Random Variables

Continuous Random Variables

- PMF → PDF
  - This now represents the relative likelihood of a value
  - Actual probabilities are ranges (because any single value is theoretically 0)
  - Visualize this as a smooth curve, rather than bar chart of mass
- Summations → Integrals
- What does this mean for expectations? Variances? Standard deviations?

Part 3

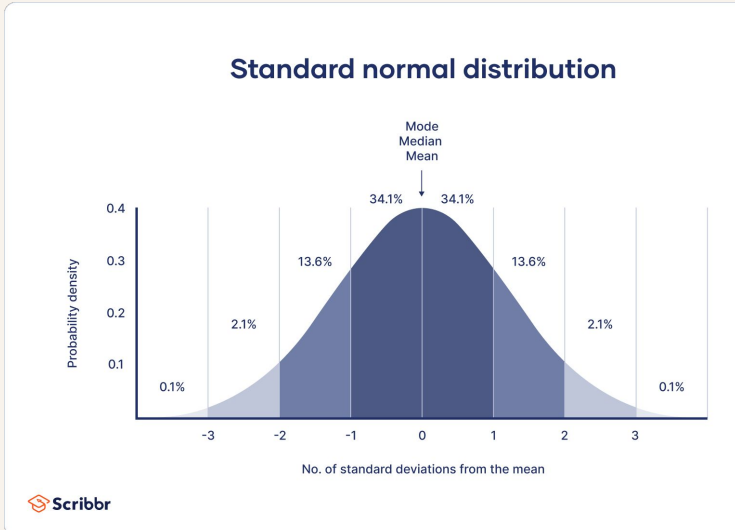
Random Variables

Common Random Variables (continuous)

- Uniform
- Exponential
- Normal
- T
- ...

Use wikipedia to get a feel for how these behave

Some (like Normal) have no closed-form representation, so we must use R or look-up tables for calculations



Part 3

Class Activity: Think-Pair-Share

Assume X is a RV that follows some continuous distribution with pdf f(x).

How would you calculate the probability that X is greater than c?

a. Generally
b. U[10,14], c = 13
c. Exp(1/3), c = 6

For each, draw a picture too
Do your answers change for X greater than or equal to c?

Practice Problems

1. Is $g(x) = x/4$ for $0 \leq x \leq 1$ a valid pdf?

2. Prove that the exponential distribution is "memoryless": $P(X > s+t \mid X > s) = P(X > t)$ for $s,t > 0$
   a. Can you interpret this in non-technical terms?

3. Is the memoryless property also true for a U[0,1] variable where $0 < s + t < 1$?

Part 3

If you claim your data follows an exponential RV, you are assuming a memoryless process

- Imagine a new surgical implant **with a constant failure rate**. If a patient's implant has been working for five years, its probability of surviving at least another year is the same as surviving at least the first year (of a brand-new implant).
- You want to model the flow of patients into an emergency waiting room. The memoryless property says that the probability that a new patient will arrive in more than X minutes does not depend on how long you've already been waiting (since the last patient arrived)

With exponential distribution, it's easier to think in terms of $P(X>t)$ ("survive at least t units until some event"), with a constant underlying event rate. $P(X>t) = \exp(-\lambda t)$

Find the expected number of *different* faces that appear in the first *n* rolls of an S-sided die

WHY DID I INTRODUCE THIS PROBLEM?

- A hospital lab tests for S different respiratory pathogens. Over a flu season, they run n tests. The expected number of different pathogens detected helps gauge if their testing protocol is broad enough to capture the diversity of circulating viruses.
- A pathogen has S known strains. A new vaccine is developed. How many people need to be vaccinated and monitored to observe the expected number of different strains that the vaccine protects against? This is a simplified way to model the diversity of immune responses and the breadth of a vaccine's protection in a population.
- A new educational campaign targets S different health behaviors. You survey n people. How many different behaviors would you expect to see changed? This helps to evaluate if the campaign is having a broad impact or if it is only affecting a small subset of the target behaviors.

Part 4

Public Health Examples

Daily Screen Time	Normal (or Log-Normal if skewed right)	Continuous measurement, often clustered around a mean with natural variation.
Whether a parent plans to vaccinate their child against a new virus	Bernoulli	A single trial with two possible results.
Steps per Day	Poisson or Negative Binomial Normal or Log-Normal	Count that may be positively skewed
How many times a specific piece of health misinformation is shared on a platform in an hour.	Poisson	Counts of relatively rare events in a fixed interval.
Time until next doctor’s appointment	Exponential	Time until the next event in a process where events occur at a constant average rate.

Part 4

Public Health Examples

Mental Wellness App Usage: The number of times a university student opens a mental wellness app in a week.
Gene Editing Success: Whether a specific gene editing attempt in a lab experiment is successful
Number of Positive tests: The count of positive diagnoses within a small community of 105 villagers
Proportion of Masks: The proportion of people observed wearing masks correctly in a grocery store.

Part 4

Exit Ticket*

Air Quality Index (AQI): The daily AQI reading in a large metropolitan area (based on 6 pollutants; score of 0-500).

Propose a distribution and write a short justification

*ANSWER is ungraded; only participation is graded

Taking a step back

- Why do we care about probability and random variables?
- What are some key (theoretical) learnings from this course so far? How can you use these in your research?

Where are we going?

- If you're willing to assume your outcome of interest follows a poisson(lambda), how might you figure out lambda?
- What are statistics?

Lab Recap

Q&A

Homework posted tomorrow

After this week's classes, along with the required readings (SPEEGLE Chapter 3; SPEEGLE Chapter 4; SPEEGLE Chapter 5, Sections 5.1-5.3), you should be able to:

- Define a random variable and distinguish between discrete and continuous types
- Calculate the expected value and variance using a probability mass function
- Solve for probabilities associated with Binomial, Poisson, and Normal random variables
- Appropriately select and justify probability distribution models for public health variables

TOPICS TO CONSIDER

Calculus Pre-Reqs

Introduction to R

Simulations

Basic Probability (Notation)

Conditional/Marginal/Joint Probabilities

Contingency Tables & Metrics

Random Variables (Discrete)

Random Variables (Continuous)

Expectations & Variance

Next:

Manipulating Data with R

https://www.rforecology.com/post/how-to-use-pipes/

Install and load the tidyverse package

UG/MPH INFORMATION SESSIONS 2025

Interested in our 5 year Undergraduate/Master of Public Health (UG/MPH) program?

Join one (or both—who's stopping you?) of our information sessions to learn more and meet:

- Liz Tobin-Tyler, Professor of Health Services, Policy and Practice, Associate Director of Dual Degree Programs
- Will Goedel, Associate Professor of Epidemiology, Interim Director of the Undergraduate Public Health Concentration

Virtual Information Session
Tuesday, October 7, 12pm-1pm
RSVP Here!

Pizza will be served!

In-Person Information Session
Thursday, October 30, 5pm-6pm
Smith-Buonanno Hall, Room 207
Scan QR code to register!

Stats Open House

Friday, October 3rd
Salomon Center Room 202

3:00 **Intro to the Concentration with Prof Paul:** What's changed with the new curriculum? what's the difference between stats and biostats? how do I double concentrate?

3:30 **Q&A with Current Concentrators:** example concentration plans, favorite courses, capstone/thesis projects

4:00 **Open Hours with the Stats DUG:** come ask any remaining questions about registration or the concentration

Coffee and Donuts!