工程概率统计
Probability and Statistics for Engineering

第三章 联合分布

Chapter 3 Joint Distributions

Chapter 3 Joint Distributions

- 3.1 Random Vector and Joint Distribution
- 3.2 Relationship between Two Random Variables
- 3.3 Function of Multiple Random Variables
- 3.4 Multivariate Normal Distribution

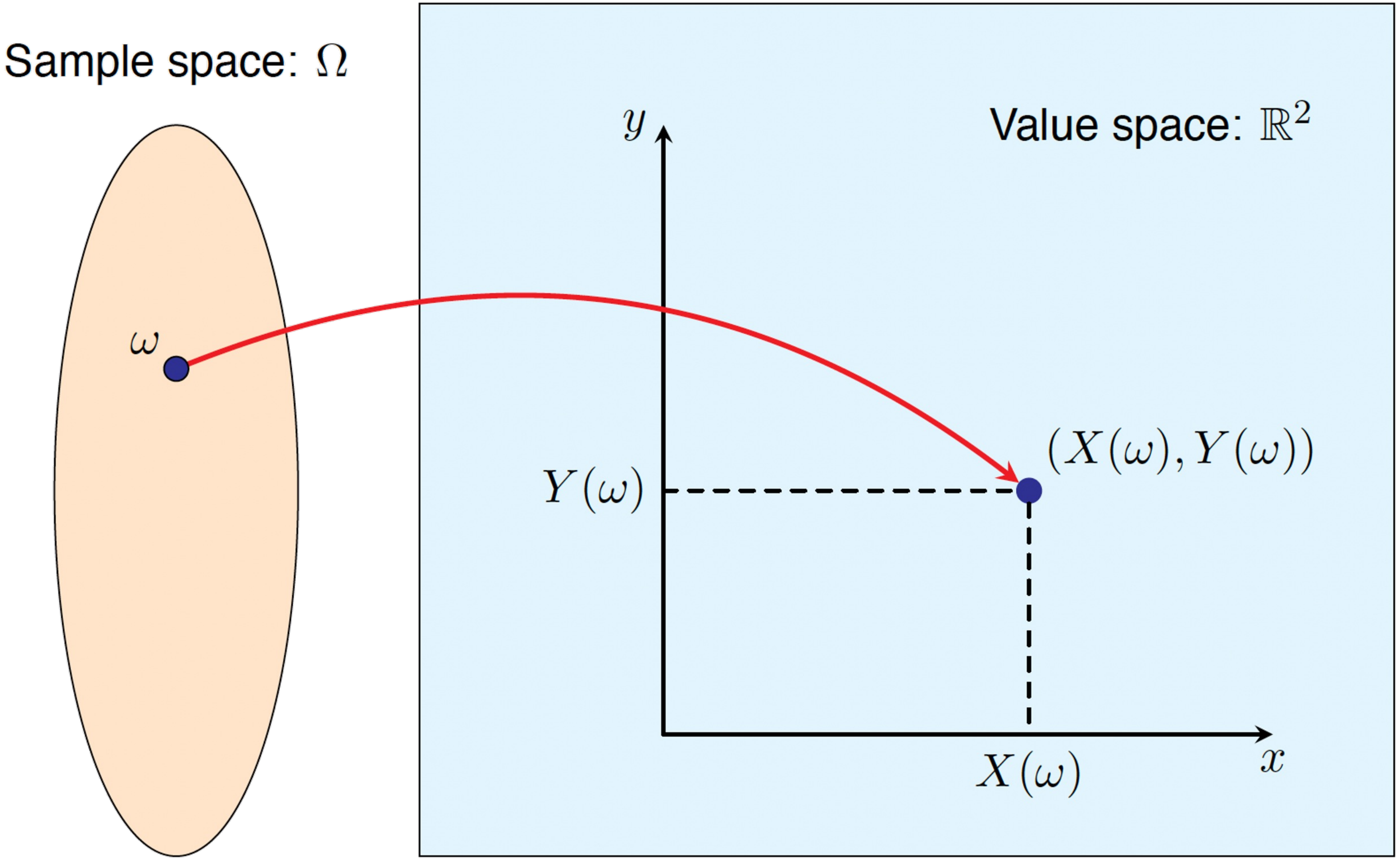
3.1 Random Vector and Joint Distribution

- We have been considering a single random variable each time, however, often we need to deal with several random variables. E.g.,
  - The height and weight of a randomly selected person.
  - The temperature, humidity, wind, precipitation of Singapore on a randomly selected day.
- Since different metrics of the same object are typically associated with each other, they should be considered simultaneously.
- Here we will talk about how to study two random variables X and Y simultaneously, all the concepts can be extended to n random variables $X_1, X_2, \dots, X_n$.

