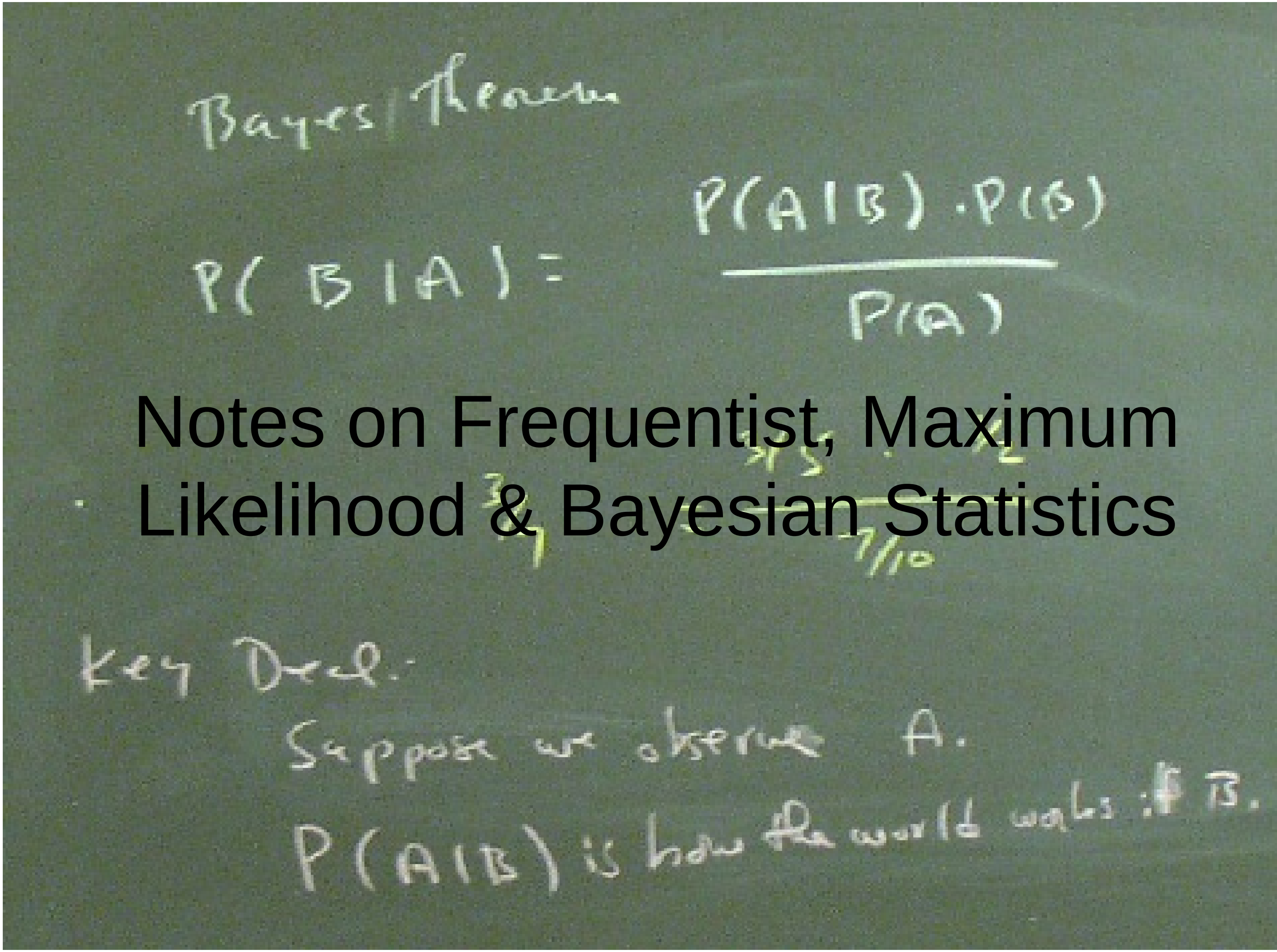


Notes on Frequentist, Maximum Likelihood & Bayesian Statistics

Statistical Methods

- Frequentist
- Likelihood
- Bayesian

- What are they?
- How are the different?

