MA204: Mathematical Statistics

Tutorial 1

T1.1 Mutual Independency $\Rightarrow$ Pairwise Independency

1.1.1 Independency for events

Three events, $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ are mutually independent *if and only if* (iff)

(1) $\Pr(\mathbb{A}_1 \cap \mathbb{A}_2) = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2)$, i.e., $\mathbb{A}_1$ and $\mathbb{A}_2$ are independent;

(2) $\Pr(\mathbb{A}_2 \cap \mathbb{A}_3) = \Pr(\mathbb{A}_2) \times \Pr(\mathbb{A}_3)$, i.e., $\mathbb{A}_2$ and $\mathbb{A}_3$ are independent;

(3) $\Pr(\mathbb{A}_1 \cap \mathbb{A}_3) = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_3)$, i.e., $\mathbb{A}_1$ and $\mathbb{A}_3$ are independent;

(4) $\Pr(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2) \times \Pr(\mathbb{A}_3)$. $\blacksquare$

Example T1.1 (Pairwise independent but not mutually independent). Give an example such that $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ are pairwise independent but not mutually independent.

Solution: Suppose a box contains 4 tickets labeled as $\{112, 121, 211, 222\}$. Let's choose one ticket at random, and consider the following three events:

$$\begin{aligned} \mathbb{A}_1 &= \{1 \text{ occurring at the first place}\}, \\ \mathbb{A}_2 &= \{1 \text{ occurring at the second place}\}, \\ \mathbb{A}_3 &= \{1 \text{ occurring at the third place}\}. \end{aligned}$$

So we obtain

$$\Pr(\mathbb{A}_1) = \frac{1}{2}, \quad \Pr(\mathbb{A}_2) = \frac{1}{2}, \quad \Pr(\mathbb{A}_3) = \frac{1}{2}.$$

Since

$$\mathbb{A}_1 \cap \mathbb{A}_2 = \{112\}, \quad \Pr(\mathbb{A}_1 \cap \mathbb{A}_2) = \tfrac{1}{4} = \Pr(\mathbb{A}_1)\Pr(\mathbb{A}_2),$$
$$\mathbb{A}_2 \cap \mathbb{A}_3 = \{211\}, \quad \Pr(\mathbb{A}_2 \cap \mathbb{A}_3) = \tfrac{1}{4} = \Pr(\mathbb{A}_2)\Pr(\mathbb{A}_3),$$
$$\mathbb{A}_1 \cap \mathbb{A}_3 = \{121\}, \quad \Pr(\mathbb{A}_1 \cap \mathbb{A}_3) = \tfrac{1}{4} = \Pr(\mathbb{A}_1)\Pr(\mathbb{A}_3),$$

we have the conclusion that $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ are pairwise independent. On the other hand, note that $\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 = \emptyset$. then,

$$\Pr(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = 0 \neq \frac{1}{8} = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2) \times \Pr(\mathbb{A}_3).$$

So $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ are not mutually independent. $\|$

Example T1.2 (Not pairwise independent but satisfying condition (4)). Give an example satisfying

$$\Pr(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2) \times \Pr(\mathbb{A}_3)$$

but $\mathbb{A}_1$, $\mathbb{A}_2$ and $\mathbb{A}_3$ are not pairwise independent.

Solution: Toss two different standard dice. The sample space $\mathbb{S}$ of the outcomes consists of all the ordered pairs:

$$\mathbb{S} = \left\{ \begin{array}{llll} (1,1), & (1,2), & \cdots, & (1,6) \\ (2,1), & (2,2), & \cdots, & (2,6) \\ (3,1), & (3,2), & \cdots, & (3,6) \\ (4,1), & (4,2), & \cdots, & (4,6) \\ (5,1), & (5,2), & \cdots, & (5,6) \\ (6,1), & (6,2), & \cdots, & (6,6) \end{array} \right\}.$$

Each point in $\mathbb{S}$ has a probability of $1/36$. Consider the following three events:

$$\begin{aligned} \mathbb{A}_1 &= \{\text{first die shows 1 or 2 or 3}\}, \\ \mathbb{A}_2 &= \{\text{first die shows 3 or 4 or 6}\}, \\ \mathbb{A}_3 &= \{\text{sum of two faces is 9}\}. \end{aligned}$$

So we have

$$\Pr(\mathbb{A}_1) = \frac{1}{2}, \quad \Pr(\mathbb{A}_2) = \frac{1}{2}, \quad \Pr(\mathbb{A}_3) = \Pr\{(3,6),(4,5),(5,4),(6,3)\} = \frac{1}{9}.$$

Note that $\mathbb{A}_1 \cap \mathbb{A}_2 = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$, so

$$\Pr(\mathbb{A}_1 \cap \mathbb{A}_2) = \frac{6}{36} = \frac{1}{6} \neq \frac{1}{4} = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2).$$

