Random Variables

STAT 240 - Fall 2025

Robert Sholl

Definitions

Random Variable

A rule for assigning a numeric value to each outcome of a random experiment.

$$X = \{\text{The number of heads in three coin flips}\}$$

$$Y = \{\text{The sum of two dice rolls}\}$$

EXPERIMENT: ANY TRIAL WHERE OUTCOME IS DETERMINED BY CHANCE

Support

The set of possible values a random variable can be.

3 COINS $S_X = \{0, 1, 2, 3\}$

2 DICE $S_Y = \{2, 3, \ldots, 11, 12\}$

Notation

X = 2

- X represents the proposed random variables
  - It has possible values, but no actual value
- x represents a realization of the random variable X

$$P(X = x) \Rightarrow \text{The probability that r.v. X realizes to x}$$

$$P(X > x) \Rightarrow \text{The probability that r.v. X realizes greater than x}$$

Discrete Random Variables

Distribution

Formally: The function that dictates the generation and moments of a random element. VARIABLE

← REPRESENT R.V.

- Think of all of our measures of center, spread, and position
- Those *uniquely* identify a grouping of data
- Something fundamentally *controls* what those measures should be

Discrete Distributions

- Discrete random variables have **mass** to their probability
- **Probability mass functions** satisfy the following

PMF

$$0 \leq P(X = x) \leq 1$$

$$\sum_x P(X = x) = 1$$

} NECESSARY / REQUIRED

SUFFICIENT: A FEATURE

Moments

Measures related to the shape of a function's graph

We care about two moments right now

$$\mathbb{E}X \Rightarrow \text{Expectation}$$

$$\mathbb{V}X \Rightarrow \text{Variance}$$

Discrete Expectation

For discrete random variables:

$$\mathbb{E}X = \sum_x xP(X = x)$$

- The sum of all values of the r.v.
  - *Weighted* by their probabilities

Discrete Variance

For discrete random variables:

$$\mathbb{V}X = \sum_x (x - \mathbb{E}X)^2 P(X = x)$$

- The sum of the squared differences between all values and the expectation
  - *Weighted* by their probabilities

Inference

MEAN

- Expectation: *long-run* average, as $n \longrightarrow \infty$ the expectation is the center of the r.v.

SPREAD

- Variance: *long-run* spread, as $n \longrightarrow \infty$ the variance is the squared spread of the r.v.
  - We can make variance more interpretable (and useful overall)

$$\sqrt{\mathbb{V}X} = \sigma_X$$

FIX UNITS

- Standard deviation functions the same as before

In Practice

Let X represent the number of caffeinated beverages anyone in the class consumes daily

- Determine the distribution of X
- Find the expected value of X
- Find the variance and standard deviation of X

0	1	2	3	4	5	6	7	X
5	13	7	3	4	0	0	1	33
.15	.40	.21	.09	.12	0	0	.03	1.00

$\mathbb{E}X = \sum_x x \, P(X=x)$

$0(.15) + 1(.40) + 2(.21) + 3(.09) + 4(.12) + 5(0) + 6(0) + 7(.03)$ $\underline{1.78}$

$\mathbb{V}X = \sum_x (x - \mathbb{E}X)^2 P(X=x)$

$$((0-1.78)^2(.15)) + ((1-1.78)^2(.40)) + ((2-1.78)^2(.21)) + ((3-1.78)^2(.09)) + ((4-1.78)^2(.12)) + ((7-1.78)^2(.03))$$

$\mathbb{V}X = 2.27$

Exam

↳ Biased

↳ Correct

1) Why you got it wrong
2) Why is it wrong
3) What's right
4) Why

Before Bonus & Quiz

$\bar{X} = 65$ MED = 68

$S = 13.5$

Add Bonus

$\bar{X} = 70$ MED = 72

$S = 15.5$

Litter Size	2	3	4	5	
Freq	1	1	1	6	= 30
Rel. Freq	1/30				

Exercise: Simple

Problems: Complex

$$\frac{\text{Freq}_i}{n} = \text{Rel. Freq}_i) \quad \Sigma = 1$$

Correlation ≠ Causation

Sample → Statistics → Parameters (inference) Population

$$\bar{X} = (2*1)+(3*1)+(4*1)+(5*6)$$

$$S^2 = \frac{\Sigma (X_i - \bar{X})^2}{n-1}$$

← Average of total squared difference

$$z = \frac{X - \bar{X}}{S}$$

$$(Z_{x_i} * Z_{y_i})$$

$$S = \sqrt{S^2}$$

$$r = \frac{1}{n-1} \Sigma \left(\left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_i - \bar{y}}{S_y} \right) \right)$$

Your Turn

- Determine the distribution of Y
- Find the expected value of Y
- Find the variance and standard deviation of Y

Continuous Random Variables

Definition

Continuous random variable: The **support** consists of all numbers in an **interval** of the **real number line** and are *uncountable infinite*.

