Lecture 5: Random Variable, Part III

Yi, Yung (이융)

EE210: Probability and Introductory Random Processes
KAIST EE

June 12, 2021

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs.

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

(7) Random number of sum of random variables

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs.

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

(7) Random number of sum of random variables

Derived Distribution: $Y = g(X)$

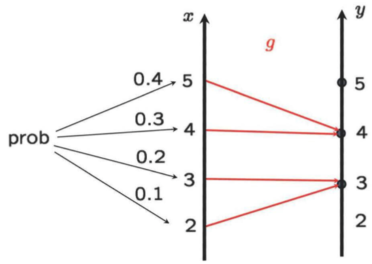
- Given the PDF of X, What is the PDF of $Y = g(X)$?
- Wait! Didn't we cover this topic? No. We covered just $\mathbb{E}[g(X)]$.
- Examples: $Y = X$, $Y = X + 1$, $Y = X^2$, etc.
- What are easy or difficult cases?
- Easy cases
  - Discrete
  - Linear: $Y = aX + b$

Discrete Case

- Take all values of x such that $g(x) = y$, i.e.,

$$p_Y(y) = \mathbb{P}(g(X) = y) = \sum_{x:g(x)=y} p_X(x)$$

$$p_Y(3) = p_X(2) + p_X(3) = 0.1 + 0.2 = 0.3$$
$$p_Y(4) = p_X(4) + p_X(5) = 0.3 + 0.4 = 0.7$$



Linear: $Y = aX + b$, $a \neq 0$, X: Continuous

If $a > 0$, $\quad F_Y(y) = \mathbb{P}(aX + b \le y) = \mathbb{P}(X \le \frac{y-b}{a}) = F_X(\frac{y-b}{a})$

$$\to f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

If $a < 0$, $\quad F_Y(y) = \mathbb{P}(aX + b \le y) = \mathbb{P}(X \ge \frac{y-b}{a}) = 1 - F_X(\frac{y-b}{a})$

$$\to f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Therefore,

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)}$$

Linear: $Y = aX + b$, when X is exponential

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if} \quad x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{\lambda}{|a|} e^{-\lambda (y-b)/a}, & \text{if} \quad (y-b)/a \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

- If $b = 0$ and $a > 0$, Y is exponential with parameter $\frac{\lambda}{a}$, but generally not.

Linear: $Y = aX + b$, when X is normal

- Remember? Linear transformation preserves normality. Time to prove.

$$\boxed{\text{If } X \sim \mathcal{N}(\mu, \sigma^2), \text{ then for } a \neq 0 \text{ and } b,\ Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2).}$$

- Proof.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|}\frac{1}{\sqrt{2\pi}} \exp\left\{-\left(\frac{y-b}{a} - \mu\right)^2 / 2\sigma^2\right\}$$
$$= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left\{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right\}$$

Generally, $Y = g(X)$, X: Continuous

Step 1. Find the CDF of Y:
$F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y)$

Step 2. Differentiate: $f_Y(y) = \frac{dF_Y}{dy}(y)$

Ex1. $Y = X^2$.

$$F_Y(y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), \quad y \ge 0$$

Ex2. $X \sim \mathcal{U}[0,1]$. $Y = \sqrt{X}$.

$$F_Y(y) = \mathbb{P}(\sqrt{X} \le y) = \mathbb{P}(X \le y^2) = y^2$$
$$f_Y(y) = 2y, \quad 0 \le y \le 1$$

Ex3. $X \sim \mathcal{U}[0,2]$. $Y = X^3$.

$$F_Y(y) = \mathbb{P}(X^3 \le y) = \mathbb{P}(X \le \sqrt[3]{y}) = \frac{1}{2} y^{1/3}$$
$$f_Y(y) = \frac{1}{6} y^{-2/3}, \quad 0 \le y \le 8$$

When $Y = g(X)$ is monotonic, a general formula can be drawn (see the textbook at pp 207)

Functions of multiple rvs: $Z = g(X, Y)$ (1)

Basically, follow two-step approach: (i) CDF and (ii) differentiate.

Ex1. $X, Y \sim \mathcal{U}[0,1]$, and $X \perp\!\!\!\perp Y$. $Z = \max(X, Y)$.