That is, $\mathbb{A}_1$, $\mathbb{A}_2$, and $\mathbb{A}_3$ are not pairwise independent. However, $\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3 = \{(3,6)\}$, we obtain

$$\Pr(\mathbb{A}_1 \cap \mathbb{A}_2 \cap \mathbb{A}_3) = \frac{1}{36} = \Pr(\mathbb{A}_1) \times \Pr(\mathbb{A}_2) \times \Pr(\mathbb{A}_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{9}. \qquad \|$$

1.1.2 Independency for random variables

Example T1.3 (Mutual independence implying pairwise independence). Let three continuous random variables X_1, X_2, X_3 be independent, i.e.,

$$F_{123}(x_1, x_2, x_3) = F_1(x_1) \times F_2(x_2) \times F_3(x_3), \tag{T1.1}$$

where $F_{123}(x_1, x_2, x_3)$ is the joint cdf of $(X_1, X_2, X_3)^\top$ and $F_i(\cdot)$ is the cdf of X_i $(i = 1, 2, 3)$. Then X_1, X_2, X_3 are pairwise independent.

Proof: Note that $F_3(\infty) = 1$, we have

$$\begin{aligned}
F_{12}(x_1, x_2) &= \Pr(X_1 \leqslant x_1, X_2 \leqslant x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{12}(x, y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \left[\int_{-\infty}^{\infty} f_{123}(x, y, z)\,\mathrm{d}z\right] \mathrm{d}x\mathrm{d}y \\
&= \Pr(X_1 \leqslant x_1, X_2 \leqslant x_2, X_3 \leqslant \infty) \\
&= F_{123}(x_1, x_2, \infty) \overset{(\mathrm{T1.1})}{=} F_1(x_1) \times F_2(x_2) \times F_3(\infty) \\
&= F_1(x_1) \times F_2(x_2),
\end{aligned}$$

indicating that $X_1 \perp\!\!\!\perp X_2$. By symmetry, we can prove that $X_1 \perp\!\!\!\perp X_3$ and $X_2 \perp\!\!\!\perp X_3$. $\square$

Example T1.4 (Discrete r.v.'s: Pairwise independence not implying mutual independence). Let $X, Y \overset{\text{iid}}{\sim} \text{Bernoulli}(0.5)$ and $Z = X + Y - 2XY$. Show that X, Y, Z are pairwise independent but not mutually independent.

Proof: (i) We first show that $Z \sim \text{Bernoulli}(0.5)$. Note that X, Y take values 0 and 1, then $Z = Y \cdot I(X = 0) + (1 - Y) \cdot I(X = 1)$ only takes 0 and 1. Since

$$\begin{aligned} \Pr(Z = 1) &= \Pr(X + Y - 2XY = 1) = \Pr(X = 1, Y = 0) + \Pr(X = 0, Y = 1) \\ &= \Pr(X = 1) \times \Pr(Y = 0) + \Pr(X = 0) \times \Pr(Y = 1) \\ &= 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5, \end{aligned}$$

indicating that $Z \sim \text{Bernoulli}(0.5)$.

(ii) We second show that $X \perp\!\!\!\perp Z$. Since

$$\begin{aligned} \Pr(X = 0, Z = 0) &= \Pr(X = 0, X + Y - 2XY = 0) = \Pr(X = 0, Y = 0) \\ &= 0.5 \times 0.5 = \Pr(X = 0) \times \Pr(Z = 0), \\ \Pr(X = 0, Z = 1) &= \Pr(X = 0, X + Y - 2XY = 1) = \Pr(X = 0, Y = 1) \\ &= 0.5 \times 0.5 = \Pr(X = 0) \times \Pr(Z = 1), \\ \Pr(X = 1, Z = 0) &= \Pr(X = 1, X + Y - 2XY = 0) = \Pr(X = 1, Y = 1) \\ &= 0.5 \times 0.5 = \Pr(X = 1) \times \Pr(Z = 0), \\ \Pr(X = 1, Z = 1) &= \Pr(X = 1, X + Y - 2XY = 1) = \Pr(X = 1, Y = 0) \\ &= 0.5 \times 0.5 = \Pr(X = 1) \times \Pr(Z = 1), \end{aligned}$$

indicating that $X \perp\!\!\!\perp Z$. By symmetry, we can prove that $Y \perp\!\!\!\perp Z$. Therefore, X, Y, Z are pairwise independent.

