# Statistical Power and Bayesian Inference

# Today’s objectives and plan

Morning:

- Review the hypothesis testing and type I and type II errors; which error is more serious?
- Compute power; understand relationship and trade-offs against other factors (significance level, effect size, variance, sample size) that influence power
- Morning assignment: Compute power and sample size

Afternoon:

- Review conditional probability
- Bayes' rule
- Bayesian inference
- Pair programming: Bayesian analysis, verification by simulation

# Morning: Statistical Power

# Hypothesis Testing: Possible Outcomes

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Fail to reject $H_0$ | Correct Decision<br>$(1 - \alpha)$ | Type II Error<br>$(\beta)$ |
| Reject $H_0$ | Type I Error<br>$(\alpha)$ | Correct Decision<br>$(1 - \beta = \pi)$ |

# Type I and Type II Errors

Type I Error

$= P(\text{Reject } H_0 \mid H_0 \text{ is true})$

$= \alpha$ (alpha, the significance level)

Type II Error

$= P(\text{Fail to reject } H_0 \mid H_0 \text{ is false})$

$= \beta$ (beta) $= 1 - \pi$ (one minus power)

# Which Type of Errors is More Serious?

# Example 1: You have been put on trial for murder

$H_0$: You are innocent (presumption of innocence)

$H_A$: You are guilty

The question for the jury is whether it finds enough evidence that you are guilty

# Example 1: You have been put on trial for murder (cont.)

| | $H_0$ is true<br>You are **truly** innocent | $H_0$ is false<br>You are **truly** guilty |
|---|---|---|
| Fail to reject $H_0$<br>Did not found enough evidence that you are guilty: You are **declared** innocent | Correct | Type II Error<br>You are guilty but set free |
| Reject $H_0$<br>Found enough evidence that you are guilty: You are **declared** guilty | Type I Error<br>You are found guilty of a murder that you did not commit (and the murderer is still free…)<br>**The American justice system puts a lot of emphasis on avoiding type I errors** | Correct |

# Example 2: You are being screened for a disease

$H_0$: You don’t have the disease

$H_A$: You have the disease

The question for the test is whether it finds enough evidence (e.g., markers) that you have the disease

# Example 2: You are being screened for a disease (cont.)

| | $H_0$ is true<br>You **truly** are disease-free | $H_0$ is false<br>You **truly** have the disease |
|---|---|---|
| Fail to reject $H_0$<br>Did not found enough evidence that you are sick: You are **declared** to be disease-free | Correct | Type II Error<br>You'll have the incorrect assurance that you don't have the disease<br>**As a result, you won't be treated for your disease** |
| Reject $H_0$<br>Found enough evidence that you are sick: You are **declared** to have the disease | Type I Error<br>You'll be anxious for a while but this will lead to other testing procedures. Eventually, you'll discover that the initial test was incorrect | Correct |

# Example 3: You are working on a new drug

$H_0$: The new drug is not more effective than current drugs

$H_A$: The new drug more effective than current drugs

Underpowered study:

- With a type II error, you would fail to reject that the new drug is not more effective than current drugs while said drug being really more effective