* $\mathbb{P}(X \le z) = \mathbb{P}(Y \le z) = z, \ z \in [0,1]$.

$$F_Z(z) = \mathbb{P}(\max(X,Y) \le z) = \mathbb{P}(X \le z, Y \le z) = \mathbb{P}(X \le z)\mathbb{P}(Y \le z) = z^2 \qquad \text{(from } X \perp\!\!\!\perp Y\text{)}$$

$$f_Z(z) = \begin{cases} 2z, & \text{if } 0 \le z \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Functions of multiple rvs: $Z = g(X, Y)$ (2)

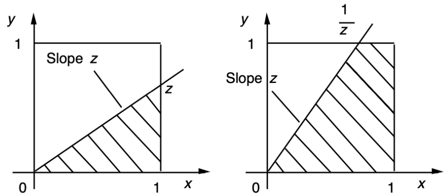
Basically, follow two step approach: (i) CDF and (ii) differentiate.

Ex2. $X, Y \sim \mathcal{U}[0,1]$, and $X \perp\!\!\!\perp Y$. $Z = Y/X$. VIDEO PAUSE

$$F_Z(z) = \mathbb{P}(Y/X \le z) = \begin{cases} z/2, & 0 \le z \le 1 \\ 1 - 1/2z, & z > 1 \\ 0, & \text{otherwise} \end{cases}$$

$$f_Z(z) = \begin{cases} 1/2, & 0 \le z \le 1 \\ 1/(2z^2), & z > 1 \\ 0, & \text{otherwise} \end{cases}$$

- Depending on the value of z, two cases need to be considered separately.



(Note) Sometimes, the problem is tricky, which requires careful case-by-case handing. :-)

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs.

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

(7) Random number of sum of random variables

Functions of multiple rvs: $Z = X + Y$, $X \perp\!\!\!\perp Y$ (1)

KAIST EE

- Sum of two independent rvs
- A very basic case with many applications
- Assume that $X, Y \in \mathbb{Z}$

$$p_Z(z) = \mathbb{P}(X+Y=z) = \sum_{\{(x,y):x+y=z\}} \mathbb{P}(X=x, Y=y) = \sum_x \mathbb{P}(X=x, Y=z-x)$$

$$= \sum_x \mathbb{P}(X=x)\mathbb{P}(Y=z-x) = \sum_x p_X(x)p_Y(z-x)$$

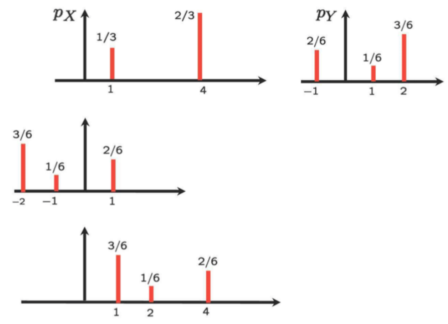
- $p_Z(z)$ is called $\boxed{\text{convolution}}$ of the PMFs of X and Y.

Functions of multiple rvs: $Z = X + Y$, $X \perp\!\!\!\perp Y$ (2)

KAIST EE

- Convolution: $p_Z(z) = \sum_x p_X(x)p_Y(z-x)$
- Interpretation for a given z:
  - (i) Flip (horizontally) the PMF of Y ($p_Y(-x)$)
  - (ii) Put it underneath the PMF of X
  - (iii) Right-shift the flipped PMF by z ($p_Y(-x+z)$)

Example. $z = 3$



$Y = X + Y$, $X \perp\!\!\!\perp Y$: Continuous

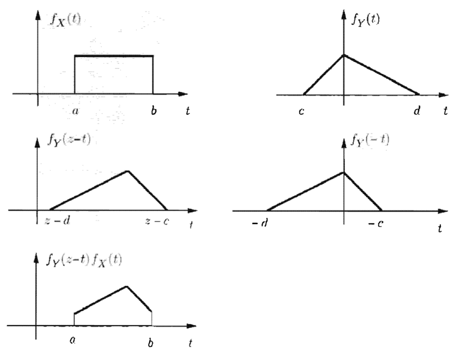
KAIST EE

- Same logic as the discrete case

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

- Youtube animation for convolution: `https://www.youtube.com/watch?v=C1N55M1VD2o`

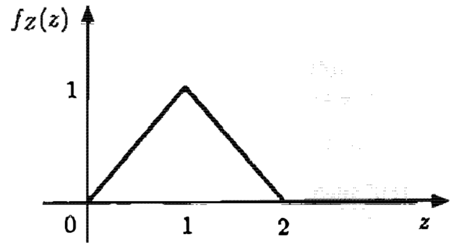
For a fixed z,



Example

KAIST EE

- Example. $X, Y \sim \mathcal{U}[0,1]$ and $X \perp\!\!\!\perp Y$. What is the PDF of $Z = X + Y$? Draw the PDF of Z.



