Lecture 6

Continuous Random Variables

Today's Updates / Reminders

- Lab 2 is due Friday – come see me or Makenna for help!
- Homework 3 is due, end of the day, next time we meet.
- Exam 1 is about 3 weeks away – Lectures 1-9 – more info on this coming soon
- Stat Hub is now open for business. Free help/tutoring. SAS1101, M-Th, 5:30-8:00pm

Cumulative Binomial

- When the problem says, "What is the probability of **exactly** x successes?" we use the previous formula.
- What if we want to know the probability of **greater than or equal to** x successes?
- We add up all the probabilities from x to n.
- If it said **less than** x, we would count from 0 to x - 1.
- Recall most calculators do the probability of ≤ x.

Practice

- You are taking a 10-question test with just True/False answers. You have no idea what any of the right answers are so you guess. You need at least a 60% to pass. What is the probability that you pass?

- This year's Super Bowl had a 43.5 Nielsen rating, meaning 43.5% of households watched it. There are 12 houses on your street. What is the probability that at least one household had the game on?

- The sun shines 300 days a year in Phoenix. What is the probability that there are 52 non-sunny days in Phoenix this year?

Example

- Airlines oversell flights. If the daily 3:15pm flight from RDU to DFW has a historic track record where 95% of passengers actually show up to the flight, and the plane seats 88, what is the probability that they will oversell the flight if they sell 91 tickets?

Mean and Standard Deviation

- Recall we use the terms "mean," μ, and "expected value," E(X), interchangeably.

$$E(X) = \mu = np$$

- We also use V(X) or σ^2 for variance.

$$V(X) = \sigma^2 = npq$$

- And standard deviation is the square root of variance.

$$\sigma_X = \sqrt{npq}$$

Example

- One in 20 people have a food allergy. Suppose I have a classroom with 90 students in it. What is the expected value of students with allergies? What is the standard deviation?
- If you took the mean $\pm$ standard deviation, what is the range of the number of students?

Example

- When you donate blood, they draw samples to test for various diseases. Pooling samples is a way to shorten screening time and reduce the number of tests needed and was used during the height of the COVID pandemic. Let's say the incidence of COVID in the population is 5%. If we pool samples in groups of 8, what is the probability that the pooled sample would test positive?

Chapter 3.6

Poisson Probability Distribution

Definition

- A discrete random variable X is said to have a Poisson distribution with parameter μ (μ>0) if the pmf of X is:

$$p(x;\ \mu) = \frac{e^{-\mu} \cdot \mu^{x}}{x!}, where\ x = 0, 1, 2, 3, \ldots$$

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391	0.047553	0.049041	0.049787
1	0.141304	0.147070	0.148609	0.149361
2	0.227656	0.225153	0.224414	0.224042
3	0.236088	0.227474	0.225170	0.224042
4	0.177066	0.170606	0.168877	0.168031
5	0.102305	0.101308	0.100985	0.100819
6	0.047363	0.049610	0.050153	0.050409
7	0.018043	0.020604	0.021277	0.021604
8	0.005764	0.007408	0.007871	0.008102
9	0.001565	0.002342	0.002580	0.002701
10	0.000365	0.000659	0.000758	0.000810

Comparing the Poisson and Three Binomial Distributions

The Mean and Variance of X

- Since binomial approaches the Poisson as $n \rightarrow \infty$, $p \rightarrow 0$, $np \rightarrow \mu$, the mean and variance of a binomial variable should approach those of a Poisson variable. These limits are $np \rightarrow \mu$ and $npq \rightarrow \mu$.
- Therefore, if X has a Poisson distribution with parameter μ, then:

$$E(X) = V(X) = \mu$$

Explaining Terms

- Other versions of the formula use λ instead of μ. And it is often called the rate.
- Also binomial doesn't model events that occur at the same time. The successes occur one after the other.
- When you have a large number of events with a low probability, then the number of events that occur in a fixed time interval follows a Poisson distribution.
- Poisson can be used to describe the probability of a number of events that occur at a specific rate, λ, and within a specific time interval.

