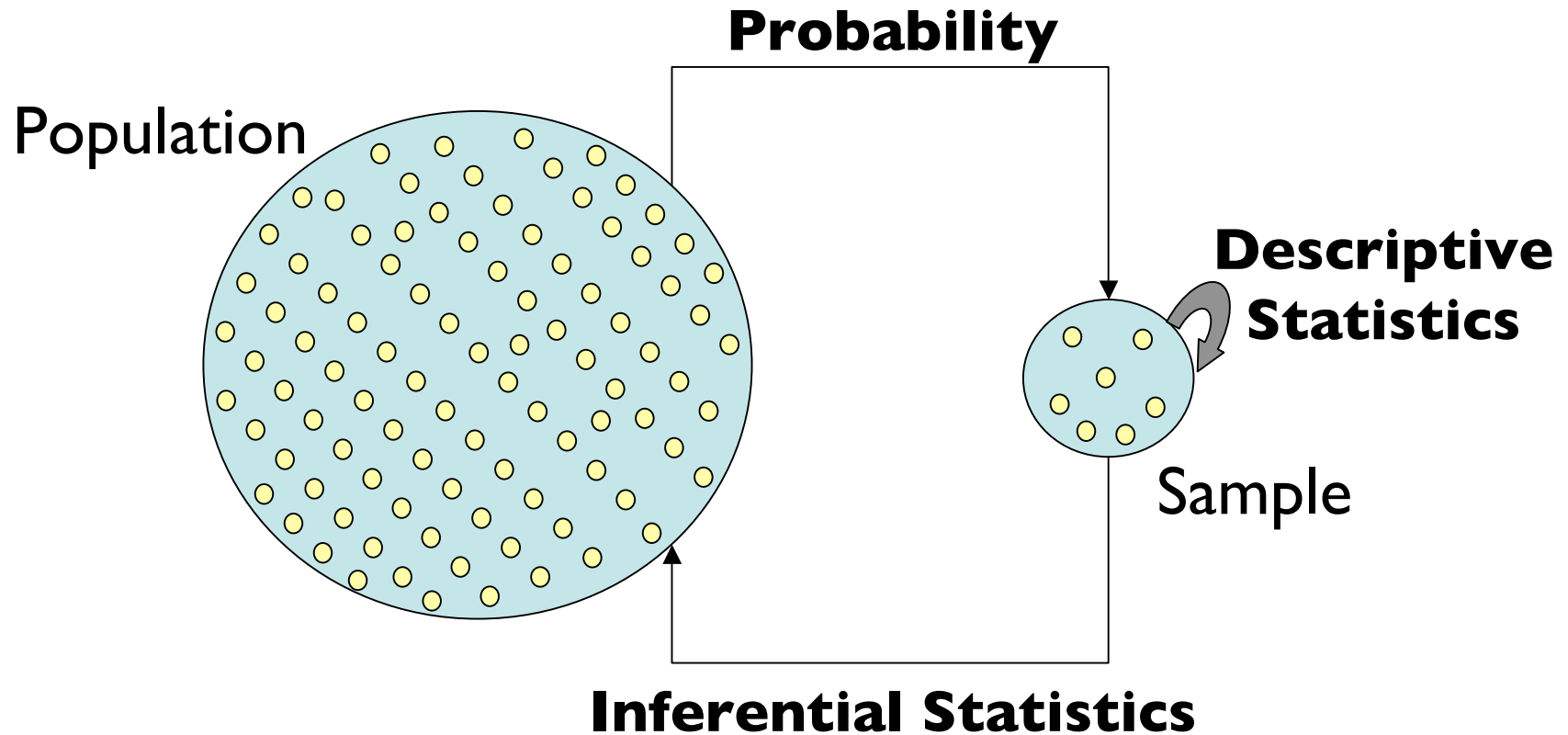


# **Lecture 5: Estimation**

# Goals

- Basic concepts of estimation
- Statistical approaches for estimating parameters
- Parametric interval estimation
- Nonparametric interval estimation (bootstrap)

# “Central Dogma” of Statistics



# Estimation

- **Estimator:** Statistic whose calculated value is used to estimate a population parameter,  $\theta$
- **Estimate:** A particular realization of an estimator,  $\hat{\theta}$
- **Types of Estimators:**
  - point estimate: single number that can be regarded as the most plausible value of  $\theta$
  - interval estimate: a range of numbers, called a confidence interval indicating, can be regarded as likely containing the true value of  $\theta$

# Properties of Good Estimators

- In the ***Frequentist*** world view parameters are fixed, statistics are rv and vary from sample to sample (i.e., have an associated sampling distribution)
- In theory, there are many potential estimators for a population parameter
- What are characteristics of good estimators?