Random Vector

We say that (X, Y) is a random vector (随机向量) if $\omega \in \Omega \longmapsto (X(\omega), Y(\omega))$ is a function valued on $\mathbb{R}^2$. (X, Y) can also be called a two-dimensional random variable (二维随机变量).

3.1 Random Vector and Joint Distribution



3.1 Random Vector and Joint Distribution

- The major concepts that will be introduced for random vectors are:
  - (Joint) CDF/PMF/PDF;
  - Marginal CDF/PMF/PDF;
  - Conditional PMF/PDF.
- First, for a discrete/continuous random vector, we can describe its distribution using the (joint) CDF (cumulative distribution function).

Joint Cumulative Distribution Function

For a random vector (X, Y), either discrete or continuous, its cumulative distribution function (CDF, 累积分布函数) is defined as

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

$F(x, y)$ is also called the joint CDF (联合累积分布函数) of X and Y.

- You can treat $\{X \leq x\}$ as an event A, and $\{Y \leq y\}$ as an event B, then

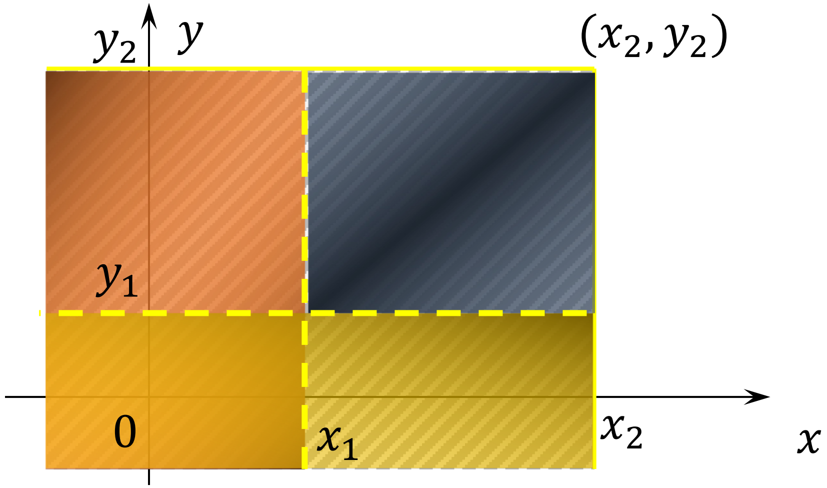
$$F(x, y) = P(A \cap B).$$

- It follows that for $\forall x, y \in \mathbb{R}$, we have $0 \leq F(x, y) \leq 1$ and

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

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

3.1 Random Vector and Joint Distribution



- The joint CDF can be used to compute $P(x_1 < X \leq x_2, y_1 < Y \leq y_2)$ for any $-\infty < x_1 < x_2 < \infty, -\infty < y_1 < y_2 < \infty$.

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

- With the joint CDF, it is straight forward to define the marginal CDF.

Marginal CDF

Let $F(x, y)$ be the joint CDF of (X, Y), $F_X(x)$ be the CDF of X without considering Y, and $F_Y(y)$ be the CDF of Y without considering X, then for $\forall x, y \in \mathbb{R}$:

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

F_X (F_Y) is also called the marginal CDF (边缘累积分布函数) of X (Y).

Note: While the joint CDF uniquely determines the marginal CDFs, the reverse is not true.

3.1 Random Vector and Joint Distribution

Example 3.1

- Suppose that the joint CDF of random vector (X, Y) is (a, b, c are constants)

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

- 1. Determine the value of a, b, c. 2. Calculate $P(-2 < X \leq 2, -2 < Y \leq 2)$.
- 3. Obtain the marginal CDFs of X and Y.

Solution

3.1 Random Vector and Joint Distribution

Solution

3.1 Random Vector and Joint Distribution

- Then for discrete random vectors, we introduce the joint and marginal PMF.

Joint and Marginal PMF for Discrete Random Vector

- For a random vector (X, Y), let $S_X = \{x_1, x_2, \ldots\}$ and $S_Y = \{y_1, y_2, \ldots\}$ be the support of X and Y, respectively. Then the joint PMF (联合概率质量函数) of (X, Y) is defined as

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

- The marginal PMF (边缘概率质量函数) of X is the PMF of X without considering Y:

$$p_X(x_i) = P(X = x_i) = \sum_{j=1}^{\infty} p_{ij} \triangleq p_{i\cdot}, i = 1, 2, \cdots.$$

- Similarly, the marginal PMF of Y is

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

- The joint PMF satisfies:
- Non-negativity: $p_{ij} \geq 0, i, j = 1, 2, \ldots;$
- Normalization: $\sum_i \sum_j p_{ij} = 1.$

- The joint PMF is typically displayed in a tabular format:

$X\backslash Y$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$

3.1 Random Vector and Joint Distribution

Example 3.2

- Two dice are tossed independently. Let X be the smaller number of points and Y be the larger number of points. If both dice show the same number, say, z points, then $X = Y = z$.
- 1. Find the joint PMF of (X, Y); 2. Find the marginal PMF of X.

