DCP 1206: Probability
Lecture 8 — Expectation, Variance, and Higher Moments

Ping-Chun Hsieh

October 9, 2019

Announcement

- HW2 is posted on E3 (Due: 10/23 in class)

Powerball Lottery and Fortune Cookies

- One day, in 2005, <u>110 lucky people</u> in the US won the same Powerball lottery.

C^{69}_{5} 26

- <u>Powerball</u>: Pick 5 numbers from 1~69 + 1 number from 1~26
  - What is the probability of winning the lottery?

https://www.nytimes.com/2005/05/11/nyregion/who-needs-giacomo-bet-on-the-fortune-cookie.html

Quick Review

- Geometric ($X \sim \text{Geometric}(p)$)

- Negative Binomial ($X \sim \text{NB}(p, r)$)

- Discrete Uniform ($X \sim \text{Unif}(a, b)$)

- Hypergeometric ($X \sim \text{HG}(N, D, n)$)

This Lecture

1. Expected Value

2. Variance and Moments

3. Special Discrete R.V.: Expected Value and Variance

- Reading material: Chapter 4.4-4.5 and 5.1-5.3

7. Hypergeometric Random Variables (Formally)

total # of students

交清

Hypergeometric Random Variables: A random variable X is hypergeometric with parameters (N, D, n) $(N, D, n \in \mathbb{N}$ with $D < N$ and $n \leq \min(D, N-D))$, if its PMF is given by

$$P(X = k) = \frac{C_k^D C_{n-k}^{N-D}}{C_n^N}, \ k = 0,1,2,\cdots,n$$

$$\sum_{k=0}^{n} \frac{C_k^D \cdot C_{n-k}^{N-D}}{C_n^N} = 1$$

- Do we have $\sum_{k=0}^{n} P(X = k) = 1$?

$$(1+x)^N = (1+x)^D \cdot (1+x)^{N-D}$$

$$x^n \qquad x^n$$

Why Is It Called "Hypergeometric"?

more

$$P(X=k)=\frac{C_k^D C_{n-k}^{N-D}}{C_n^N},\ k=0,1,2,\cdots,n$$

$1,\ \frac{1}{2},\ \left(\frac{1}{2}\right)^2,\ \left(\frac{1}{2}\right)^3,\ \dots$

$a_1,\ a_2,\ a_3,\ a_4$

$a_k=\left(\frac{1}{2}\right)^{k-1}$

$\frac{a_{k+1}}{a_k}=\frac{1}{2}$

- Can we find $\dfrac{P(X=k+1)}{P(X=k)}$?

$$\frac{\dfrac{C_{k+1}^{D}C_{n-(k+1)}^{N-D}}{C_n^N}}{\dfrac{C_k^D C_{n-k}^{N-D}}{C_n^N}}=\frac{\dfrac{D(D-1)\cdots(D-k)}{(k+1)!}\cdot\dfrac{\overbrace{(N-D)\cdots}^{n-(k+1)\text{ terms}}}{(n-(k+1))!}}{\dfrac{D(D-1)\cdots(D-k+1)}{k!}\cdot\dfrac{\underbrace{(N-D)\cdots}_{(n-k)\text{ terms}}}{(n-k)!}}$$

$$=\frac{(D-k)\cdot(n-k)}{(k+1)\cdot(N-D-(n-k)+1)}$$

$f(k)$

$g(k)$

Hypergeometric: A Limiting Case

$$P(X = k) = \frac{C_k^D C_{n-k}^{N-D}}{C_n^N},\ k = 0,1,2,\cdots,n$$

- What if $D/N = p$ and $D, N \gg n$? $\Rightarrow$ Binomial

支清收径

with probability $p \rightarrow$ type A

$1-p \rightarrow$ type B

1. Expected Value

Motivation: Guidelines for Decision Making?

- Suppose we are given 2 options:

$5 \times 10^8 \times \frac{1}{1000} = 500000 = 500k$

Option A

6

10^6

Win 500 millions with prob. 0.1%

Lose 50k dollars, otherwise

risk-seeking

Option B

70~

Win 50k dollars, with prob. 1

risk-averse

100k

- Which option will you choose?

- Could you come up with a creative way to get a reward higher than B but with a lower risk than A?

Why Expected Value?

- Example: Imagine you are an investor of Shinemood
  - $X = $ # of waffles sold by Shinemood today
  - Suppose $X \sim$ Poisson($\lambda = 150$, $T = 1$ day)

- Question 1: How many waffles are expected to be sold today?

expected value

- Question 2: In the coming year, how many waffles will be sold per day on average?

$$\frac{X_1 + X_2 + \cdots + X_{365}}{365}$$

empirical average

- Q1 and Q2 are closely related: Law of Large Numbers

Example: Expected Value and Poisson

- Example: $X = \#$ of waffles sold by Shinemood today
  - Suppose $X \sim$ Poisson$(\lambda = 150, T = 1 \text{ day})$
  - PMF of X? How to define the expected value of X?