Example

- Suppose a restaurant receives an average of 10 customers per hour. We wish to know the probability of exactly 15 customers coming.
  - Note why binomial won't apply here.
- So this means μ = 10 and x = 15.

- What about the probability of greater than or equal to 15? We must use tables. Let's go to the textbook in WebAssign.

Let's try Matlab…
poisspdf
poisscdf

Traffic Engineering

- You are designing an efficient traffic signal system. The number of vehicles arriving at an intersection in a given time interval can be modeled using a Poisson distribution.
- Terms:
  - X = the number of vehicles arriving at the intersection per unit of time, let's say per minute.
  - λ = the average rate (same as μ), based on historical data or a preliminary study. Let's say the average is 2.5 vehicles per minute. So $\lambda = 2.5$.
  - Key assumption: the arrival of vehicles are independent of each other. That is, one car's arrival doesn't influence when the next car will arrive.
- What is the probability that exactly 3 cars will arrive in the next minute?

- What is the probability that more than 5 cars will arrive in the next minute?

Challenge Problem

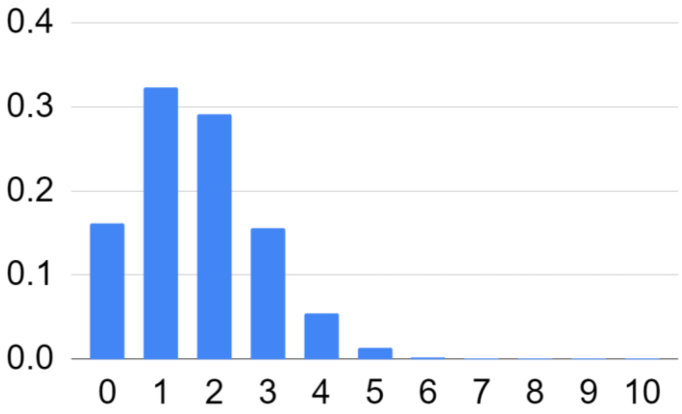
- An automatic car wash takes exactly 5 minutes to wash a car. On average, 10 cars per hour arrive at the car wash. Suppose that 1 hour before closing time, 10 cars are in line. If a new car can be washed as soon as the previous car is finished, is it likely anyone will be in line at closing time?

Chapter 4.1 and 4.2

Probability Density Functions, Cumulative Distribution Functions, Expected Value

Binomial Probability Review

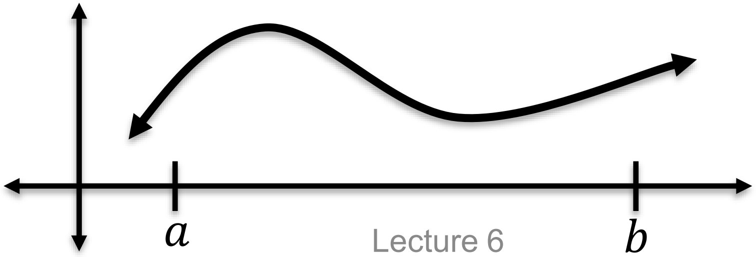
- If you roll a die 10 times, what is the probability of rolling a 4? We did this last class. It produces this distribution. The y-axis corresponds to the probability for each discrete event.
- To find cumulative, you sum the heights of bars or use technology.



Google Sheets Example

Distribution of Continuous Variables

- Continuous variables can land anywhere on the number line. In other words, they can be decimal or fractions.
- When we speak of the distribution of these variables, we call it a probability density function (pdf).
- For notation, we say $P(a \leq X \leq b)$. This represents the area under the curve between a and b.



