Recall

Let $X, \widetilde{X}, Y$ and Z be random variables.

Conditional expectation

Given $y \in \mathbb{R}$, $E[X|Y = y]$ is the expectation of X with respect to the conditional probability $P\{X \in \cdot\, |Y = y\}$. As y varies, we obtain a function $f\colon y \mapsto E[X|Y = y]$. Then the *conditional expectation* $E[X|Y]$ is a random variable $f(Y)$. Hence $E[\,\cdot\,|Y]$ maps a random variable X to another random variable $E[X|Y]$.

Some basic properties of the map $E[\,\cdot\,|Y]$:

(1) (linear) $\forall\, \alpha, \beta \in \mathbb{R},\ E[\alpha X + \widetilde{X} + \beta|Y] = \alpha E[X|Y] + E[\widetilde{X}|Y] + \beta$.

(2) (monotone) If $X \le Z$, then $E[X|Y] \le E[Z|Y]$.

(3) In most cases, for a function $g\colon \mathbb{R} \to \mathbb{R}$, we have $E[g(Y)X|Y] = g(Y)E[X|Y]$.
Since for $y \in \mathbb{R}$, $E[g(y)X|Y = y] = g(y)E[X|Y = y]$.

(4) In particular, $E[E[X|Y]|Y] = E[X|Y]$ by (3).

(5) $E[X] = E[E[X|Y]]$. This allows us to compute expectations by conditioning.

(6) We take X in (5) to be the indicator variable χ_E for an event E. Note that $E[\chi_E] = P(E)$ and $E[\chi_E|Y = y] = P\{E|Y = y\}$. Then we can compute probabilities by conditioning,

$$P(E) = \begin{cases} \sum_y P\{E|Y = y\}P\{Y = y\} & \text{if } Y \text{ discrete} \\ \int_{-\infty}^{\infty} P\{E|Y = y\}f_Y(y)dy & \text{if } Y \text{ continuous.} \end{cases}$$

In particular, if $Y = \sum_{i=1}^n i\chi_{F_i}$ for some partition $F_1, \dots, F_n$ of the sample space, then the law of total probability is recovered.

Moment generating functions

For a random variable X, the *moment generating function* (MGF) is $M_X(t) \coloneqq E[e^{tX}]$ for $t \in \mathbb{R}$ whenever $E[e^{tX}]$ exists. Note $M_X(t) > 0$. The following facts make MGF useful:

- $E[X^n] = M_X^{(n)}(0)$ for $n \in \mathbb{N}$ (if $E[X^n] < \infty$).
- If there exists $t_0 > 0$ such that $M_X(t) = M_Y(t)$ for $t \in (-t_0, t_0)$, then $F_X = F_Y$.
- If X, Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

A table about MGFs of common distributions can be found in the textbook.

Examples

Example 1. Let X, Y be random variables and $g\colon \mathbb{R} \to \mathbb{R}$ be a function. Show that

(i) $\operatorname{Cov}(X, E[Y|X]) = \operatorname{Cov}(X, Y)$.

(ii) $E[(X - E[X|Y])^2] = E[X^2] - E[E[X|Y]^2]$.

(iii) $E[(X - g(Y))^2] \geq E[(X - E[X|Y])^2]$.

Proof. (i) It follows from (3) that $XE[Y|X] = E[XY|X]$. Then by (5),

$$\begin{aligned}\operatorname{Cov}(X, E[Y|X]) &= E[XE[Y|X]] - E[X]E[E[Y|X]] \\ &= E[E[XY|X]] - E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \\ &= \operatorname{Cov}(X, Y).\end{aligned}$$

(ii) By (5) and (3), we have

$$E[XE[X|Y]] = E[E[XE[X|Y]|Y]] = E[E[X|Y]E[X|Y]] = E[E[X|Y]^2].$$

Hence

$$\begin{aligned}E[(X - E[X|Y])^2] &= E[X^2] - 2E[XE[X|Y]] + E[E[X|Y]^2] \\ &= E[X^2] - 2E[E[X|Y]^2] + E[E[X|Y]^2] \\ &= E[X^2] - E[E[X|Y]^2].\end{aligned}$$

(iii) By (5), it suffices to prove $E[(X - g(Y))^2|Y] \geq E[(X - E[X|Y])^2|Y]$.

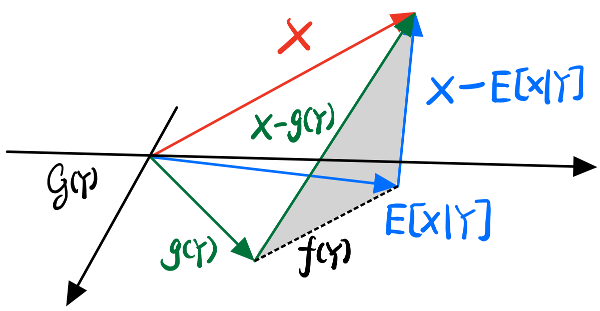

Figure 1: A possible intuition about (iii)