PMF of X:

$$P(X=n) = \frac{e^{-\lambda T} \cdot (\lambda T)^{n}}{n!}$$

$$1 \cdot P(X=1) + 2 \cdot P(X=2) + \vdots = \sum_{n=1}^{\infty} n \cdot P(X=n).$$

Expected Value of a Discrete R.V. (Formally)

Expected Value (or Mean / Expectation):

Let X be a discrete random variable with

- the set of possible values S $= \{1, 2, 3, 4, 5, \ldots\}$
- PMF of X is $p_X(x)$

The expected value of X is defined as

$$E[X] := \sum_{x \in S} x \cdot p_X(x)$$

(high school def.)

- Sometimes we use the notation: $\mu_X \equiv E[X]$

Example: Expected Value

- Example: Suppose X has a PMF: $p_X(n) = \frac{1}{n(n+1)}$, $n \geq 1$
  - What is $E[X]$?

$n \in \mathbb{N}$

$\overset{\shortparallel}{P(X=n)}$

$$E[X] = \sum_{n=1}^{\infty} n \cdot P_X(n)$$

$$= \sum_{n=1}^{\infty} \cancel{n} \cdot \frac{1}{\cancel{n}(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$$

$E[X]$ is not well-defined.

$\int_1 \frac{1}{x} dx$

$= \log x$

$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}$

$\geq \frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}_{\frac{1}{2}} + \underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{\frac{1}{2}} + \cdots$

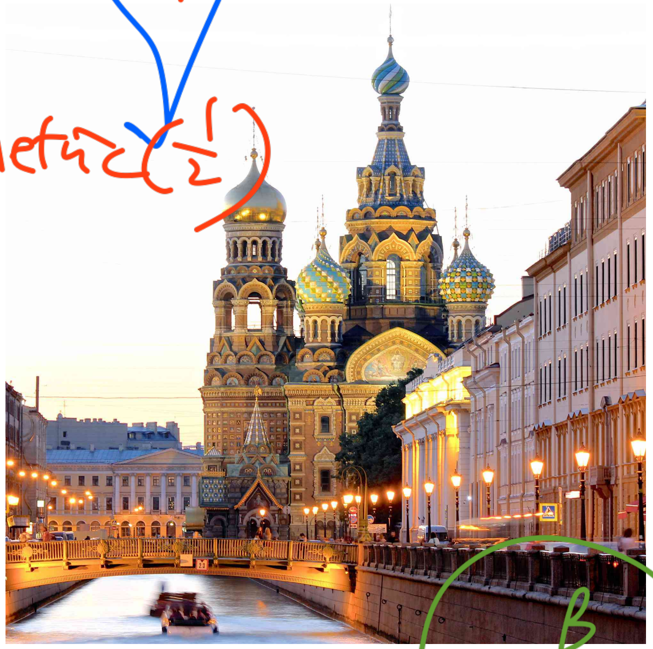
$X \in \underbrace{\{a_1, a_2, a_3, \ldots, a_{100}\}}$

$\underbrace{a_1 < a_2 < a_3 < \cdots < a_{100}}$

$E[X] \leq a_{100}$

Example: St. Petersburg Paradox 悖論

- Example: We are asked to pay 10000 dollars to play a game.
  - We can keep flipping a fair coin until a head is observed.
  - If the 1st head occurs at n-th toss, then we get a prize of 2^n dollars and the game is over.
  - Shall we play this game?



$X \sim \text{Geometric}(\frac{1}{2})$

$X =$ 1st head occurs at n-th toss

10

$Y =$ prize we get

$$P(Y=k) = \begin{cases} \frac{1}{2} & k = \cancel{2}^1\ \beta^1 \\ (\frac{1}{2})^2 & k = \cancel{2}^2\ \beta^2 \\ (\frac{1}{2})^3 & k = \cancel{2}^3\ \beta^3 \\ \vdots & \vdots \\ (\frac{1}{2})^k & k = \cancel{2}^k\ \beta^k \\ & \vdots \end{cases}$$

$$E[Y] = \sum_y y \cdot P(Y=y) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \cdot \beta^k = \sum_{k=1}^{\infty} \left(\frac{\beta}{2}\right)^k = \frac{\frac{\beta}{2}}{1-\frac{\beta}{2}} \geq 10000$$

Example: St. Petersburg Paradox

- Example: We are asked to pay <u>10000 dollars</u> to play a game.
  - We can keep flipping a <u>fair</u> coin <u>until a head is observed</u>.
  - If the 1st head occurs at n-th toss, then we get a prize of <u>β^n dollars</u> and the game is over.
  - Under what β shall we play this game?