Solution

3.1 Random Vector and Joint Distribution

- Then for continuous random vectors, we introduce the joint PDF.

Joint PDF for Continuous Random Vector

- (X,Y) is said to be a continuous random vector if there exists a non-negative function $f(x,y)$, defined for all $(x,y)\in\mathbb{R}^2$, satisfies that for any $D\subset\mathbb{R}^2$,

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

$f(x,y)$ is called the joint PDF (联合概率密度函数) of (X,Y)

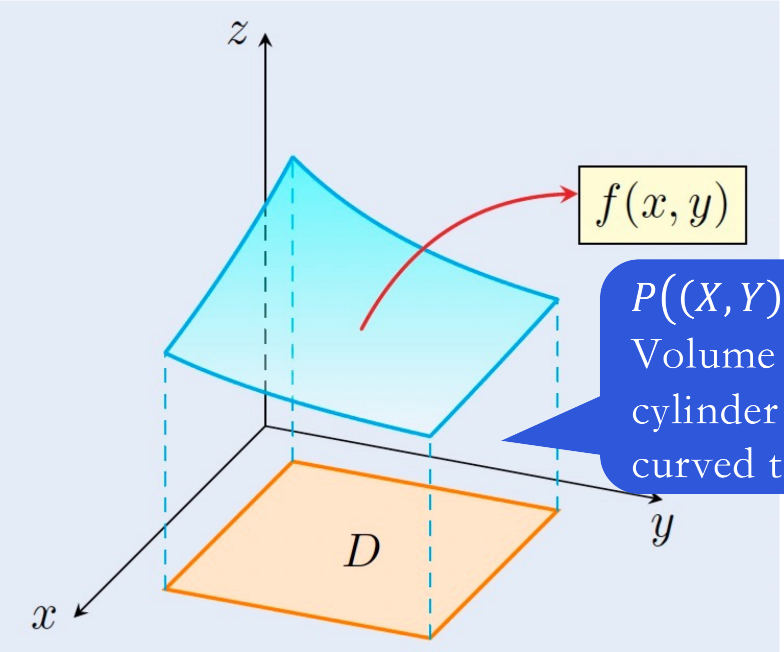
- Particularly, we have the joint CDF of (X,Y) to be

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

- It follows that

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

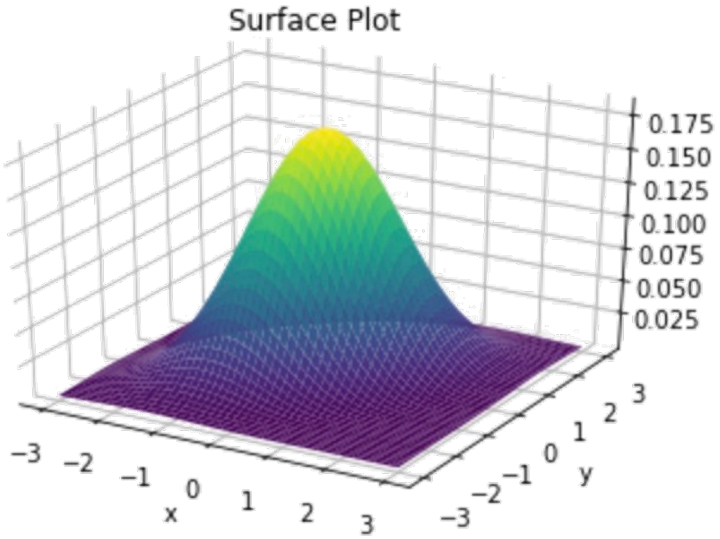
- The joint PDF satisfies:
- Non-negativity: $f(x,y)\ge 0, \forall (x,y)\in\mathbb{R}$;
- Normalization: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy=1$.