Convolution in Image Processing

https://www.youtube.com/watch?v=MQm6ZP1F6ms

$Y = X + Y$, $X \perp\!\!\!\perp Y$, Normal (1)

- Very special, but useful case
  - X and Y are normal.

Sum of two independent normal rvs

$X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_x, \sigma_x^2)$ Then, $X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

- Why normal rvs are used to model the sum of random noises.
- Extension. The sum of finitely many independent normals is also normal.

$Y = X + Y$, $X \perp\!\!\!\perp Y$, Normal (2)

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\} \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{(z-x-\mu_y)^2}{2\sigma_y^2}\right\} dx
\end{aligned}
$$

- The details of integration is a little bit tedious. :-)

$$
f_Z(z) = \frac{1}{\sqrt{2\pi(\sigma_x^2 + \sigma_y^2)}} \exp\left\{-\frac{(z-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}\right\}
$$

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

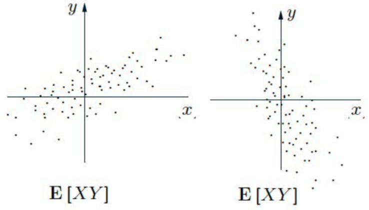
(7) Random number of sum of random variables

Making a Metric of Dependence Degree

- Goal: Given two rvs X and Y, assign some number that quantifies the degree of their dependence.
  - feeling/weather, university ranking/annual salary,
- Requirements
  - **R1.** Increases (resp. decreases) as they become more (resp. less) dependent. 0 when they are independent.
  - **R2.** Shows the 'direction' of dependence by + and -
  - **R3.** Always bounded by some numbers (i.e., dimensionless metric). For example, $[-1, 1]$
- Good engineers: Good at making good metrics
  - Metric of how our society is economically polarized
  - Cybermetrics in MLB (Major League Baseball): `http://m.mlb.com/glossary/advanced-stats`

OK. Let's Design!

- Simple case: $\mathbb{E}[X] = \mu_x = 0$ and $\mathbb{E}[Y] = \mu_Y = 0$
- Dependent: Positive (If $X \uparrow$, $Y \uparrow$) or Negative (If $X \uparrow$, $Y \downarrow$)
- What about $\mathbb{E}[XY]$? Seems good.
  - $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$ when $X \perp\!\!\!\perp Y$
  - More data points (thus increases) when $xy > 0$ (both positive or negative)
  - $|\mathbb{E}[XY]|$ also quantifies the amount of spread.



(Q) What about $\mathbb{E}[X + Y]$?
- When they are positively dependent, but have negative values?

What If $\mu_X \neq 0, \mu_Y \neq 0$?

- Solution: Centering. $X \to X - \mu_X$ and $Y \to Y - \mu_Y$

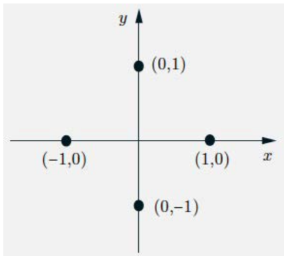
Covariance

$$\mathrm{cov}(X, Y) = \mathbb{E}\Big[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\Big]$$

- After some algebra, $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $X \perp\!\!\!\perp Y \implies \mathrm{cov}(X, Y) = 0$
- $\mathrm{cov}(X, Y) = 0 \implies X \perp\!\!\!\perp Y$? NO.
- When $\mathrm{cov}(X, Y) = 0$, we say that X and Y are $\boxed{\text{uncorrelated.}}$

Example: $\mathrm{cov}(X, Y) = 0$, but not independent

- $p_{X,Y}(1, 0) = p_{X,Y}(0, 1) = p_{X,Y}(-1, 0) = p_{X,Y}(0, -1) = 1/4$.
- $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, and $\mathbb{E}[XY] = 0$. So, $\mathrm{cov}(X, Y) = 0$
- Are they independent? No, because if $X = 1$, then we should have $Y = 0$.