(iii) **Three discrete r.v.'s X, Y, Z are said to be mutually independent** if for all choices of x_i, y_j, z_k, we have

$$\Pr(X = x_i, Y = y_j, Z = z_k) = \Pr(X = x_i) \times \Pr(Y = y_j) \times \Pr(Z = z_k).$$

In this example, it is easy to verify that

$$\Pr(X=1,Y=1,Z=1)=0\neq\frac{1}{8}=\Pr(X=1)\times\Pr(Y=1)\times\Pr(Z=1),$$

implying that X,Y,Z cannot be mutually independent r.v.'s. $\square$

Example T1.5 (Continuous r.v.'s: Pairwise independence not implying mutual independence). For three continuous r.v.'s X_1,X_2,X_3, let $X_1 \perp\!\!\!\perp X_2$, $X_1 \perp\!\!\!\perp X_3$ and $X_2 \perp\!\!\!\perp X_3$. Given an example such that

$$f_{123}(x_1,x_2,x_3)\neq f_1(x_1)\times f_2(x_2)\times f_3(x_3),$$

where $f_{123}(x_1,x_2,x_3)$ is the joint pdf of $(X_1,X_2,X_3)^\top$ and $f_i(\cdot)$ is the pdf of X_i $(i=1,2,3)$.

Solution: Let the joint pdf of the random vector $\mathbf{x}=(X_1,X_2,X_3)^\top$ be defined by

$$f_{123}(x_1,x_2,x_3)=2\phi(x_1)\phi(x_2)\phi(x_3)\cdot I(\boldsymbol{x}\in\mathbb{D}),$$

where $\phi(x)=(1/\sqrt{2\pi})\mathrm{e}^{-x^2/2}$ denotes the pdf of $N(0,1)$, $\boldsymbol{x}=(x_1,x_2,x_3)^\top$, $\mathbb{D}\mathrel{\hat{=}}\cup_{i=1}^4\mathbb{D}_i$,

$$\begin{aligned}
\mathbb{D}_1 &= \{\boldsymbol{x}\colon\ x_1\geqslant 0,\ x_2\geqslant 0,\ x_3\geqslant 0\},\\
\mathbb{D}_2 &= \{\boldsymbol{x}\colon\ x_1< 0,\ x_2< 0,\ x_3\geqslant 0\},\\
\mathbb{D}_3 &= \{\boldsymbol{x}\colon\ x_1< 0,\ x_2\geqslant 0,\ x_3< 0\},\quad \text{and}\\
\mathbb{D}_4 &= \{\boldsymbol{x}\colon\ x_1\geqslant 0,\ x_2< 0,\ x_3< 0\}.
\end{aligned}$$

In addition, we define

$$\begin{aligned}
\mathbb{P}_1 &= \{(x_1,x_3)^\top\colon\ x_1\geqslant 0,\ x_3\geqslant 0\},\\
\mathbb{P}_2 &= \{(x_1,x_3)^\top\colon\ x_1< 0,\ x_3\geqslant 0\},\\
\mathbb{P}_3 &= \{(x_1,x_3)^\top\colon\ x_1< 0,\ x_3< 0\},\quad \text{and}\\
\mathbb{P}_4 &= \{(x_1,x_3)^\top\colon\ x_1\geqslant 0,\ x_3< 0\}.
\end{aligned}$$

(i) We first prove that $f_{13}(x_1, x_3) = \phi(x_1)\phi(x_3)$ for all $x_1, x_3 \in \mathbb{R} = (-\infty, \infty)$.
Case 1: If $(x_1, x_3)^\top \in \mathbb{P}_1 \cup \mathbb{P}_3$, we have

$$f_{123}(x_1, x_2, x_3) = 2\phi(x_1)\phi(x_2)\phi(x_3) \cdot I(x_2 \geqslant 0)$$

so that

$$f_{13}(x_1, x_3) = \int_{-\infty}^{\infty} f_{123}(x_1, x_2, x_3)\,\mathrm{d}x_2 = \phi(x_1)\phi(x_3) \int_0^{\infty} 2\phi(x_2)\,\mathrm{d}x_2 = \phi(x_1)\phi(x_3).$$

Case 2: If $(x_1, x_3)^\top \in \mathbb{P}_2 \cup \mathbb{P}_4$, we have

$$f_{123}(x_1, x_2, x_3) = 2\phi(x_1)\phi(x_2)\phi(x_3) \cdot I(x_2 < 0)$$

so that

$$f_{13}(x_1, x_3) = \int_{-\infty}^{\infty} f_{123}(x_1, x_2, x_3)\,\mathrm{d}x_2 = \phi(x_1)\phi(x_3) \int_{-\infty}^{0} 2\phi(x_2)\,\mathrm{d}x_2 = \phi(x_1)\phi(x_3).$$

By combining Case 1 and Case 2, we have

$$f_{13}(x_1, x_3) = \phi(x_1)\phi(x_3)$$

for all $x_1, x_3 \in \mathbb{R}$, indicating that $X_1, X_3 \stackrel{\text{iid}}{\sim} N(0, 1)$.