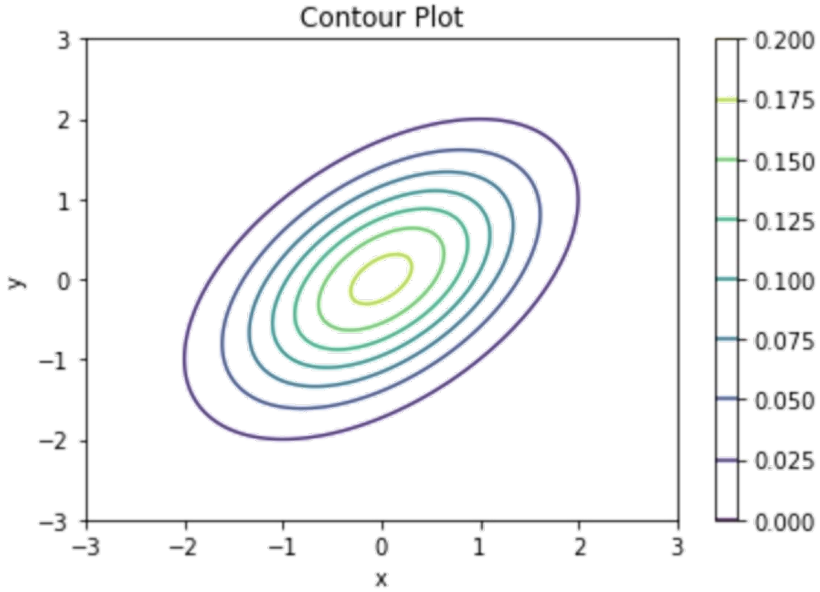


$P\big((X,Y)\in D\big)=$ Volume of the cylinder with a curved top

3.1 Random Vector and Joint Distribution

- Similar to the PDF of a single r.v., the joint PDF $f(x,y) \neq P(X=x, Y=y)$. Instead, it reflects the degree to which the probability is concentrated around (x,y).
- The joint PDF are typically visualized with the surface plot (曲面图) or the contour plot (等高线图) which help in intuitively understanding the distribution and relationship between X and Y.
  - Surface plot: a 3D plot where the height represents the value of the joint PDF at each point (x,y).
  - Contour plot: a 2D plot showing level curves where the joint PDF has constant values.



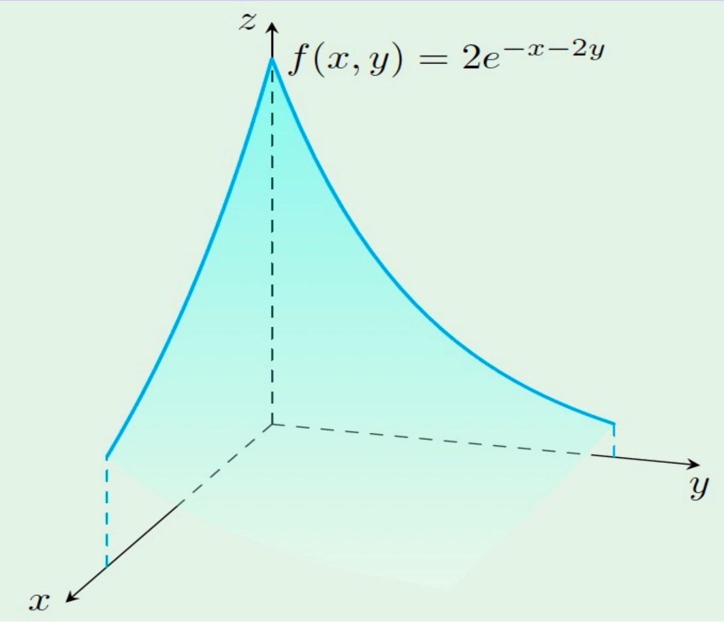


Example 3.3

- The lifetime (in years) of two electronic components of a randomly selected machine is denoted by r.v.s X and Y, which has a joint PDF

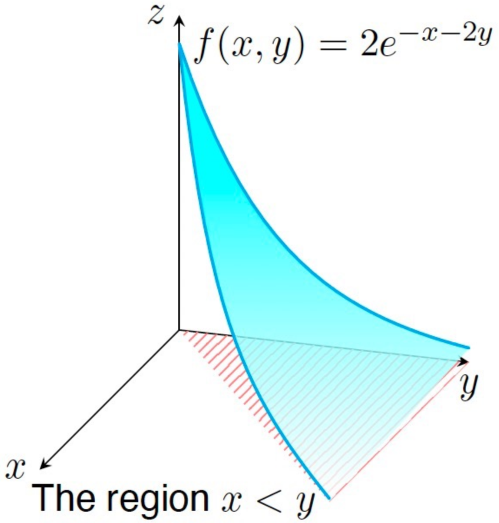
$$f(x,y)=\begin{cases}2e^{-x-2y}, & 0<x,y<\infty\\ 0, & \text{otherwise}\end{cases}.$$

- Compute: 1. $P(X<1,Y<1)$; 2. $P(X<Y)$.



3.1 Random Vector and Joint Distribution

Solution



3.1 Random Vector and Joint Distribution

- With the joint PDF, the marginal PDF can be determined.
- The marginal PDF mirrors the definition of the marginal PMF for the discrete case, except with sums replaced by integrals and the joint PMF replaced by the joint PDF.