Based on the above intuition, we first establish that $X - E[X|Y]$ is 'orthogonal' to the 'plane'. For any function $f\colon \mathbb{R} \to \mathbb{R}$, by (3) and (4) we have

$$\begin{aligned}E[(X - E[X|Y])f(Y)|Y] &= f(Y)E[X - E[X|Y]|Y] \\ &= f(Y)(E[X|Y] - E[E[X|Y]|Y]) \\ &= f(Y)(E[X|Y] - E[X|Y]) \\ &= 0.\end{aligned}$$

Next we focus on the shaded 'right triangle'. By viewing $E[X|Y] - g(Y)$ as $f(Y)$,

$$\begin{aligned}
&E[(X - g(Y))^2|Y] \\
&= E[(X - E[X|Y] + E[X|Y] - g(Y))^2|Y] \\
&= E[(X - E[X|Y])^2|Y] + 2E[(X - E[X|Y])(E[X|Y] - g(Y))|Y] + E[(E[X|Y] - g(Y))^2|Y] \\
&= E[(X - E[X|Y])^2|Y] + 0 + E[(E[X|Y] - g(Y))^2|Y] \\
&\geq E[(X - E[X|Y])^2|Y],
\end{aligned}$$

where the last inequality follows from $E[(E[X|Y] - g(Y))^2|Y] \geq 0$.

□

Example 2. Let $X \sim U(-1/2, 1/2)$ and $I \sim Bern(1/2)$. Suppose that X, I are independent. Define

$$Y := \begin{cases} X & \text{if } I = 0 \\ -X & \text{if } I = 1. \end{cases}$$

Find $\text{Cov}(X, Y)$. Are X, Y independent?

Solution. Since X, I are independent, we have X^2, I are independent. Thus $E[X^2|I = i] = E[X^2]$ for $i = 0, 1$. Note $E[X] = 0$. Then

$$\begin{aligned}
\text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[E[XY|I]] - 0 && \text{by (5)} \\
&= E[XY|I = 0]P\{I = 0\} + E[XY|I = 1]P\{I = 1\} \\
&= \frac{1}{2}E[X^2|I = 0] - \frac{1}{2}E[X^2|I = 1] && \text{by def. of } Y \\
&= \frac{1}{2}(E[X^2] - E[X^2]) && \text{by independence of } X^2, I \\
&= 0.
\end{aligned}$$

Next we focus on the dependence of X, Y. Let $A, B \subset \mathbb{R}$. Then by conditioning on I,

$$\begin{aligned}
P\{Y \in A\} &= P\{Y \in A,\ I = 0\} + P\{Y \in A,\ I = 1\} \\
&= P\{X \in A,\ I = 0\} + P\{-X \in A,\ I = 1\} \\
&= \frac{1}{2}P\{X \in A\} + \frac{1}{2}P\{X \in -A\} \\
&= P\{X \in A\}
\end{aligned}$$

where the last equality follows from $P\{X \in A\} = P\{X \in -A\}$. Similarly,

$$\begin{aligned}
P\{X \in A,\ Y \in B\} &= P\{X \in A,\ Y \in B,\ I = 0\} + P\{X \in A,\ Y \in B,\ I = 1\} \\
&= P\{X \in A,\ X \in B,\ I = 0\} + P\{X \in A,\ -X \in B,\ I = 1\} \\
&= P\{X \in A \cap B,\ I = 0\} + P\{X \in A \cap (-B),\ I = 1\} \\
&= \frac{1}{2}P\{X \in A \cap B\} + \frac{1}{2}P\{X \in A \cap (-B)\}.
\end{aligned} \tag{1}$$

Let $A = B = [1/8, 1/4]$. Then $A \cap B = A$ and $A \cap (-B) = \emptyset$,

$$P\{X \in A,\ Y \in B\} = \frac{1}{2}P\{X \in A\} = \frac{1}{2} \times \frac{1}{8} \neq \frac{1}{8} \times \frac{1}{8} = P\{X \in A\}P\{Y \in B\}.$$

This shows that X and Y are not independent. □

Remark. Example 2 is another example showing that $\text{Cov}(X, Y) = 0 \not\Rightarrow$ independence. It is interesting to describe the join distribution of X, Y in a way more explicit than Equation (1).

— *THE END OF MAIN CONTENT* —

Limit theorems

This section is included only for the completeness but without examples.

Inequalities

Proposition 3 (Markov inequality). *Let X be a non-negative random variable. Then for $\varepsilon > 0$,*

$$P\{X \geq \varepsilon\} \leq \frac{E[X]}{\varepsilon}.$$

Proposition 4 (Chebyshev inequality). *Let X be a random variable with finite mean μ and variance σ^2. Then for $\varepsilon > 0$,*

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

Limit theorems

Theorem 5 (Weak Law of Large Numbers). *Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with finite mean μ. Then for $\varepsilon > 0$,*

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{as } n \to \infty.$$

Theorem 6 (Strong Law of Large Numbers). *Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with finite mean μ. Then*

$$P\left\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right\} = 1.$$

Theorem 7 (Central Limit Theorem). *Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with finite mean μ and variance σ^2. Then for $t \in \mathbb{R}$,*

$$P\left\{\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leq t\right\} \to \Phi(t) \quad \text{as } n \to \infty$$

where Φ denotes the CDF of the standard normal random variable.

Remark. There are some simulation experiments for limit theorems by clicking here.