(ii) Similarly, we can prove that $X_1, X_2 \stackrel{\text{iid}}{\sim} N(0, 1)$, and $X_2, X_3 \stackrel{\text{iid}}{\sim} N(0, 1)$. However, X_1, X_2, X_3 are not mutually independent standard normal r.v.'s, since $f_{123}(x_1, x_2, x_3) \neq \phi(x_1)\phi(x_2)\phi(x_3)$ for any $x_1, x_2, x_3 \in \mathbb{R}$. ‖

T1.2 Expectation, Variance, Quantile, Median, Chebyshev's and Jensen's Inequalities

1.2.1 Expectation, variance, quantile and median

- Let X be a discrete (or continuous) r.v. with pmf (or pdf) $f(x)$, and $g(x)$ be an arbitrary function. Then $g(X)$ is also a r.v. and the expectation of $g(X)$ is defined by

$$E[g(X)] = \begin{cases} \displaystyle\sum_x g(x)f(x), & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x, & \text{if } X \text{ is continuous,} \end{cases}$$

provided that $E[|g(X)|] < \infty$.

- The *expectation* and *variance* of X are defined by

$$\mu \mathrel{\hat{=}} E(X) \quad \text{and} \quad \sigma^2 \mathrel{\hat{=}} \mathrm{Var}(X) = E(X-\mu)^2 = E(X^2) - \mu^2.$$

- If X is continuous, then the q-th quantile of X, denoted by ξ_q, is defined as the smallest real number ξ satisfying $F(\xi_q) = \Pr(X \leqslant \xi_q) = q$ or

$$\xi_q = F^{-1}(q), \qquad q \in (0, 1). \tag{T1.2}$$

Especially, $\xi_{0.5} \mathrel{\hat{=}} \mathrm{med}(X)$ is the median of X, satisfying

$$\int_{-\infty}^{\mathrm{med}(X)} f(x)\,\mathrm{d}x = 0.5 = \int_{\mathrm{med}(X)}^{\infty} f(x)\,\mathrm{d}x.$$

For example, if $X \sim N(\mu, \sigma^2)$, then $\mathrm{med}(X) = \mu$. In fact, for any density which is symmetric on the population mean μ, then $\mathrm{med}(X) = \mu$.

- If X is discrete, the median of X satisfies

$$\Pr\{X \leqslant \mathrm{med}(X)\} \geqslant 0.5 \quad \textbf{and} \quad \Pr\{X \geqslant \mathrm{med}(X)\} \geqslant 0.5. \tag{T1.3}$$

Example T1.6 (Exponential distribution). Let X follow the exponential distribution with pdf $f(x) = \beta \mathrm{e}^{-\beta x}$ for $x \geqslant 0$ and $\beta > 0$. Find ξ_q and $\mathrm{med}(X)$ of X.

Solution: The cdf of X is $F(x) = \int_0^x f(t)\,\mathrm{d}t = 1 - \mathrm{e}^{-\beta x}$ for $x \geqslant 0$. Let $F(\xi_q) = q$, then

$$\xi_q = F^{-1}(q) = -\frac{\log(1-q)}{\beta}, \quad 0 \leqslant q < 1.$$

Thus, $\mathrm{med}(X) = \xi_{0.5} = \log(2)/\beta$. ‖

Example T1.7 (Standard Laplace distribution). Let X follow the standard Laplace distribution (or the double exponential distribution) with pdf $f(x) = 0.5\mathrm{e}^{-|x|}$ for $x \in \mathbb{R} \mathrel{\hat{=}} (-\infty, \infty)$. Find ξ_q and $\mathrm{med}(X)$ of X.

Solution: The cdf of X is

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0.5\mathrm{e}^{x}, & \text{if } x < 0, \\ 1 - 0.5\mathrm{e}^{-x}, & \text{if } x \geqslant 0. \end{cases}$$

Let $F(\xi_q) = q$, then

$$\xi_q = F^{-1}(q) = \begin{cases} \log(2q), & \text{if } 0 < q < 0.5, \\ -\log\{2(1-q)\}, & \text{if } 0.5 \leqslant q < 1. \end{cases}$$

Thus, $\mathrm{med}(X) = \xi_{0.5} = 0$. ‖

Example T1.8 (Finite discrete distribution). Find the median $\mathrm{med}(X)$ of the discrete random variable X with pmf

X	2	3	4	5	6	7	8	9	10	11	12
$p_i = \Pr(X = i+1)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

for $i = 1, \ldots, 11$.

Solution: We have $\mathrm{med}(X) = 7$ because

$$\begin{aligned} \Pr(X \leqslant 7) &= \frac{1+2+3+4+5+6}{36} = \frac{21}{36} \approx 0.583 \geqslant 0.5 \quad \text{and} \\ \Pr(X \geqslant 7) &= \frac{6+5+4+3+2+1}{36} = \frac{21}{36} \approx 0.583 \geqslant 0.5. \end{aligned}$$

‖

1.2.2 Chebyshev's inequality

Let X be an r.v. and c be a positive constant, then $\Pr(|X - \mu| \geqslant c\sigma) \leqslant 1/c^2$.

