

CSE 312

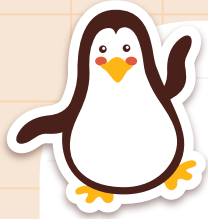


SECTION 5

ZOO OF RANDOM VARIABLES

– Welcome back, everyone! –





AGENDA

- 01 ANNOUNCEMENTS
- 02 VARIANCE
- 03 INDEPENDENT RV
- 04 ZOO OF RANDOM VARIABLES

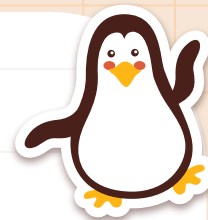




01

ANNOUNCEMENTS





SCHEDULE REMINDERS

PSET3 GRADES WERE RELEASED

(regrade requests open and close after a week)

PSET4 DUE YESTERDAY

PSET5 WAS RELEASED

Coding: task 7 due Feb 9th

Other tasks due Feb 7th



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”

$$\text{Var}(X) = E[(X - E(X))^2] = \sum_x (P(X=x) * (x - E(X))^2)$$

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

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

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



03

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 –

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

$$\text{Var}(X + Y) = \text{Var}[X] + \text{Var}[Y]$$

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** –

$$P(X=k) = P(Y=k) \quad \text{for any value } k$$

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



04

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

MODELS SITUATIONS WHERE EACH
OUTCOME IS EQUALLY LIKELY

$X \sim \text{Uniform}(a, b)$ if X is equally likely
to take on any value between a and b


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

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

$X \sim \text{Uniform}(1, 6)$


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



BERNOULLI (INDICATOR)

MODELS SITUATIONS WHERE THE RV CAN
TAKE ON 0 OR 1 (WHETHER SUCCESS OR NOT)

$X \sim \text{Bernoulli}(p)$ if X is 1 with
probability of p


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

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

$X \sim \text{Bernoulli}(3/6)$

probability of 3/6 for "success"



BINOMIAL

**MODELS SITUATIONS WHEN WE COUNT THE
TIMES AN EVENT OCCURS IN n TRIES**

$X \sim \text{Binomial}(n, p)$ means X represents the number of times an event with probability p happens after n trials

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

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

$X \sim \text{Binomial}(\frac{1}{6}, 9)$


probability of success (rolling a 6) on a single dice roll is $\frac{1}{6}$, and 9 trials (rolls)



Geometric

**MODELS SITUATIONS WHEN WE COUNT
THE # TRIALS UNTIL SOME EVENT OCCURS**

$X \sim \text{Geometric}(p)$ means X represents the number of trials before success (an event with probability p happens)


$$p_X(k) = (1 - p)^{k-1} p,$$

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1-p}{p^2}$$

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

$X \sim \text{Geometric}(\frac{1}{6})$

on a single dice roll, there's a probability of $\frac{1}{6}$ for success (that it rolls a 6)



NEGATIVE BINOMIAL

(RELATED TO GEOMETRIC)

MODELS SITUATIONS WHERE WE COUNT #
TRIALS TO GET SOME NUMBER OF SUCCESSSES

$X \sim \text{NegBin}(r,p)$ means X represents the number of trials to get r successes (probability of success on a single trial is p)


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

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

$X \sim \text{NegBin}(4, 1/6)$


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



POISSON

MODELS SITUATIONS WITH *TIME* - HOW MANY SUCCESSSES IN A UNIT OF TIME

$X \sim \text{Poisson}(\lambda)$ means X represents the number of success in a unit of time, where λ is average rate of successes per unit of time


$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\text{Var}(X) = \lambda$$

X represents number of people born during a particular minute

$X \sim \text{Poisson}(\lambda)$

where λ represents the average birth rate per minute



HYPERGEOMETRIC

- **MODELS SITUATIONS WITH CHOOSING - HOW MANY “SUCCESSES” DO YOU GET WHEN CHOOSING WITHOUT REPLACEMENT**

Number of ways you can choose n items with k successes

$$p_X(k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

$X \sim \text{HypGeo}(N, K, n)$ means X represents the number of successes out of n draws from N items with K successes

$$\mathbb{E}[X] = n \frac{K}{N} \quad \text{Var}(X) = n \cdot \frac{K(N-K)(N-n)}{N^2(2N-1)}$$

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

$X \sim \text{HypGeo}(100, 10, 30)$

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





LET'S TRY IT!

Let's identify these distributions in
some real examples!

THANKS!