Marginal PDF for Continuous Random Vector

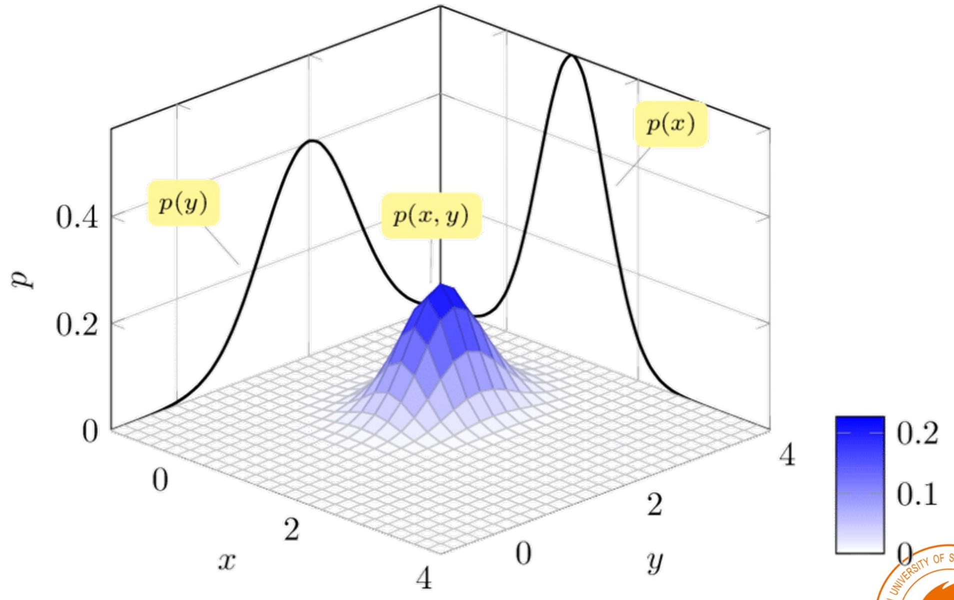
- (X, Y) a continuous random vector with joint PDF $f(x, y)$, then the **marginal PDF (边缘概率密度函数)** of X, i.e., the PDF of X without considering Y is

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

- Likewise, the marginal PDF of Y is

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

- Note that a joint PDF uniquely defines the marginal PDFs, however, the reverse is not true.



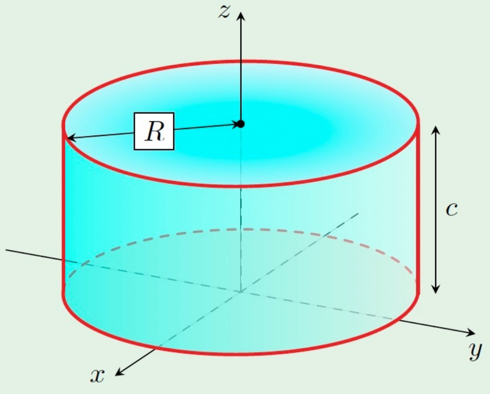
3.1 Random Vector and Joint Distribution

Example 3.4

- Suppose that the joint PDF of a random vector (X, Y) is given by

$$f(x,y) = \begin{cases} c, \text{if } x^2 + y^2 \leq R^2 \\ 0, \quad\quad \text{otherwise} \end{cases}$$

- 1. Determine the constant c;
- 2. Find the marginal PDF of X.



Solution

3.1 Random Vector and Joint Distribution

- Finally, we talk about the conditional PMF/PDF.

Conditional PMF/PDF

- For a **discrete** random vector (X, Y) with joint PMF $p(x, y)$, the conditional PMF (条件概率质量函数) of X given $Y = y$ is defined as

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

for all values of y such that $p_Y(y) > 0$.

- For a **continuous** random vector (X, Y) with joint PDF $f(x, y)$, the conditional PDF (条件概率密度函数) of X given $Y = y$ is defined as

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

for all values of y such that $f_Y(y) > 0$.

- The conditional PDF mirrors the definition of the conditional PMF for the discrete case, except with the joint/marginal PMF replaced by the joint/marginal PDF.
- The conditional PMF/PDFs also satisfy the non-negativity and normalization properties.
- Question: for the continuous case, $P(Y = y) = 0$, so that $P(X = x|Y = y)$ or $P(X \leq x|Y = y)$ is not defined. Then how to understand the conditional PDF of X given $Y = y$?

