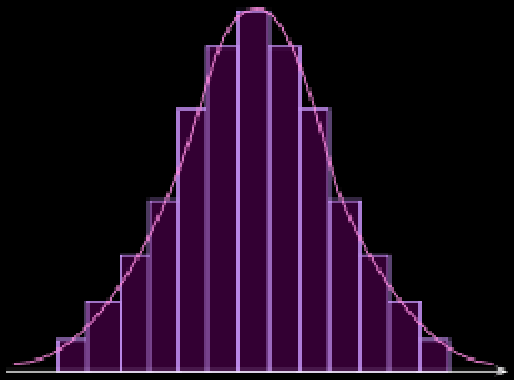
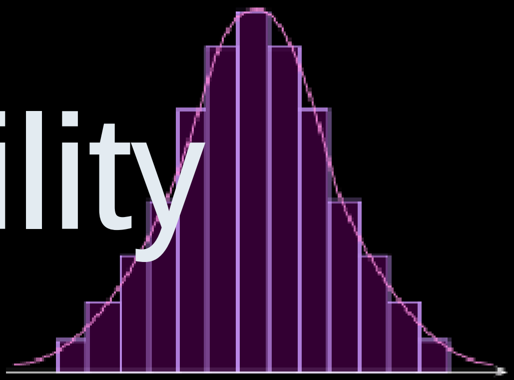


Elements of Probability & Statistics

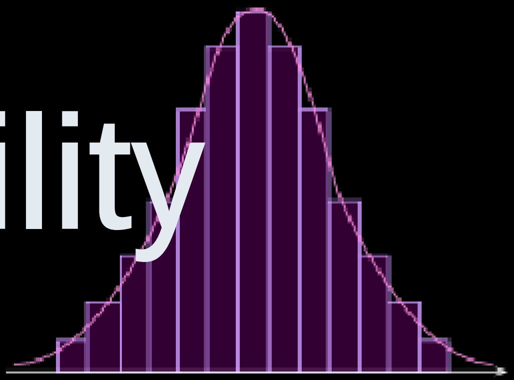
RET-Mathematics
Summer 2006

Laws of Probability



- Addition Rule
- Law of Total Probability
- Multiplication Rule
- Bayes' Theorem

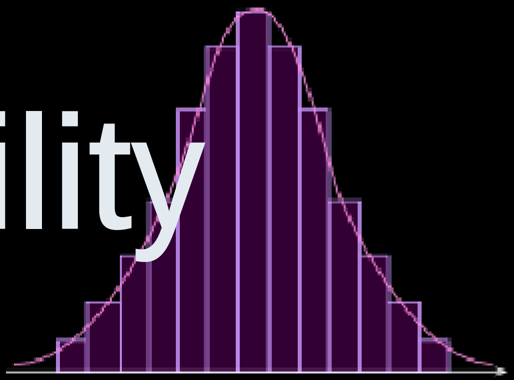
Laws of Probability



- Addition Rule

  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Law of Total Probability
- Multiplication Rule
- Bayes' Theorem

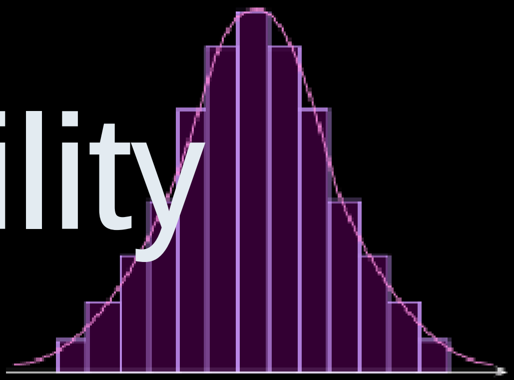
Laws of Probability



- Addition Rule
- Law of Total Probability

  $P(A) = P(A \cap B) + P(A \cap B^c)$
- Multiplication Rule
- Bayes' Theorem

Laws of Probability

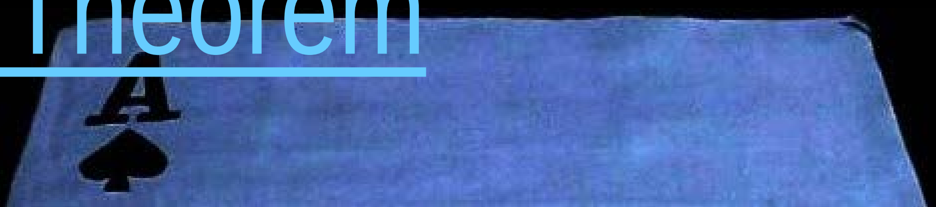


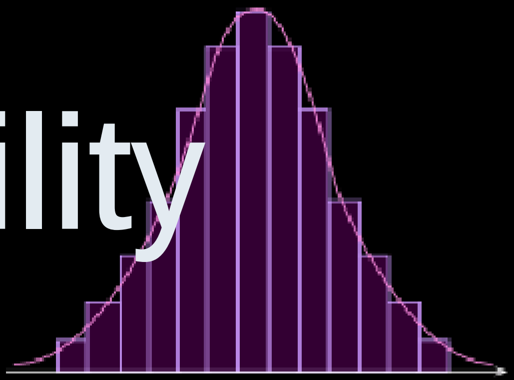
- Addition Rule
- Law of Total Probability
- Multiplication Rule

