Problem Session 1: Probability Review

AA 228 / CS 238

Fall 2023

Stanford University

Today's Objectives

1 Solidify fundamental concepts of **random variables** and **probability distributions**.

2 Refresh understanding of **joint probability** and **conditional probability**.

3 Lay the groundwork for exploring **Bayesian Networks** later this week and next.

What is Probability?

- A Formal Definition…
  - "The **probability** of an event is the proportion of times the event occurs in many repeated trials of a random phenomenon." – *ECON 102A*
  - "The **probability** of the event occurring, P(Event), is the ratio of trials that result in the event, written as count(Event), to the number of trials performed, *n*." – *CS 109*
  - "$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$" – *CS 109*

    - ***"P(E) is a measure of the chance of E occurring."*** – *Universally*

Properties to Know

- Axioms of Probability
  - **Axiom 1**: $0 \leq P(E) \leq 1$, where E = event space
  - **Axiom 2**: $P(S) = 1$, where S = sample space
  - **Axiom 3**: If E and F are mutually exclusive $P(E \cap F = \emptyset)$, then $P(E) + P(F) = P(E \cup F)$

- Provable Identities
  - $P(E^c) = 1 - P(E)$
  - If $E \subseteq F$, then $P(E) \leq P(F)$
  - $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

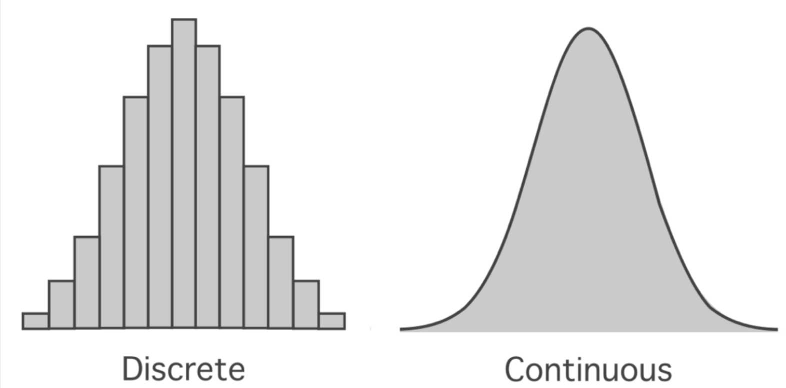
Random Variables

- **Random variable** (r.v.): a variable that **probabilistically** takes on different values
- Aka… a function that maps *outcomes* to *numerical values*
- E.g. rolling two die; define r.v. Y as the sum of the die
  - P(Y = 0) = 0/36
  - P(Y = 1) = 0/36
  - P(Y = 2) = 1/36
  - P(Y = 3) = 2/36
  - …

- Random variables can be **discrete** or **continuous**
- **Discrete** = finite number of values
  - E.g. the number of students in this session {0, 1, 2, …., 500}
- **Continuous** = infinitely many possible values
  - E.g. the time this problem session will run in minutes {59.9, 59.99, 59.999…}

Probability Distributions

- **Probability distribution**: "assigns probabilities to different outcomes"
  - Shows how the probabilities of outcomes are distributed over different values of the r.v.
- **Semantics**: X follows a distribution D; X is a "D random variable"
  - E.g., Gaussian, Uniform, Categorical
- **Syntax**: $X \sim D$
- Represented in different ways depending on whether the r.v. described is discrete or continuous

