CSE 312

SECTION 5

ZOO OF RANDOM VARIABLES

– *Welcome back, everyone!* –

AGENDA

01 ANNOUNCEMENTS

02 Independent RV

03 Zoo of Random Variables

04 Practice Problems

01

ANNOUNCEMENTS

SCHEDULE REMINDERS

PSET3 GRADES WERE RELEASED

(regrade requests open and close after a week)

PSET4 WAS RELEASED

due next Wednesday

02

LOE REMINDER

LOE

When working with linearity of expectation, remember to

first define the RVs and the summation relationships
don't worry how the individual RVs are distributed

then apply linearity of expectation and find each value

02

VARIANCE

Variance is a another property of RVs (like expectation) that measures how much the values in the RV “vary”

VARIANCE - how “different” are values from the expectation “on average”

$$Var(X) = E[(X-E(X))^2] = \Sigma_x (P(X=x) * (x-E(X))^2)$$

expected value of the squared distance between each RV outcome and the expected value of RV

add up all the squared distances weighted by their probabilities

Properties

$$Var(a \cdot X + b) = a^2 \cdot Var(X)$$

$$Var(X) = E[X^2] - (E[X])^2$$

02

Independent RV

What does independence mean for random variables?

Random variables X and Y are **independent** if –

$$P(X=x, Y=y) = P(X=x) \cdot P(Y=y)$$

Knowing the value of X doesn't help "guess" what Y is

it's a useful property! if X and Y are independent random variables then —

$E(X \cdot Y) = E[X] \cdot E[Y]$

$Var(X + Y) = Var[X] + Var[Y]$

Random variables X and Y are **independent** if –

$$P(X=x,\ Y=y) = P(X=x) \cdot P(Y=y)$$

Knowing the value of X doesn't help "guess" what Y is

Additionally, there's **independent and identically distributed (aka, "i.i.d.")** random variables

Identically distributed means the random variables **have the same pmf –**

$$P(X=k) = P(Y=k)$$ for any value k

For example, rolling a die twice, where X is the first roll number and Y is the second roll number

03

ZOO OF RV'S

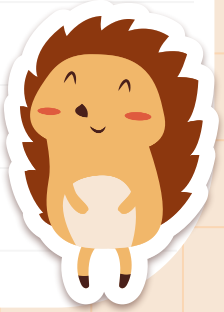
zoo of discrete random variables!

Zoo of Discrete Random Variables

Random variables allow us to represent different random experiments/situations

We've seen how tedious computing pmfs, expectations, and variances can be.

There are some *common situations* that call for a random variable that is too complex to analyze, so we derive pmfs, expectations, and variance for this *"zoo" of RVs*.



UNIFORM

EACH OUTCOME IS EQUALLY LIKELY

X ~ Uniform(a, b) if X is equally likely
to take on any value between a and b

$$p_X(k) = \frac{1}{b-a+1} \qquad \mathbb{E}[X] = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

A random variable X representing the outcome of rolling a fair 6 sided dice

X~Uniform(1, 6)

choosing a random value between 1 and 6 with each outcome equally likely

Bernoulli (Indicator)

RV is either 0 or 1 (i.e. success or not)

X ~ Bernoulli(p) if X is 1 with probability of p

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases} \qquad \mathbb{E}[X] = p \qquad \mathrm{Var}(X) = p(1-p)$$

X represents whether outcome of rolling a fair 6 sided dice is even (1) or not (0)

X~Bernoulli(3/6)

probability of 3/6 for "success"

Binomial

count # of times event occurs in n tries

X ~ Binomial(n, p) means X represents the number of times an event with probability p happens in n trials

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \mathbb{E}[X] = np \qquad \mathrm{Var}(X) = np(1-p)$$

X represents the number of times the dice rolled to a 6 during 9 dice rolls

X~Binomial(9,⅙)

probability of success (rolling a 6) on a single dice roll is ⅙, and 9 trials (rolls)

Geometric

Count # of trials until event occurs

X ~ Geometric(p) means X represents the number of trials before success (an event with probability p happens)

$$p_X(k) = (1-p)^{k-1}p \qquad \mathbb{E}[X] = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

X represents the number of times we roll a 6 sided die, before it rolls a 6

X~Geometric(⅙)

on a single dice roll, there's a probability of ⅙ for success (that it rolls a 6)

Poisson

How many successes in a unit of time

X ~ Poisson(λ) means X represents the number of success in a unit of time, where λ is average rate of successes per unit of time

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad \mathbb{E}[X] = \lambda \qquad \mathrm{Var}(X) = \lambda$$

X represents number of people born during a particular minute

X~Poisson(λ)

where λ represents the average birth rate per minute

negative binomial

(related to geometric)

models situations where we count # trials to get some number of successes

X ~ NegBin(r,p) means X represents the number of trials to get r successes (probability of success on a single trial is p)

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad \mathbb{E}[X] = \frac{r}{p} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

X represents number of dice rolls before we get 4 rolls with a 6

X~NegBin(4, 1/6)

because we want to have 4 trials (rolls) with success, and a success (rolling 6) has probability 1/6

HYPERGEOMETRIC

Models situations with *choosing* - how many "successes" do you get when choosing without replacement

X ~ HypGeo(N,K,n) means X represents the number of successes out of n draws from N items with K successes

Number of ways you can choose n items with k successes

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \qquad \mathbb{E}[X] = n\frac{K}{N} \qquad Var(X) = n \cdot \frac{K(N-K)(N-n)}{N^2(2N-1)}$$

Number of ways you can choose n items from N

X represents number of Kit-Kats we will get when drawing 30 candies from a bowl of 100 candies that contain 10 Kit-Kats

X~HypGeo(100, 10, 30)

because we draw 30 from 100 items with 10 successes (Kit-Kats)

LET'S TRY IT!