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$ if *A* and *B* are dependent

$$P(A \cap B) = P(A) \cdot P(B)$$ if *A* and *B* are independent

- Bayes' Theorem



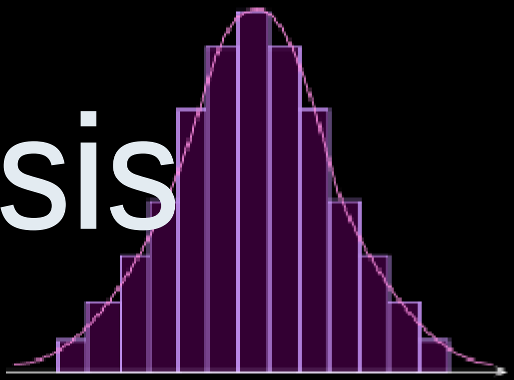


Laws of Probability

- Addition Rule
- Law of Total Probability
- Multiplication Rule
- Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid A^c) \cdot P(A^c)}$$

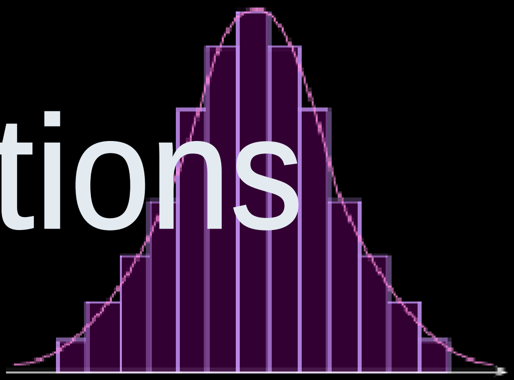
Probability Calculator



Bayesian Analysis

- Used to update the probability of an event, given new information based on the prior probability of the event

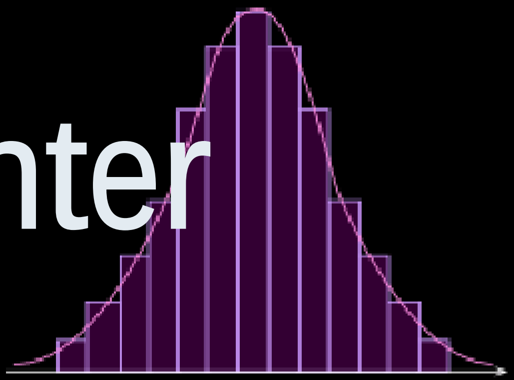
Bayesian Applications



- Medical testing
- Spam filters
- Search engines
- Trial evidence
- Nuclear plant operation analysis

To Switch or Not

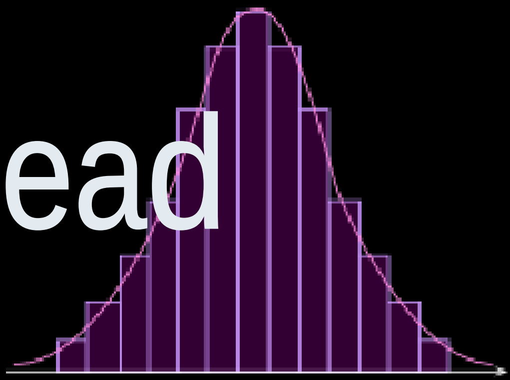
Measures of Center



- Mean = “average” value

$$\mu = \frac{\sum x_i}{n}$$

- Median = “middle” of sorted data
- Mode = most often occurring x_i
- Midrange = mean of max and min



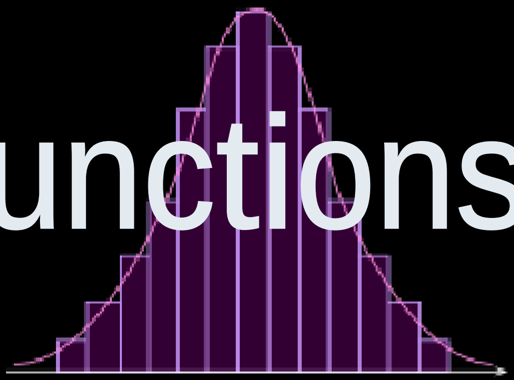
Measures of Spread

- Standard deviation
  = mean deviation from the mean

$$\sigma = \sqrt{\frac{\sum \left(x_i - \mu\right)^2}{n}}$$

- Variance = square of standard deviation
- Range = “length” of data (x_{max} - x_{min})