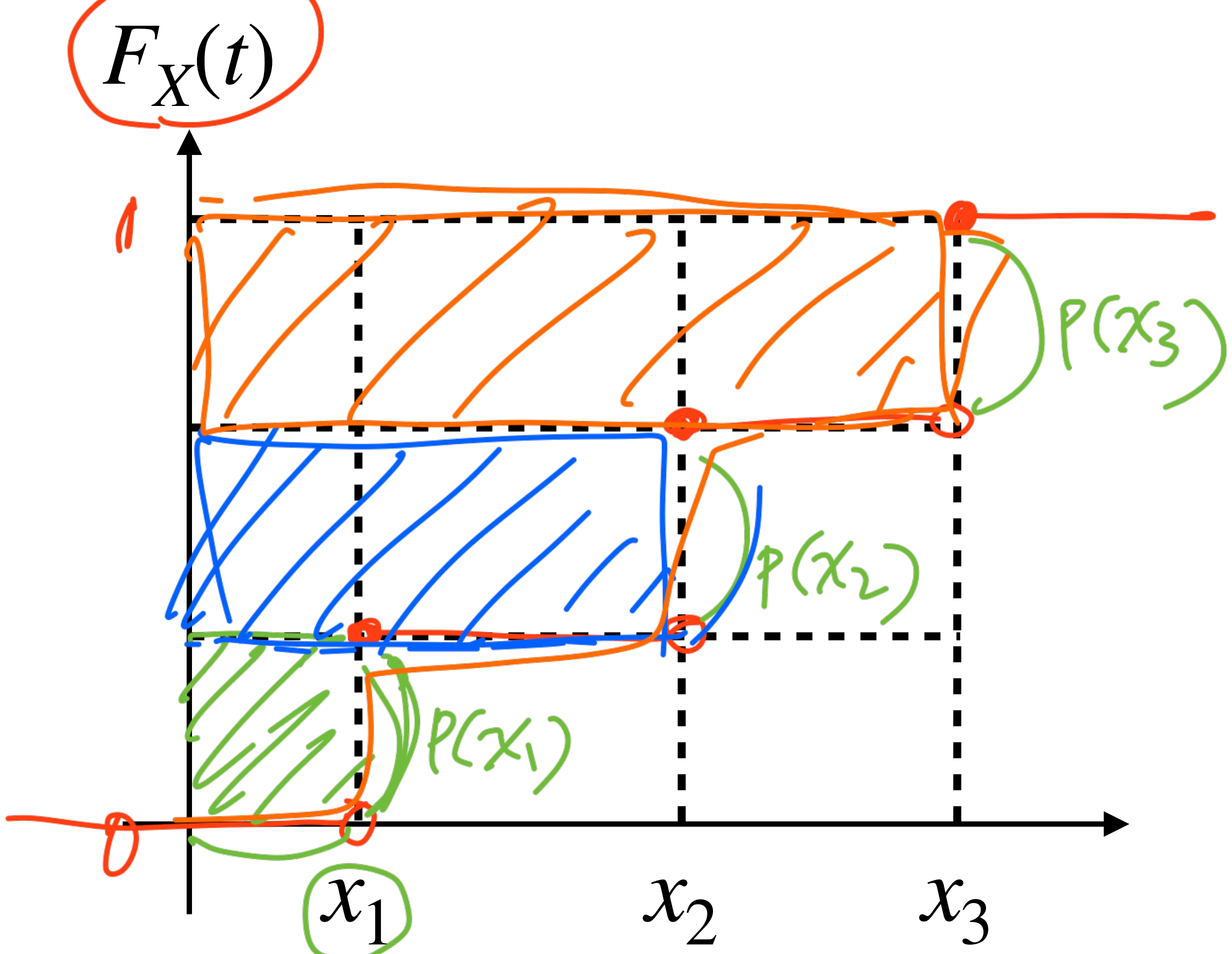
Visualize the Expected Value Using CDF

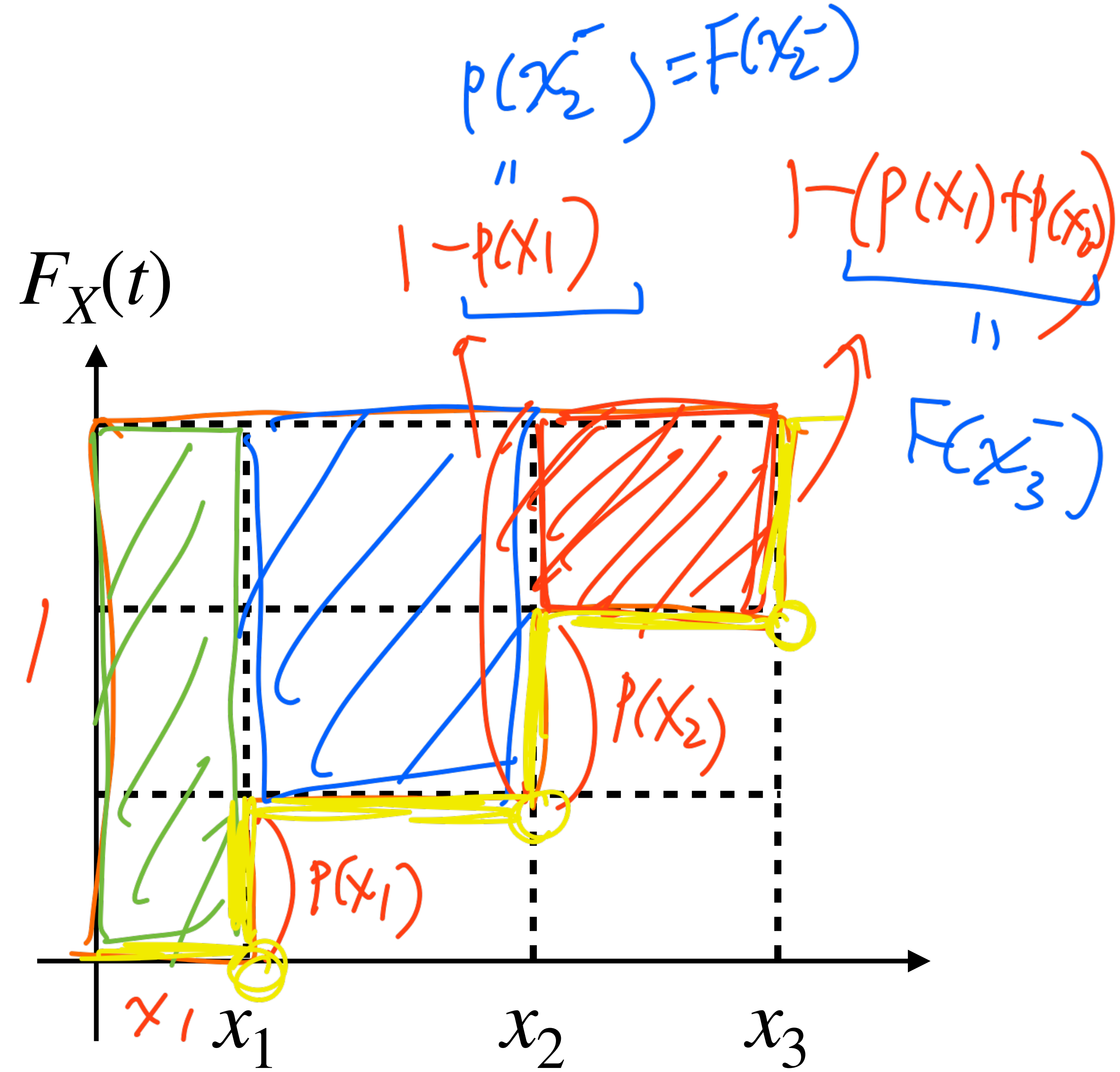
$X \geq 0$

- Suppose X is a non-negative discrete random variable with
  - The set of possible values $\{x_1, x_2, x_3, \cdots\}$ (assume $x_i < x_{i+1}$)
  - Denote $x_0 = 0$

CDF

$E[X] = \sum x \cdot P(x)$





Expected Value of a Discrete Random Variable: An Alternative Expression

Expected Value (or Mean / Expectation):

Let X be a non-negative discrete random variable with

- the set of possible values $S = \{x_1, x_2, x_3 \cdots\}$
- CDF of X is $F_X(t)$

Denote $x_0 = 0$. The expected value of X is

$$E[X] = \sum_{i=1}^{\infty} (x_i - x_{i-1}) \cdot \left(1 - F_X(x_i^-)\right)$$

- What if $S = \{1,2,3 \cdots\}$?

$$E[X] = \sum_{i=1}^{\infty} 1 \cdot P(X > i)$$ (next week)

- How about continuous cases?

$$E[X] = \int_0^{\infty} P(X > t)\, dt$$

Example: Using the Alternative Expression

- Example: Suppose X is a discrete random variable
  - For X, the set of possible values $A = \{2, 4, 6, 8 \cdots\}$
  - The CDF of X is $F_X(t) = 1 - \frac{1}{t^2}$, $t \in A$
  - What is $E[X]$?