Some Properties

$$\text{cov}(X, X) = \text{var}(X)$$

$$\text{cov}(aX + b, Y) = \mathbb{E}[(aX + b)Y] - \mathbb{E}[aX + b]\mathbb{E}[Y] = a \cdot \text{cov}(X, Y)$$

$$\text{cov}(X, Y + Z) = \mathbb{E}[X(Y + Z)] - \mathbb{E}[X]\mathbb{E}[Y + Z] = \text{cov}(X, Y) + \text{cov}(X, Z)$$

$$\text{var}[X + Y] = \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 = \text{var}[X] + \text{var}[Y] + 2\text{cov}(X, Y)$$

$$\text{var}\left[\sum X_i\right] = \sum \text{var}[X_i] + \sum_{i \neq j} \text{cov}(X_i, X_j)$$

Example: The hat problem in Lecture 3. Remember?

- n people throw their hats in a box and then pick one at random
- X: number of people with their own hat
- (Q) $\text{var}[X]$
- Key step 1. Define a rv $X_i = 1$ if i selects own hat and 0 otherwise. Then, $X = \sum_{i=1}^n X_i$.
- Key step 2. Are X_is are independent?
- $X_i \sim \text{Bern}(1/n)$. Thus, $\mathbb{E}[X_i] = 1/n$ and $\text{var}[X_i] = \frac{1}{n}(1 - \frac{1}{n})$

◦ For $i \neq j$,

$$\begin{aligned}\text{cov}(X_i, X_j) &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] \\ &= \mathbb{P}(X_i = 1 \text{ and } X_j = 1) - \frac{1}{n^2} \\ &= \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1 | X_i = 1) - \frac{1}{n^2} \\ &= \frac{1}{n}\frac{1}{n-1} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}\end{aligned}$$

$$\begin{aligned}\text{var}[X] &= \text{var}\left[\sum X_i\right] \\ &= \sum \text{var}[X_i] + \sum_{i \neq j} \text{cov}(X_i, X_j) \\ &= n\frac{1}{n}(1 - \frac{1}{n}) + n(n-1)\frac{1}{n^2(n-1)} = 1\end{aligned}$$

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

(7) Random number of sum of random variables

Bounding the metric: Correlation Coefficient

- Reqs. **R1** and **R2** are satisfied.
  R3. Always bounded by some numbers (dimensionless metric)
- How? $\boxed{\text{Normalization,}}$ but by what?

Correlation Coefficient

$$\rho(X, Y) = \mathbb{E}\left[\frac{(X - \mu_X)}{\boxed{\sigma_X}} \cdot \frac{(Y - \mu_Y)}{\boxed{\sigma_Y}}\right] = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}[X]\text{var}[Y]}}$$

- Theorem.
  1. $-1 \leq \rho \leq 1$ (proof at the next slide)
  2. $|\rho| = 1 \Leftrightarrow X - \mu_X = c(Y - \mu_Y)$ for some constant c ($c > 0$ when $\rho = 1$ and $c < 0$ when $\rho = -1$). In other words, linear relation, meaning VERY related.

1. $-1 \le \rho \le 1$

KAIST EE

- Cauchy-Schwarz inequaility. For any rvs X and Y, $(\mathbb{E}(XY))^2 \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$
- Proof of $-1 \le \rho \le 1$:
  Let $\tilde{X} = X - \mathbb{E}(X)$ and $\tilde{Y} = Y - \mathbb{E}(Y)$. Then, $\left(\rho(X,Y)\right)^2 = \dfrac{\left(\mathbb{E}[\tilde{X}\tilde{Y}]\right)^2}{\mathbb{E}(\tilde{X}^2)\mathbb{E}(\tilde{Y}^2)} \le 1$
- Proof of CSI: For any constant a,
  $0 \le \mathbb{E}\left[(X-aY)^2\right] = \mathbb{E}\left[X^2 - 2aXY + a^2Y^2\right] = \mathbb{E}(X^2) - 2a\mathbb{E}(XY) + a^2\mathbb{E}(Y^2)$
  Now, choose $a = \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}$. Then,
  $$\mathbb{E}(X^2) - 2\frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}\mathbb{E}(XY) + \frac{(\mathbb{E}[XY])^2}{(\mathbb{E}[Y^2])^2}\mathbb{E}(Y^2) = \mathbb{E}(X^2) - \frac{(\mathbb{E}[XY])^2}{\mathbb{E}(Y^2)} \ge 0$$