Probability Density Functions

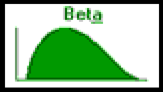


A function *f*(*x*) will model a ***pdf*** if

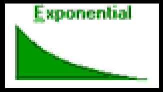
- *f*(*x*) is nonnegative
- the total area under *f*(*x*) = 1

Types of Distribution

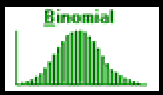
Uniform



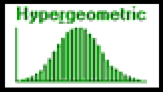
Beta



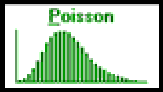
Exponential



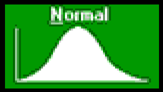
Binomial



Hypergeometric



Poisson

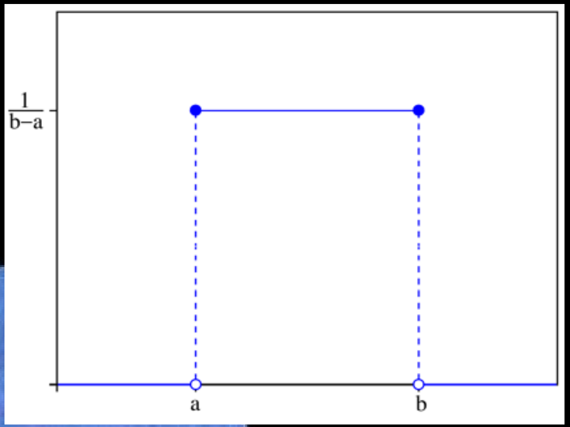


Normal & Standard Normal

Uniform Distribution

- Description: Events are equally likely
- Applications:
  - Although not commonly found in nature, it is particularly useful for sampling from arbitrary distributions
  - Used by computers for random number generation
- Formula (parameters *a* and *b* are endpoints):

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$



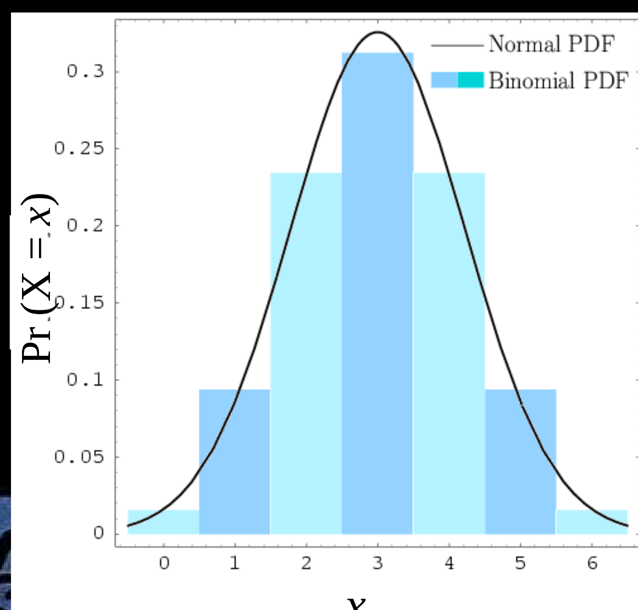
Binomial Distribution

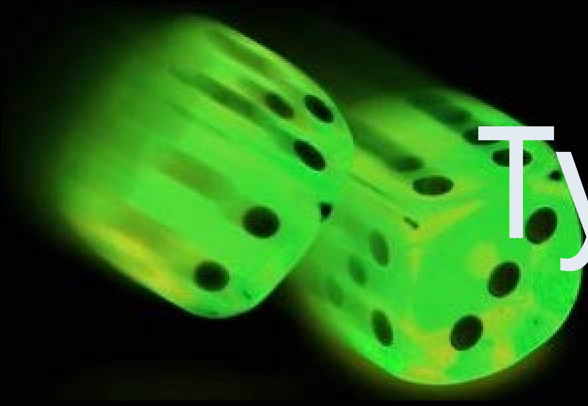
- <u>Description</u>: Multiple trials of an event with two outcomes, often modeled by yes/no situations
- <u>Applications</u>:
  - Polling, taste tests, randomized response for sensitive questions
- <u>Formula</u> (parameters n = number of trials and p = probability of success):

$$\text{binomialpdf}(n,p,x) = \left({}_nC_x\right)p^x(1-p)^{n-x}$$

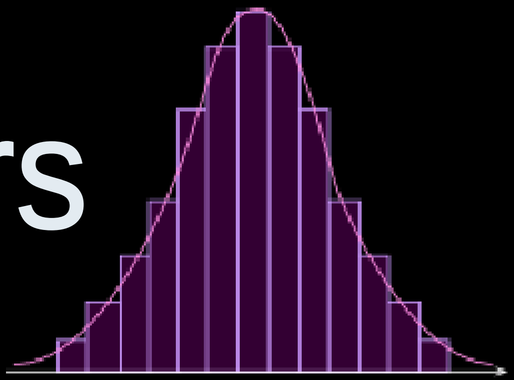
Binomial Distribution

- <u>Notes</u>:
  - Most important discrete distribution
  - Binomial distribution approximates a hypergeometric distribution when the sample size is “small” with respect to the population