# Minimizing the type II error ($\beta$)

Is equivalent to maximize $\pi$

$= 1 - \beta$

$= 1 - P(\text{Fail to reject } H_0 \mid H_0 \text{ is false})$

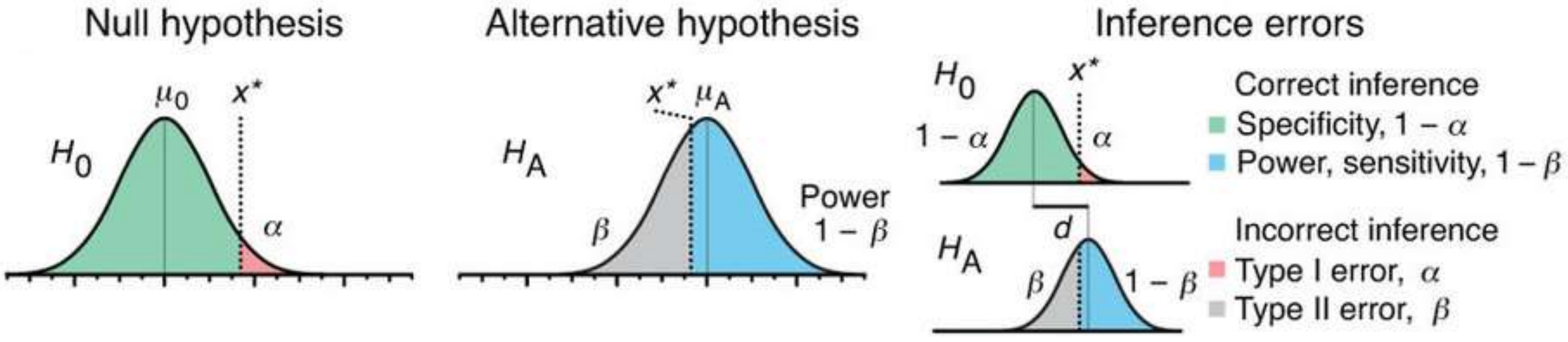
$= P(\text{Reject } H_0 \mid H_0 \text{ is false})$

$= P(\text{Reject } H_0 \mid H_A \text{ is true})$

# Hypothesis Testing: Possible Outcomes

| | $H_0$ is true | $H_0$ is false |
|---|---|---|
| Fail to reject $H_0$ | Correct Decision<br>($1 - \alpha$) | Type II Error<br>($\beta$) |
| Reject $H_0$ | Type I Error<br>($\alpha$) | Correct Decision<br>($1 - \beta = \pi$) |

# Hypothesis Testing: Outcomes and Regions



# Compute Statistical Power of a Test

Example: One-sample Test of Mean

$H_0$: $\mu_1 - \mu_0 = 0$ (e.g., the test finds no evidence of the better effectiveness of the new drug)

$H_A$: $\mu_1 - \mu_0 > 0$ (e.g., the test finds evidence of the better effectiveness of the new drug)

# Compute Statistical Power of a Test (cont.)

First, we reject $H_0$ when:

$$Z_\alpha \le Z = \frac{X - \mu_0}{s/\sqrt{n}} \ or \ \mu_0 + Z_\alpha \frac{s}{\sqrt{n}} = X^* \le X$$

Then, we find the corresponding cut-off of this value under $H_A$ is:

$$X^* = \mu_1 - Z_\beta \frac{s}{\sqrt{n}}$$

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

# Statistical Power and its relationship with other Factors

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

The equation above links the following variables:

- $\alpha$ (type I error; significance level)
- $\beta$ (type II error; $\pi = 1 - \beta$, the statistical power)
- $\mu_1 - \mu_0$ (effect size)
- s (standard deviation)
- n (sample size)

Define 4 of these variables and solve for the remaining one

# A. Type I Error and Type II Error Trade-Off

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If $\alpha \downarrow$ (Type I Error increases):

?

# A. Type I Error and Type II Error Trade-Off (cont.)

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If $\alpha \downarrow$ (Type I Error increases):

- $Z_\alpha \uparrow$, $Z_\beta \downarrow$
- $\beta \uparrow$ (Type II error increases)
- $1 - \beta \downarrow$ (Power decreases)

# B. Power and Effect Size Trade-Off

Assuming all other variables are fixed

$$Z_{\alpha} + Z_{\beta} = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If 1 - β ↑ (Power increases):

?

# B. Power and Effect Size Trade-Off (cont.)

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If $1 - \beta \uparrow$ (Power increases):

- $\beta \downarrow$
- $Z_\beta \uparrow$
- $\mu_1 - \mu_0 \uparrow$ (Effect size increases)

# C. Power and Sample Size Trade-Off

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If 1 - β↑ (Power increases):

?

