Lecture 5: Random Variable, Part III

Yi, Yung (이융)

EE210: Probability and Introductory Random Processes
KAIST EE

MONTH DAY, 2021

- Famous discrete random variables used in the community
  - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- Summarizing a random variable: Expectation and Variance
- Functions of a single random variable, Functions of multiple random variables
- Conditioning for random variables, Independence for random variables
- Continuous random variables
  - Normal, Uniform, Exponential, etc.
- Bayes' rule for random variables

- (Derived) Distribution of $Y = g(X)$ or $Z = g(X, Y)$

- Quantifying the degree of dependence between two rvs.

- Conditional expectation/variance

- (Random) Sum of random variables

- 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
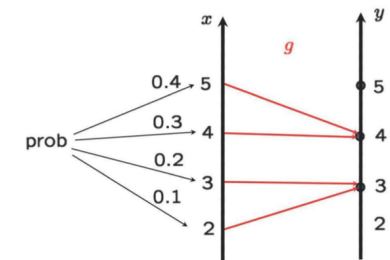 - Discrete
  - Linear: $Y = aX + b$

- 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: $Y = aX + b$, $a \neq 0$

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

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

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

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

Therefore, $f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$

Special case. X is normal. Then, Y is also normal, i.e., $Y \sim N(a\mu + b, a^2\sigma^2)$

Generally, $Y = g(X)$

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

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

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

Ex1. $X \sim uniform[0,1]$. $Y = \sqrt{X}$.

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

Ex2. $X \sim uniform[0,2]$. $Y = X^3$.

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

Ex3. X with $f_X(x)$. $Y = X^2$.

$$F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$
$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) +$$
$$\frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), \quad y \geq 0$$

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

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

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

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

$$F_Z(z) = \mathbb{P}(\max(X,Y) \leq z) = \mathbb{P}(X \leq z, Y \leq z)$$
$$= \mathbb{P}(X \leq z)\mathbb{P}(Y \leq z) = z^2$$
$$f_Z(z) = \begin{cases} 2z, & \text{if } 0 \leq z \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

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

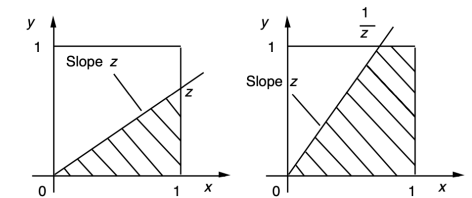Basically, follows two step approach: (i) CDF and (ii) differentiate.

Ex2. $X, Y \sim uniform[0,1]$, and $X \perp\!\!\!\perp Y$. $Z = Y/X$.

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

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

- Depending on the value of z, two cases need to be considered separately.
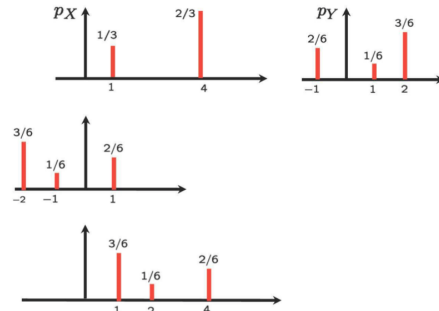
Functions of multiple rvs: $Z = X + Y$, $X \perp\!\!\!\perp Y$

- 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.

- Interpretation (for a given z)
(i) Flip (horizontally) $p_Y(y)$ $(p_Y(-x))$
(ii) Put it underneath $p_X(x)$ $(p_Y(-x + z))$

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

- Same logic as the discrete case

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

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

> **Sum of two independent normal rvs**
> $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_x, \sigma_x^2)$
> Then, $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$

- Why normal rvs are used to model the sum of random noises.

- (Extension) The sum of finitely many independent normals is also normal.

Roadmap

- Famous discrete random variables used in the community
  - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- Summarizing a random variable: Expectation and Variance
- Functions of a single random variable, Functions of multiple random variables
- Conditioning for random variables, Independence for random variables
- Continuous random variables
  - Normal, Uniform, Exponential, etc.
- Bayes' rule for random variables

- (Derived) Distribution of $Y = g(X)$ or $Z = g(X, Y)$

- Quantifying the degree of dependence between two rvs.