$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots$

$x_1 \quad x_2 \quad x_3 \quad x_4$

$$E[X] = \sum_{i=1}^{\infty} \underbrace{(x_i - x_{i-1})}_{=\,2,\ \text{for all } i} \cdot \underbrace{\left(1 - F_X(x_i^-)\right)}$$

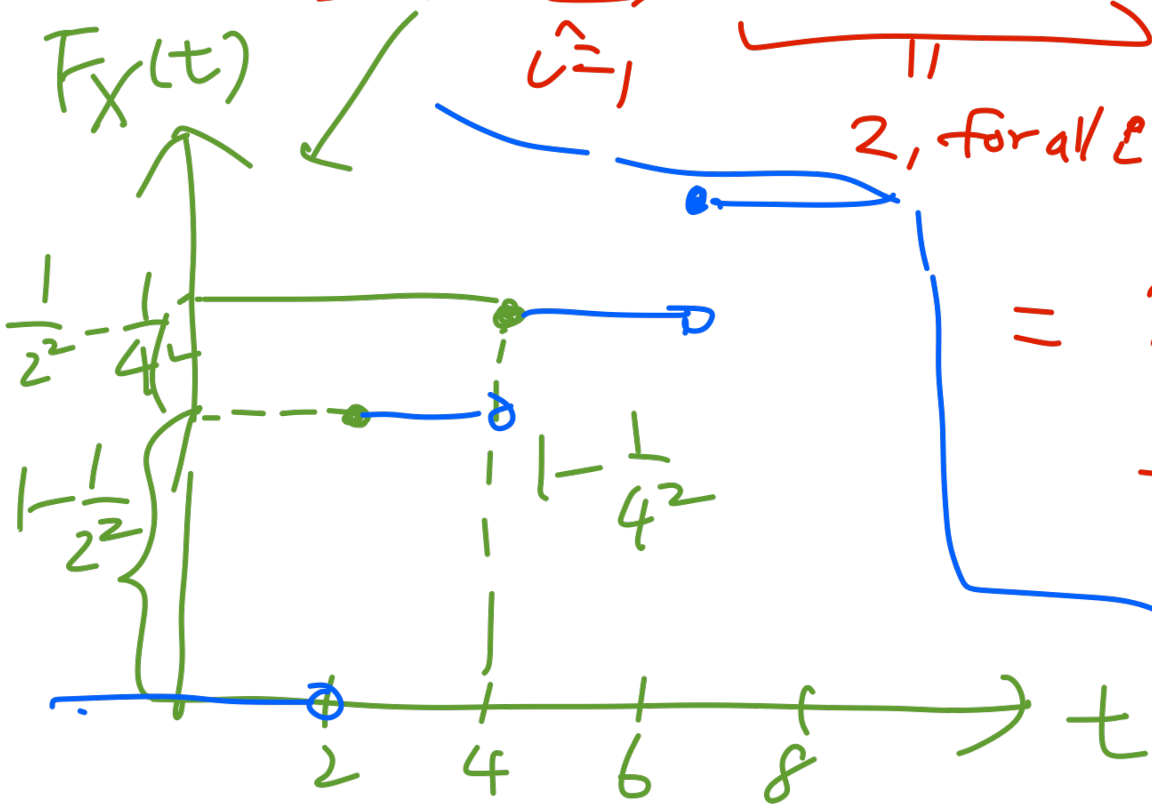
$$= 2 \cdot (1 - 0) + \sum_{i=2}^{\infty} 2 \cdot \left(\frac{1}{(2(i-1))^2} \right)$$

$$= 2 + \frac{1}{2} \sum_{i=2}^{\infty} \frac{1}{(i-1)^2} \qquad = \frac{\pi^2}{6}$$

$F_X(2^-) = 0$

$F_X(4^-) = 1 - \frac{1}{2^2}$

$F_X(6^-) = 1 - \frac{1}{4^2}$



A Property of Expected Value

Theorem (Expectation of a Function of r.v.):

1. Let X be a discrete random variable with
   - the set of possible values S
   - PMF of X is $p_X(x)$
2. Let $g(\cdot)$ be a real-valued function