Example T1.9 (Uniform distribution). Let the pdf of X be given by

$$f(x) = \begin{cases} \dfrac{1}{2\sqrt{3}}, & \text{if } -\sqrt{3} < x < \sqrt{3}, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Calculate $\Pr(|X| \geqslant \frac{3}{2})$.

(b) Check the answer by the Chebyshev inequality.

Solution: (a) According to definition, we calculate

$$\begin{aligned}
\Pr\left(|X| \geqslant \frac{3}{2}\right) &= \Pr\left(X \geqslant \frac{3}{2} \text{ or } X \leqslant -\frac{3}{2}\right) = 1 - \Pr\left(-\frac{3}{2} \leqslant X \leqslant \frac{3}{2}\right) \\
&= 1 - \int_{-3/2}^{3/2} f(x)\, \mathrm{d}x = 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}}\, \mathrm{d}x \\
&= 1 - \frac{1}{2\sqrt{3}}\left[\frac{3}{2} - \left(-\frac{3}{2}\right)\right] = 1 - \frac{\sqrt{3}}{2} \approx 0.134.
\end{aligned}$$

(b) The mean of variance of X are given by

$$\begin{aligned}
\mu &= E(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x}{2\sqrt{3}}\, \mathrm{d}x = 0 \quad \text{and} \\
\sigma^2 &= E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - 0 = \int_{-\sqrt{3}}^{\sqrt{3}} \frac{x^2}{2\sqrt{3}}\, \mathrm{d}x = 1.
\end{aligned}$$

We want to check if

$$\Pr(|X - \mu| \geqslant c\sigma) = \Pr(|X| \geqslant c) \leqslant \frac{1}{c^2}$$

for some positive constant c. In fact, from (a), we have $\Pr(|X| \geqslant 3/2) \approx 0.134$. For $c = 3/2$,

$$\Pr\left(|X| \geqslant \frac{3}{2}\right) \approx 0.134 \leqslant \frac{1}{(\frac{3}{2})^2} \approx 0.44,$$

so the Chebyshev inequality holds. ‖

1.2.3 Convex function

Example T1.10 (Equivalence of two definitions of a convex function). Show that the two definitions (see page 20 of the Textbook) of a convex function are equivalent.

Proof: Let $g(x)$ be a convex function defined on $\mathbb{S} \subseteq \mathbb{R}$. If $g''(x) \geqslant 0$, by applying the second–order Taylor expansion of $g(x)$ around $x_0 \in \mathbb{S}$, we have

$$\begin{aligned} g(x) &= g(x_0) + (x - x_0)g'(x_0) + \frac{1}{2}(x - x_0)^2 \underbrace{g''(x^*)}_{\geqslant 0} \\ &\geqslant g(x_0) + (x - x_0)g'(x_0) \ \hat{=} \ \ell(x), \end{aligned}$$

where x^* is a point between x and x_0, $\ell(x)$ is the tangent line going through the point $(x_0, g(x_0))$. □

1.2.4 Jensen's inequality

Example T1.11 (Equivalent statements for a convex function). If g is a twice differentiable function defined on a convex set $\mathbb{S}$, then the following statements are equivalent:

(1) g is convex.

(2) $g''(x) \geqslant 0$.

(3) $\forall\, x, x_0 \in \mathbb{S}$, we have $g(x) \geqslant g(x_0) + (x - x_0)g'(x_0)$, which is called the *supporting hyperplane inequality*.

(4) $\forall\, x_1, x_2 \in \mathbb{S}$ and $\alpha \in [0, 1]$, we have $g[\alpha x_1 + (1 - \alpha)x_2] \leqslant \alpha g(x_1) + (1 - \alpha)g(x_2)$.

(5) Let $\alpha_i \geqslant 0$ and $\sum_{i=1}^n \alpha_i = 1$, then $g(\sum_{i=1}^n \alpha_i x_i) \leqslant \sum_{i=1}^n \alpha_i g(x_i)$, which is called the *discrete version* of Jensen's inequality.

Proof: (1) $\Rightarrow$ (5). It is equivalent to deriving (5) from (1.23) in Textbook. Define a discrete random variable X as follows:

X	$x_1, \ldots, x_i, \ldots, x_n$
$\Pr(X = x_i)$	$\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n$

where the probabilities $\alpha_i > 0$ and $\sum_{i=1}^n \alpha_i = 1$. It is clear that

$$E(X) = \sum_{i=1}^n x_i \Pr(X = x_i) = \sum_{i=1}^n \alpha_i x_i \tag{T1.4}$$

and

$$E\{g(X)\} = \sum_{i=1}^n \alpha_i g(x_i). \tag{T1.5}$$

Therefore, Eqn (1.23) implies

$$g\left(\sum_{i=1}^n \alpha_i x_i\right) \overset{\text{(T1.4)}}{=} g(E(X)) \leqslant E\{g(X)\} \overset{\text{(T1.5)}}{=} \sum_{i=1}^n \alpha_i g(x_i).$$

(5) $\Rightarrow$ (4). Simply taking $n = 2$ in (5), we obtain (4).