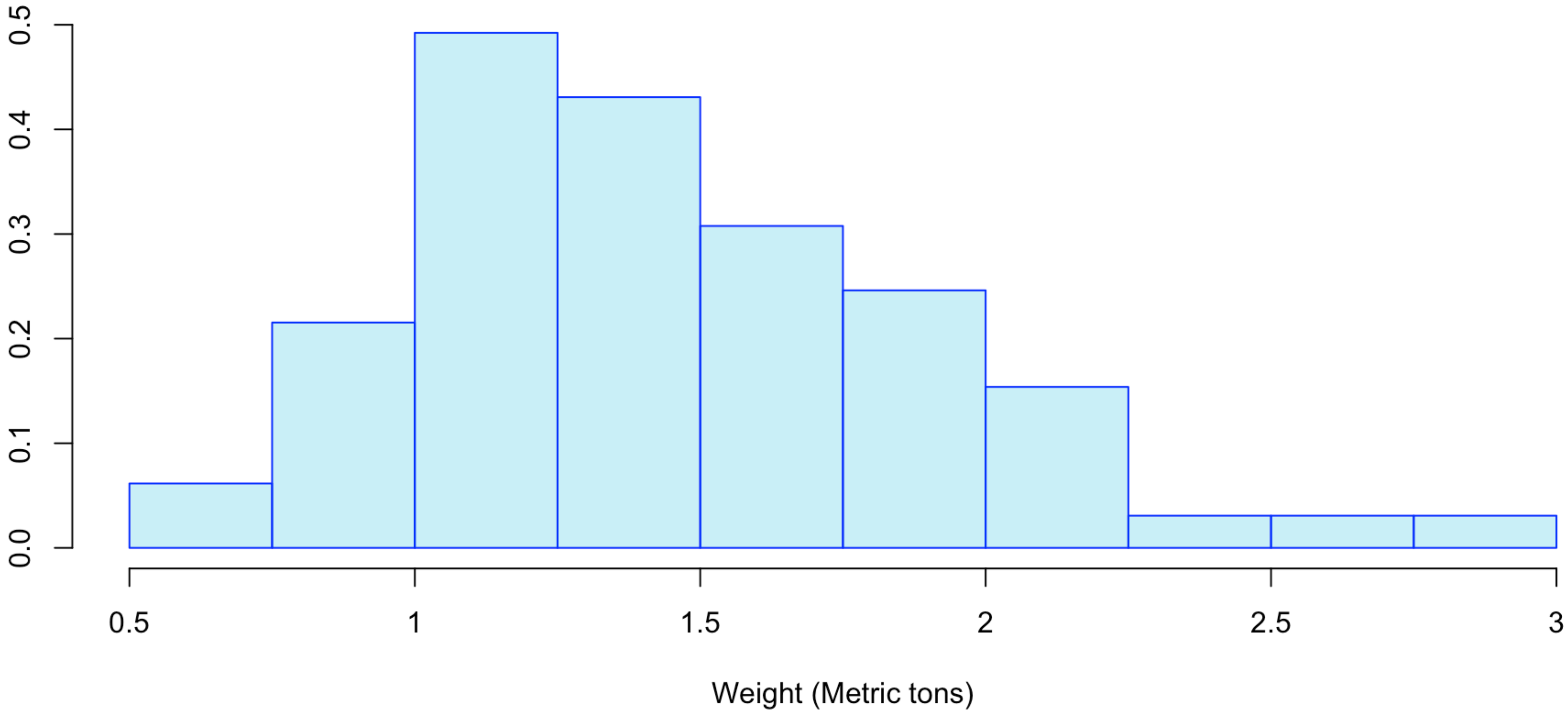
$$\text{Let } X = \{\text{the body mass of a dairy cow}\}$$

