L9 10/3

Expectation of Random variable

- a numerial measure for the center of a probability distribution
- Motivation (of the def)

E.g. dataset 4, 3, 4, 2, 5, 1, 6, 6, 5, 2.

mean of the data set $= \frac{\text{sum}}{\#\text{ variables}} = 3.8$.

Alternatively: mean $= \sum \text{val} \times \text{freq}$

$1 \times \frac{1}{10} + \cdots\cdots$.

mean of probability distribution $= \sum$ value $\times$ probability

Def: The expectation (expected value, or mean) of a random variable x, is the mean of the probability distribution of x, denoted by $E(x)$ or μ.

$$E(x) = \mu = \sum_{\text{all } x} x p(x)$$

$$= \sum_{i=1}^{k} x_i p(x_i)$$

$E(x)$ is the weighted (probability) average of values of x

E.g. x: # produced by rolling a dice

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Find $E(x)$

Sol: $E(x) = \sum_{\text{all } x} x p(x)$

$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} \cdots\cdots + 6 \times \frac{1}{6} = 3.5$

Eg 2 X - sum of points in 2 tosses of a dice

x	2	3	4	...	12
$p(x)$	$1/36$	$2/36$	...		$1/36$

$E(x) = 2 \cdot \frac{1}{36} + 3 \times \frac{2}{36} + \cdots + 12 \times \frac{1}{36} = 7$ ← coinsidence ? No.

Additional Rule of Expectation.

- For any 2 random variables X, Y (not necessarily independent)

  $E(X+Y) = E(X) + E(Y)$

- In general: $X_1, X_2 \cdots X_n$

  $E(X_1 + X_2 + \cdots X_n) = E(X_1) + E(X_2) + \cdots E(X_n)$

Eg. $X = X_1 + X_2$. (2 tosses)

X_1: 1st toss, X_2: 2nd toss

$E(X) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$.

L10 10/8

Practice Exam 21c $\frac{48}{52} \times \frac{47}{51}$.

Recall $E(x) = \sum_{i=1}^{k} x_i p(x_i)$

Interpretation of E(X)

1) Long-Run Average. $E(X)$ is approximately the long-run average of x, if probabilities are approximate long-run relative frequencies.

E.g. X - # produced by rolling a dice.

$E(X) = 3.5$.

meaning: · roll the dice $n=2$ times

$\underline{2}$ $\underline{6}$ $\rightarrow$ mean $\frac{2+6}{2} = 4$.

· roll the dice $n=10$ times.

1 5 4 2 3 6 1 2 5 2 $\rightarrow$ mean $\frac{1+5+4+\cdots+2}{10} = 3.1$

· roll the dice $n=1000$ times.

$\underbrace{- - - - - - -}_{1000 \text{ numbers}} \Rightarrow = 3.4$

the average number of rolls will be close to $E(X)$ as n increases.

2) Gambling Interpretation

If $X denote the return, then $E(X) is the fair price to pay.

E.g. Consider the game of roulette. Each of 38 numbers.

00, 0, 1, ----, 36 are equally likely. Suppose we bet $1 on black.

Let X denote the return

$$X = \begin{cases} 2 & \text{if black} \\ 0 & \text{otherwise.} \end{cases}$$

Find the expected return:

Sol: $E(X) = 2 \cdot P(X=2) + 0 \cdot P(X=0)$

$= 2 \times \frac{18}{38} + 0 \cdot \frac{20}{38} = \$0.947$ → expected return.

$P(X=2) = P(\# \text{ is black}) = \frac{18}{38}$

$P(X=0) = \frac{20}{38}$

What does this mean?

Over the long run, this prize makes wins and loses tend to concel out over the long run.

· play 1 times, \$2.

\$0.947 / game.

we know, pay \$1 to play the game

expected return - \$ $= 0.947 - 1 = -0.053$.

↳ over the long run, player lose a little bit more than 5 cents per game.

For every 2 random variables X, Y,

$E(X+Y) = E(X) + E(Y)$ (Not necessarily independent)

In general, $E(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} E(X_i)$

E.g. Let Y be the sum of points in n tosses of a dice. Find $E(Y)$

Sol: $E(Y) = ?$

Let X_i - the #produced in i^{th} roll, $i = 1, 2, \cdots, n$

$Y = X_1 + X_2 + \cdots + X_n$

$E(Y) = E(X_1) + E(X_2) + \cdots + E(X_n) = 3.5n$

E.g. A construction company submits bids for 2 projects. Listed have are the profit and the probability of winning each project.

Assume that the outcomes of the 2 bids are independent.

Project	Profit	Probability of bid
A	\$175,000	.5
B	\$220,000	.65.

Let x denote the company's total profit out of the 2 contracts. Find $E(x)$, the expected total profit.

Sol: x_1: the profit out of project A

X_2: - - - - - - - - - - B.