(4) $\Rightarrow$ (3). We rewrite $g(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha g(x_1) + (1-\alpha)g(x_2)$ as

$$\frac{g(x_2 + \alpha(x_1 - x_2)) - g(x_2)}{\alpha} \leqslant g(x_1) - g(x_2),$$

where $\alpha \in (0, 1)$. Without loss of generality, let $x_1 \neq x_2$ and let $\alpha \to 0$, we have

$$(x_1 - x_2) \lim_{\alpha \to 0} \frac{g(x_2 + \alpha(x_1 - x_2)) - g(x_2)}{\alpha(x_1 - x_2)} \leqslant g(x_1) - g(x_2),$$

or $(x_1 - x_2)g'(x_2) \leqslant g(x_1) - g(x_2)$, which implies (3).

(3) $\Rightarrow$ (2). From (3), we have

$$g(x) \geqslant g(x+\varepsilon) - \varepsilon g'(x+\varepsilon) \quad \text{and} \quad g(x+\varepsilon) \geqslant g(x) + \varepsilon g'(x),$$

so that

$$\varepsilon g'(x+\varepsilon) \geqslant g(x+\varepsilon) - g(x) \quad \text{and} \quad -\varepsilon g'(x) \geqslant g(x) - g(x+\varepsilon).$$

By combining the two inequalities, we obtain

$$\varepsilon g'(x+\varepsilon) - \varepsilon g'(x) \geqslant 0. \tag{T1.6}$$

Therefore,

$$g''(x) = \lim_{\varepsilon \to 0} \frac{g'(x+\varepsilon) - g'(x)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{\varepsilon g'(x+\varepsilon) - \varepsilon g'(x)}{\varepsilon^2} \overset{\text{(T1.6)}}{\geqslant} 0,$$

which implies (2). $\square$

T1.3 Conditional Expectation and Conditional Variance

1.3.1 Conditional expectation

Let X and Y be r.v.'s and $f(x|y)$ be the conditional pmf (or pdf) of X given $Y = y$. Then the conditional expectation of $g(X)$ given $Y = y$ is:

$$E[g(X)|Y = y] = \begin{cases} \displaystyle\sum_x g(x)f(x|y), & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f(x|y)\,\mathrm{d}x, & \text{if } X \text{ is continuous.} \end{cases}$$

Note that $E[g(X)|Y]$ is a function of the r.v. Y and we can similarly define the *conditional expectation* and *conditional variance* as in the unconditional case. ■

Example T1.12 (Distribution in square). Suppose that the conditional pdf of (X, Y) given the r.v. Z is

$$f(x, y|z) = [z + (1 - z)(x + y)]I_{(0,1)}(x)I_{(0,1)}(y),$$

for $0 < z < 2$, and the density of Z is $f(z) = \frac{1}{2}I_{(0,2)}(z)$, where $I_{\mathbb{A}}(x)$ denotes the indicator function, i.e., $I_{\mathbb{A}}(x) = 1$ if $x \in \mathbb{A}$ and $I_{\mathbb{A}}(x) = 0$ if $x \notin \mathbb{A}$.

(a) Find the expectation $E(X + Y)$.

(b) Determine whether X and Y are independent or not.

(c) Determine whether X and Z are independent or not.

Solution: (a) Note that $E(X + Y) = E[E(X + Y|Z)]$. We first calculate

$$\begin{aligned} & E(X + Y|Z = z) \\ = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x + y)[z + (1 - z)(x + y)]I_{(0,1)}(x)I_{(0,1)}(y)\,\mathrm{d}x\,\mathrm{d}y \\ = & \int_0^1\int_0^1(x + y)[z + (1 - z)(x + y)]\,\mathrm{d}x\,\mathrm{d}y \end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \int_0^1 xz + yz + (1-z)x^2 + 2(1-z)xy + (1-z)y^2 \,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1 \left[\frac{x^2 z}{2} + xyz + \frac{(1-z)x^3}{3} + (1-z)x^2 y + (1-z)xy^2 \right]\Bigg|_0^1 \mathrm{d}y \\
&= \int_0^1 \frac{2+z}{6} + y + (1-z)y^2 \,\mathrm{d}y \\
&= \left[\frac{(2+z)y}{6} + \frac{y^2}{2} + \frac{(1-z)y^3}{3} \right]\Bigg|_0^1 = \frac{7-z}{6},
\end{aligned}$$