Frequentist Methods

- Probability is a long-term frequency statement about the data
  - if repeated, what proportion of the time would you see that result or a more-extreme one

- Test theoretical value against observed
  - i.e. flip coin 10 times and get 3 heads
    (obs ratio = 3/10)

    theoretical (based upon long-term frequency) = 5/10

- Can only test one hypothesis at a time (H_o and H_a)

Maximum Likelihood

- Maximum Likelihood:
  The best model is the model that maximizes the probability of seeing the observed data

- Important Information:
  - The data
  - The hypothesis (model) – describes the probability of observing the data

Likelihood in Action

Example:

4 sided die and 6 sided die in a bag

Blindly pull out one die and roll a 5…

Prob of rolling a 5 on a 4-sided die $P(5|4) = 0$

Prob of rolling a 5 on a 6-sided die $P(5|6) = 1/6$

Prob of rolling a 5 if it were a 6-sided die

Prob of the **data** (X) given the **hypothesis** (H)

(Conditional)

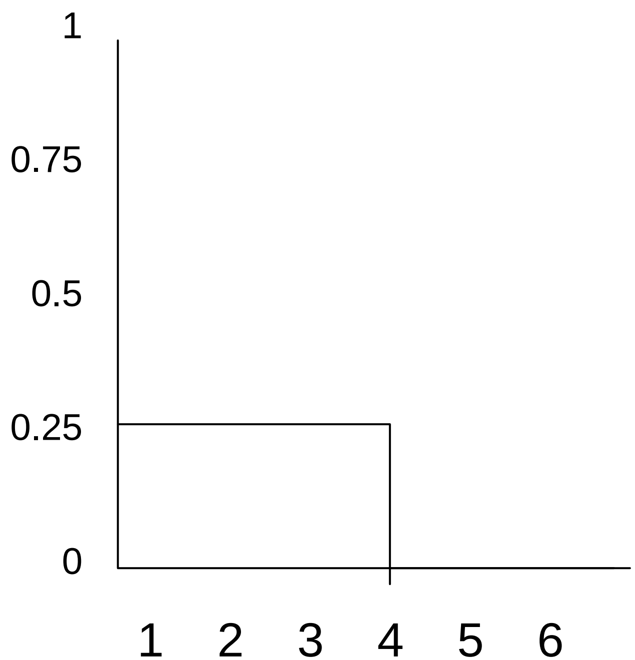
$$P(X|H)$$

Picture and some of dice example adapted from McCoy 2004

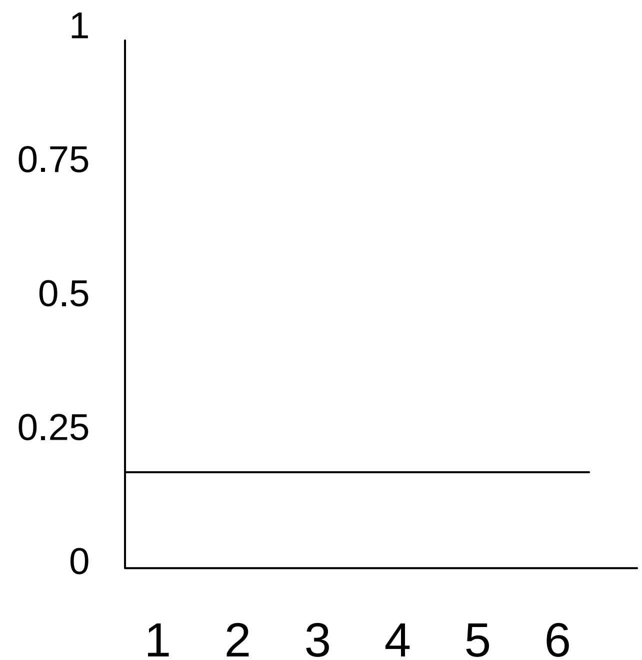
Likelihood and Dice

Probability Distribution

4-sided die



6-sided die



Maximum Likelihood

- Many point estimates from likelihood distribution – searching for the max

- Likelihood ratio

$$\frac{\text{6-sided die } (H_o)}{\text{not 6-sided die } (H_a)} \qquad \frac{P(5|6)}{P(5|4)} = \frac{1/6}{0} = \text{big \#}$$

(high confidence)

- Log Likelihood Ratio – simplifies large numbers

Bayes' Theorem

prior

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$

Likelihood

Posterior Probability

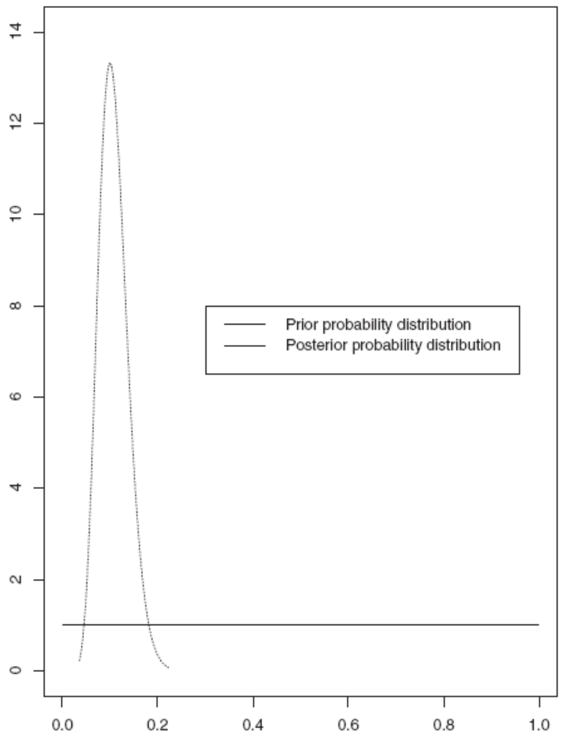
- Still has those important parts – the model (hypothesis) and data
- BUT Likelihood was P (X|H), Bayesian is P (H|X)
  - Incorporation of Prior