3.1 Random Vector and Joint Distribution

- Here we talk about how to understand the conditional PDF of X given $Y = y$.
- Conditioning on $Y = y$ can be understood as conditioning on $\{y \leq Y \leq y + \varepsilon\}$ where $\varepsilon \to 0$.
- Consider the conditional CDF:

$$P\{X \leq x | y < Y \leq y + \varepsilon\} = \frac{P\{X \leq x, y < Y \leq y + \varepsilon\}}{P\{y < Y \leq y + \varepsilon\}} = \frac{\int_{-\infty}^{x} \int_{y}^{y+\varepsilon} f(u, v) dv du}{\int_{y}^{y+\varepsilon} f_Y(y) dy}$$

$$= \frac{\varepsilon \int_{-\infty}^{x} f(u, y_\varepsilon) du}{\varepsilon f_Y(\tilde{y}_\varepsilon)} \to \int_{-\infty}^{x} \frac{f(u, y)}{f_Y(y)} du \quad (\varepsilon \to 0).$$

By the mean value theorem of integrals (积分中值定理)

The conditional PDF

- By the definition of the conditional PDF, we have

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

- Take the integration w.r.t. y on both sides, the marginal distribution of X can be expressed as

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

The law of total probability under the continuous case

3.1 Random Vector and Joint Distribution

Example 3.5

- For a randomly selected person in an automobile accident, let X be his/her extent of injury and Y be the type of safety restraint he/she was wearing at the time of the accident. The joint PMF of X and Y is

$X\backslash Y$	1 (None)	2 (Belt Only)	3 (Belt and Harness)	$p_X(x)$
1 (None)	0.065	0.075	0.060	0.20
2 (Minor)	0.165	0.160	0.125	0.45
3 (Major)	0.145	0.10	0.055	0.30
4 (Death)	0.025	0.015	0.010	0.05
$p_Y(y)$	0.40	0.35	0.25	1.00

- 1. What is the PMF of extent of injury for a randomly selected person with no restraint?
- 2. What is the PMF of extent of injury for a randomly selected person with belt and harness?

Solution

3.1 Random Vector and Joint Distribution

Example 3.4 (Continued)

- Determine the conditional PDF of X given $Y = y$ (where $|y| \leq R$).

Solution

Chapter 3 Joint Distributions

- 3.1 Random Vector and Joint Distribution
- 3.2 Relationship between Two Random Variables
- 3.3 Function of Multiple Random Variables
- 3.4 Multivariate Normal Distribution

3.2 Relationship Between Two Random Variables

- In the end of Section 3.1, we mentioned the concept of independence between random variables.
- Recall the independence between random events, the independence between random variables can be defined similarly: the value of Y does not affect the distribution of X or vice versa.
- For example, for the continuous case:

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

or

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

By $f(x,y) = f_X(x)f_Y(y)$, we have

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

3.2 Relationship Between Two Random Variables

Independence of Random Variables

- Let $F(x_1, x_2, \dots x_n)$ be the joint CDF of $(X_1, X_2, \dots, X_n)$, $F_{X_i}(x_i)$ be the marginal CDF of X_i, then if for $\forall x_1, x_2, \dots, x_n \in \mathbb{R}$ we have
$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n),$$
then we say that random variables $X_1, X_2, \cdots, X_n$ are **(mutually) independent (相互独立)**.
- For discrete random variables $X_1, X_2, \cdots, X_n$, if they are independent, then the PMF satisfies
$$p(x_1, x_2, \cdots, x_n) = P(X_1 = x_1, \dots, X_n = x_n) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_n}(x_n).$$
- For continuous random variables $X_1, X_2, \cdots, X_n$, if they are independent, then the PDF satisfies
$$f(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

- **Question:** the definition of independence between three or more random events require multiple equations, why the independence between three or more random variables only require one?

3.2 Relationship Between Two Random Variables

Example 3.6

- The PDF of a standard normal random variable Z is $f(z) = ce^{-z^2/2}$, $-\infty < z < \infty$.
- We already know that $c = 1/\sqrt{2\pi}$, however, how is this value obtained?
- Surprisingly, the easiest way to determine c is to define two independent standard normal random variables and use the fact that their joint PDF must integrate to 1.

3.2 Relationship Between Two Random Variables

Solution

3.2 Relationship Between Two Random Variables

Example 3.7

- Suppose that the number of people who enter a shopping mall on a randomly selected weekday follows a Poisson distribution with parameter λ.
- If each person who enters the shopping mall is a male with probability 0.2 and a female with probability 0.8.
- Show that the number of males and females entering the shopping mall are independent Poisson random variables with parameters 0.2λ and 0.8λ, respectively.

Solution

3.2 Relationship Between Two Random Variables

Solution

3.2 Relationship Between Two Random Variables

- If X and Y are independent random variables, then, for any functions g and h, we have

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

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

Proof: Without loss of generality, show the case for the continuous case.

Suppose that X and Y have joint density $f(x,y)$, then

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

$$= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy = \mathrm{E}[g(X)]\mathrm{E}[h(Y)].$$

Let $\mathrm{E}[g(X)] = a$ and $\mathrm{E}[h(Y)] = b$, then

$$= \mathrm{E}[g(X) - a]\mathrm{E}[h(Y) - b] = 0$$

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

3.2 Relationship Between Two Random Variables

- Special case: $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ and $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X and Y are independent.
- What if X and Y are not independent?
- Think about the difference between $\mathrm{Var}(X+Y)$ and $\mathrm{Var}(X) + \mathrm{Var}(Y)$:

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

- So, if $\mathrm{E}\big[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\big] \neq 0$, X and Y cannot be independent.
- Therefore, $\mathrm{E}\big[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\big]$ can be used to measure the relationship between X and Y.

