CS 2810: Mathematics of Data Models, Section 1

Spring 2022 — Felix Muzny

Probability Mass Functions and distributions

What is the expected value of a weighted die that has a uniform chance of rolling a 1 - 5 and a 40% chance of rolling a 6?

$E[X]$ = avg. value after infinite trials/experiments

$= \sum$ prob of the outcome * value of the outcome

$= .12(1) + .12(2) + \ldots + .12(5) + .4(6) = 4.2$

Law of Large Numbers

- The **law of large numbers** says that as we get closer to infinite trials from now to the end of time, the average value of the trials will approach the **expected value**.

experimental avg

- Example: A six-sided die has an expected value of 3.5

# exps.	experimental avg
1	1
5	2.4
100	3.45
1000	3.52
10 000	3.509

in general

1) more python
2) less/same python

Independent Variables

- Say that we have two independent random variables: a 6-sided die and a 4-sided die. We can write their probability distributions like....

X

Y

X

1	1/6
2	
3	
⋮	

Y

1	1/4
2	
3	
4	

- What is the probability of rolling a 9 with these two dice?

$$P(X+Y=9) = P(X=5, Y=4) + P(X=6, Y=3)$$

$$= (1/6 * 1/4) + (1/6 * 1/4)$$

Dependent Variables

- Say that we have two **dependent** random variables: the number of donuts that Felix eats in a given morning and the time of day that they need to take a nap....

X

Y

	donuts = 0	donuts = 1	donuts = 2
nap = 12pm	0	0	0.4
nap = 1pm	0	0.3	0
nap = 2pm	0	0.1	0
nap = 3pm	0.2	0	0

Dependent Variables

- Say that we have two **dependent** random variables: the number of donuts that Felix eats in a given morning and the time of day that they need to take a nap....

- Joint probability distributions for random variables have **marginal distributions**—a fancy way of talking about the probability distribution for each individual random variable

↳ prob. that Felix eats 2 donuts

↳ prob. of 12 pm nap

Dependent Variables

- Joint probability distributions for random variables have **marginal distributions**—a fancy way of talking about the probability distribution for each individual random variable

X

Y

	donuts = 0	donuts = 1	donuts = 2
nap = 12pm	0	0	0.4
nap = 1pm	0	0.3	0
nap = 2pm	0	0.1	0
nap = 3pm	0.2	0	0

$P(X=0) = 0.2$

$P(X=1) = 0.3 + 0.1$

$P(X=2) = 0.4$

$P(Y=12pm) = 0.4$

$P(Y=1pm) = 0.3$

$\vdots$

sum is 1

Dependent Variables

- Say that we have two **dependent** random variables: the number of donuts that Felix eats in a given morning the number of donuts that Felix eats in a given afternoon

X morning

Y

	donuts = 0	donuts = 1	donuts = 2
donuts = 0	0.5	0	0.05
donuts = 1	0	0.25	0
donuts = 2	0.2	0	0

Expected Value - ICA Question 1

What is the expected value of the total number of donuts that Felix eats in a given day?

$E[X] = P(x=0) = 0.5 \quad P(X=3) = 0$

$P(x=1) = 0+0 \quad P(x=4) = 0$

$P(x=2) = 0.2 + 0.25 + 0.05$

$0.5(0) + 0.5(2)$

$= \boxed{1}$

What about the variance?

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

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

an alternate formula that you might find convenient at times

	donuts = 0	donuts = 1	donuts = 2
donuts = 0	0.5	0	0.05
donuts = 1	0	0.25	0
donuts = 2	0.2	0	0

Distributions

- We've already talked about **probability distributions**—we can think of these as a set of numbers that sum to one whose job is to characterize the chances of a set of outcomes....

- the prob. distribution of tomorrow's weather: {cold: 0.6, warm: 0.1, cool: 0.3}