- Conditional expectation/variance

- (Random) 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
- Reqs.
  a) Increases (resp. decreases) as they become more (resp. less) dependent.
  b) 0 when they are independent.
  c) Shows the direction of dependence by $+$ and $-$
  d) Always bounded by some numbers, e.g., $[-1, 1]$
- Good engineers: Good at making good metrics
  - Metric of how our society is economically polarized
  - A lot of metrics in our professional sports leagues (baseball, basketball, etc)
  - Cybermetrics in MLB (Major League Baseball):
    http://m.mlb.com/glossary/advanced-stats

- 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)
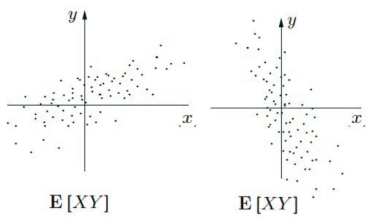
(Q) What about $\mathbb{E}[X + Y]$?

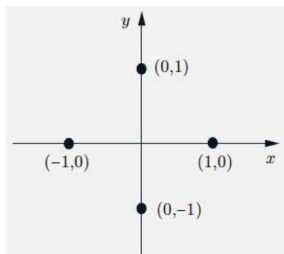- Solution: Centering. $X \to X - \mu_X$ and $Y \to Y - \mu_Y$

> **Covariance**
>
> $$\text{cov}(X, Y) = \mathbb{E}\Big[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\Big]$$

- After some algebra, $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

- $X \perp\!\!\!\perp Y \implies \text{cov}(X, Y) = 0$

- $\text{cov}(X, Y) = 0 \implies X \perp\!\!\!\perp Y$? NO.

- When $\text{cov}(X, Y) = 0$, we say that X and Y are $\boxed{\text{uncorrelated.}}$

- $p_{X,Y}(1, 0) = p_{X,Y}(0, 1) = p_{X,Y}(-1, 0) = p_{X,Y}(0, -1) = 1/4$.

- $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, and $\mathbb{E}[XY] = 0$. So, $\text{cov}(X, Y) = 0$

- Are they independent? No, because if $X = 1$, then we should have $Y = 0$.

$$\text{cov}(X, X) = 0$$

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

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

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

- n people throw their hats in a box and then pick one at random
- X: number of people with their own hat
- (Q) 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?
- $X_i \sim Bernoulli(1/n)$. Thus, $\mathbb{E}[X_i] = 1/n$ and var$[X_i] = \frac{1}{n}(1 - \frac{1}{n})$

○ For $i \neq j$,

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

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

- Reqs. a), b), and c) satisfied.
  d) Always bounded by some numbers, e.g., $[-1, 1]$
- Dimensionless metric. How? $\boxed{\text{Normalization,}}$ but by what?

Correlation Coefficient

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

- $-1 \leq \rho \leq 1$
- $|\rho| = 1 \implies X - \mu_X = c(Y - \mu_Y)$ (linear relation, VERY related)

- ○ Famous discrete random variables used in the community
  - Bernoulli, Uniform, Binomial, Geometric, Poisson, etc.
- ○ Summarizing a random variable: Expectation and Variance
- ○ Functions of a single random variable, Functions of multiple random variables
- ○ Conditioning for random variables, Independence for random variables
- ○ Continuous random variables
  - Normal, Uniform, Exponential, etc.
- ○ Bayes' rule for random variables

- ○ (Derived) Distribution of $Y = g(X)$ or $Z = g(X, Y)$
- ○ Quantifying the degree of dependence between two rvs.
- ○ Conditional expectation/variance
- ○ (Random) Sum of random variables

- 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, such that

$$g(y) = \mathbb{E}[X|Y = y] = \begin{cases} 3, & \text{if } y = 0 \\ 8, & \text{if } y = 1 \\ 9, & \text{if } y = 2 \end{cases}$$

- 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 notate it with some fancy one.

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

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 Conditional Expectation

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

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

- 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
  $\mathbb{E}[\text{revised forecast}] = \text{original one}$

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

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

Conditional Variance

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

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

Law of total variance

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

Proof.

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

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

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

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

- 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?

KAIST EE

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.