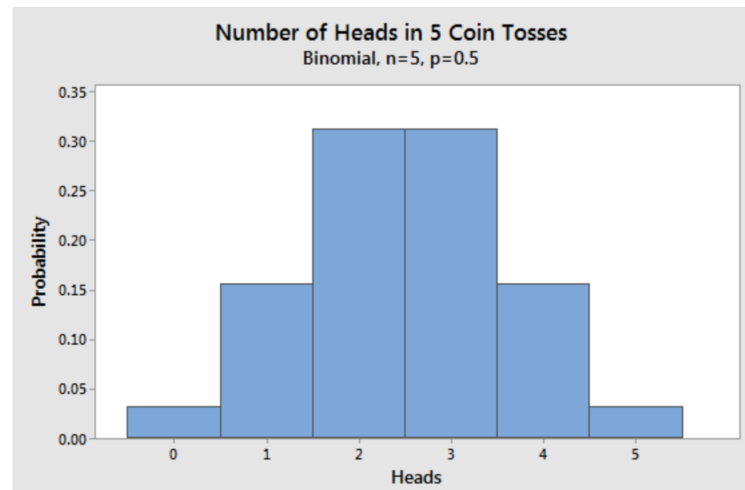


$$\sum_{x \in \{\text{All possible outcomes}\}} p(x) = 1$$

Probability Distributions: Discrete, PMF

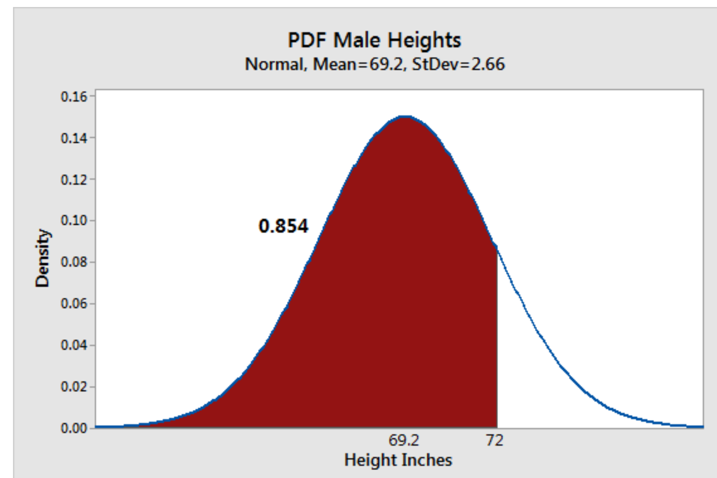
- Distributions over discrete outcomes have a **probability mass function (PMF)**
- **PMF**: a function that maps possible outcomes of a discrete random variable to the corresponding probabilities

$$\sum_{x \in \{\text{All possible outcomes}\}} p(x) = 1$$



Probability Distributions: Continuous, PDF

- Distributions over continuous outcomes have a **probability density function (PDF)**
- **PDF**: a function that maps outcomes of a continuous r.v. to ***relative likelihoods***
  - $P(X = a) = \int_a^a f(x)dx = 0$
- To compute a *probability* from a PDF, integrate!
  - $P(a \leq X \leq b) = \int_a^b f(x)dx$



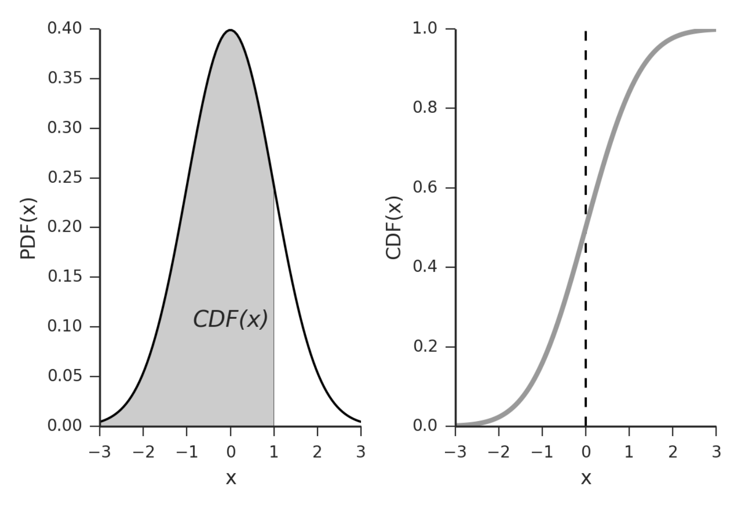
PDF vs. CDF

- **Cumulative Distribution Function (CDF):** a function that yields the probability that a r.v. is less than or equal to x, denoted as $F(x)$