$$S_X = (0, \infty)$$

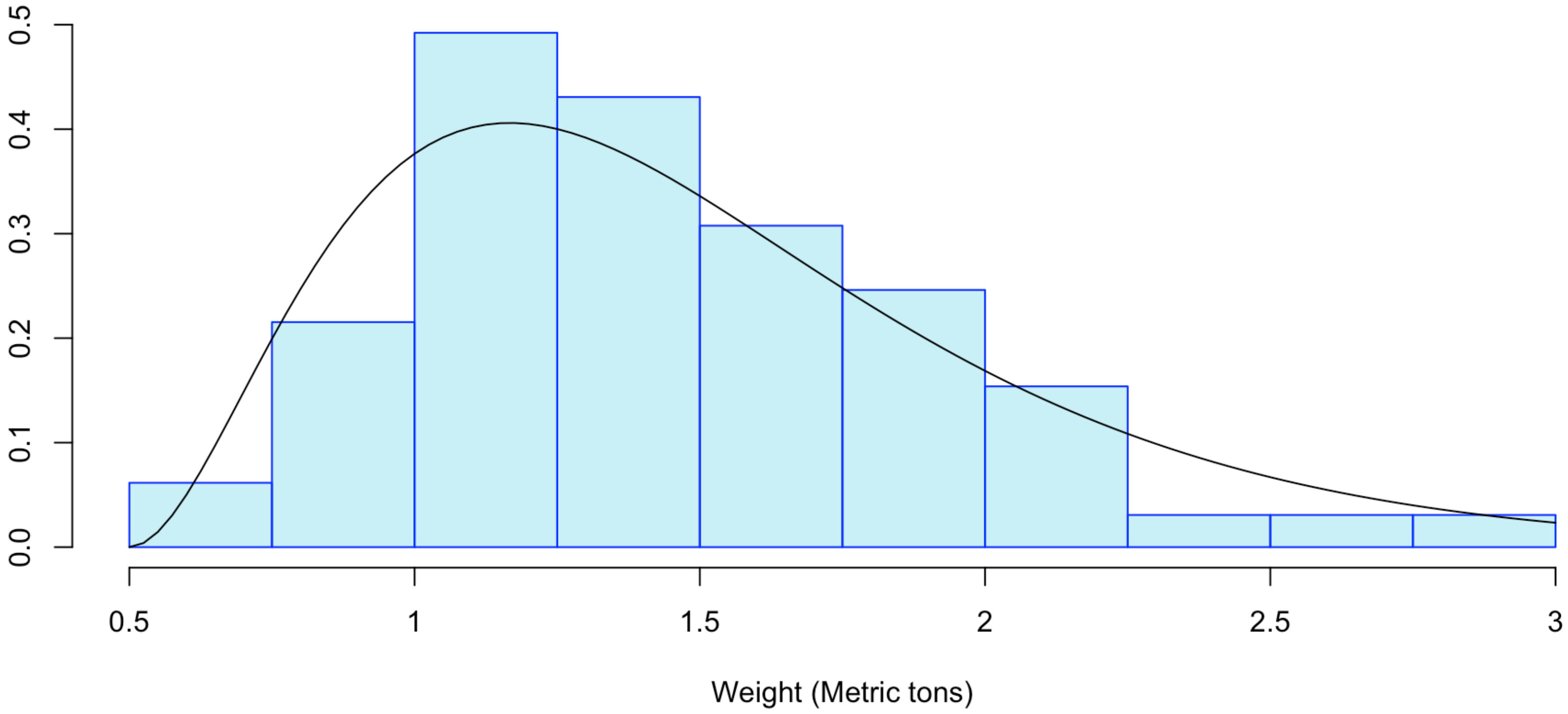
$[0, 100]$

$S_\theta = \{1, 2, \ldots\}$

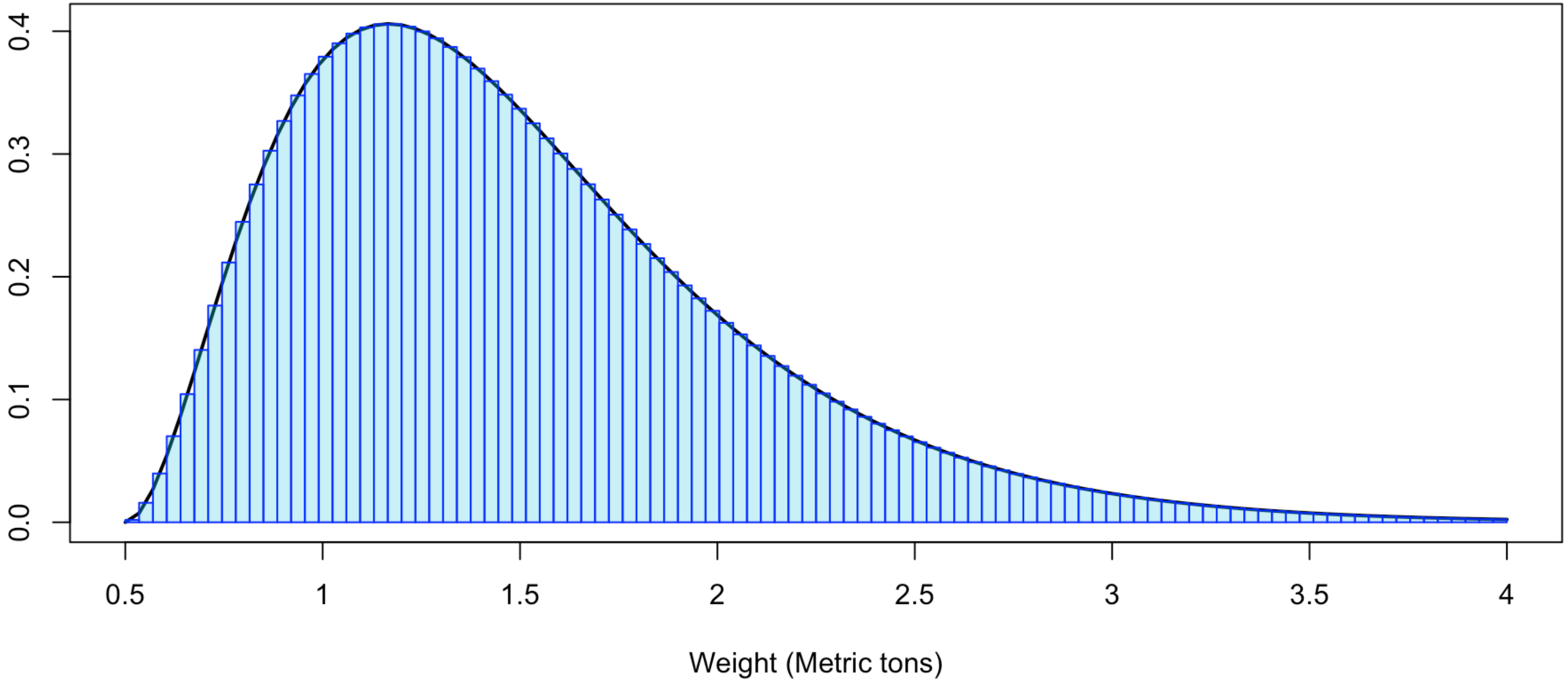
Theory



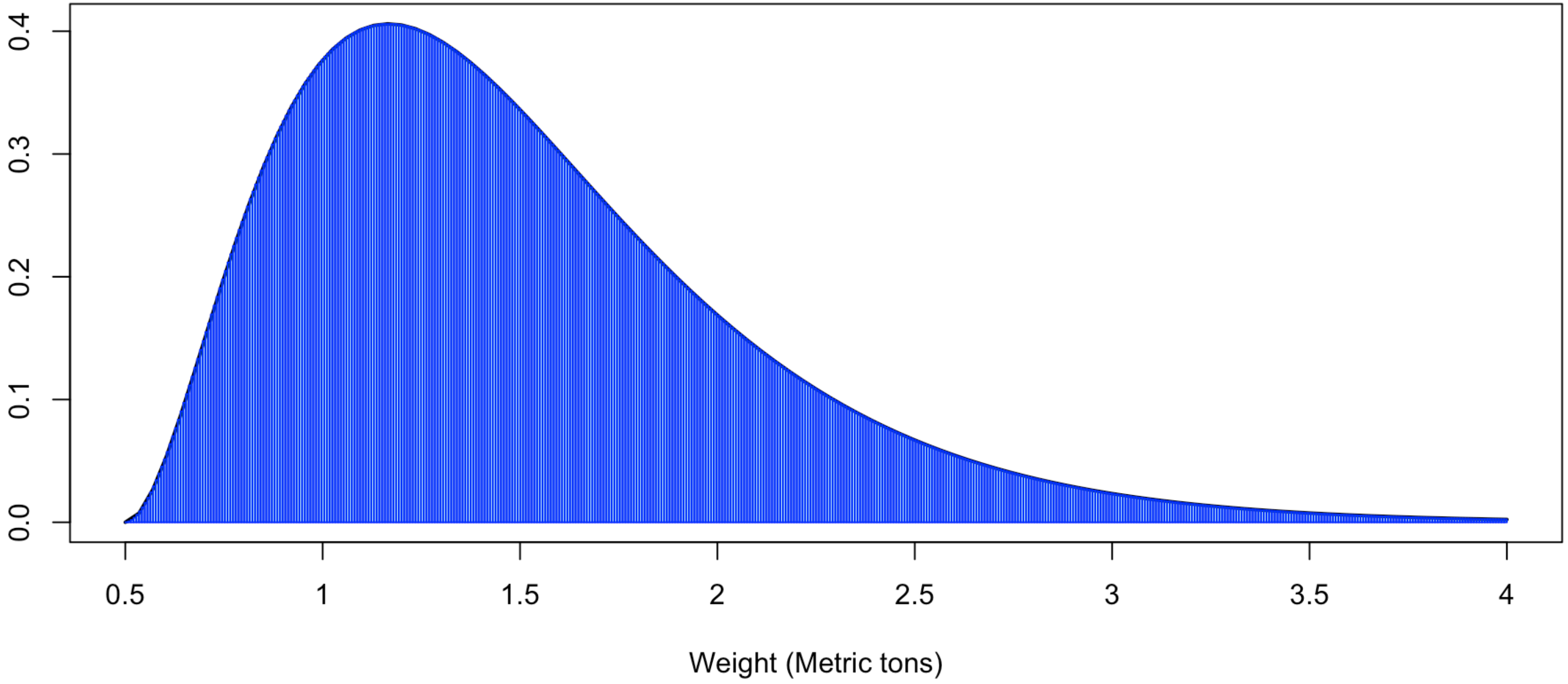
Theory



Theory



Theory



Continuous Distributions

- Continuous random variable probabilities are discussed as **densities**
- **Probability density functions** must satisfy similar rules to PMFs
  - There's an issue though

PROBABILITY $0 < P < 1$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$0 \leq \int_{a}^{b} f(x)dx \leq 1$$