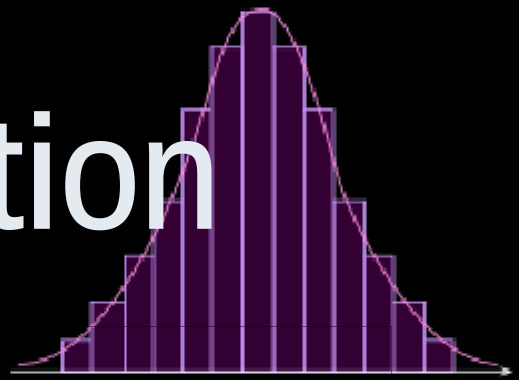


Types of Errors



- Type I (Over-reaction) = Believing that an effect exists when it does not
- Type II (Missed opportunity) = Believing that an effect does not exist when it does
- The two errors are at war with each other (inversely related)

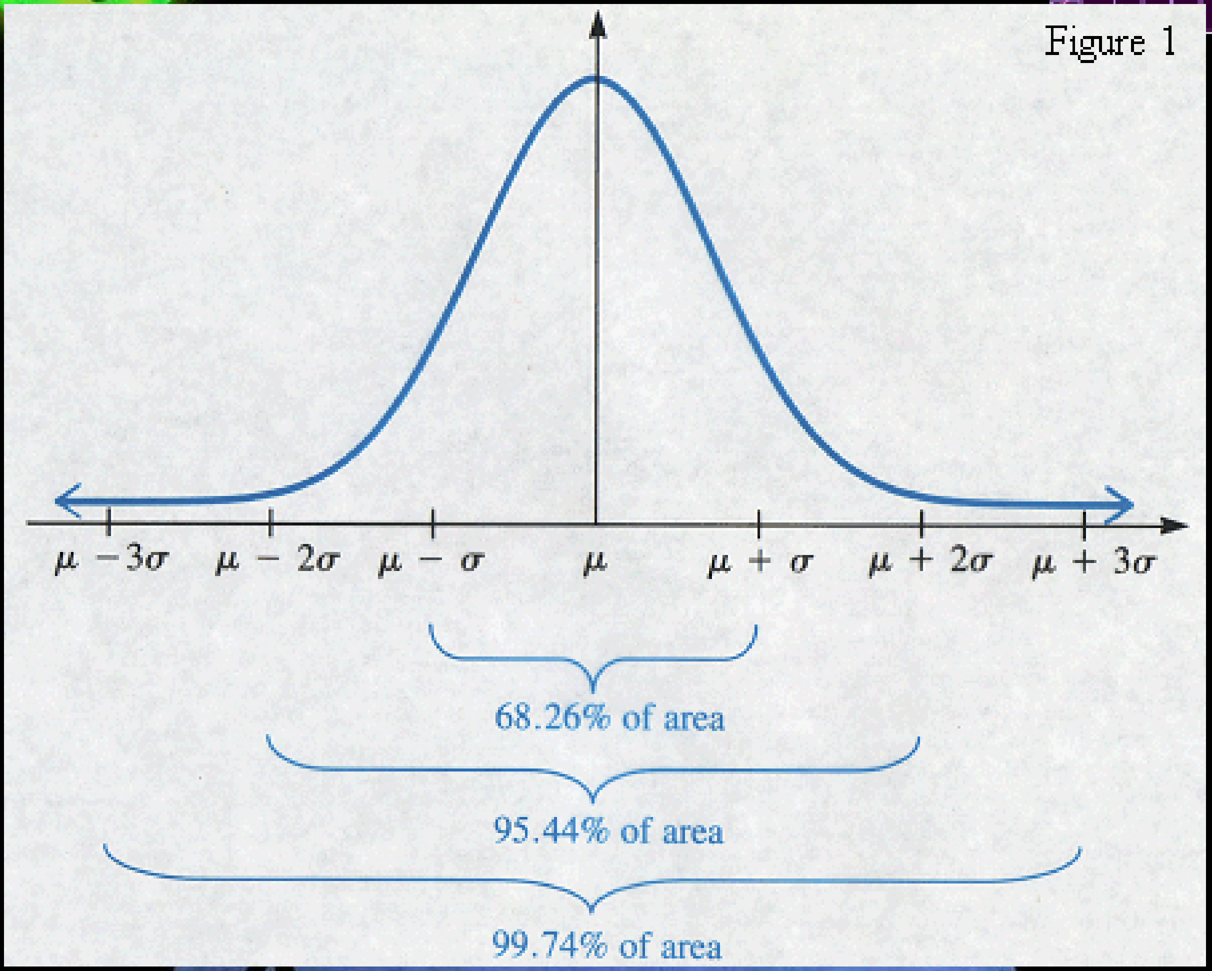
Normal Distribution



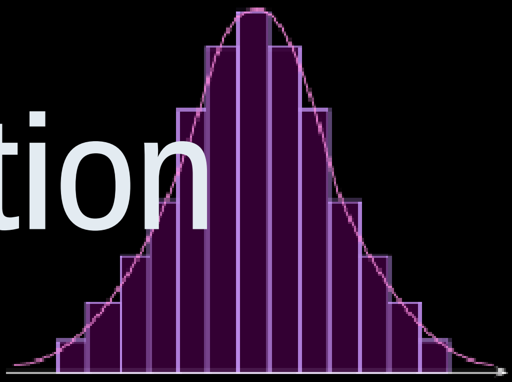
- Description: “bell” curve
- Applications:
  - physical data; test scores
  - approximates other distributions
- Formula (parameters *μ* = mean and *σ* = standard deviation):

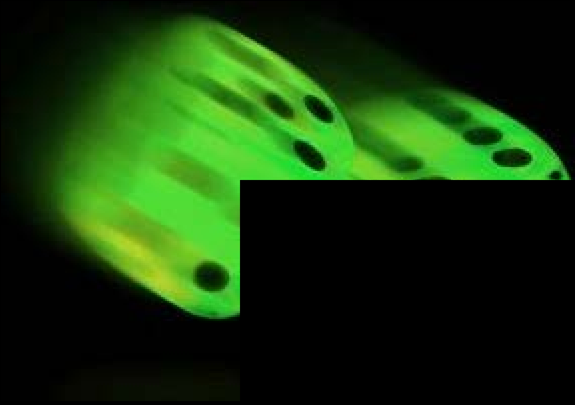