# C. Power and Sample Size Trade-Off (cont.)

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If 1 - β ↑ (Power increases):

- β ↓
- $Z_\beta$ ↑
- $\frac{\mu_1-\mu_0}{s/\sqrt{n}}$ ↑
- $s/\sqrt{n}$ ↓
- n ↑ (Sample size increases)

# D. Sample Size and Effect Size Trade-Off

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If n ↑ (Sample size increases):

?

# D. Sample Size and Effect Size Trade-Off (cont.)

Assuming all other variables are fixed

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If n $\uparrow$ (Sample size increases):

- $s/\sqrt{n} \downarrow$
- $\mu_1$ - $\mu_0 \uparrow$ (Effect size increases)

Power analysis allows you to determine the sample size needed to detect a particular effect

# Morning assignment

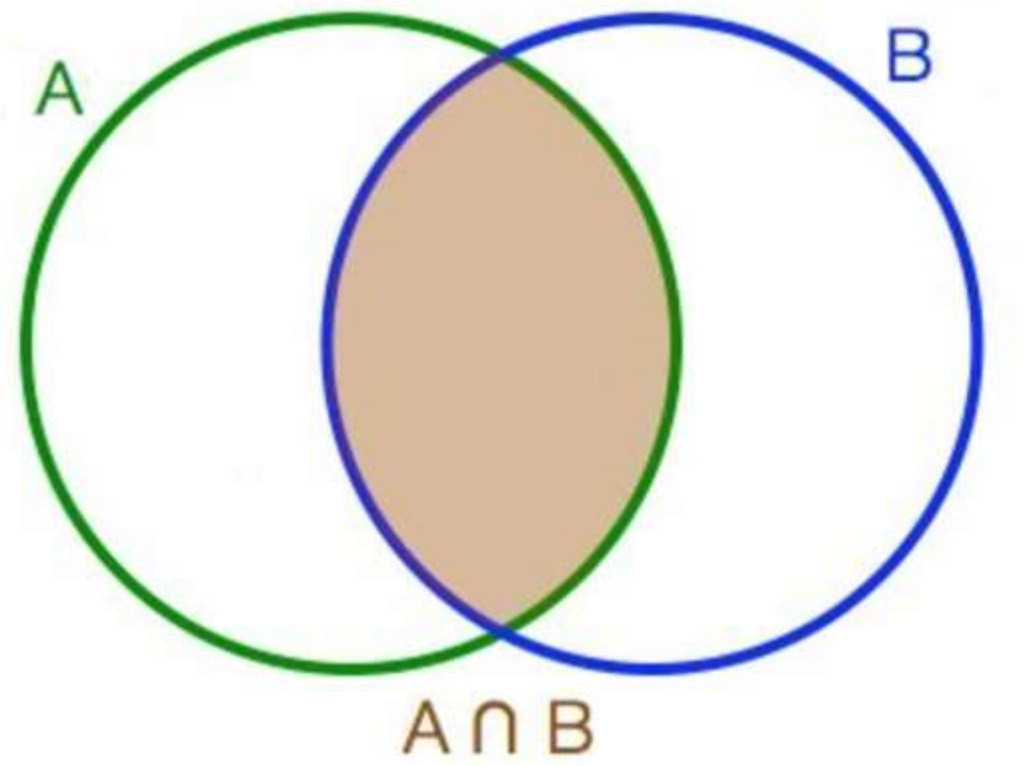
# Afternoon: Bayesian Inference

# Bayes’ Rule: Motivation

- How to relate conditional probabilities between two events
- How to incorporate prior knowledge and belief into interpretation of data