# Statistical Jargon for Good Estimators

## Good Estimators Are:

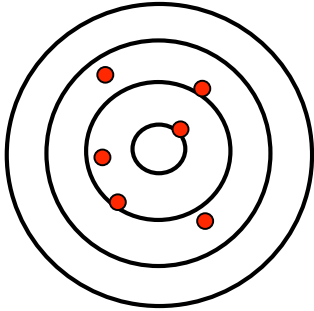
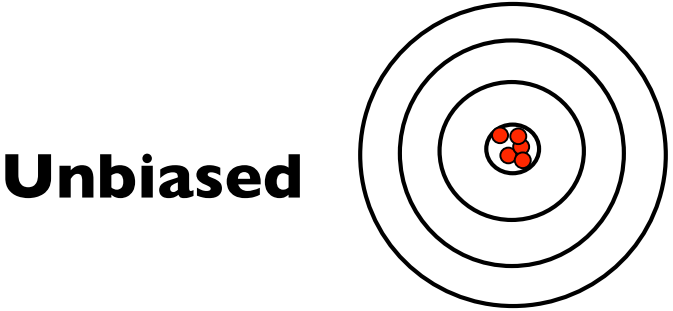
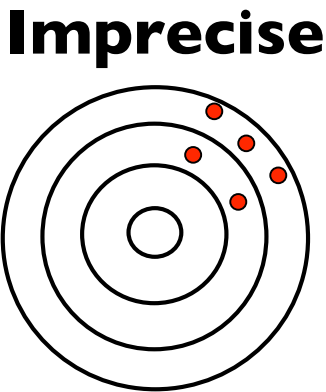
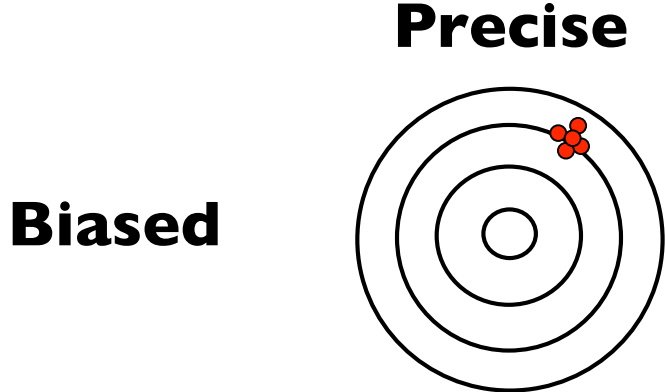
- Consistent: As the sample size increases  $\hat{\theta}$  gets closer to  $\theta$

$$\lim_{n \rightarrow \infty} P\left(\left|\hat{\theta} - \theta\right| > \varepsilon\right) = 0$$

- Unbiased:  $E[\hat{\theta}] = \theta$

- Precise: Sampling distribution of  $\hat{\theta}$  should have a small standard error

# Bias Versus Precision



# Methods of Point Estimation

**1. Method of Moments**

**2. Maximum Likelihood**

**3. Bayesian**



# Methods of Moments

- **Advantage:** simplest approach for constructing an estimator
- **Disadvantage:** usually are not the “best” estimators possible
- **Principle:**

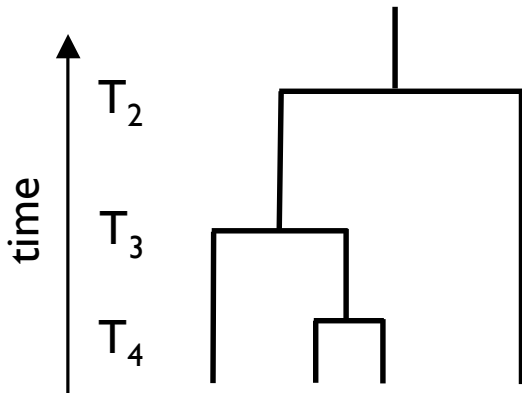
Equate the  $k^{\text{th}}$  population moment  $E[X^k]$  with the  $k^{\text{th}}$  sample moment  $\frac{1}{n} \sum_{i=1}^n X_i^k$  and solve for the unknown parameter

# Method of Moments Example

- How can I estimate the scaled population mutation rate:

$$\theta = 4N_e\mu$$

- Brief (very brief) expose of coalescent theory:



Coalescent times follow a geometric distribution

$$E[T_i] = \frac{4N}{i(i-1)}$$

$$T_c = \sum_{i=2}^n iT_i$$

# Method of Moments Example

$$E[T_c] = \sum_{i=2}^n iE[T_i]$$

# Method of Moments Example

$$E[T_c] = \sum_{i=2}^n iE[T_i] = \sum_{i=2}^n \frac{4Ni}{i(i-1)}$$

# Method of Moments Example

$$E[T_c] = \sum_{i=2}^n iE[T_i] = \sum_{i=2}^n \frac{4Ni}{i(i-1)} = 4N \sum_{i=2}^n \frac{1}{i-1}$$

# Method of Moments Example

$$E[T_c] = \sum_{i=2}^n iE[T_i] = \sum_{i=2}^n \frac{4Ni}{i(i-1)} = 4N \sum_{i=2}^n \frac{1}{i-1}$$

$$E[S_n] = \mu E[T_c]$$

$$E[S_n] = \mu \cdot 4N \sum_{i=2}^n \frac{1}{i-1}$$

$$E[S_n] = \theta \sum_{i=2}^n \frac{1}{i-1} \quad \xrightarrow{\text{mom}} \quad \hat{\theta} = \frac{S_n}{\sum_{i=2}^n \frac{1}{i-1}}$$

# Methods of Point Estimation

1. Method of Moments

**2. Maximum Likelihood**

3. Bayesian

# Introduction to Likelihood

- Before an experiment is performed the outcome is unknown. Probability allows us to predict **unknown** outcomes based on **known** parameters:

$$P(\text{Data} \mid \theta)$$

- For example:

$$P(x \mid n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$



# Introduction to Likelihood

- After an experiment is performed the outcome is **known**. Now we talk about the **likelihood** that a parameter would generate the observed data:

$$L(\theta \mid Data)$$

$$L(\theta \mid Data) = P(Data \mid \theta)$$

- For example:

$$L(p \mid n, x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- Estimation proceeds by finding the value of  $\theta$  that makes the observed data most likely

# Let's Play T/F

- True or False: The maximum likelihood estimate (mle) of  $\theta$  gives us the probability of  $\hat{\theta}$
- **False - why?**
- True or False: The mle of  $\theta$  is the most likely value of  $\hat{\theta}$
- **False - why?**
- True or False: Maximum likelihood is cool

# Formal Statement of ML

- Let  $x_1, x_2, \dots, x_n$  be a sequence of  $n$  observed variables

- Joint probability:

$$P(x_1, x_2, \dots, x_n | \theta) = P(X_1=x_1)P(X_2=x_2)\dots P(X_n=x_n)$$

$$= \prod_{i=1}^n P(X_i = x_i)$$

- Likelihood is then:

$$L(\theta | x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(X_i = x_i)$$

$$\text{Log } L(\theta | x_1, x_2, \dots, x_n) = \sum_{i=1}^n \log[P(X_i = x_i)]$$

# MLE Example

- I want to estimate the recombination fraction between locus A and B from 5 heterozygous (AaBb) parents. I examine 30 gametes for each and observe 4, 3, 5, 6, and 7 recombinant gametes in the five parents. What is the mle of the recombination fraction?

Probability of observing  $X = r$  recombinant gametes for a single parent is binomial:

$$P(X = r) = \binom{n}{r} \theta^r (1 - \theta)^{n-r}$$

# MLE Example: Specifying Likelihood

## Probability:

$$P(r_1, r_2, \dots, r_n \mid \theta, n) = P(R_1 = r_1)P(R_2 = r_2)\dots P(R_5 = r_5)$$

$$P(r_1, r_2, \dots, r_n \mid \theta, n) = \binom{n}{r_1} \theta^{r_1} (1 - \theta)^{n-r_1} \cdot \binom{n}{r_2} \theta^{r_2} (1 - \theta)^{n-r_2} \cdot \dots \cdot \binom{n}{r_5} \theta^{r_5} (1 - \theta)^{n-r_5}$$

## Likelihood:

$$L(\theta \mid r_1, r_2, \dots, r_n, n) = \prod_{i=1}^5 \binom{n}{r_i} \theta^{r_i} (1 - \theta)^{n-r_i}$$

$$\text{Log } L = \sum_{i=1}^5 \log \binom{n}{r_i} + r_i \cdot \log \theta + (n - r_i) \log(1 - \theta)$$