$$F(x) = \mathrm{P}[X \leq x]$$

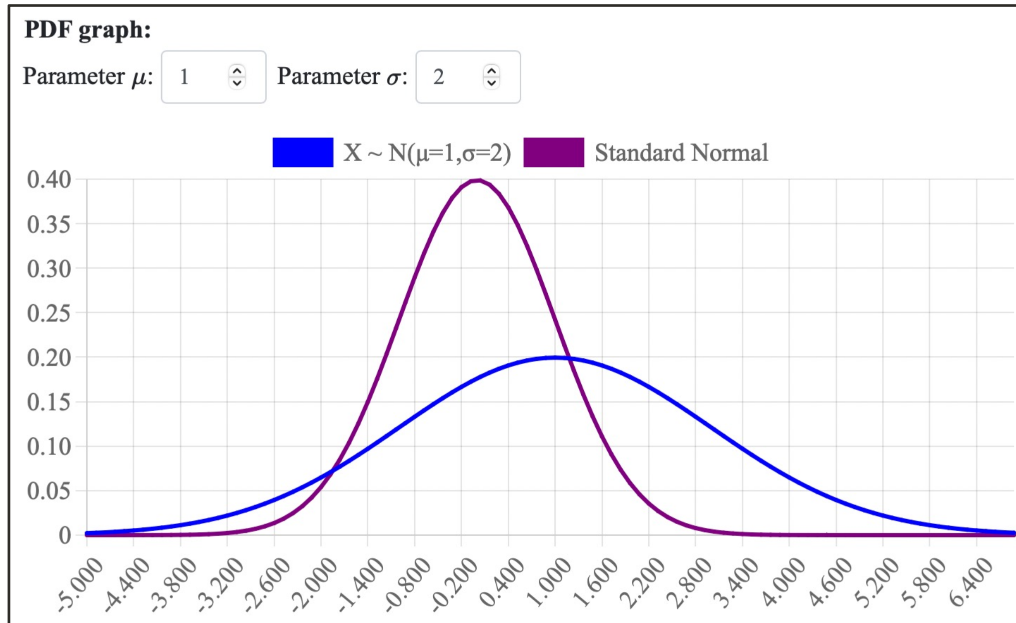
- CDF, in terms of PDF:

$$\mathrm{cdf}_X(x) = P(X \leq x) = \int_{-\infty}^{x} p(x')dx'$$



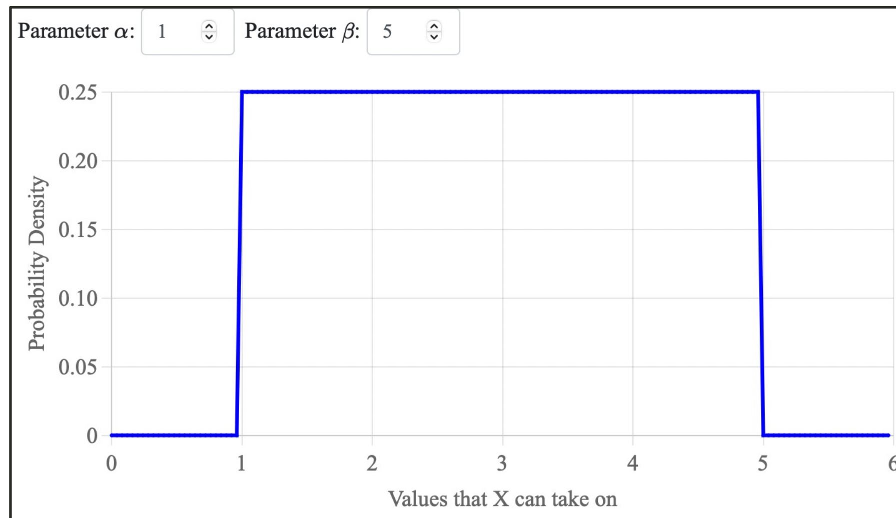
Continuous Probability Distributions: Gaussian