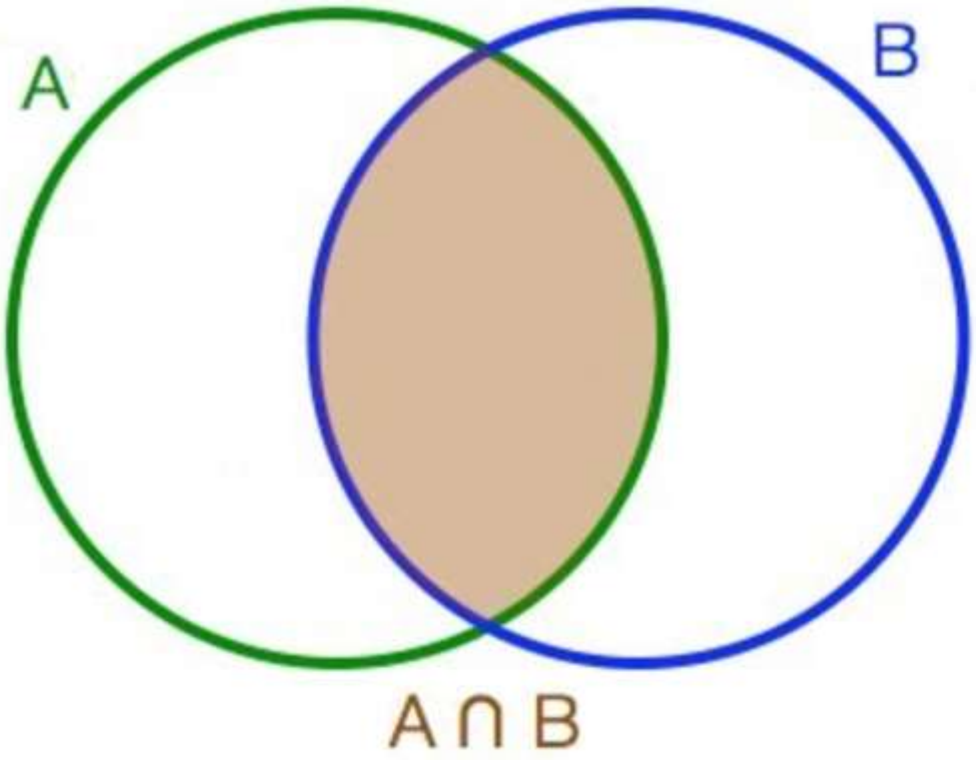
# Conditional Probability Review

$$P(B|A) = \frac{P(A \cap B)}{P(B)}$$



# Poll: What's the probability of rolling a dice with a value less than 4 knowing that the value is odd

$$P(B|A) = \frac{P(A \cap B)}{P(B)}$$

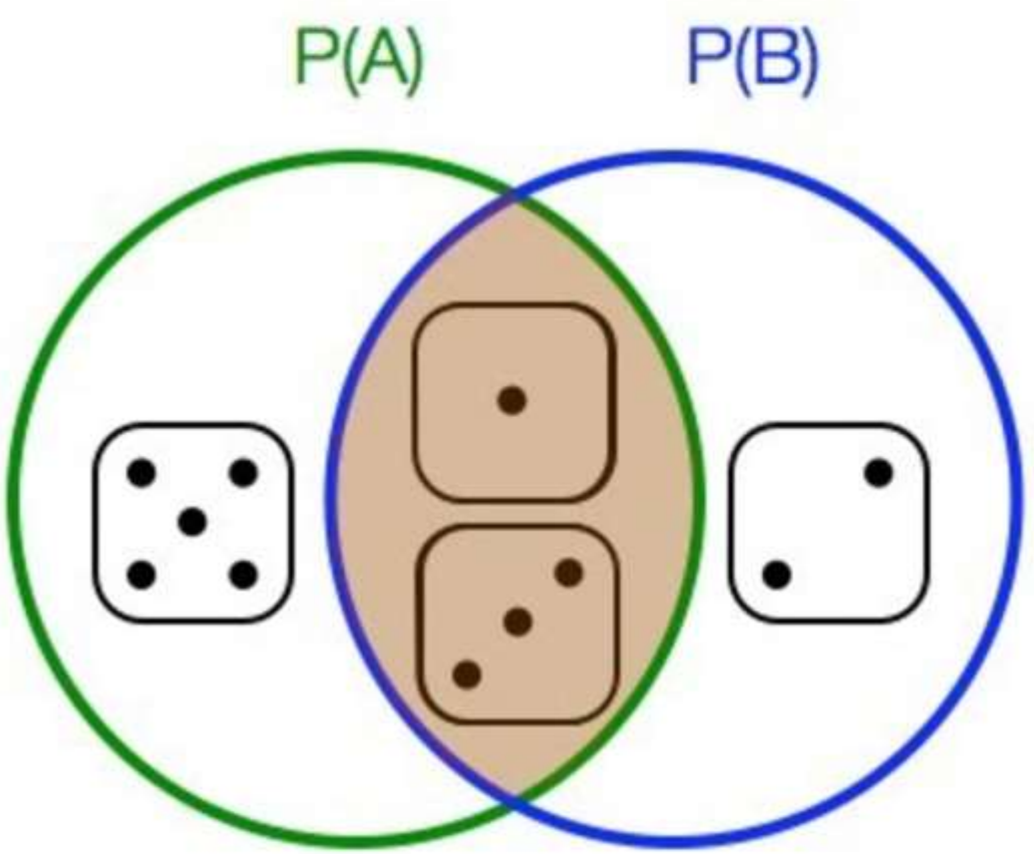


Poll: What's the probability of rolling a dice with a value less than 4 knowing that the value is odd

B = dice with a value less than 4

A = dice with an odd number

$$P(B|A) = \frac{P(A \cap B)}{P(B)} = \frac{2}{3}$$



# Bayes' Rule

likelihood of
the evidence
"E" if the
hypothesis "H"
is true

prior probability
of "H"

posterior
probability of
"H" given the
evidence

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

prior
probability that
the evidence
itself is true

# Bayes’ Rule (cont.)

$$P(H|E) = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)P(H^c)}$$

# Poll: You are planning a picnic today

You are planning a picnic today, but the morning is cloudy. What is the chance that it will rain during the day knowing that:

- 50% of all rainy days start off cloudy
- Cloudy mornings are common (40% of days start cloudy)
- This month is usually a dry month (only 3 of 30 days tend to be rainy)

# Poll: You are planning a picnic today

$$\mathrm{P(rainy\ day\ |\ cloudy\ morning)} =$$

$$\frac{\mathrm{P(cloudy\ morning\ |\ rainy\ day)P(rainy\ day)}}{\mathrm{P(cloudy\ morning)}} =$$