Distribution of Continuous Variables

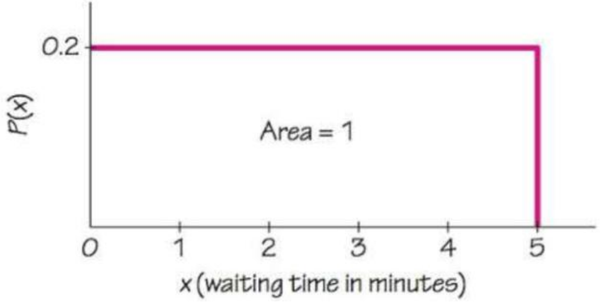
- Continuous variables do not have a probability **distribution** function, but rather a probability **density** function, denoted f(x).
- The y-axis is NOT probability. The value is determined so that the area under the curve equals 1.
- Probabilities are represented by area under the curve. The area under the entire curve equals 1. Remember, $0 \leq \sum P(x) \leq 1$
- $\int_{-\infty}^{\infty} f(x) = 1$ which is the calculus way of saying area = 1.
- $P(x = a) = 0$ or said another way, there MUST be an interval
- $P(a \leq x \leq b)$ is the same as saying $P(a < x < b)$.

Discrete vs. Continuous

- Recall that for discrete, we could find the pmf or the cdf.
  - What is the probability of exactly 3 heads on 10 flips? (pmf)
  - What is the probability of more than 3 heads on 10 flips? (cdf)
- For continuous
  - What is the probability that someone in this classroom is 5.1483 feet tall? (pmf) – **the answer here is 0**.
  - What is the probability that someone in this classroom is less than 5.1483 feet tall? (cdf)

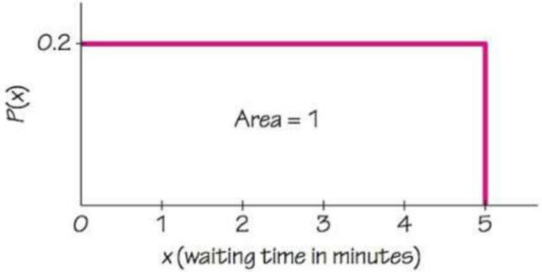
Uniform Distribution

- Values are spread evenly over the range of possibilities. The graph of a uniform distribution results in a rectangular shape.
- For example, if the Wolfline passes by your stop every 5 minutes, what is the probability of the various wait times, between 0 and 5 minutes?

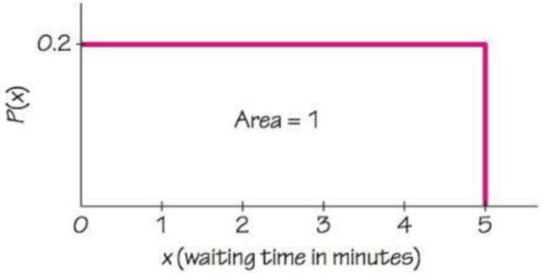


Uniform Distribution

- Find the probability of waiting more than 4 minutes.



- Find the probability of waiting $1 < x \leq 3$ minutes.



Example

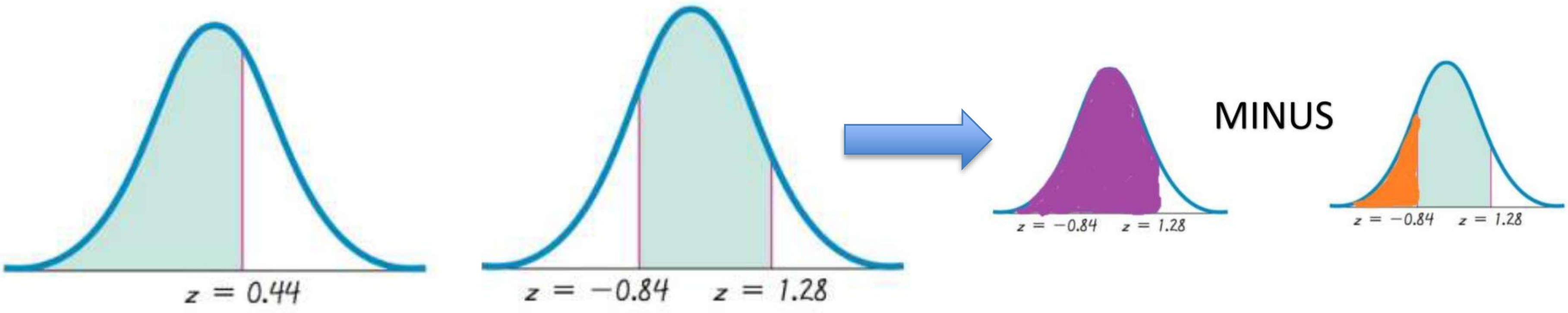
- Suppose the reaction temperature (in Celsius) in a certain chemical process has a uniform distribution between 3.5 and 4.0 degrees C. Draw and label the axes of the uniform distribution and find $P(3.57 < X \leq 3.91)$.