- Arguably the single most important r.v. type is the Gaussian r.v.
  - $X \sim N(\mu, \sigma^2)$
- PDF for a Normal X:
  - $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
- By definition:
  - $E[X] = \mu$
  - $Var(X) = \sigma^2$



Continuous Probability Distributions: Uniform

- The most basic continuous r.v. type; equally likely to take on any value in its range (α, β)
  - $X \sim \text{Uni}(\alpha, \beta)$
- PDF for a Uniform X:
  - $f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha \le x \le \beta \\ 0, & \text{otherwise} \end{cases}$
- By definition:
  - $E[X] = \frac{1}{2}(\alpha + \beta)$
  - $Var(X) = \frac{1}{12}(\beta - \alpha)^2$



Exercise 2.1, Chapter 2 from Textbook

Exercise 2.1. Consider a continuous random variable X that follows the *exponential distribution* parameterized by λ with density $p(x \mid \lambda) = \lambda \exp(-\lambda x)$ with nonnegative support. Compute the cumulative distribution function of X.

Exercise 2.1, Chapter 2 from Textbook

- Definition of CDF:

$$\mathrm{cdf}_{\mathrm{X}}(x) = \int_{-\infty}^{x} p(x')dx'$$

- Nonnegative support indicates that there is no probability mass in the interval $(-\infty, 0)$; can plug-in the exponential distribution:

$$\mathrm{cdf}_{\mathrm{X}}(x) = \int_{0}^{x} \lambda e^{-\lambda x'} dx'$$

- Compute the integral:

$$\mathrm{cdf}_{\mathrm{X}}(x) = -e^{-\lambda x'} \Big|_{0}^{x}$$

$$\mathrm{cdf}_{\mathrm{X}}(x) = \boldsymbol{1 - e^{-\lambda x}}$$

Joint Probability

- What happens when a problem involves not just one random variable, but several that may influence each other?
  - E.g., determining the probability you get into a bike crash and are wearing a helmet
- Joint function for two discrete r.v.:
  - $p(X = x, Y = y), p(x, y)$
- Joint density function if there is at least one continuous r.v.:
  - $f(X = x, Y = y)$
- Same idea extends if there are many r.v. …
  - $p(x_1, x_2, \dots, x_n)$
  - $p(\vec{x})$

Marginal Distribution via Law of Total Probability

- **Marginal distribution**: the probability distribution of the r.v. contained in some subset of all r.v.

$$p(x) = \sum_y p(x, y)$$

or

$$p(x) = \int f(x, y)\, dy$$

Conditional Probability

- **Conditional probability**: the probability of E given that (i.e., conditioned on) event F already happened

$$P(E \mid F) = \frac{P(E \text{ and } F)}{P(F)}$$

- Leads us to the **Chain Rule**:

$$P(E \text{ and } F) = P(E \mid F) * P(F)$$

- Conditioning on multiple events…

$$P(E \mid F, G) = \frac{P(E \text{ and } F \mid G)}{P(F \mid G)}$$

Bayes Theorem

- **Bayes' theorem:** presents a way to convert a conditional probability from one direction to the other direction
- E.g., from $P(E \mid F)$ to $P(F \mid E)$

$$P(E \mid F) = \frac{P(E,F)}{P(F)}$$

- (Conditional Probability)

$$= \frac{P(F \mid E) * P(E)}{P(F)}$$

- (Chain rule)

$$p(x \mid y) = \frac{p(y \mid x) * p(x)}{p(y)}$$

$$= \frac{p(y \mid x) * p(x)}{\sum_x p(x, y)}$$

Exercise, Courtesy CS 109 Course Reader

- In this problem we are going to calculate the probability that a patient has an illness given a positive test result for the illness. A positive test result means the test thinks the patient has the illness. You know the following information:
    - The natural occurrence of breast cancer is 8%.
    - The mammogram test returns a positive result 95% of the time for patients who have breast cancer.
    - The test returns a positive result 7% of the time for people who do not have breast cancer.
- *What is the probability that the patient has breast cancer given a positive mammogram result?*