PROBLEM 1: CONTENT REVIEW

TRUE OR FALSE

$Var(X + Y) = Var(X) + Var(Y)$ for all random variables X and Y.

False. This property only hold if X and Y are independent.

What is $\mathrm{Var}(3X + 4)$?

- ○ $3\mathrm{Var}(X) + 4$
- ○ $3\mathrm{Var}(X)$
- ○ $9\mathrm{Var}(X)$
- ○ $\mathrm{Var}(X)$

$9\mathrm{Var}(X)$ by the scaling property of variance that states:

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

TRUE OR FALSE

$E[X + Y] = E[X] + E[Y]$ for all random variables X and Y.

True. This is by the linearity of expectation. X and Y do not have to be independent.

What is $\mathbb{E}[3X + 4]$?

- ○ $3\mathbb{E}[X] + 4$
- ○ $3\mathbb{E}[X]$
- ○ $9\mathbb{E}[X]$
- ○ $\mathbb{E}[X]$

$3E[X] + 4$ by the linearity of expectation.

Problem 2: Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B+R+G=N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

a) How many of the next 10 fish I catch are blue, if I catch and release them

- ○ $$\text{Bin}\left(10, \frac{B}{N}\right)$$
- ○ $$\text{Ber}\left(\frac{B}{N}\right)$$
- ○ $$\text{Bin}\left(1, \frac{B}{N}\right)$$

Problem 2: Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B+R+G=N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

b) How many fish I had to catch until my first green fish, if I catch and release them

- ○ $\text{Ber}\left(\frac{G}{N}\right)$
- ○ $\text{Bin}\left(1, \frac{G}{N}\right)$
- ○ $\text{Geo}\left(\frac{G}{N}\right)$

Problem 2: Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B+R+G=N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

c) How many red fish I catch in the next five minutes, if I catch on average r red fish per minute

- ○ $\text{Poi}(5R)$
- ○ $\text{Bin}\left(5, \frac{R}{N}\right)$
- ○ $\text{Poi}(5r)$

Problem 2: Pond Fishing

Suppose I am fishing in a pond with B blue fish, R red fish, and G green fish, where $B+R+G=N$. Each fish is equally likely to be caught. For each of the following scenarios, identify the most appropriate distribution (with parameter(s)):

d) Whether or not my next fish is blue

- ○ $\text{Poi}(5B)$
- ○ $\text{Bin}\left(1, \frac{R}{N}\right)$
- ○ $\text{Ber}\left(\frac{B}{N}\right)$

Problem 4: Variance of a Product

Let X, Y, Z be independent random variables with means μ_X, μ_Y, μ_Z and variances $\sigma_X^2, \sigma_Y^2, \sigma_Z^2$, respectively. Find $\mathrm{Var}(XY - Z)$.

PROBLEM 5:
TRUE OR FALSE?

TRUE OR FALSE

a) For any random variable X, we have $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.

True, using the fact that variance must be non-negative.

TRUE OR FALSE

b) Let X, Y be random variables. Then, X and Y are independent if and only if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

False. The forward implication is true but the reverse is not.

TRUE OR FALSE

c) Let $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ be independent. Then, $X + Y \sim \text{Binomial}(n + m, p)$.

True.

TRUE OR FALSE

d) Let $X_1, \ldots, X_{n+1}$ be independent $\mathsf{Bernoulli}(p)$ random variables. Then, $\mathbb{E}[\sum_{i=1}^{n} X_i X_{i+1}] = np^2$.

True.

TRUE OR FALSE

e) Let $X_1, ..., X_{n+1}$ be independent $\mathsf{Bernoulli}(p)$ random variables. Then, $Y = \sum_{i=1}^{n} X_i X_{i+1} \sim \mathsf{Binomial}(n, p^2)$.

False.

TRUE OR FALSE

f) If $X \sim \text{Bernoulli}(p)$, then $nX \sim \text{Binomial}(n, p)$.

False.

TRUE OR FALSE

g) If $X \sim \text{Binomial}(n, p)$, then $\frac{X}{n} \sim \text{Bernoulli}(p)$.

False.

TRUE OR FALSE

h) For any two independent random variables X, Y, we have $\mathrm{Var}(X - Y) = \mathrm{Var}(X) - \mathrm{Var}(Y)$.

False.

EXTRA: PROBLEM 3 BEST COACH EVER!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

a) How many matches do you expect to fight until you win 10 times and what kind of random variable is this?

EXTRA: PROBLEM 3 BEST COACH EVER!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

b) You only get to play 12 matches every year. To win a spot in the Annual Boxing Championship, a boxer needs to win at least 10 matches in a year. What is the probability that you will go to the Championship this year and what kind of random variable is the number of matches you win out of the 12?

EXTRA: PROBLEM 3 BEST COACH EVER!!

You are a hardworking boxer. Your coach tells you that the probability of your winning a boxing match is 0.2 independently of every other match.

c) Let p be your answer to part (b). How many times can you expect to go to the Championship in your 20 year career?

Section AA

THANKS!