Experiment Time

- Open this Sheets file. If you don't have a laptop or tablet to click this link, point at the QR code.
- Find your name on the roster by last name and first initial
- Roll your die 5 times and record the number you roll each time.

Normal Distribution

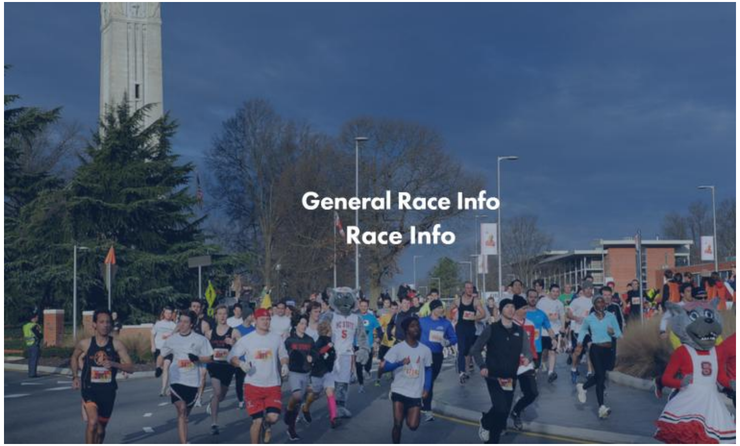
- In this class, we will complete this type of exercise frequently… finding the area under the normal curve. There is no closed form integral of the normal probability function. We must use numerical methods (series). For the normal curve, we will use technology or software. Sheets, MATLAB, R, etc., …. More on this next lecture…



Example

- The Krispy Kreme Challenge is an awesome NC State tradition. Try it in February (a week from Saturday!) You run 2.5 miles from the Belltower to Krispy Kreme on Peace St., eat a dozen doughnuts, and run back to the Belltower. Getting to KK is no problem. It's the getting back that's the problem. Trash cans are stationed throughout the course. The organizers are trying to decide how to distribute the trash cans. They have collected data on when participants are most likely to "fall ill" and need your help to assess the probability of a sick participant needing a trash can. The probability can be modeled using a quartic function.

Photos



Example

- Using curve-fitting techniques for a higher order polynomial, they have found that the probability density function of sickness occurring at various mile markers, x, to be:

$$f(x) = \frac{6}{625}(x^4 - 10x^3 + 25x^2)$$

- Find the probability of sickness on the interval from 0 to 5. Or,

$$\int_0^5 \frac{6}{625}(x^4 - 10x^3 + 25x^2)\,dx$$

$$F'(x) = f(x)$$

$$F(x) = \int f(x)\,dx$$

Percentiles and Median

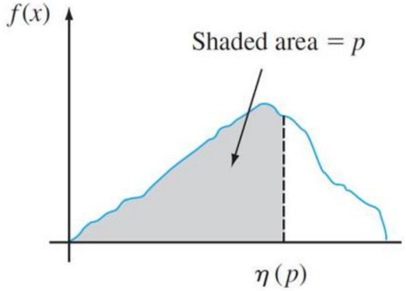
$$f(x) = pdf$$

$$F(x) = cdf$$

- The previous example found the area under the curve between 0 and 5 to be 1.
- Using percentiles, we can reverse the process. Let's say we wish to know where to place the first 40% of the trash cans. In other words, when will 40% of the sickness occur?