$Y = g(X)$

$E[Y] = \sum_y y \cdot P_Y(y)$

The expectation of $g(X)$ is

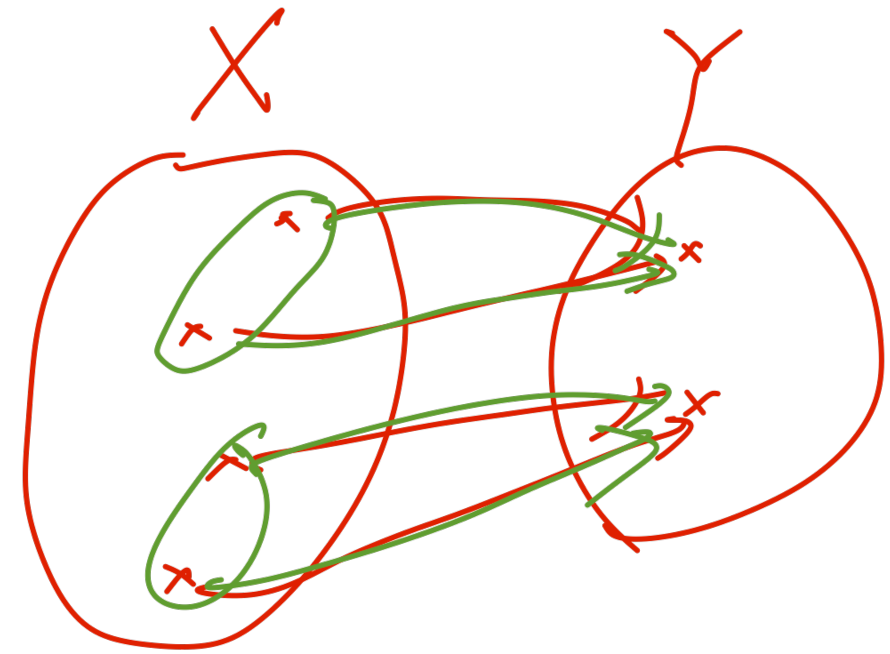
$$\underbrace{E[g(X)]}_{E[X]} = \sum_{x \in S} \overset{x}{\overset{\downarrow}{g(x)}} \cdot \underbrace{p_X(x)}_{\text{PMF}}$$

- Is this intuitive? Do we need a proof?
- Also called Law of the unconscious statistician (LOTUS)

Proof of Law of the Unconscious Statistician

$$E[g(X)] := \sum_{x \in S} g(x) \cdot p_X(x)$$

Define $Y = g(X)$



$$E[g(X)] = E[Y]$$

$$= \sum_y y \cdot p_Y(y)$$

$$p_Y(y) = \left(\sum_{x: g(x)=y}\right) p_X(x)$$

$$= \sum_x g(x) \cdot p_X(x)$$

Linearity of Expected Values (I)

Linearity Property (I):

Let X be a discrete random variable and α, β be real numbers. Then, we have

$$E[\alpha X + \beta] = \alpha \cdot E[X] + \beta$$

- How to show this?

$g(X) = \alpha X + \beta$

LOTUS: $E[\alpha X + \beta] = E[g(X)] = \sum_x g(x) \cdot P_X(x)$

$= \sum_x (\alpha x + \beta) P_X(x)$

$= \alpha \cdot E[X] + \beta$

Linearity of Expected Values (II)

Linearity Property (II):

Let X be a discrete random variable and $g(\cdot), h(\cdot)$ be real numbers. Then, we have

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]$$

- How to show this?

$$= \sum_x (g(x) + h(x)) \cdot P_X(x)$$

$$= \sum_x g(x) \cdot \_\_ + \sum h(x) \cdot \_\_$$

$$= E[g(X)] + E[h(X)]$$

#

Conditional Expectation

- Example: Roll a fair 6-sided die once
  - Define $X =$ the number that we observe
  - Given that $X \geq 4$, what is the expected value of X?

$\{1,2,3,4,5,6\}$

$P(X=4 \mid X\geq 4) = \frac{1}{3}$

$P(X=5 \mid X\geq 4) = \frac{1}{3}$

$P(X=6 \mid \cdots) = \frac{1}{3}$

Conditional Expectation:

Let X be a discrete random variable with the set of possible values $S = \{x_1, x_2, x_3 \cdots\}$. Let A be an event. The expected value of X conditioned on A

$$E[X \mid A] := \sum_{x \in S} x \cdot P(X = x \mid A)$$

Example: Taiwan Receipt Lottery

invoice

- Example: Suppose we have a receipt at hand
  - Define X = the prize we get
  - What is $E[X]$?
  - Given that the last digit is 9, what is the expected value of X?

108年 7 - 8 月 統一發票開獎		
特別獎	45698621	與左欄號碼相同者獎金1,000萬元
特獎	19614436	與左欄號碼相同者獎金200萬元
頭獎	96182420 47464012 62781818	頭獎 與頭獎號碼完全相同者獎金20萬元 二獎 與頭獎末7碼相同者各得獎金4萬元 三獎 與頭獎末6碼相同者各得獎金1萬元 四獎 與頭獎末5碼相同者各得獎金4千元 五獎 與頭獎末4碼相同者各得獎金1千元 六獎 與頭獎末3碼相同者各得獎金2百元
增開六獎	928 899	末3碼與增開六獎號碼相同者各得獎金2百元