# MLE Example: Maximizing the Likelihood

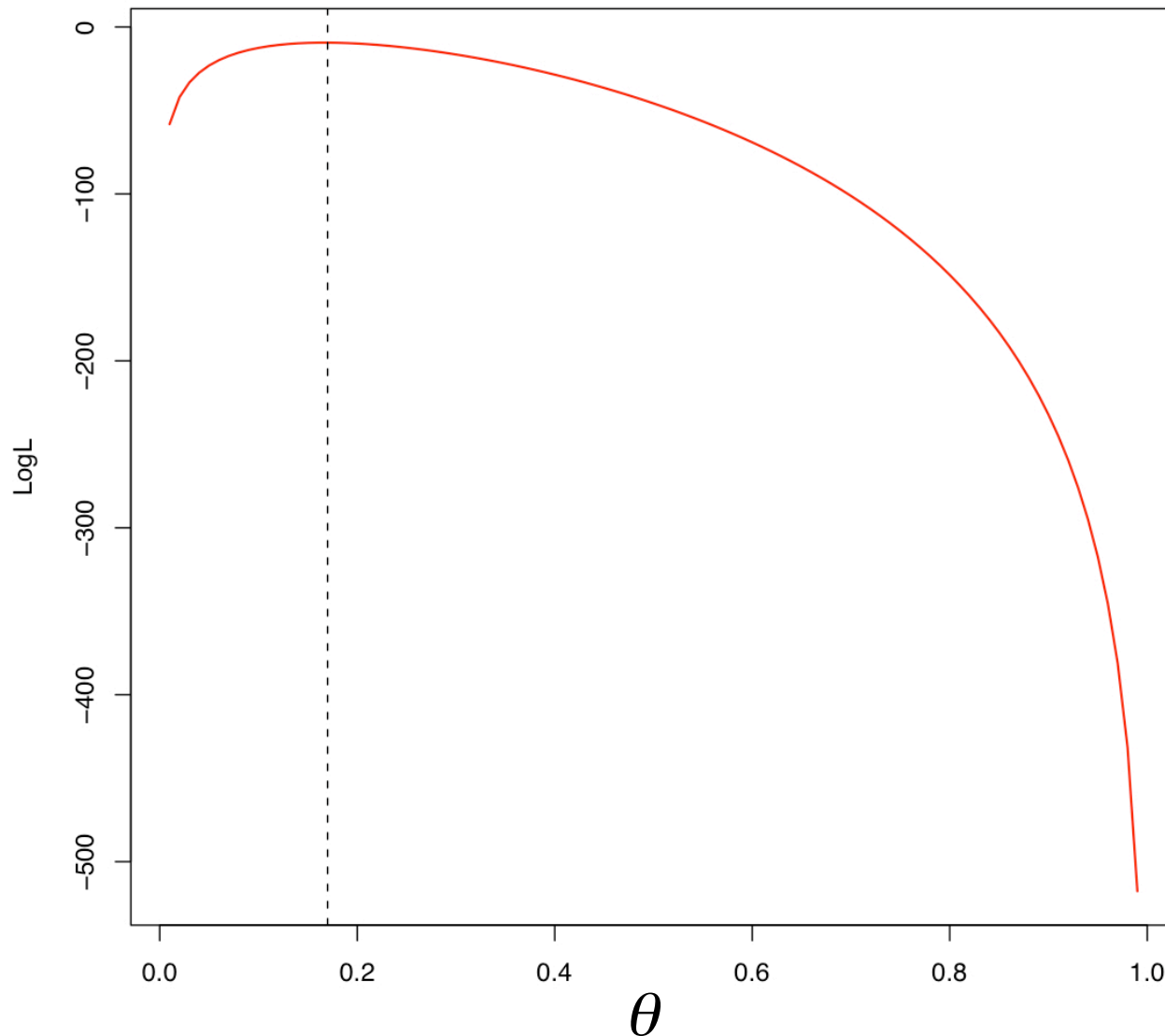
- **Want to find  $p$  such that  $\text{Log } L$  is maximized**

$$\text{Log } L = \sum_{i=1}^5 \log \binom{n}{r_i} + r_i \cdot \log \theta + (n - r_i) \log(1 - \theta)$$

- **How?**

1. Graphically
2. Calculus
3. Numerically

# MLE Example: Finding the mle of $p$



# Methods of Point Estimation

1. Method of Moments

2. Maximum Likelihood

**3. Bayesian**



# World View According to Bayesian's

- The classic philosophy (frequentist) assumes parameters are **fixed** quantities that we want to estimate as precisely as possible
- Bayesian perspective is different: parameters are **random variables** with probabilities assigned to particular values of parameters to reflect the degree of evidence for that value

# Revisiting Bayes Theorem

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

**Discrete**

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

**Continuous**

$$P(B) = \int P(B | A)P(A)dA$$

# Bayesian Estimation

- In order to make probability statements about  $\theta$  given some observed data,  $D$ , we make use of Bayes theorem

$$f(\theta|D) = \frac{f(\theta)f(D|\theta)}{f(D)} = \frac{f(\theta)L(\theta|D)}{\int_{\theta} f(\theta)f(D|\theta)d\theta}$$

Posterior  $\propto$  likelihood  $\times$  prior

The **prior** is the probability of the parameter and represents what was thought before seeing the data.