$$\text{normalpdf}(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

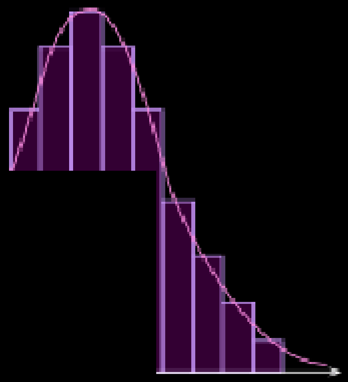
Normal Distribution



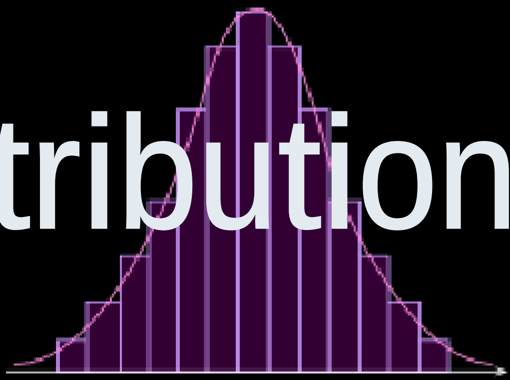
Normal Distribution








A

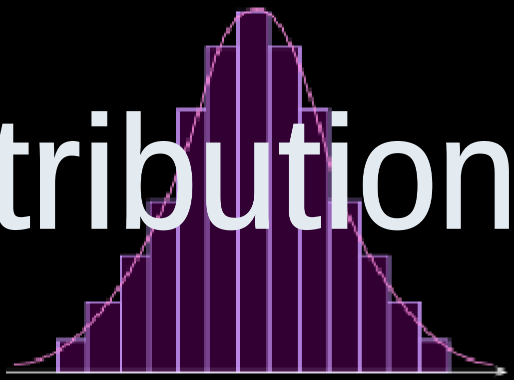


Standard Normal Distribution

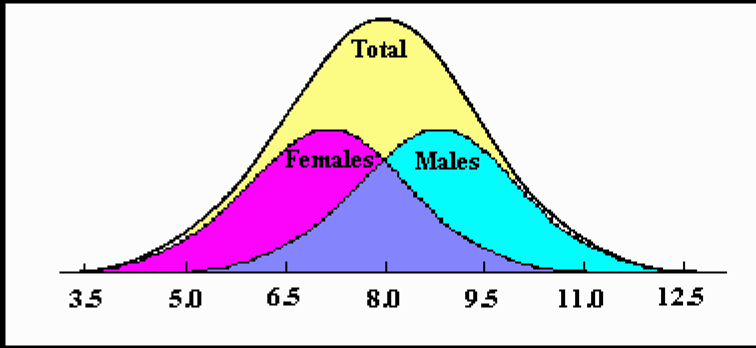
- <u>Description</u>: Special size "bell" curve centered at 0 so that positives represent numbers above the mean and negatives below
- <u>Applications</u>:
  - standardized test scores, evaluating program progress (NCA)
- <u>Formula</u> (normal with $\mu = 0$ and $\sigma = 1$):