$x = X_1 + X_2$.

$E(x) = E(X_1) + E(X_2)$

$= 175000 \times 0.5 + 220{,}000 \times 0.65 = 230{,}500$

Method II. (obtain prob distribution of $\underline{x}$)

4 outcomes of bidding 2 projects.				Probability.	
$W_A W_B$	$175000 + 222\,000$	$= \$395\,000$	$P(x=395000) =$	0.5×0.65	$= 0.325$
$W_A W_B^c$	$175000 \quad 0$	175000		0.5×0.35	$= 0.175$
$W_A^c W_B$	$0 \quad 222\,000$	222000		0.5×0.65	$= 0.325$
$W_A^c W_A^B$	$0 \quad 0$	0		0.5×0.35	$= 0.175$

$E(x) = 0.325 \times 395{,}000 + 0.175 \times 175{,}000 + 0.325 \times 222{,}000 + 0.175 \times 0$

$= 230{,}500$

Condition to use addition rule:

If x is possible to be written as sum of random variables and the expectations of these random variables are easy to calculate.

E.g. A trip insurance policy pay \$3000 to the customer in case of a loss due to theft or damage on a five-day trip. If the risk of such a loss is 1 in 200, what is the expected cost, per customer, to cover?

Sol: X - cost / payment of the insurance company to customer

	x	$P(x)$
$X \Rightarrow$	\$3000	$\frac{1}{200}$
	0	$\frac{199}{200}$

$E(X) = 3000 \times \frac{1}{200} + 0 \times \frac{199}{200} = 15$

This is the fair premium the insurance company choose.

$\hookrightarrow$ over the long run, the insurance company neither make a profit or lose money.

In practice, the insurance company will set a premium higher than \$E(X)

Indicator of an event.

For any events A, the indicator of A,

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

Properties:

① I_A is a random variable

$P(I_A = 1) = P(A)$, $P(I_A = 0) = 1 - P(A)$.

② The expection of I_A, $E(I_A) = P(A)$.

$\hookrightarrow = 1 \cdot P(A) + 0 \cdot (1 - P(A))$.

If X is the number of events that occur among some collection of events $A_1, A_2 \cdots$ And the possible value of X are $\{0, 1, \cdots, n\}$.

The random variable X is called a countary variable.

The expectation of x $E(x) = P(A_1) + \cdots + P(A_n)$.

Reason: x - # events occur.

$I_{A_1}, I_{A_2} \cdots\cdots I_{A_n}$

If A_1, A_3 occur, the others do not occur.

$x = 2$ I_{A_1} I_{A_3}.

$1 + 0 + 1 + \cdots\cdots + 0 = 2$.

$E(x) = E(I_{A_1}) + E(I_{A_2}) + \cdots + E(I_{A_n}) = P(A_1) + \cdots P(A_n)$.

E.g. Two boys and three girls meet in a row, when a boy is next to a girl, we will call this a meeting point. When 5 kids are seated there may be only one meeting point. (e.g. BBGGG), or maybe as many as 4 meeting points. (GBGBG) Suppose the 5 kids are randomly seated, what is the expected number of meeting points?

Sol: X - denotes the # meeting points.

$E(x) = ?$ find the prob distribution of X.

5 kids: possible # arrangement of 5 kids (2 boys) $= \binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10$.

1 meeting point: BBGGG, GGGBB. $\times 2$

2 : BGGGB, GBBGG, GGBBG. $\times 3$

3 : BGBGG, BGGBG, GBGGB, GGBGB. $\times 4$

4 : GBGBG $\times 1$

$E(x) = E(X_1) + E(X_2) + E(X_3) + E(X_4)$

$= 1 \times \frac{1}{10} + 2 \times \frac{3}{10} + 3 \times \frac{4}{10} + 4 \times \frac{1}{10} = \frac{24}{10} = 2.4$

L10 Exam

L11 10/15

Recall: $E(X) = P(A_1) + P(A_2) + \cdots + P(A_n)$

if X - the # events occur among $A_1, A_2, \cdots A_n$.

E.g. 5 kids in a row. 2 boy 3 girls.

X - # of meeting points.

If 200 boys, 300 girls, what is $E(X)$.

$\binom{500}{200}$ # possible outcomes. $\to$ hard to find probability distribution.

Use method of indicator.

In general, suppose b boys and g girls are seated in a row. $E(X) = ?$

1 2 ... $g+b$.

Define A_1: the 1^{st} and 2^{nd} are of different genders.

A_2: the 2^{nd} and 3^{rd} ---

$\vdots$

A_i: the i^{th} and $(i+1)^{th}$ --- $i = 1, 2, \cdots, g+b-1$

meeting points = # event occur among $\{A_1, \cdots A_{b+g-1}\}$.