Exercise, Courtesy CS 109 Course Reader

- Let B be the event that the patient has breast cancer
- Let P be the event that the mammogram result is positive
- Want $\mathrm{P}(B|P)$
- Know $\mathrm{P}(P|B) = 0.95$, $\mathrm{P}(P|B^c) = 0.07$, $\mathrm{P}(B) = 0.08$
- Apply Bayes' Theorem with Total Probability:

$$\mathrm{P}(B|P) = \frac{\mathrm{P}(P|B)\mathrm{P}(B)}{\mathrm{P}(P|B)\mathrm{P}(B) \; + \; \mathrm{P}(P|B^c)\mathrm{P}(B^c)}$$

$$= \frac{(0.95)(0.08)}{(0.95)(0.08) + (0.07)(1 - 0.08)} = \mathbf{0.5413}$$

The Monty Hall Problem

- Suppose you're on a game show, and you're given the choice of three doors: Behind one door is an amazing prize; behind the others, homework.
- You pick a door, say Door 1, and the host, who knows what's behind the doors, opens another door, say Door 3, which has a homework.
- The host then says to you, "Do you want to pick Door 2?"
- *Is it to your advantage to switch your choice?*

The Monty Hall Problem (Solved!)

- Assume the prize is behind any door
  - $P(\text{prize}_i) = \frac{1}{3}$ for $i = 1, 2, 3$
- You chose **Door 1**; the host opened **Door 3** and there was homework ☹
- **Scenario 1**: prize is behind Door 1, and host opened Door 2 or Door 3 with equal probability
  - $P(\text{prize}_1, \text{host}_3) = \frac{1}{3} * \frac{1}{2} = \frac{1}{6}$
  - $\boldsymbol{P(\text{host}_3)} = P(\text{host}_3|\text{prize}_1)P(\text{prize}_1) + P(\text{host}_3|\text{prize}_2)\,P(\text{prize}_2) +$ $P(\text{host}_3|\text{prize}_3)\,P(\text{prize}_3) = \frac{1}{2} * \frac{1}{3} + 1 * \frac{1}{3} + 0 = \frac{1}{2}$
  - $P(\text{keep Door 1 and win}) = P(\text{prize}_1|\ \text{host}_3) = \frac{P(\text{prize}_1, \text{host}_3)}{P(\text{host}_3)} = (\frac{1}{6})/(\frac{1}{2}) = \boldsymbol{\frac{1}{3}}$

The Monty Hall Problem (Solved!)

- **Scenario 2**: prize is behind Door 2, and host opened Door 3
  - $P(\text{prize}_2, \text{host}_3) = \frac{1}{3} * 1 = \frac{1}{3}$
  - $P(\text{keep Door 1 and lose}) = P(\text{prize}_2 | \text{ host}_3) = \frac{P(\text{prize}_2, \text{host}_3)}{P(\text{host}_3)} = (\frac{1}{3})/(\frac{1}{2}) = \mathbf{\frac{2}{3}}$

- *You are twice as likely to win (2/3 vs. 1/3) by switching!*

Questions?

Additional Resources

- Chapter 2, *Algorithms for Decision Making*
  - Link: https://algorithmsbook.com/files/chapter-2.pdf
- Course Reader for *CS 109: Probability for Computer Scientists*
  - Link: https://chrispiech.github.io/probabilityForComputerScientists/en/
- Unit 7: Probability, Khan Academy
  - Link: https://www.khanacademy.org/math/statistics-probability/probability-library

On the Horizon

- **Thursday, Sept. 28th**: Week 1, Lecture 2 on representing uncertainty
  - My O.H. are 6:45-8:45pm tomorrow, so feel free to stop by!
- **Friday, Sept. 29th**: Project 0 and Quiz 0 due
- **Wednesday, Oct. 4th**: Bayesian Networks Problem Session

Thank You