$$\text{normalpdf}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Standard Normal Distribution

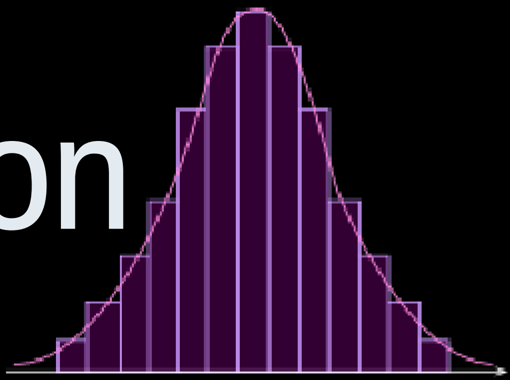


- <u>Notes</u>:
  - $z = (x - \mu) / \sigma$
  - Best used to compare two sets of data



- <u>Example</u>:
  - Shoe size for men (μ = 9, σ = 1.25) and women (μ = 7, σ = 1.25)
  - A woman and a man both wear size 8 shoes have very different *z*-scores (+0.8 for the woman and -0.8 for the man), meaning the woman is in the 79th percentile, whereas the man is in the 21st percentile

Beta Distribution

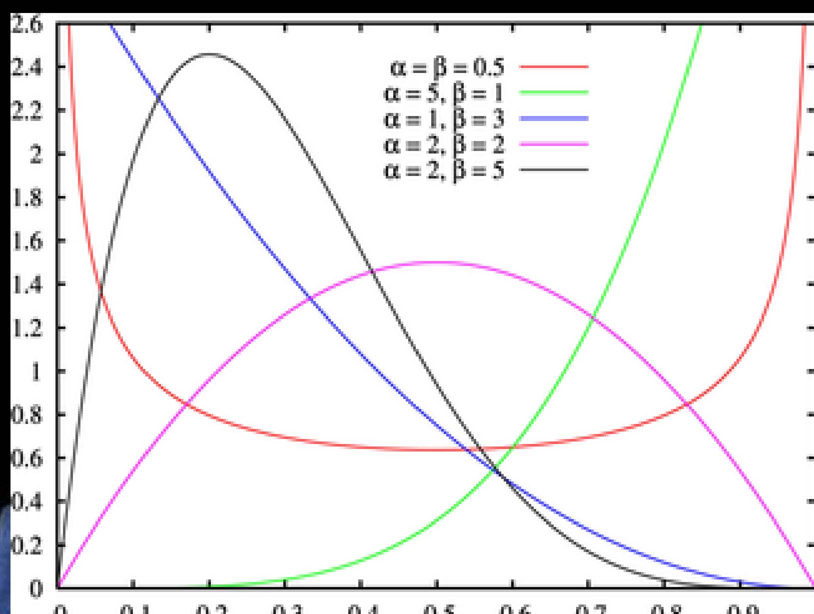


- Description: Probability on a constrained interval
- Applications:
  - Bayesian statistics
  - Project management to model events which take place within an interval defined by a minimum and maximum value
- Formula (parameters *α* and *β* are positive, constant *k* ensures a total probability of 1):

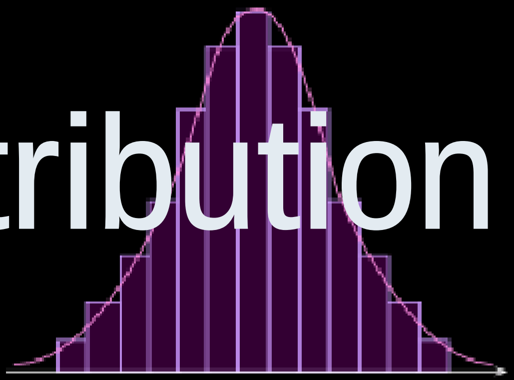
$$f(x) = kx^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } 0 \le x \le 1$$

Beta Distribution

- <u>Notes</u>:
  - The special case of the beta distribution when $\alpha = 1$ and $\beta = 1$ is the uniform distribution
  - If $\alpha = \beta$, the pdf is symmetric about 0.5 (red and purple plots)