$g+b-2$ / $b-1$ boys.

i $i+1$

$X = I_{A_1} + I_{A_2} + \cdots + I_{A_{g+b-1}}$.

$E(X) = P(A_1) + P(A_2) + \cdots P(A_{g+b-1})$.

How to find $P(A_i) = \dfrac{\binom{2}{1}\binom{g+b-2}{b-1}}{\binom{g+b}{b}} = \dfrac{2\times \cancel{(g+b-2)(g+b-3)\cdots}(g)\times(b)\cancel{(b-1)\cdots 1}}{(g+b)(g+b-1)\cancel{\cdots(g+1)}\times\cancel{(b-1)(b-2)\cdots 1}}$

$= \dfrac{2bg}{(g+b)(g+b-1)}$

$E(X) = \sum_{i=1}^{g+b-1} P(A_i) = (g+b-1)\dfrac{2bg}{(g+b)(g+b-1)}$

$= \dfrac{2bg}{g+b}$

200 boys, 300 girls $E(X) = \frac{200 \times 300 \times 2}{200+300} = 240.$

Condition to use methods of indicators to find $E(X)$

- $P(X=x)$ are known but it makes $E(X) = \sum x p(x)$ too complex to simplify or $p(X=x)$ is hard to find
- X can be represented by a "sum of indicators". or X is easy to be intepreted as # events occur among $\{A_1, A_2, \cdots, A_n\}$.

E.g. Suppose in the DNA sequence of length N, each position is randomly selected as T, C, G, A. Let X be the number of matches with these consecutive T's or three consecutive G's. Find $E(X)$.

Sol. X: # of TTT's or GGG's.

$\underbrace{\_\ \_\ \_\ \cdots\ \_\ \_\ \_}_{N}$ possible three consecutive position : $N-2$.

Def. A_i = the i^{th} position and next two position are three T's or G's.

for $i = 1, 2, \cdots, N-2$.

$P(A_i) = 2 \times \left(\frac{1}{4}\right)^3 = \frac{1}{32}.$

X : # of events that occur among $\{A_1, A_2, \cdots A_{N-2}\}$.

$E(X) = \sum_{i=1}^{N-2} P(A_i) = \frac{N-2}{32}$

Expetation of Binormial $(n, p) = np$. $\{0, 1, 2, \cdots, n\}$

Proof: i) $E(X) = \sum_{\text{all } x} x p(x) = \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x \cdot q^{n-x} = np.$

ii) (using methods of indicators).

n trails.

X - # successes among n trails

Define: (A_i) i^{th} trail is a success. $i=1,2,\cdots n$.

$P(A_i) = p$.

X - # events occur among $\{A_1, A_2, \cdots A_n\}$.

$E(X) = \sum_{i=1}^{n} A(i) = np$.

· ~~Tail Sum formula~~ for expectation of a counting random variables.

X $\{0,1,\cdots,n\}$

For a r.v. X with possible values $\{0,1,2,\cdots,n\}$.

$$E(X) = \sum_{j=1}^{n} P(X \geq j)$$

Proof: i) $E(X) = \sum_{x=0}^{n} x P(X=x)$

$$= \cancel{0 \times P(X=0)} + 1 \times \overset{P_1}{P(X=1)} + \cdots + n \cdot \overset{P_n}{P(X=n)}.$$

$$= p_1 + 2p_2 + \cdots + np_n.$$

$$= p_1 + p_2 + \cdots P_n. \rightarrow = P(X \geq 1).$$

$$+ p_2 + p_3 + \cdots P_n$$

$$+ \cdots$$

$$+ P_n$$

$$= P(X \geq 1) + \cdots + P(X \geq n) = \sum_{j=1}^{n} P(X \geq j).$$

ii) X - counting variable $\{0,1,\cdots,n\}$.

Define: $A_j = \{X \geq j\}$, $j=1,\cdots,n$.

X - # events occur among $\{1, A_2 \cdots A_n\}$.

$$E(X) = \sum_{i=1}^{n} P(A_i) = \sum_{j=1}^{n} P(A \geq j)$$

E.g. Suppose 4 dice are rolled, Let m be the minimum of 4 numbers rolled. Find $E(m)$

Sol. $\_\ \_\ \_\ \_ = 6^4$ possible outcome.

$P(m=i)$ possible value : $1 \sim 6$.

$$P(M=3) = \frac{4^4}{6^4}$$

Tail prob $P(M \geq i) = ?$

L12 10/17

$$E(M) = \sum_{j=1}^{6} P(M \geq j)$$

↳ tail sum formula.

$$P(M \geq j) = P(X_1 \geq j) P(X_2 \geq j) P(X_3 \geq j) P(X_4 \geq j)$$

$$= \left(\frac{6-j+1}{6}\right)^4, \quad j = 1, \cdots 6$$

$$E(M) = \sum_{j=1}^{6} \left(\frac{7-j}{6}\right)^4 = \left(\frac{6}{6}\right)^4 + \left(\frac{5}{6}\right)^4 + \left(\frac{4}{6}\right)^4 + \cdots \left(\frac{1}{6}\right)^4$$

$$\approx 1.755.$$

E.g. Let S be the sum of the largest three numbers among the four number in above example, Find $E(S)$.