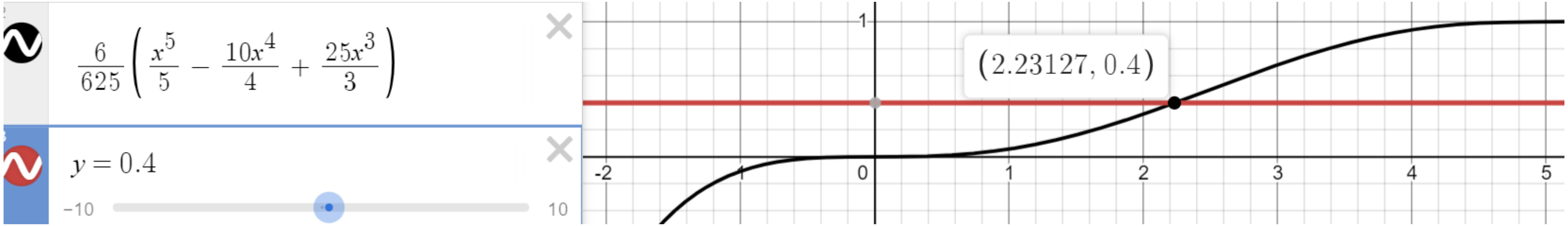
$$F(x) = \int_0^x \frac{6}{625}(x^4 - 10x^3 + 25x^2)$$

- Set F(x) equal to your percentile and solve for x.
- This often involves numerical methods.



40^{th} percentile of ill runners

- Graph F(x)
- Find the intersection with percentile.
- Or if you want to get ChatGPT to ask you for a premium subscription, ask it to find the solution. 🤑
- How would you go about finding the median here?



Mean and Variance of a Continuous R.V.

- Mean
  - For discrete: $\mu_x = E(x) = \sum[x \cdot P(x)]$
  - For continuous: $\mu_x = E(x) = \int x \cdot f(x)\, dx$
  - Note this mean is also called expected value.
- Variance
  - For discrete: $\sigma_x^2 = \sum(x - \mu_x)^2 \cdot P(x)$
  - For continuous: $\sigma_x^2 = \int_{-\infty}^{\infty}(x - \mu_x)^2 \cdot f(x)\, dx$
  - Note again that standard deviation, $\sigma_x = \sqrt{\sigma_x^2}$

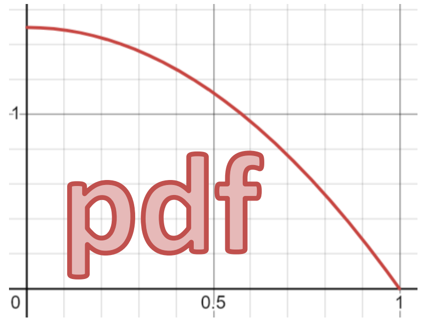
Expected Value

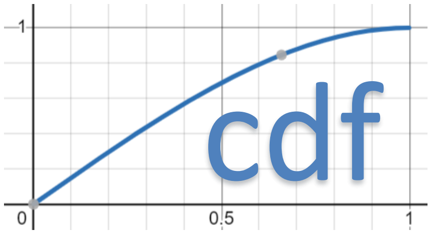
- The pdf of weekly gravel sales X is:

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

- Therefore, we end up with E(x) as:

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)dx$$





Variance

- For variance, we can use the "shortcut" method (see Lec5, slide 9):

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)dx$$

The Normal Distribution

- This is the first day of the rest of your life, right here
- A continuous rv X is said to have a **normal distribution** with parameters μ and σ, where -∞ < μ < ∞ and 0 < σ if the pdf of X is:

$$f(x;\ \mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

Simpler Representation

The mathematical expression for the height of the normal distribution is:

Gauss (1777-1855)

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Given the values of the mean and the standard deviation, this is a function of x.