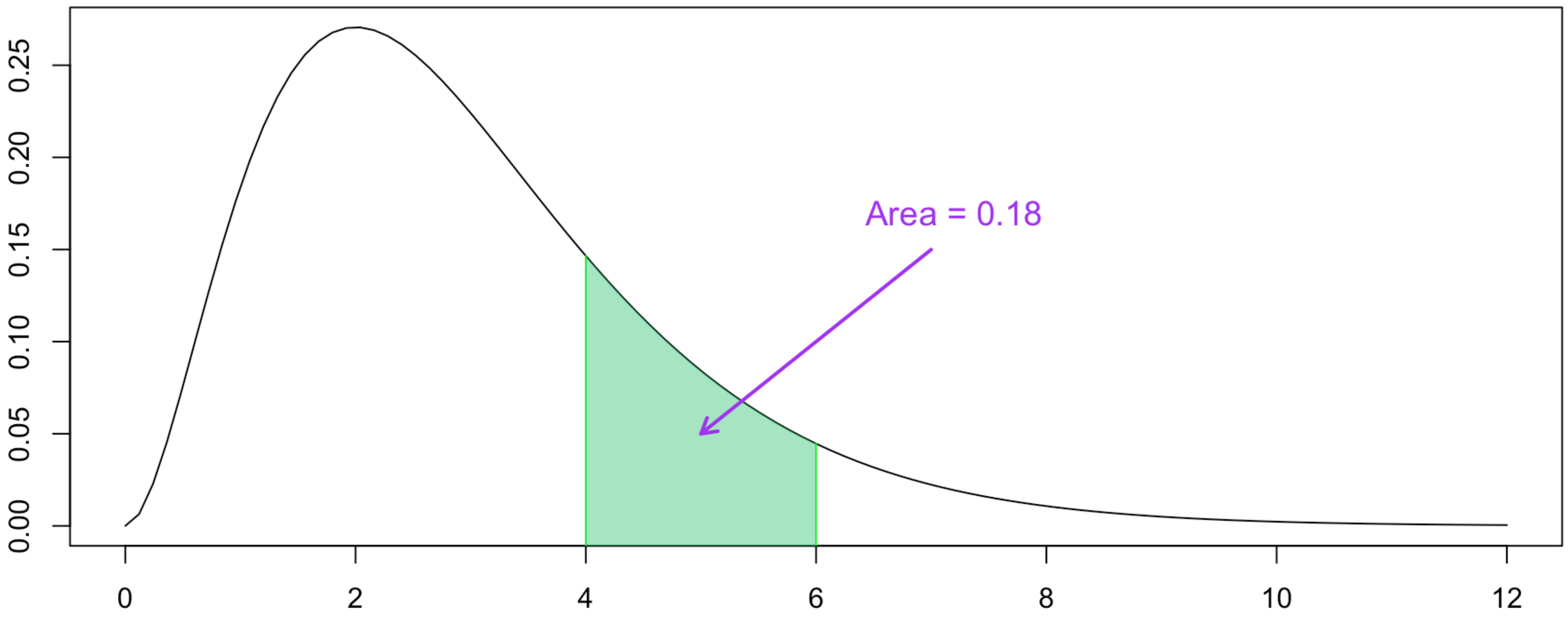
Probability Density

By definition continuous probabilities are the *area under the curve* between two values within the distribution.

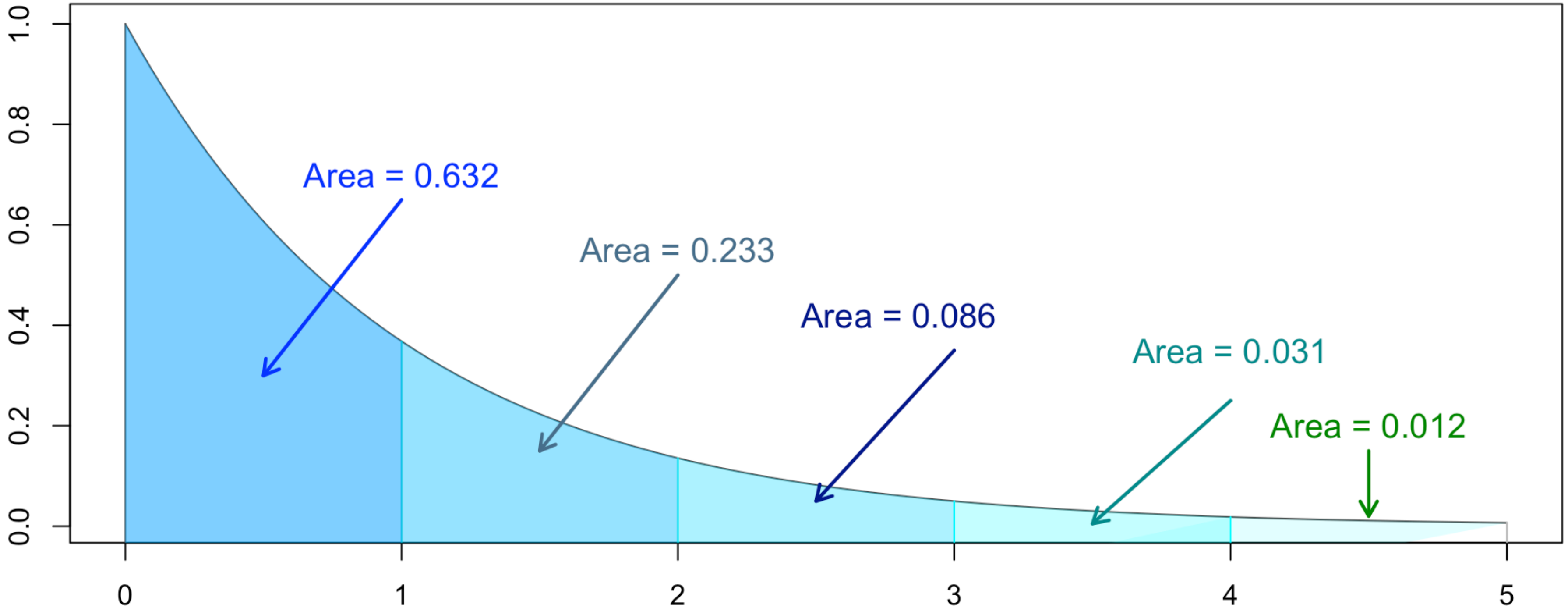
I have two rules for my class:

1. I won't invite the formal practice of calculus into my classroom
2. I'll burn at the stake before I use trigonometry

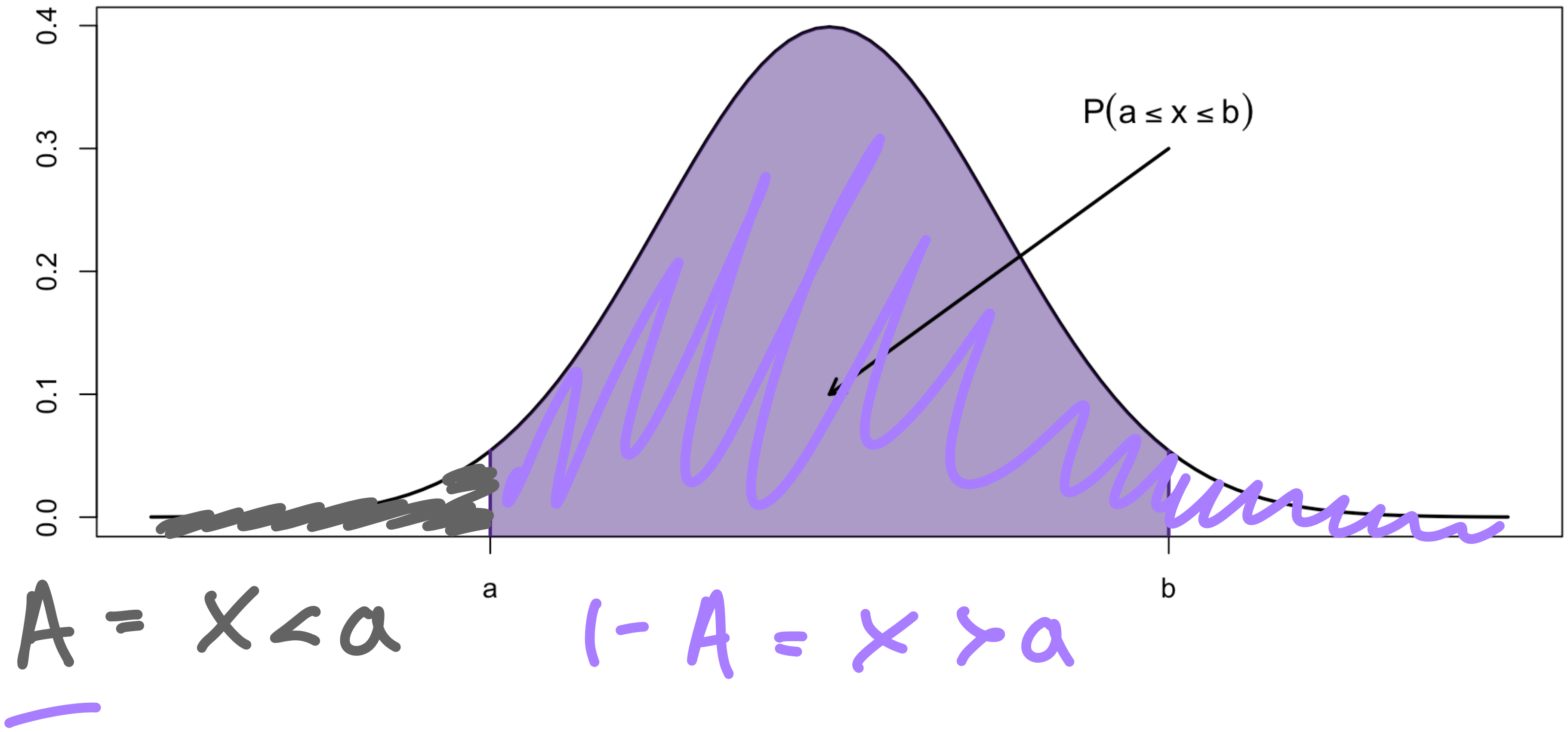
Probability Density



Probability Density

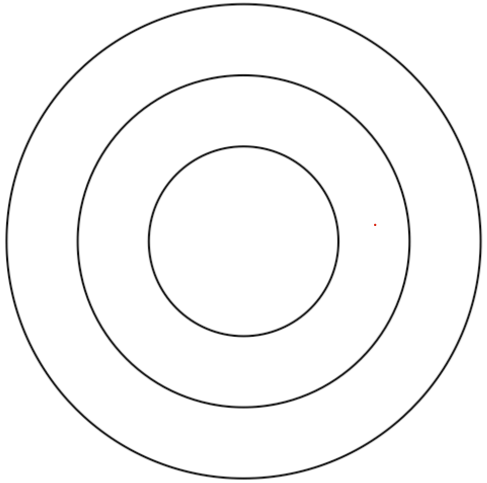


Continuous Probabilities

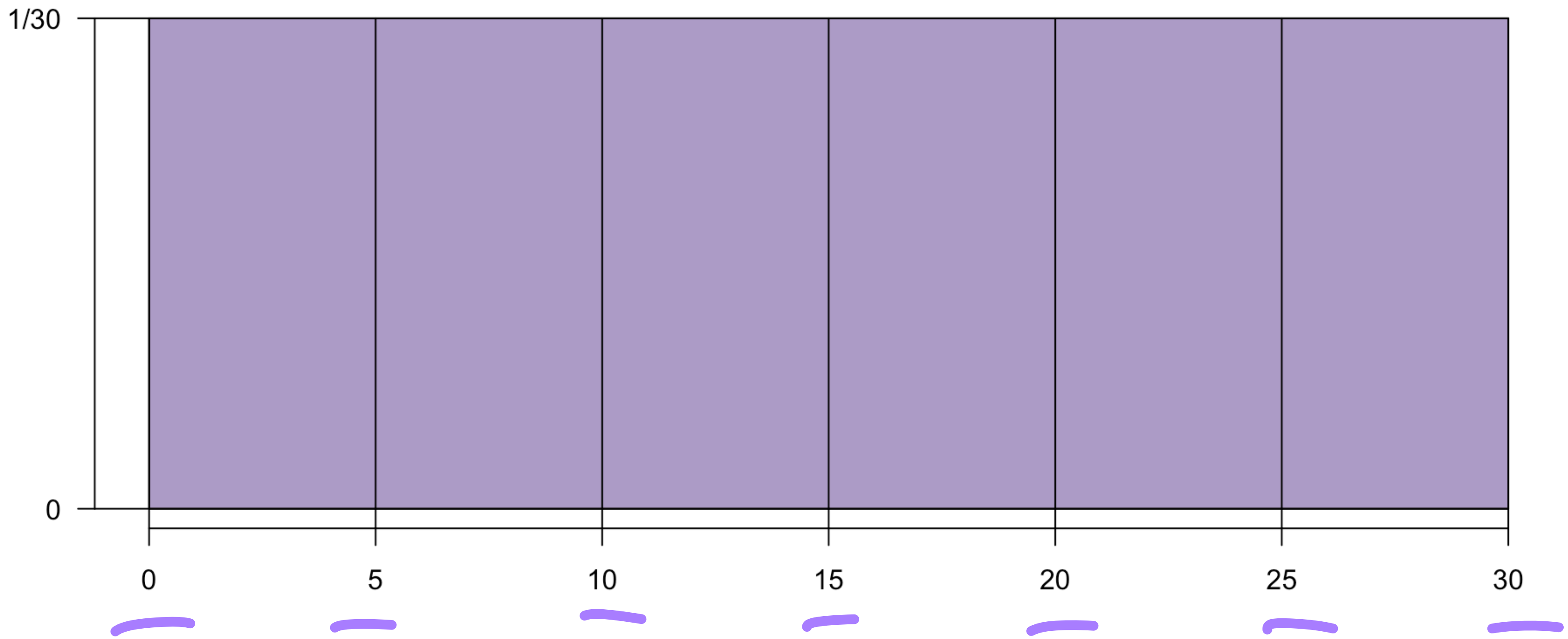


Equality

$P(X = x)$



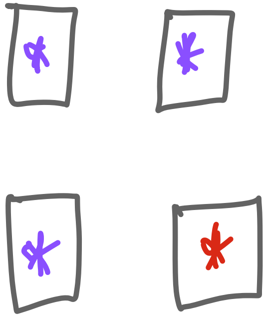
Uniform Distributions



Moments

EXPECTED VALUE

↳ GAMBLER'S RUIN



$p = 1/4 = .25$

$$\mathbb{E}X = \int_{-\infty}^{\infty} x\, f(x)dx$$

$$\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2\, f(x)dx$$

$$\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

Alternate Methods

> If you take STAT 341 (Biometrics II) you'll have to be able to calculate these things. But that class doesn't require advanced calculus.

- Sample interpolation
  - Guess exclusively from samples
- Numerical Quadrature
  - Draw boxes underneath the curve and calculate their areas
- Monte Carlo Simulation

Distribution Theory

Another alternate option would be to calculate everything **once** and then *force* the data to look like our calculations

- This is the motivation behind the field of Distribution Theory
- There's three important distributions we'll talk about:
  - Symmetric curves → NORMAL ←
  - Coin flips → BERNOULLI
  - Uniform bars → UNIFORM

EMPIRICAL RULE

→ MEAN = MED = MODE

NORMAL DISTRIBUTION

$X \sim N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma}$$

STANDARD NORMAL

$Z \sim N(0, 1)$

Go Away

1) WHY YOU GOT IT WRONG
2) WHY IS IT WRONG
3) WHAT'S RIGHT
4) WHY