Hypergeometric Distribution

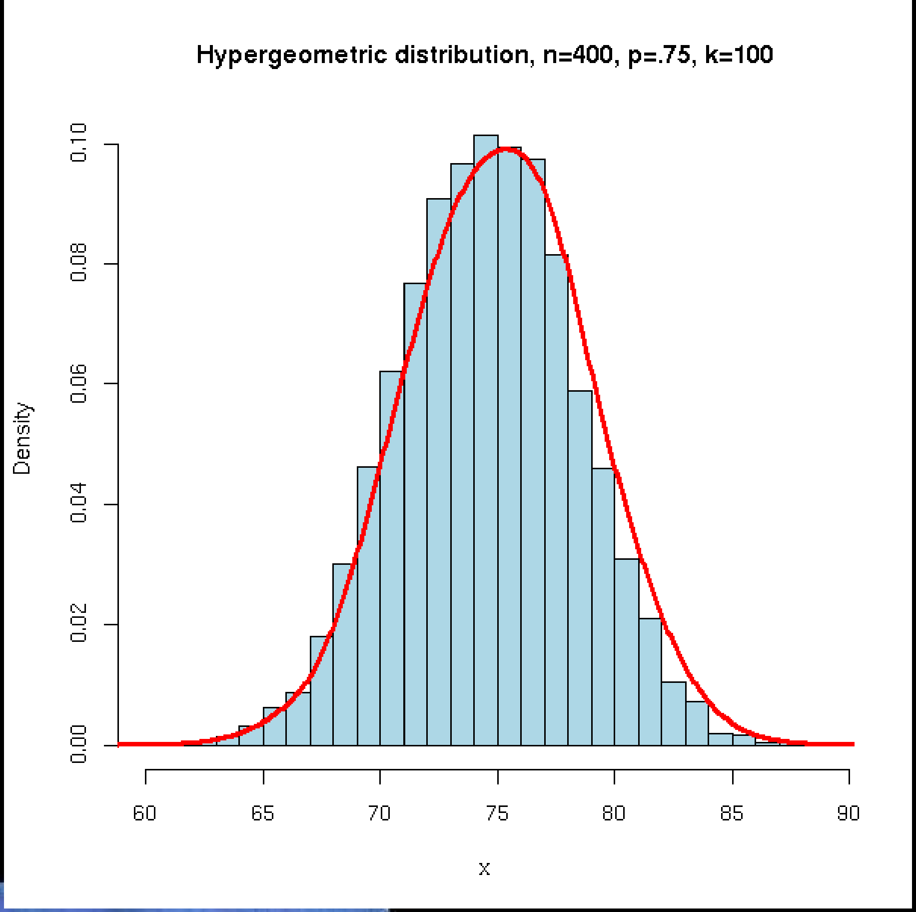


- <u>Description</u>: Taking random samples from a population that consists of objects of two types
- <u>Applications</u>:
  - choosing a committee of five politicians from a legislature of 52 Republicans and 48 Democrats
- <u>Formula</u> (parameters N = population size, t = sample size, r = desired population size):

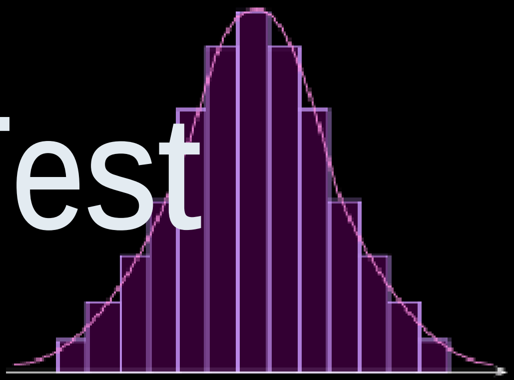
$$f(x) = \frac{\left({}_{r}C_{x}\right)\left({}_{n-r}C_{t-x}\right)}{\left({}_{n}C_{t}\right)}$$

Hypergeometric Distribution

- <u>Notes</u>:
  - Although samples are defined as being taken without replacement, for large populations this is much less important, as the chances of a repetition are small
  - For small populations, hypergeometric is appropriate; for large populations, binomial will suffice



Fisher's Exact Test

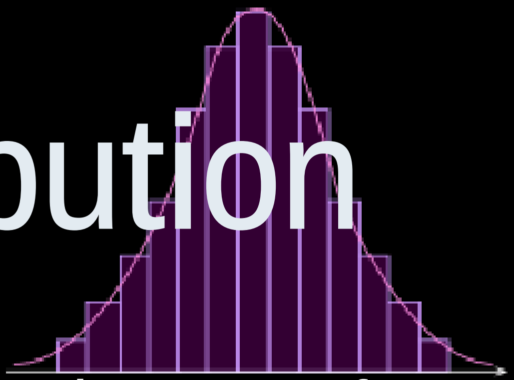


- A test which measures the "extremeness" of data
- Examples:
  - Brain surgery
  - Endangered populations
  - Microscopic observations