The **likelihood** is the probability of the data given the parameter and represents the data now available.

The **posterior** represents what is thought given both prior information and the data just seen.

# Bayesian Estimation: “Simple” Example

- I want to estimate the recombination fraction between locus A and B from 5 heterozygous ( $AaBb$ ) parents. I examine 30 gametes for each and observe 4, 3, 5, 6, and 7 recombinant gametes in the five parents. What is the mle of the recombination fraction?
- Tedious to show Bayesian analysis. Let's simplify and ask what the recombination fraction is for parent three, who had 5 observed recombinant gametes.

# Specifying The Posterior Density

$$f(\theta | n = 30, r = 5) = \frac{f(\theta) f(r = 5 | \theta, n = 30)}{\int_0^{0.5} f(r = 5 | \theta, n = 30) f(\theta) d\theta}$$

**prior** =  $f(\theta) = \text{uniform}[0, 0.5] = 0.5$

**likelihood** =  $P(r = 5 | \theta, n = 30) = \binom{30}{5} \theta^5 (1 - \theta)^{25}$

**normalizing constant** =  $\int_0^{0.5} P(r = 5 | \theta, n = 30) f(\theta) d\theta$   
=  $0.5 \cdot \binom{30}{5} \int_0^{0.5} \theta^5 (1 - \theta)^{25} d\theta \approx 6531$

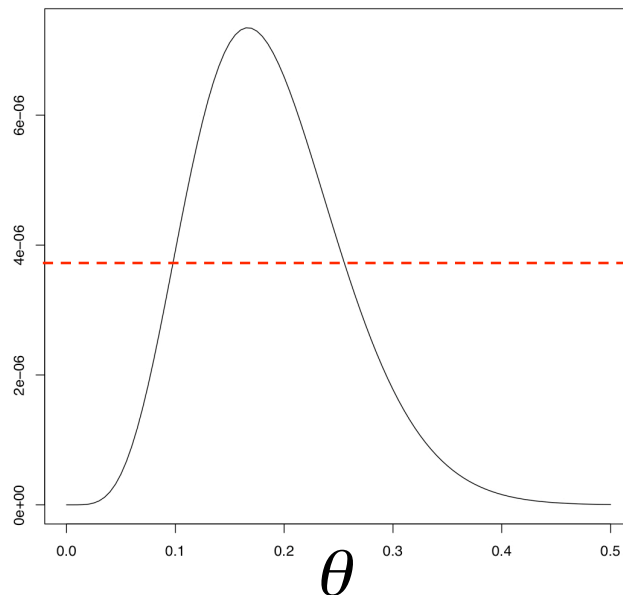
# Specifying The Posterior Density

$$f(\theta | n = 30, r = 5) = \frac{f(\theta) f(r = 5 | \theta, n = 30)}{\int_0^{0.5} f(r = 5 | \theta, n = 30) f(\theta) d\theta}$$

$$f(\theta | n = 30, r = 5) = \frac{0.5 \cdot \binom{30}{5} \theta^5 (1 - \theta)^{25}}{6531}$$

**Ta da...**

$f(\theta | n = 30, r = 5)$

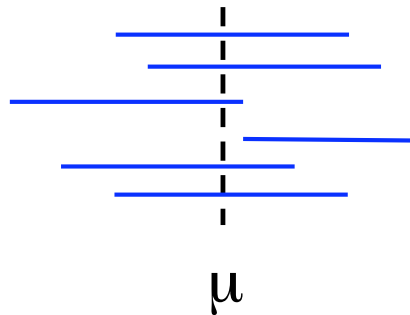


# Interval Estimation

- In addition to point estimates, we also want to understand how much uncertainty is associated with it
- One option is to report the standard error
- Alternatively, we might report a confidence interval
- **Confidence interval:** an interval of plausible values for the parameter being estimated, where degree of plausibility specified by a “confidence level”

# Interpreting a 95% CI

- We calculate a 95% CI for a hypothetical sample mean to be between 20.6 and 35.4. Does this mean there is a 95% probability the true population mean is between 20.6 and 35.4?
- **NO!** Correct interpretation relies on the long-rang frequency interpretation of probability



- Why is this so?