正確資訊請以財政部提供為準 中央社祝您幸運中獎

00~99

$\underbrace{000\cdots0}_{8} \rightarrow \underbrace{999\cdots9}_{8}$

$\frac{1}{10^8} \times (1000萬)$

$\frac{1}{10^8} \times (200萬)$

$\vdots$

$\frac{2}{1000} \times (200)$

$E(X \mid \text{last digit}=9)$

$= 200 \times \frac{1}{100}$

2. Variance and Moments

Moments and Others

X^2

$$E[g(X)] := \sum_{x \in S} g(x) \cdot p(x)$$

- Example: $g(X) = X^2$ $\rightarrow$ 2nd moment of X
- Example: $g(X) = X^n$ $\rightarrow$ n-th moment
- Example: $g(X) = (X - \mu_X)^2$ $\rightarrow$ Variance (2nd "central" moment)
- Example: $g(X) = (X - \mu_X)^n$ $\rightarrow$ n-th "central" moment
- Example: $g(X) = e^{tX}$ $\rightarrow$ Moment Generating function (MGF)

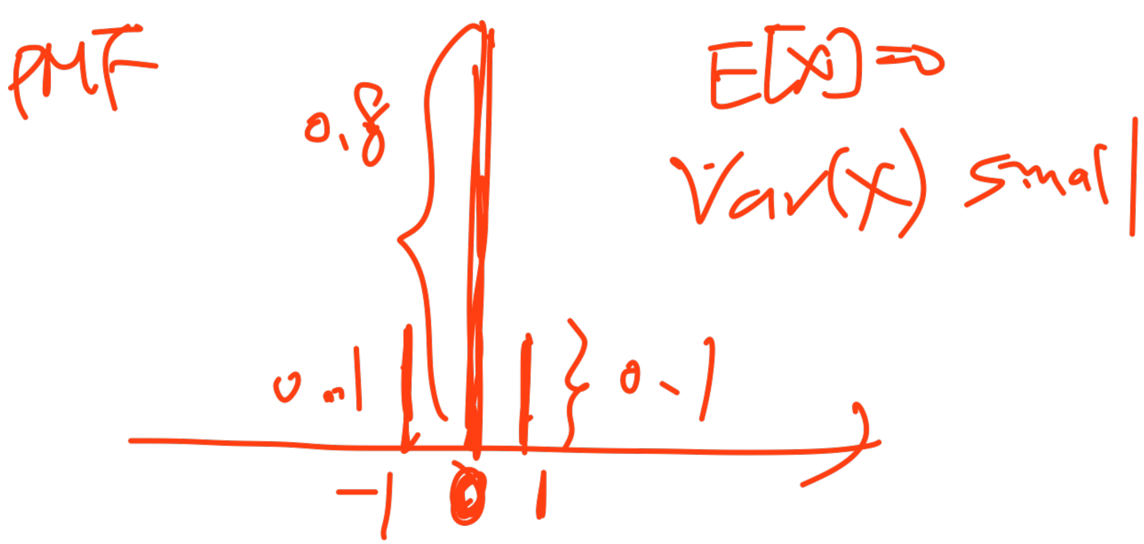
Variance

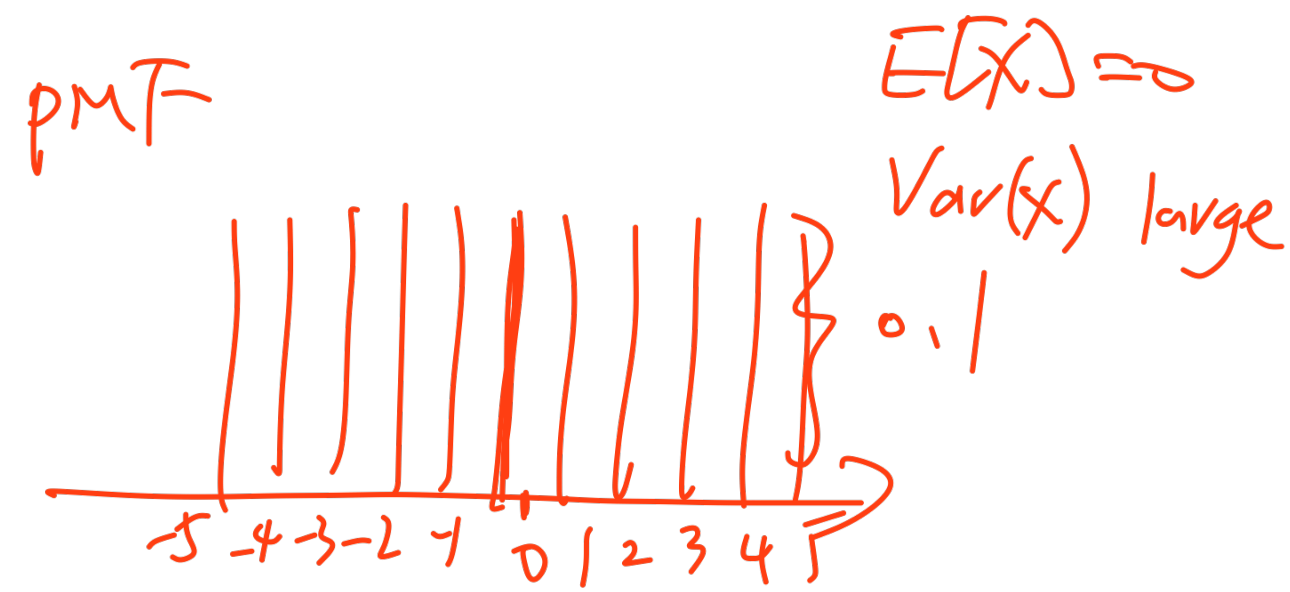
Variance (2nd central moment):