so that $E(X+Y|Z) = (7-Z)/6$. Since $E(Z)=1$, we have

$$E(X+Y) = E[E(X+Y|Z)] = \frac{7-E(Z)}{6} = 1.$$

(b) Since

$$f(x,y,z) = f(x,y|z)f(z) = \frac{1}{2}[z + (1-z)(x+y)] I_{(0,1)}(x) I_{(0,1)}(y) I_{(0,2)}(z),$$

we have

$$\begin{aligned}
f(x,y) &= \int_{-\infty}^{\infty} f(x,y,z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{1}{2}[z + (1-z)(x+y)] I_{(0,1)}(x) I_{(0,1)}(y) I_{(0,2)}(z)\,\mathrm{d}z \\
&= \int_0^2 \frac{1}{2}[z + (1-z)(x+y)] I_{(0,1)}(x) I_{(0,1)}(y)\,\mathrm{d}z \\
&= \frac{1}{2} I_{(0,1)}(x) I_{(0,1)}(y) \int_0^2 x + y + (1-x-z)z\,\mathrm{d}z \\
&= \frac{1}{2} I_{(0,1)}(x) I_{(0,1)}(y) \left[(x+y)z + \frac{(1-x-y)z^2}{2} \right]\Bigg|_0^2 = I_{(0,1)}(x) I_{(0,1)}(y).
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \int_{-\infty}^{\infty} I_{(0,1)}(x)I_{(0,1)}(y)\,\mathrm{d}y \\
&= \int_0^1 I_{(0,1)}(x)\,\mathrm{d}y = I_{(0,1)}(x), \\
f(y) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \int_{-\infty}^{\infty} I_{(0,1)}(x)I_{(0,1)}(y)\,\mathrm{d}x \\
&= \int_0^1 I_{(0,1)}(y)\,\mathrm{d}x = I_{(0,1)}(y).
\end{aligned}
$$

Therefore, we obtain $f(x,y) = f(x)f(y)$, i.e., X and Y are independent.

(c) Note that

$$
\begin{aligned}
f(x,z) &= \int_{-\infty}^{\infty} f(x,y,z)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \frac{1}{2}[z+(1-z)(x+y)]I_{(0,1)}(x)I_{(0,1)}(y)I_{(0,2)}(z)\,\mathrm{d}y \\
&= \int_0^1 \frac{1}{2}[z+(1-z)(x+y)]I_{(0,1)}(x)I_{(0,2)}(z)\,\mathrm{d}y \\
&= \frac{1}{2}I_{(0,1)}(x)I_{(0,2)}(z)\int_0^1 z+(1-z)x+(1-z)y\,\mathrm{d}y \\
&= \frac{1}{2}I_{(0,1)}(x)I_{(0,2)}(z)\left[yz+(1-z)xy+\frac{(1-z)y^2}{2}\right]\Big|_0^1 \\
&= \frac{1+2x+z-2xz}{4}I_{(0,1)}(x)I_{(0,2)}(z) \\
&\neq f(x)f(z) = \frac{1}{2}I_{(0,1)}(x)I_{(0,2)}(z),
\end{aligned}
$$

then, X and Z are not independent. $\|$

1.3.2 Calculation formulae of expectation and variance

It can be shown that

$$\begin{aligned} E(X) &= E[E(X|Y)] = \int_{-\infty}^{\infty} E(X|Y=y) f(y)\,\mathrm{d}y, \\ \mathrm{Var}(X) &= E[\mathrm{Var}(X|Y)] + \mathrm{Var}[E(X|Y)]. \end{aligned}$$

■

Example T1.13 (Mixture distribution). Let $X_1, X_2 \overset{\text{iid}}{\sim} N(0,1)$, $U \sim U(0,1)$, and U be independent of (X_1, X_2). Define $Z = UX_1 + (1-U)X_2$.

(a) Find the conditional distribution of Z given $U = u$.

(b) Find $E(Z)$ and $\mathrm{Var}(Z)$.

(c) Find the distribution of Z.

Solution: (a) $Z|(U=u) = uX_1 + (1-u)X_2 \sim N(0, u^2 + (1-u)^2)$. Hence,

$$Z|U \sim N(0, U^2 + (1-U)^2)$$

so that $E(Z|U) = 0$ and $\mathrm{Var}(Z|U) = U^2 + (1-U)^2$.