Sol: $E(T) = 4 \times E(X) = 4 \times 3.5 = 14.$

$E(S) = 14 - 1.755 = 12.245.$

E.g. if X be # of heads in 3 tosses a coin. Find $E(X^2)$

Sol: Is X^2 a random variable?

value of x	0	1	2	3
x^2	0	1	4	9
possibility	1/8	3/8	3/8	1/8.

$$E(x^2) = 0\times\frac{1}{8} + 1\times\frac{3}{8} + 4\times\frac{3}{8} + 9\times\frac{1}{8} = 3.$$

- The expetation of a function of r.v. x $g(x)$ is

$$E(g(x)) = \mu_{g(x)} = \sum_{\text{all } x} g(x)\, p(X=x)$$

Typically $E(g(x)) \neq g(E(x))$

In particular, for $g(x) = x^k$, $k = 1, 2 \cdots$ then

$$E(X^k) = \sum_{\text{all } x} x^k\, p(X=x)$$

$E(X^k) \neq (E(X))^k$. for $k = 2, 3 \cdots$

- If $g(x) = ax + b$, where $a \in \mathbb{R}$, $b \in \mathbb{R}$, then

$$\underset{E(g(x))}{E(ax+b)} = \underset{g(E(x))}{a(E(x)) + b}.$$

If $b = 0$ $E(ax) = aE(x)$.

E.g. A contestant on a quiz show is presented with 2 questions Q_1 and Q_2. The probability that he answers Q_i correctly is P_i. and he will receive \$$V_i$ if answer is correct.

If he decides to try Q_1 first. He will be allowed to go on to the other question only if he answers Q_1 correctly, otherwise the game is over.

Which game should he answer first to max his $E(X)$.

Sol. try Q_1 first. Let X_i be the winnings.

Q_1	Q_2	V_1+V_2	P_1P_2
	Q_2^c	V_1	$P_1(1-P_2)$
Q_1^c	$\times$	0	$1-P_1$

$$E(X_1) = (V_1+V_2)P_1P_2 + V_1P_1(1-P_2) = V_2P_1P_2 + V_1P_1$$

$$E(X_2) = (V_2+V_1)P_2P_1 + V_2P_2(1-P_1) = V_1P_1P_2 + V_2P_2$$

if $E(X_1) \geq E(X_2)$ try Q_1 first, otherwise try Q_2.

Eg. A and B play the following game: A writes down either number 1 or number 2 and B must guess which one. If B guesses correctly, B will receive \$1. Otherwise B will pay \$3/4 to A. Suppose B randomizes his decision by guessing 1 with probability P and 2 with probability $(1-P)$.

a) Determine B's expected gain if A has written down #1.

b) - - - - - - - - - - - - - - - #2.

c) what value of p maximize the minimum possible value of B's expected gain? what is the value.

Sol. a) Let X_i: the gain of B if A write down #i.

if B guesses #1: $X_1 = \$1$ (P)

#2: $X_1 = \$-\frac{3}{4}$ $(1-P)$.

$$E(X_1) = P - \frac{3}{4} + \frac{3}{4}P = \frac{7}{4}P - \frac{3}{4}$$

b). $E(X_2) = 1(1-p) + (-\frac{3}{4}P) = 1 - \frac{7}{4}P$.

c) when $E(X_1) = E(X_2)$

$$\frac{7}{4}p - \frac{3}{4} = 1 - \frac{7}{4}p \quad p = \frac{1}{2}.$$

$$E(X_1) = E(X_2) = \frac{1}{8}$$

Independent random variables.

X, Y are independent, intuitively: $P(X=x)$ for some values of X is not affected by $P(X=x \mid Y=y)$.

Def. Random Variables X and Y are independent if.

$$P(X=x, Y=y) = P(X=x)\,P(Y=y). \quad \forall x, y.$$

$$P(X \in A, Y \in B) = P(X \in A)\,P(Y \in B) \quad \forall A, B.$$

· Multiplication rule for expetation.

If X, Y are independent, then

$$E(XY) = E(X)\,E(Y).$$

Proof: $$E(XY) = \sum_x \sum_y (xy)\,P(X=x, Y=y)$$
$$= \sum_x x\,P(X=x) \sum_y y\,P(Y=y)$$
$$= E(X)\,E(Y).$$

E.g. Roll 2 dice, what is the expetation of product of 2 points

$$E(X) = E(X_1 X_2)$$
$$= 3.5^2.$$