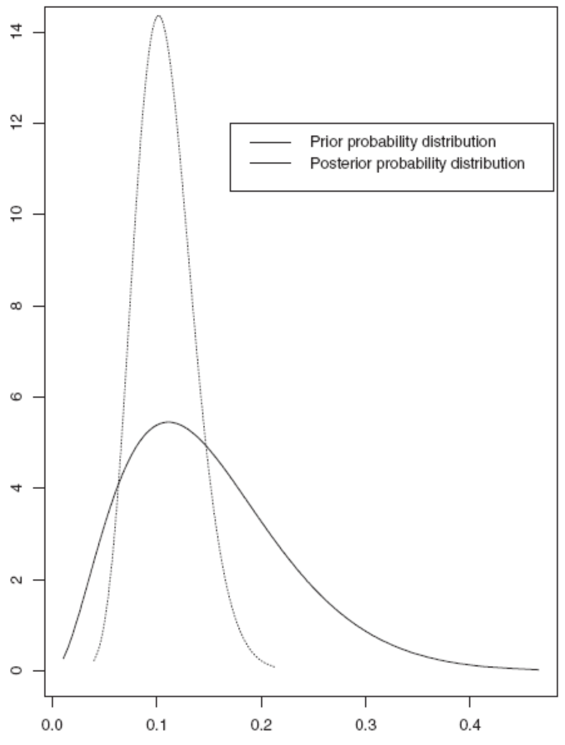
Bayesian Statistics

- The prior: *a priori* knowledge
  - From previous experiments, theory, “gut feeling,” etc.

$$P(H)$$

- Changes the probability distribution





Bayesian Statistics

- Probability is a measure of uncertainty
  - PP of 0.50

    ‘I am 50% sure the Nebraska football team will win a national championship in 2008’

    or…

    ‘I am as sure of that as I am of getting a head when I flip a fair coin one time”

- Choice of prior can influence outcome
  - Most extreme with small data set
- Subjective?

Pros - Bayesian

- Incorporate Priors
- Posterior Probability vs P-value
  - P-value influenced by sample size
  - P-value refers to unobserved data (‘results at least as extreme as that observed’)
- “more natural” – we change our opinions as we gather information

Bayesian in Practice

- Population Genetics
- Forensics
- Gene Finding
- Phylogeography
- Haplotype Inference
- Many more…

Priors – allele freq, heterozygosities, population structure, sampling location, codon bias, mutation rate, etc!

Further Comparisons

Testing Linkage

Traditional:

If the loci are separated by x cM, how unlikely would the observed data be?

Bayesian:

Given the results, what is the probability that two loci are separated by up to x cM?

Confidence

Traditional:

Indirect assessment – compare results of multiple tests to determine level of overall plausibility

Bayesian:

Directly assessed by the posterior probability

Summary

Frequentist

- Probability is a long-term statement
- Asks: do I reject/accept H_a?
- 2-hypotheses at a time
- P-value may be confounded by sample size
- No prior
- Less computationally demanding

Max Likelihood

- $P(X|H)$ – prob of data given the hypothesis
- Uses many point estimates to try to find max
- Choice of model/hypothesis important
- No prior
- Computationally demanding

Bayesian

- $P(H|X)$ – posterior prob of hypothesis given the data
- Includes priors
- Posterior prob is statement of confidence
- Influence of prior confounded by sample size
- Choice of prior subjective
- Computationally demanding

- Bayesian resources
  - Beaumont and Rannala. 2004. The Bayesian Revolution in Genetics. Nature Reviews Genetics. 5:251
  - Shoemaker et al. 1999. Bayesian Statistics in Genetics: A Guide for the Uninitiated. Trends in Genetics. 15:354.
  - Austin et al. 2002. Bayeswatch: An Overview of Bayesian Statistics. J. of Evaluation in Clinical Practice. 8:277.
  - Beerli. 2006. Comparison of Bayesian and Maximum Likelihood inference of population genetic parameters. Bioinformatics. 23:341.
  - McCoy. 2004. Liking Likelihood. Acta Crystallogr. 60:2169.

Quick Comparison

Example from Peter Austin et al. (2002)

- Frequentist – The name of the author on the paper is a constant
  - Can not make probability statements about it
  - No other set of values it can assume
- Bayesian – Not a constant
  - Chance that author is in the witness protection program, or a spy, or there was a typo
  - BUT, you meet Peter at a conference and may now be more certain (prior)