2. $|\rho| = 1 \Leftrightarrow X - \mu_X = c(Y - \mu_Y)$

KAIST EE

($\Rightarrow$) Suppose that $|\rho| = 1$. In the proof of CSI,
$$\mathbb{E}\left[\left(\tilde{X} - \frac{\mathbb{E}(\tilde{X}\tilde{Y})}{\mathbb{E}(\tilde{Y}^2)}\tilde{Y}\right)^2\right] = \mathbb{E}(\tilde{X}^2) - \frac{(\mathbb{E}[\tilde{X}\tilde{Y}])^2}{\mathbb{E}(\tilde{Y}^2)} = \mathbb{E}(\tilde{X}^2)(1-\rho^2) = 0$$
$$\tilde{X} - \frac{\mathbb{E}(\tilde{X}\tilde{Y})}{\mathbb{E}(\tilde{Y}^2)}Y = 0 \leftrightarrow \tilde{X} = \frac{\mathbb{E}(\tilde{X}\tilde{Y})}{\mathbb{E}(\tilde{Y}^2)}\tilde{Y} = \rho\sqrt{\frac{\mathbb{E}(\tilde{X}^2)}{\mathbb{E}(\tilde{Y}^2)}}\tilde{Y}$$
($\Leftarrow$) If $\tilde{Y} = c\tilde{X}$, then
$$\rho(X,Y) = \frac{\mathbb{E}(\tilde{X}c\tilde{X})}{\sqrt{\mathbb{E}[\tilde{X}^2]\mathbb{E}[(c\tilde{X})^2]}} = \frac{c}{|c|}$$

Roadmap

KAIST EE

(1) Derived distribution of $Y = g(X)$ or $Z = g(X,Y)$

(2) Derived distribution of $Z = X + Y$

(3) Covariance: Degree of dependence between two rvs

(4) Correlation coefficient

(5) Conditional expectation and law of iterative expectations

(6) Conditional variance and law of total variance

(7) Random number of sum of random variables

A Special Random Variable

KAIST EE

- Consider a rv Y, such that
$$Y = \begin{cases} 0, & \text{w.p. } 1/4 \\ 1, & \text{w.p. } 1/4 \\ 2, & \text{w.p. } 1/2 \end{cases}$$
- If $h(y) = y^2$, then a new rv $h(Y)$ is:
$$h(Y) = \begin{cases} 0, & \text{w.p. } 1/4 \\ 1, & \text{w.p. } 1/4 \\ 4, & \text{w.p. } 1/2 \end{cases}$$
- Consider other rv X, which, we assume, has:
$$g(y) = \mathbb{E}[X|Y=y] = \begin{cases} 3, & \text{if } y = 0 \\ 8, & \text{if } y = 1 \\ 9, & \text{if } y = 2 \end{cases}$$
- Then, a rv $g(Y)$ is:
$$g(Y) = \begin{cases} 3, & \text{w.p. } 1/4 \\ 8, & \text{w.p. } 1/4 \\ 9, & \text{w.p. } 1/2 \end{cases}$$

- The rv $g(Y)$ looks special, so let's give a fancy notation to it.

- What about? $X_{exp}(Y)$, $\mathbb{E}[X_Y]$, $\mathbb{E}_X[Y]$?

Conditional Expectation $\mathbb{E}[X|Y]$

KAIST EE

Conditional Expectation

A random variable $g(Y) = \boxed{\mathbb{E}[X|Y]}$, called $\boxed{\text{conditional expectation of } X \text{ given } Y}$, takes the value $g(y) = \mathbb{E}[X|Y=y]$, if Y happens to take the value y.

- A function of Y
- A random variable
- Thus, having a distribution, expectation, variance, all the things that a random variable has.
- Often confusing because of the notation.