- the prob. distribution of a fair coin: {H: 0.5, T: 0.5}

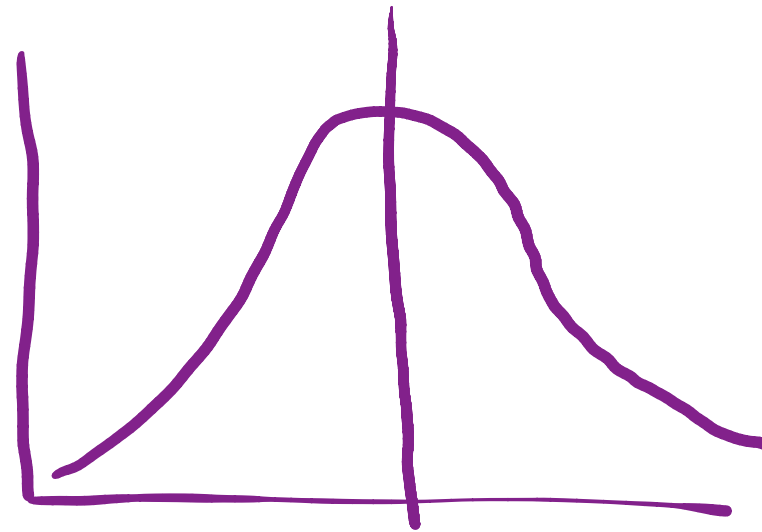
  ↳ uniform

- the prob. distribution of an unfair coin: {H: 0.51, T: 0.49}

Dist. graphs

uniform

height



normal
(next Mon)

Distributions

- It's helpful for us to categorize the kinds of probability distributions our random variables are associated with because that tells us what kind of math we can do with them.

- All distributions have a **probability mass function**. This tells us how the mass is distributed across outcomes.

Uniform distributions

- All distributions have a **probability mass function**. This tells us how the mass is distributed.

- Example: a **uniform distribution** has the PMF of $f(k; n) = P(X = k) = \dfrac{1}{n}$ where k is one of the possible outcomes and n is the number of outcomes

a fair 6-sided die: $f(1;6) = P(X=1) = \frac{1}{6}$

$f(\underset{4}{3}; 57) = P(X=3) = \frac{1}{57}$

this is a valid outcome

Binomial distributions

- Turtle has two moods:
- 
- 

Binomial distributions

- On every day, turtle has a .85 chance of being happy (a natural optimist) and a <u>0.15</u> chance of being sad.

  1 − 0.85

- What are the chances of turtle being happy today? 0.85

- What are the chances of turtle being happy today and tomorrow?

  0.85 * 0.85 = 0.72

- What are the chances of turtle being happy today and tomorrow and tomorrow's tomorrow?

  0.85 * 0.85 * 0.85 = 0.614

- What are the chances of turtle being happy exactly 6 of the next 10 days?

  ↳ tricky

Expected Value - ICA Question 2

What is the probability of Turtle only being happy for the next 10 days?

$$0.85^{10} = 0.197$$

What is the probability of Turtle being happy for the first **k** of the next **n** days?

(and is unhappy for the rest of the days)

$$0.85^{k} * (0.15)^{n-k}$$

to think about: prob of being happy for 6 (k) days of 10? (n)

Binomial distributions

- A **binomial distribution** is for random variables that have **two** outcomes

  ↳ turtle, coin, happening or not

- We can write equations for:
  - The probability of outcomes in a specific order
  - The probability of outcomes in any order
  - The probability of "at least" x outcomes

Binomial distributions

- A **binomial distribution** is for random variables that have **two** outcomes
- $P(k; n, p) = P(X = k) = \underline{\binom{n}{k} p^k (1-p)^{n-k}}$ the same as what we were doing
- Where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ "n choose k"
- k is the number of "successes" – Happy
- n is the number of "trials" – days – 2
- p is the probability of a "success" – 0.85

$p(1; 2, 0.85) =$

$\binom{2}{1} 0.85^1 (1-0.85)^1$

enumerate my combos
H,S or S,H

$$P(1; 2, 0.85)$$

$$= \frac{2!}{1!(2-1)!} \underbrace{0.85(1-0.85)}$$

$$= \frac{2 * 1}{1(1)}$$

$$= 2 (0.85)(0.15)$$

$$P(3; 5, 0.85)$$

$$\binom{5}{3} = \frac{5 * 4 * 3 * 2 * 1}{3!(5-3)!} = \frac{120}{3 * 2(2)} = \frac{120}{12}$$

Binomial distributions - ICA Question 3

What is the probability that turtle is happy for **exactly** 6 of the next ten days?

$$f(k;n,p) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$k \to 6$, $n \to 10$, $p \to 0.85$

$$= \binom{10}{6} 0.85^6 (1-0.85)^4$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial distributions - ICA Question 4

What is the probability that turtle is happy for **at least** 6 of the next ten days?

$$f(k;n,p) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\sum_{i=6}^{10} f(i;10,0.85)$$

Binomial

- two outcomes
- measure prob of a sequence of outcomes
  ↳ independent of order

Break: 12:54

python detour: 12:54 – 1 pm

Poisson Distributions

- Now, let's model turkeys in Boston.