(b) Method I: We need to use the following conclusion that $X \perp\!\!\!\perp Y$ iff for any functions $f(\cdot)$ and $g(\cdot)$, $f(X) \perp\!\!\!\perp g(Y)$. Since $Z = UX_1 + (1-U)X_2$, we have

$$\begin{aligned} E(Z) &= E[UX_1 + (1-U)X_2] = E(UX_1) + E[(1-U)X_2] \\ &= E(U)E(X_1) + E(1-U)E(X_2) = 0, \\ E(Z^2) &= E[U^2X_1^2 + (1-U)^2X_2^2 + 2U(1-U)X_1X_2] \\ &= E(U^2)E(X_1^2) + E[(1-U)^2]E(X_2^2) + 2E[U(1-U)]E(X_1)E(X_2) \\ &= E(U^2) + E[(1-U)^2] + 0 = \int_0^1 u^2\,\mathrm{d}u + \int_0^1 (1-u)^2\,\mathrm{d}u \\ &= \frac{1}{3} + \frac{1}{3} = \frac{2}{3}. \end{aligned}$$

Method II: $E(Z) = E[E(Z|U)] = 0$ and

$$\mathrm{Var}(Z) = E[\mathrm{Var}(Z|U)] + \mathrm{Var}[E(Z|U)] = E(U^2) + E[(1-U)^2] + 0 = \frac{2}{3}.$$

(c) Method I: Let $N(x|\mu, \sigma^2)$ denote the pdf of $N(\mu, \sigma^2)$. The pdf of Z is given by

$$\begin{aligned} f_Z(z) &= \int_0^1 f_{(Z,U)}(z,u)\,\mathrm{d}u = \int_0^1 f_{(Z|U)}(z|u) \cdot f_U(u)\,\mathrm{d}u \\ &= \int_0^1 N(z|0,\ u^2 + (1-u)^2)\,\mathrm{d}u. \end{aligned}$$

Method II: Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ be the pdf of $N(0,1)$. The cdf of Z is given by

$$\begin{aligned} F_Z(z) &= \Pr(Z \leqslant z) \\ &= \int_0^1 \Pr(Z \leqslant z|U=u) \cdot f_U(u)\,\mathrm{d}u \\ &= \int_0^1 \Pr\left(\frac{Z}{\sqrt{u^2+(1-u)^2}} \leqslant \frac{z}{\sqrt{u^2+(1-u)^2}} \middle| U = u \right) \mathrm{d}u \\ &= \int_0^1 \left\{ \int_{-\infty}^{z/\sqrt{u^2+(1-u)^2}} \phi(x)\,dx \right\} \mathrm{d}u. \end{aligned}$$

$\|$

1.3.3 Calculation of probability via expectation

For convenience, the indicator function $I_{\mathbb{A}}(x)$ can be alternatively denoted by $I(x \in \mathbb{A})$. Let X be a r.v., then $I_{\mathbb{A}}(X) = I(X \in \mathbb{A})$ is a Bernoulli r.v., i.e.

$I_{\mathbb{A}}(X)$	0	1
Probability	$1 - \Pr(\mathbb{A})$	$\Pr(\mathbb{A})$

or $I_{\mathbb{A}}(X) = I(X \in \mathbb{A}) \sim$ **Bernoulli**$(\Pr(\mathbb{A}))$**.** According to that the expectation of the Bernoulli distribution is equal to its success probability, we have

$$E[I_{\mathbb{A}}(X)] = E[I(X \in \mathbb{A})] = \Pr(\mathbb{A}) = \Pr(X \in \mathbb{A}), \tag{T1.7}$$

which is the formula (1.32) on page 24 of the textbook. ■

Example T1.14 (The second definition of expectation for a positive r.v.). For a continuous positive r.v. X, its expectation can be defined by

$$E(X) = \int_0^{\infty} \Pr(X > x)\,\mathrm{d}x.$$

Solution: Method I: Noting the following identity

$$X = \int_0^X \mathrm{d}x = \int_0^{\infty} I(x < X)\,\mathrm{d}x,$$

and taking expectations on both sides, we obtain

$$\begin{aligned} E(X) &= E\left[\int_0^{\infty} I(x < X)\,\mathrm{d}x\right] \\ &= \int_0^{\infty} E[I(X > x)]\,\mathrm{d}x \\ &\overset{\text{(T1.7)}}{=} \int_0^{\infty} \Pr(X > x)\,\mathrm{d}x. \end{aligned}$$

Methid II: Note that $x = \int_0^x \mathrm{d}y$, we have

$$\begin{aligned} E(X) &= \int_0^{\infty} x f_X(x)\,\mathrm{d}x = \int_0^{\infty} \left(\int_0^x \mathrm{d}y\right) \cdot f_X(x)\,\mathrm{d}x \\ &= \int_0^{\infty} \left(\int_0^x f_X(x)\,\mathrm{d}y\right) \mathrm{d}x \\ &= \int_0^{\infty} \left(\int_y^{\infty} f_X(x)\,\mathrm{d}x\right) \mathrm{d}y \\ &= \int_0^{\infty} \Pr(X > y)\,\mathrm{d}y = \int_0^{\infty} \Pr(X > x)\,\mathrm{d}x. \end{aligned}$$

‖