Expectation of $\mathbb{E}[X|Y]$

KAIST EE

Expectation of Conditional Expectation

$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X]$, Law of iterated expectations

Proof.

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \sum_y \mathbb{E}[X|Y=y]p_Y(y) = \mathbb{E}[X]$$

Examples and Meaning

KAIST EE

- Stick of length l
- Uniformly break at point Y, and break what is left uniformly at point X.
- $\mathbb{E}[X|Y=y] = y/2$
- $\mathbb{E}[X|Y] = Y/2$
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[Y/2] = \frac{1}{2}\frac{l}{2} = l/4$

- Forecasts on sales: calculating expected value, given any available information
- X : February sales
- Forecast in the beg. of the year: $\mathbb{E}[X]$
- End of Jan. new information $Y=y$ (Jan. sales)
  Revised forecast: $\mathbb{E}[X|Y=y]$
  Revised forecast $\neq \mathbb{E}[X]$
- Law of iterated expectations
  $\mathbb{E}$[revised forecast] = original one

Example: Averaging Quiz Scores by Section

KAIST EE

- A class: n students, student i's quiz score: x_i
- Average quiz score: $m = \frac{1}{n}\sum_{i=1}^n x_i$
- Students: partitioned into sections $A_1, \ldots, A_k$ and n_s: number of students in section s
- average score in section s = $m_s = \frac{1}{n_s}\sum_{i \in A_s} x_i$
- whole average: (i) taking the average m_s of each section and (ii) forming a weighted average

$$\sum_{s=1}^k \frac{n_s}{n} m_s = \sum_{s=1}^k \frac{n_s}{n}\frac{1}{n_s}\sum_{i \in A_s} x_i = \frac{1}{n}\sum_{i=1}^n x_i = m$$

- Understanding from $\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X]$
- X: score of a randomly chosen student, Y: section of a student ($\in \{1,\ldots,k\}$)

$$m = \mathbb{E}(X) = \mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \sum_{s=1}^k \mathbb{E}(X|Y=s)\mathbb{P}(Y=s) = \sum_{s=1}^k \left(\frac{1}{n_s}\sum_{i \in A_s} x_i\right)\frac{n_s}{n} = \sum_{s=1}^k m_s \frac{n_s}{n}$$

Roadmap

Conditional Variance $\mathrm{var}[X|Y]$

$\mathrm{var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$

$g(y) = \mathrm{var}[X|Y = y] = \mathbb{E}[(X - \mathbb{E}[X|Y = y])^2|Y = y]$

$g(Y) = \mathrm{var}[X|Y] = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y]$

Conditional Variance

A random variable $g(Y) = \boxed{\mathrm{var}[X|Y]}$ and called $\boxed{\text{conditional variance of } X \text{ given } Y}$, takes the value $g(y) = \mathrm{var}[X|Y = y]$, if Y happens to take the value y.

- A function of Y
- A random variable
- Thus, having a distribution, expectation, variance, all the things that a random variable has

Expectation and Variance of $\mathbb{E}[X|Y]$ and $\mathrm{var}[X|Y]$

	$\mathbb{E}[X\|Y]$	$\mathrm{var}[X\|Y]$
Expectation	$\mathbb{E}\Big[\mathbb{E}(X\|Y)\Big]$	$\mathbb{E}\Big[\mathrm{var}(X\|Y)\Big]$
Variance	$\mathrm{var}\Big[\mathbb{E}(X\|Y)\Big]$	$\mathrm{var}\Big[\mathrm{var}(X\|Y)\Big]$

Law of Total Variance

Law of total variance (LTV)

$$\mathrm{var}[X] = \mathbb{E}\Big[\mathrm{var}(X|Y)\Big] + \mathrm{var}[\mathbb{E}(X|Y)]$$

Proof.

$$\mathrm{var}(X|Y) = \mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2$$

$$\mathbb{E}\Big[\mathrm{var}(X|Y)\Big] = \mathbb{E}[X^2] - \mathbb{E}\Big[(\mathbb{E}[X|Y])^2\Big] \quad (1)$$

$$\mathrm{var}\Big[\mathbb{E}(X|Y)\Big] = \mathbb{E}\Big[(\mathbb{E}[X|Y])^2\Big] - (\mathbb{E}\Big[\mathbb{E}(X|Y)\Big])^2 = \mathbb{E}\Big[(\mathbb{E}[X|Y])^2\Big] - (\mathbb{E}[X])^2 \quad (2)$$