- We want to know the number of turkey delays that will afflict the green line between Coolidge Corner and the St. Mary's stop on a given trip.

- This isn't a binomial distribution because while a turkey delay is a discrete yes/no event, we're asking **how many times** in a given time interval rather than "what is the probability of having a turkey delay occurring".

Poisson Distributions

- We want to know the number of turkey delays that will afflict the green line between Coolidge Corner and the St. Mary's stop on a given trip.

- Poisson distributions use the formula: $f(k;\lambda) = P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

- k is the number of occurrences (# of turkey delays)

- e is Euler's number (2.71828)

- λ is the average number of events $\underbrace{(\lambda = E[X] = Var(X))}_{\text{a feature of poisson distr.}}$

Poisson Distributions

- Poisson distributions use the formula: $f(k; \lambda) = P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

- We want to know the number of turkey delays that will afflict the green line between Coolidge Corner and the St. Mary's stop on a given trip. Assume that one turkey delay happens every twenty minutes (on average).

- What is the probability that we encounter 1 turkey in a one minute journey?

$$\lambda = \frac{1 \text{ turkey}}{20 \text{ min}} = 0.05 \text{ turkey/min}$$

$$P(X=1) = \frac{0.05^1 (e^{-0.05})}{1!} = 0.04756$$

Poisson Distributions

one time unit

- Poisson distributions use the formula: $f(k;\lambda) = P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

- We want to know the number of turkey delays that will afflict the green line between Coolidge Corner and the St. Mary's stop on a given trip. Assume that one turkey delay happens every twenty minutes (on average).

$$\frac{1}{20} = \frac{x}{30} \rightarrow 1.5 \text{ turkeys}/30 \text{ min}$$

- What is the probability that we encounter 1 turkey in a thirty minute journey?

$$f(1;1.5) = \frac{1.5^1(e^{-1.5})}{1!} = 0.335$$

Poisson Distributions

- Poisson distributions use the formula: $f(k;\lambda) = P(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

- We want to know the number of turkey delays that will afflict the green line between Coolidge Corner and the St. Mary's stop on a given trip. Assume that one turkey delay happens every twenty minutes (on average).

$\frac{1}{20} = \frac{x}{30}$ 1.5 turkey/30 min $\frac{1}{20} = \frac{x}{1} \rightarrow 0.05$

- What is the probability that we encounter 4 turkeys in a thirty minute journey?

$$f(4;1.5) = \frac{1.5^4 e^{-1.5}}{4!} = 0.05$$

Admin

- You have a test on Thursday!
    - Our TA Divya will be here conducting the test for you all with a team of TAs. She will be in contact with me the whole time. Listen to her :)
- HW 5 will be released later today/tomorrow (Prof. Higger and I are currently finalizing it).
    - When we release it, I'll make an announcement with the due date.

Schedule

Test 1 grades were released this am

Turn in ICA 13 on **Canvas (not on Gradescope)**

TEST 2 is in class on Thursday!

Send me an email if you're feeling overwhelmed! (I know that there's a lot of work in this class, we will work with you to make sure that you don't fall behind)

Mon	Tue	Wed	Thu	Fri	Sat	Sun
February 28th Lecture 13 - law of large numbers, distributions	**Felix OH Calendly**		**TEST 2 IN CLASS**			
Lecture 14 - estimators, bias **HW 5 due @ 11:59pm*** *** most likely due date**	**Felix OH Calendly**	**Felix OH Calendly**	Lecture 15 - finish topics needed for HW 6, HW 6 work day (yes, you will get ICA credit for this day) **Felix OH Calendly**			

More recommended resources on these topics

- YouTube: 3Blue1Brown Binomial distributions | Probabilities of probabilities, part 1
- YouTube: An Introduction to the Poisson Distribution | jbstatistics
- Wikipedia binomial example: https://en.wikipedia.org/wiki/Binomial_distribution#Example
- Wikipedia poisson example: https://en.wikipedia.org/wiki/Poisson_distribution#Example
- Wikipedia uniform distribution: https://en.wikipedia.org/wiki/Discrete_uniform_distribution