$$\frac{.5\ \ *3/30}{.4} = .125$$

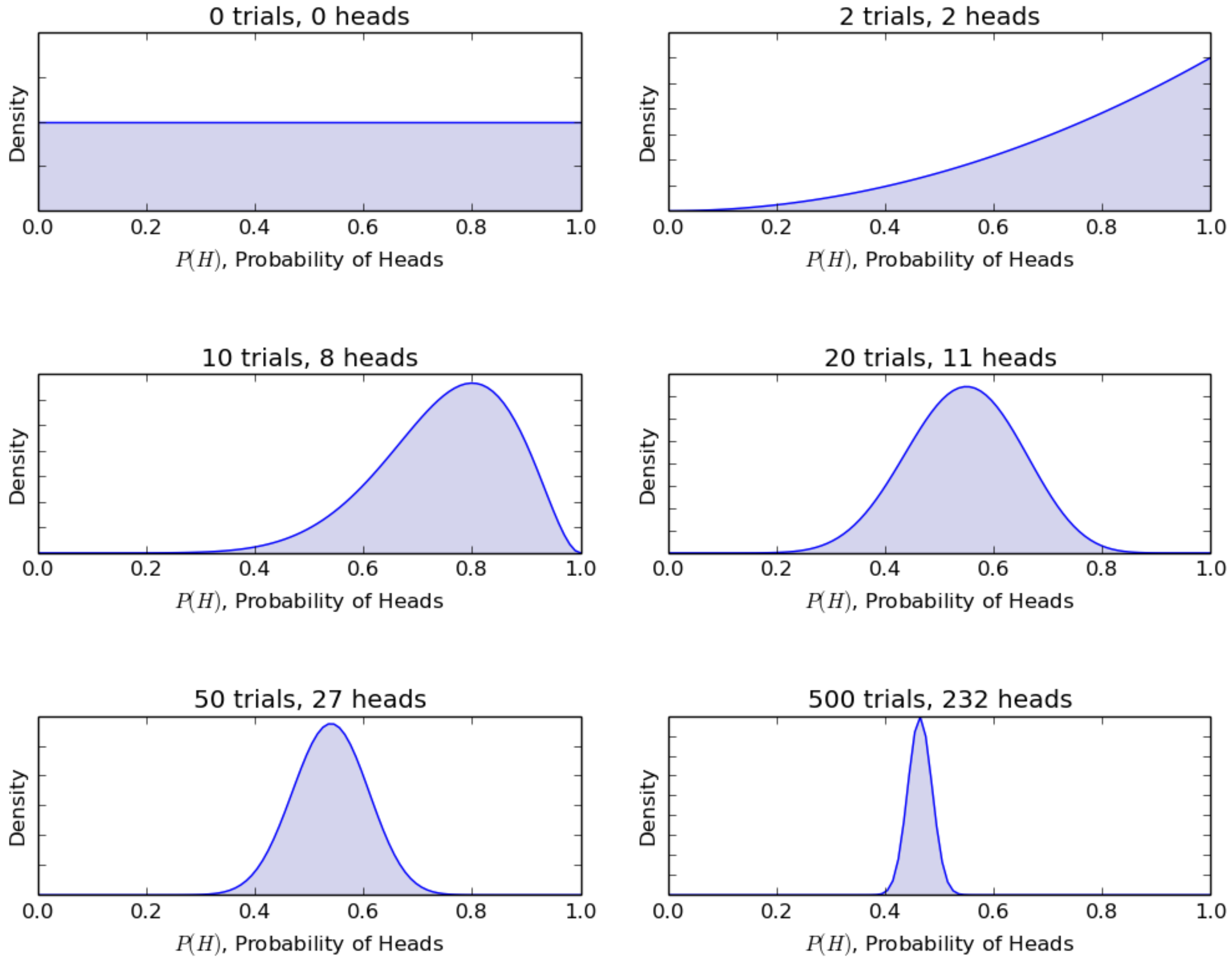
12.5%; not too bad compared with 50%…, let's have a picnic!

# Bayesian Inference

- Bayesian updates his or her beliefs after seeing evidence
  - John Maynard Keynes, a great economist and thinker, said “When the facts change, I change my mind. What do you do, sir?”

- Probability is seen as a measure of believability in events

# Bayesian Updates



# Relating Prior Knowledge/Belief to Data

You have a drawer of 100 coins, 10 of which are biased

P(head | fair) = .5

P(head | biased) = .25

You randomly choose a coin and flip it once. It comes up heads

1. What is P(fair | head)?
2. What if you flip it a second time and it comes up heads again?

# Relating Prior Knowledge/Belief to Data

After the first coin flip, E = [head]

- H = fair coin

$$P(fair \mid head) =$$

$$\frac{P(head \mid fair)P(fair)}{P(head \mid fair)P(fair) + P(head \mid biased)P(biased)} = \frac{.5 \times .9}{.5 \times .9 + .25 \times .1} = .947$$

- H = biased coin

$$P(biased \mid head) =$$

$$\frac{P(head \mid biased)P(biased)}{P(head \mid fair)P(fair) + P(head \mid biased)P(biased)} = \frac{.25 \times .1}{.5 \times .9 + .25 \times .1} = .053$$

# Relating Prior Knowledge/Belief to Data

After the second coin flip, E = [head, head]

- H = fair coin

$$P(fair \mid head) =$$

$$\frac{P(head \mid fair)P(fair)}{P(head \mid fair)P(fair) + P(head \mid biased)P(biased)} = \frac{.5 \times .947}{.5 \times .947 + .25 \times .053} = .973$$

- H = biased coin

$$P(biased \mid head) =$$

$$\frac{P(head \mid biased)P(biased)}{P(head \mid fair)P(fair) + P(head \mid biased)P(biased)} = \frac{.25 \times .947}{.5 \times .947 + .25 \times .0.53} = .027$$

# Afternoon pairing