Covariance

- The **covariance (协方差)** between X and Y, denoted by $\mathrm{Cov}(X,Y)$, is defined by

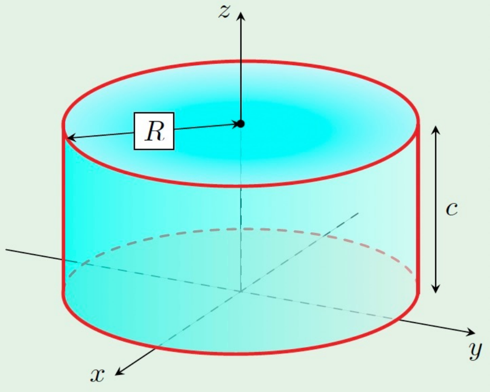
$$\mathrm{Cov}(X,Y) \triangleq \mathrm{E}\big[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\big] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

- If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$. However, if $\mathrm{Cov}(X,Y) = 0$, X and Y may not be independent, we can only say that X and Y are **uncorrelated (不相关的)**.

Example 3.4 (Continued)

- We already know that X and Y are not independent, however, show that $\mathrm{Cov}(X, Y) = 0$.

Proof



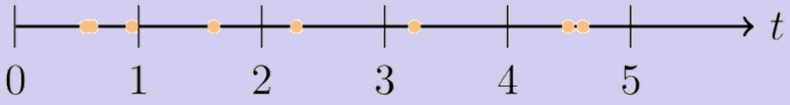
3.2 Relationship Between Two Random Variables

- Note that $\mathrm{Cov}(X,Y)$ is positive when X and Y tend to vary in the same direction and negative when they tend to vary in the opposite direction.
- The covariance has the following properties: (a, b, c are constants)
  - Covariance-variance relationship: $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.
  - Symmetry: $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
  - Constants cannot covary: $\mathrm{Cov}(X,c) = 0$.
  - Pulling out constants: $\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y)$.
  - Distributive property: $\mathrm{Cov}(X_1+X_2,Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$.
  - Bilinear property:

$$\mathrm{Cov}(a_1X_1 + a_2X_2 + \cdots a_nX_n, b_1Y_1 + b_2Y_2 + \cdots + b_mY_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\mathrm{Cov}(X_i,Y_j)\,.$$

Example 3.8

- A Geiger counter (盖革计数器) is a device used for detecting and measuring ionizing radiation (电离辐射).
- Each time it detects a radioactive particle, it makes a clicking sound. E.g., the orange points below indicates the times at which the Geiger counter detect a particle.



- Suppose that in a city, radioactive particles reach a Geiger counter according to a Poisson process at a rate of $\lambda = 0.8$ particles per second.
- The time that the first particle is detected and the time that the second particle is detected are denoted by X and Y, respectively.
- Calculate the covariance between X and Y.

3.2 Relationship Between Two Random Variables

Solution

3.2 Relationship Between Two Random Variables

- While the covariance measures the relationship between two random variables, its value depends on the unit/scale on which we measure the random variables.
  - E.g., let X (in m) and Y be the height and weight (in kg) of a randomly selected person, and $\tilde{X} = 100X$ (i..e., $\tilde{X}$ is the height measured in cm) then $\text{Cov}(\tilde{X}, Y) = 100\text{Cov}(X, Y)$.
  - Therefore, a larger covariance does not necessarily suggest a stronger relationship.
- To make the measure comparable, we need to remove the impact of unit/scale.

Correlation Coefficient

- The **correlation coefficient (相关系数)** between X and Y, denoted by $\text{Cor}(X, Y)$ or ρ_{XY}, is defined by

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

- ρ_{XY} is a normalized version of the covariance, which is a dimensionless quantity (无量纲数值).

- **Question**: what kind of relationship is ρ_{XY} measuring?

3.2 Relationship Between Two Random Variables

Example 3.9

- Calculate ρ_{XY} if the joint PDF of X and Y is $(0 < c \leq 1)$

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

Solution

3.2 Relationship Between Two Random Variables

Solution

3.2 Relationship Between Two Random Variables

- ρ_{XY} actually measure the direction and strength of **the linear relationship** between X and Y.

Proof: Consider to use a linear function of X to approximate Y, i.e., $\hat{Y} = a + bX$.

