XY} 是协方差的归一化版本，它是一个无量纲的数值。

问题：ρ_{XY} 衡量的是哪种关系？

例 3.9

如果 X 和 Y 的联合 PDF 是 $(0<c\le 1)$，计算 ρ_{XY}

$$f(x,y)=\frac{1}{2\pi c}\exp\left\{-\frac{x^2-2\sqrt{1-c^2}xy+y^2}{2c^2}\right\},\quad -\infty<x,y<\infty.$$

解答 考虑 X 的边缘 PDF：

$$\begin{aligned}f_X(x)&=\int_{-\infty}^{\infty}f(x,y)dy\\&=\frac{1}{2\pi c}\int_{-\infty}^{\infty}\exp\left\{-\frac{(y-\sqrt{1-c^2}x)^2}{2c^2}-\frac{c^2x^2}{2c^2}\right\}dy\\&=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\end{aligned}$$

所以 $X\sim N(0,1)$，类似地，我们得到 $Y\sim N(0,1)$。由此可知 $E(X)=E(Y)=0$，$\mathrm{Var}(X)=\mathrm{Var}(Y)=1$。

$$\begin{aligned}\mathrm{Cov}(X,Y)=E(XY)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(cuv+\sqrt{1-c^2}v^2)e^{-(u^2+v^2)/2}dudv\\&=\frac{c}{2\pi}\int_{-\infty}^{\infty}ue^{-u^2/2}du\int_{-\infty}^{\infty}ve^{-v^2/2}dv+\frac{\sqrt{1-c^2}}{2\pi}\int_{-\infty}^{\infty}e^{-u^2/2}du\int_{-\infty}^{\infty}v^2e^{-v^2/2}dv\\&=0+\frac{\sqrt{1-c^2}}{2\pi}\cdot\sqrt{2\pi}\cdot\sqrt{2\pi}=\sqrt{1-c^2}.\end{aligned}$$

(因为第一个积分是奇函数，第二个积分中 $\int_{-\infty}^{\infty}e^{-u^2/2}du=\sqrt{2\pi}$, $\int_{-\infty}^{\infty}v^2e^{-v^2/2}dv=\sqrt{2\pi}$)

因此，$\rho_{XY}=\mathrm{Cov}(X,Y)/\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}=\sqrt{1-c^2}$。

这是一个例子，说明边缘 PDF 不能唯一地确定联合 PDF。

ρ_{XY} 实际上衡量的是 X 和 Y 之间线性关系的方向和强度。

证明： 考虑使用 X 的线性函数来近似 Y，即 $\hat{Y}=a+bX$。

那么，近似的均方误差为

$$\begin{aligned}MSE=E\left[\left(Y-\hat{Y}\right)^2\right]&=E[(Y-a-bX)^2]\\&=E(Y^2)+b^2E(X^2)+a^2-2bE(XY)+2abE(X)-2aE(Y).\end{aligned}$$

接下来，我们希望关于 a 和 b 最小化 MSE。

$$\frac{\partial MSE}{\partial a} = 2a + 2bE(X) - 2E(Y) = 0$$
$$\frac{\partial MSE}{\partial b} = 2bE(X^2) - 2E(XY) + 2aE(X) = 0$$

解得

$$b_0 = \frac{E(XY) - E(X)E(Y)}{E(X^2) - [E(X)]^2} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$
$$a_0 = E(Y) - b_0E(X) = E(Y) - \frac{\text{Cov}(X,Y)}{\text{Var}(X)}E(X)$$

因此，

$$\begin{aligned}
\min_{a,b} MSE &= E[(Y - a_0 - b_0X)^2] = E[(Y - E(Y) - b_0(X - E(X)))^2] \\
&= \text{Var}(Y) + b_0^2\text{Var}(X) - 2b_0\text{Cov}(X,Y) \\
&= \text{Var}(Y) + \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)} - 2\frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)} \\
&= \text{Var}(Y) - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)} = \text{Var}(Y)\left[1 - \frac{[\text{Cov}(X,Y)]^2}{\text{Var}(X)\text{Var}(Y)}\right] = \text{Var}(Y)(1 - \rho_{XY}^2).
\end{aligned}$$