$(1) + (2) = \mathbb{E}[X^2] + (\mathbb{E}[X])^2 = \mathrm{var}[X]$

Example: Averarging Quiz Scores by Section

- Same setting as that in page 36
- X: score of a randomly chosen student, Y: section of a student ($\in \{1, \ldots, k\}$)
- Let's intuitively understand: $\text{var}[X] = \mathbb{E}\Big[\text{var}(X|Y)\Big] + \text{var}[\mathbb{E}(X|Y)]$
- $\mathbb{E}[\text{var}(X|Y)] = \sum_{k=1}^{s} \mathbb{P}(Y = s)\text{var}(X|Y = s) = \sum_{k=1}^{s} \frac{n_s}{n}\text{var}(X|Y = s)$
  - Weighted average of the section variances
  - average score variability within individual sections
- $\text{var}[\mathbb{E}(X|Y)]$: variability of the average of the differenct sections
  - $\mathbb{E}(X|Y = s)$: average score in section s
  - variability between sections

Example: Stick-breaking

- Stick of length l
- Uniformly break at point Y, and break what is left uniformly at point X.
- Question. $\text{var}(X)$?
- LTV: $\text{var}[X] = \mathbb{E}\Big[\text{var}(X|Y)\Big] + \text{var}[\mathbb{E}(X|Y)]$
- Fact. If a rv $X \sim \mathcal{U}[0, \theta]$, then $\text{var}(X) = \frac{\theta^2}{12}$
- Since $X \sim \mathcal{U}[0, Y]$, $\text{var}(X|Y) = \frac{Y^2}{12} \to \mathbb{E}[\text{var}[X|Y]] = \frac{1}{12}\int_0^l \frac{1}{l} y^2 dy = \frac{l^2}{36}$
- $\mathbb{E}(X|Y) = Y/2 \to \text{var}(\mathbb{E}[X|Y]) = \frac{1}{4}\text{var}[Y] = \frac{1}{4}\frac{l^2}{12} = \frac{l^2}{48}$
- $\text{var}(X) = \frac{l^2}{36} + \frac{l^2}{48} = \frac{7l^2}{144}$

Roadmap

(1) Derived distribution of $Y = g(X)$ or $Z = g(X, Y)$
(2) Derived distribution of $Z = X + Y$
(3) Covariance: Degree of dependence between two rvs
(4) Correlation coefficient
(5) Conditional expectation and law of iterative expectations
(6) Conditional variance and law of total variance
(7) Random number of sum of random variables

Sum of a random number of rvs

- N : number of stores visited (random)
- X_i: money spent in store i, independent of other X_j and N, X_is are identically distributed with $\mathbb{E}[X_i] = \mu$
- $Y = X_1 + X_2 + \ldots X_N$. What are $\mathbb{E}[Y]$ and $\text{var}[Y]$?
- $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|N]] = \mathbb{E}[N\mathbb{E}[X_i]] = \mathbb{E}[N]\mathbb{E}[X_i] = \mu\mathbb{E}[N]$
- $\text{var}[Y] = \mathbb{E}\Big[\text{var}(Y|N)\Big] + \text{var}[\mathbb{E}(Y|N)] = \mathbb{E}[N]\text{var}[X_i] + \mu^2\text{var}[N]$

$$\text{var}(\mathbb{E}[Y|N]) = \text{var}(N\mu) = \mu^2\text{var}[N]$$
$$\text{var}[Y|N] = N\text{var}[X_i]$$
$$\mathbb{E}[\text{var}(Y|N)] = \mathbb{E}[N\text{var}[X_i]] = \mathbb{E}[N]\text{var}[X_i]$$

Questions?

Review Questions

1) What are the key steps to get the derived distributions of $Y = g(X)$ or $Z = g(X, Y)$?
2) How can we compute the distribution of $Z + X + Y$ when X and Y are independent?
3) What are covariance and correlation coefficient? Why do we need them?
4) Please explain the concepts of conditional expectation and conditional variance.