Then, the mean squared error (MSE, 均方误差) of the approximation is

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

Next, we would like to minimize the MSE w.r.t. a and b.

$$\begin{cases} \dfrac{\partial \mathrm{MSE}}{\partial a} = 2a + 2b\mathrm{E}(X) - 2\mathrm{E}(Y) = 0 \\ \dfrac{\partial \mathrm{MSE}}{\partial b} = 2b\mathrm{E}(X^2) - 2\mathrm{E}(XY) + 2a\mathrm{E}(X) = 0 \end{cases} \Rightarrow \begin{cases} b_0 = \dfrac{\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)}{\mathrm{E}(X^2) - [\mathrm{E}(X)]^2} = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \\ a_0 = \mathrm{E}(Y) - b_0\mathrm{E}(X) = \mathrm{E}(Y) - \mathrm{E}(X)\dfrac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} \end{cases}$$

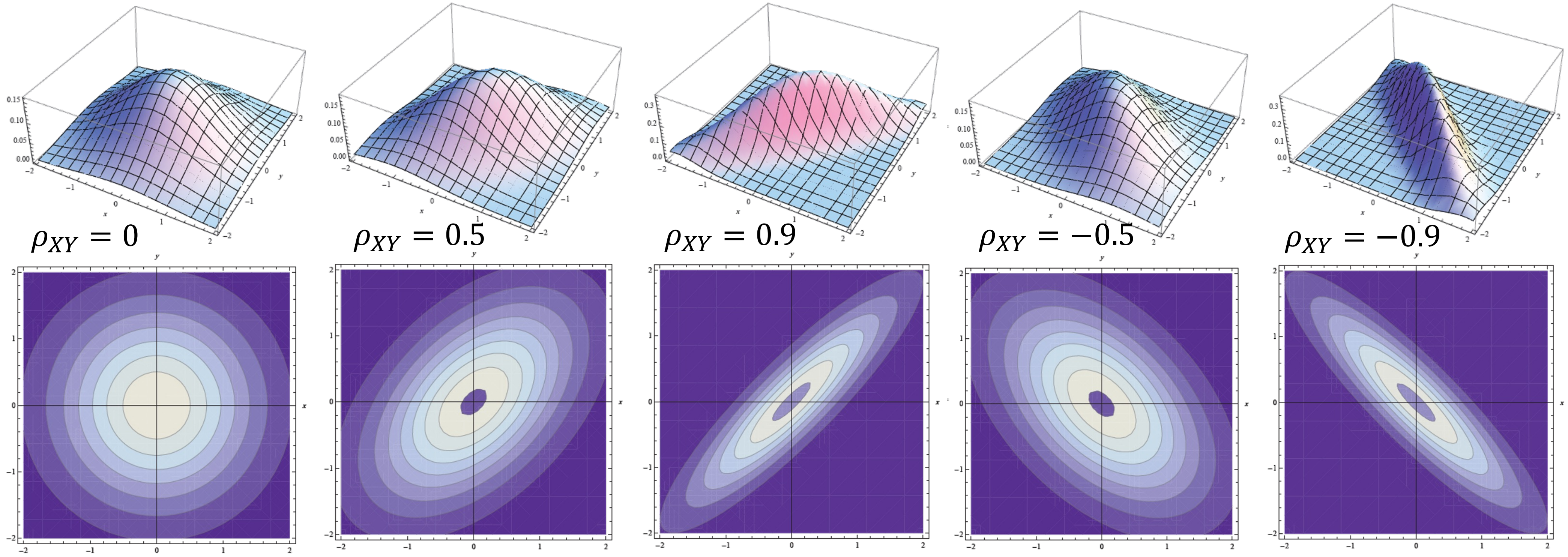
Therefore,

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

3.2 Relationship Between Two Random Variables

Not necessarily independent!

- Since $\min_{a,b} \mathrm{MSE} = \mathrm{Var}(Y)(1-\rho_{XY}^2) \geq 0$, so $\rho_{XY}^2 \leq 1 \Rightarrow -1 \leq \rho_{XY} \leq 1$.
  - $0 < \rho_{XY} \leq 1$: positively correlated; $-1 \leq \rho_{XY} < 0$: negatively correlated; $\rho_{XY} = 0$: uncorrelated.
  - When $|\rho_{XY}|$ is closer to 1, the mean squared error is smaller, i.e., the relationship between X and Y is closer to linear. Specifically, if $\rho_{XY} = \pm 1$, X and Y have **an almost perfect linear relationship**.



3.2 Relationship Between Two Random Variables

Example 3.10

- We would like to invest \$10,000 into shares of companies XX and YY.
- Shares of XX cost \$20 per share and the market analysis shows that the expected return is \$1 per share, with a standard deviation of \$0.5.
- Shares of YY cost \$50 per share, with an expected return of \$2.5 and a SD of \$1.
- What is the optimal portfolio (资产组合) consisting of shares of XX and YY, given their correlation coefficient ρ? (Note: number of shares can be any non-negative real value)

Solution

3.2 Relationship Between Two Random Variables

Solution