Let X be a discrete random variable with the set of possible values S and PMF $p_X(x)$. The variance of X is

$$\underbrace{\mathrm{Var}[X] := E[(X-\mu_X)^2]}_{\text{standard deviation}} = \sum_{x \in S} (x-\mu_X)^2 \cdot p_X(x)$$

- Sometimes we use the notation: $\sigma_X^2 \equiv \mathrm{Var}[X]$
- Variance captures the variability of a random variable





Variance: An Alternative Explanation

- Example: Suppose we are given a random variable X
  - We need to output a prediction of X (denoted by z)
  - Penalty of prediction is $(X - z)^2$
  - What is the minimum expected penalty?

Another Way for Calculating Variance

Theorem:
Let X be a random variable. Then, we have

$$\underbrace{\text{Var}[X]}_{} := \underbrace{E[X^2]}_{\text{平方的平均}} - \underbrace{(E[X])^2}_{\text{平均的平方}} \geq 0$$

- How to show this?

$$\text{Var}[X] = E\left[(X-\mu_X)^2\right]$$
$$= E\left[X^2 - 2\cdot\mu_X\cdot X + \mu_X^2\right]$$
$$= E[X^2] - \underbrace{E[2\mu_X\cdot X]}_{= 2\cdot\mu_X\cdot E[X]} + \underbrace{E[\mu_X^2]}_{\mu_X^2}$$
$$E[X] = \mu_X$$
$$= E[X^2] - \mu_X^2$$

Properties of Variance and Moments (I)

1. $\text{Var}(\underbrace{X + c}_{})$ = $\text{Var}(X)$? $\qquad E[(X-\mu_X)^2]$

   (handwritten above $X + c$: "= Y")

2. $\text{Var}(aX) = a^{2} \cdot \text{Var}(X)$?

counter-example

3. $\text{Var}(\,|X|\,) = \text{Var}(X)$?

4. $E(X^2) \geq (E(X))^2$?

5. Can $\text{Var}(X)$ be infinite? ← St. Petersburg.

When are Higher Moments Useful?

Berry-Esseen Theorem:

Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with $E[X_1] = 0$, $E[X_1^2] = \sigma^2$ and $E[|X_1|^3] < \infty$. Define $Y = (X_1 + X_2 + \cdots + X_n)/n$. Then, we have

$$|F_Y(t) - \Phi(t)| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

- Usually higher moments are used as technical conditions
  - Hence, we usually care about whether $E(X^n) < \infty$

Properties of Variance and Moments (II)

X is non-negative

$n = 1, 2, 3 \cdots$

6. If $E(X^{n+1}) < \infty$, then $E(X^n) < \infty$?

$1000^3 > 1000^2$

$|X| < 1$: Both X^{n+1} and X^n are small

$|X| \geq 1$: $|X|^{n+1} \geq X^n$

$|X|^{n+1} \leq X^n$

$E[X^n] \leq E[|X|^n] \leq E[|X|^{n+1}] < \infty$

3. Expected Value and Variance of Special Discrete Random Variables

1. Bernoulli Random Variables

- Example: $X \sim$ Bernoulli(p)
  - What is $E[X]$?
  - What is Var$[X]$?

2. Binomial Random Variables

- Example: $X \sim \text{Binomial}(n, p)$
  - What is $E[X]$?
  - What is $\text{Var}[X]$?

3. Poisson Random Variables

- Example: $X \sim \text{Poisson}(\lambda, T)$
  - What is $E[X]$?
  - What is $\text{Var}[X]$?

4. Geometric Random Variables

- Example: $X \sim$ Geometric(p)
  - What is $E[X]$?
  - What is Var$[X]$?

Next Lecture

1. Continuous random variables

2. Probability density function (PDF)

1-Minute Summary

1. Expected Value

- Definition / alternative expression / linearity / conditional expectation
- Law of the unconscious statistician (LOTUS)

2. Variance and Moments

- Definition / alternative explanation using penalty / properties

3. Special Discrete R.V.: Expected Value and Variance

- Bernoulli / Binomial / Poisson / Geometric