Poisson Distribution

- <u>Description</u>: Models the number of events occurring in a time period (0, 1, 2, 3, …)
- <u>Applications</u>:
  - traffic analysis, e.g. number of accidents at an intersection
  - ecological studies, e.g., prairie dogs per square mile
- <u>Formula</u> (parameter λ = mean number of successes in the given period):

$$\text{poissonpdf}(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$



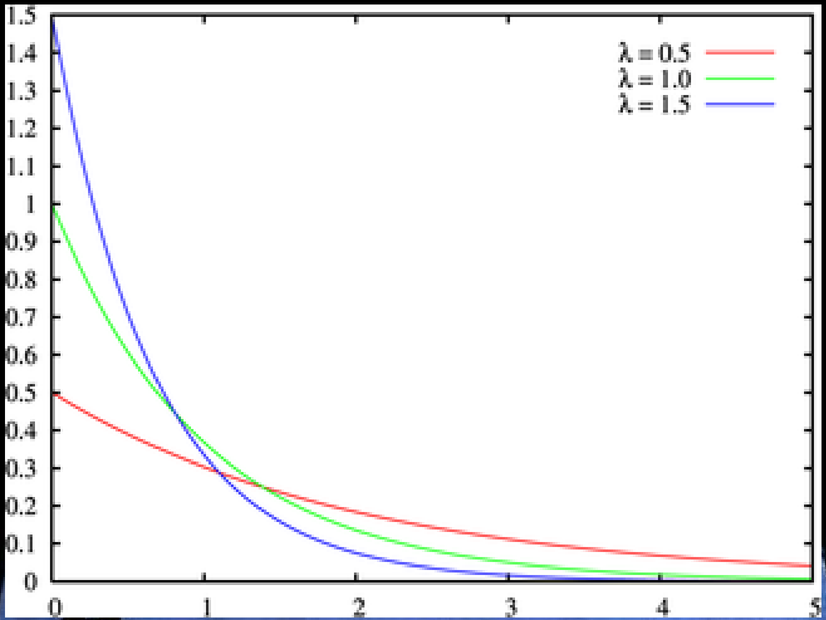
Exponential Distribution

- Description: On a given interval where the rate of change is roughly constant, the exponential distribution is a good model for the next event
- Applications:
  - Used to find the distance between mutations on a DNA strand
  - Radioactive particle decay
  - Time until your next car accident
  - Time until next phone call
- Formula (β = rate parameter):

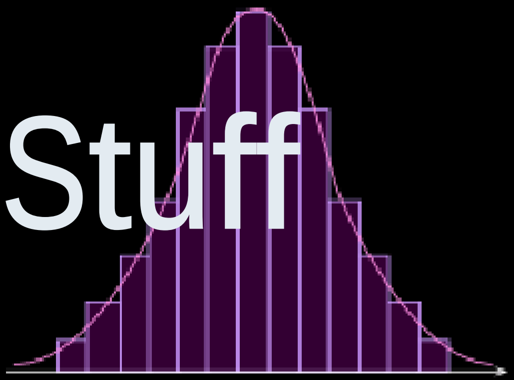
$$f(x) = \beta e^{-\beta x} \quad \text{for } x \geq 0$$

Exponential Distribution

- <u>Note</u>:
  - This distribution is said to be "memoryless" because $P(X > r + s \mid X > r) = P(X > s)$

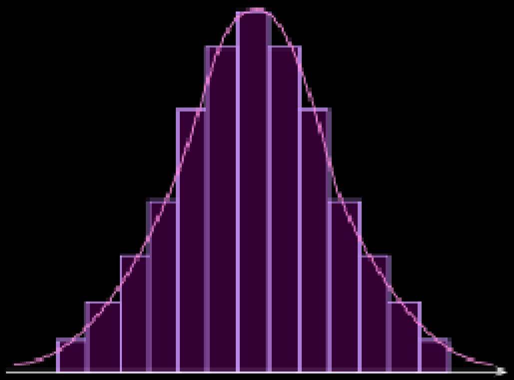


Other Interesting Stuff



- Shakespearean Monkeys
- So you want to fly in an airplane?
- Happy Birthday!
- POWERBALL
- Math can be inquiry-based (!?)

Participants



Ben Dillon
Terri Emerick
Kathy Feltz
Michael Largey
Michael Lewis
Charles Logue
Michelle McConahay

Laura McKenzie
Sean McMillan
Katy McShane
Kasey Ryan
Rich Sypel
John Treacy
Deanna Voss

Tom Edgar (Adviser)