An Introduction to Bayesian Statistics

with two examples from my research

Edward A. Roualdes

Humboldt State University

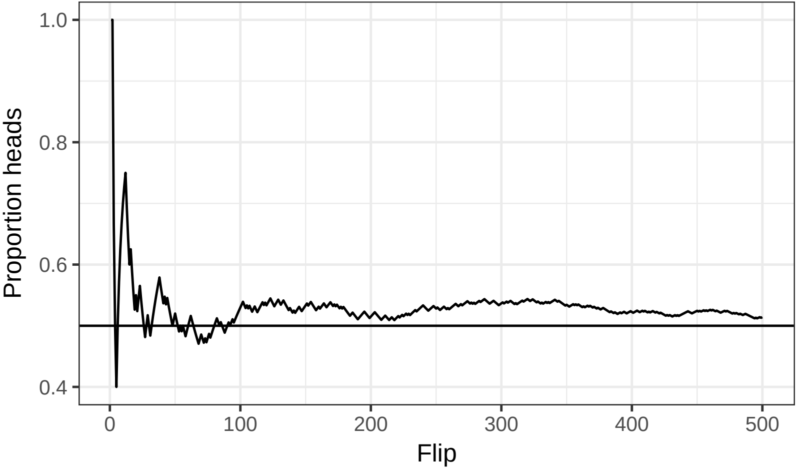
2018-04-05

Outline

outline

Flipping a coin ...

... on a computer.



Convergence in Probability

$$\lim_{N\to\infty} \mathbb{P}\left(\left|N^{-1}\sum_{n=1}^{N} f(x_n) - \mathbb{E}f(x)\right| > \epsilon\right) = 0$$

Frequentist Statistics

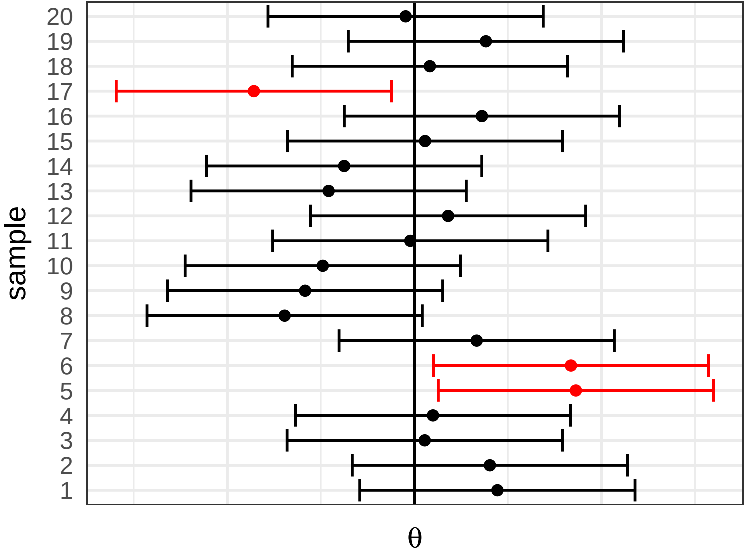
Frequentist statistics treats $\bar{X}$ as a random variable, by imagining replications of it under theoretical resampling.

$$\bar{X} \dot{\sim} \mathcal{N}(\theta, \sigma^2/N)$$

Confidence Intervals, literal translation

Imagine re-sampling R times and creating a confidence interval from each new sample of size N. Then $(1 - \alpha) * 100\%$ of those intervals would include the true population mean, θ.

Confidence Intervals, literal translation



Hypothesis Testing

Or you can evaluate a hypothesis test

$$H_0 : \theta = 0.5$$
$$H_1 : \theta \neq 0.5$$
$$\alpha = 0.05$$

with a p-value calculated from the same distribution that generated the (many) confidence intervals.

outline

Conditional Probability

Bayesian statistics emphasizes conditional probability, following from Jaynes' desiderata [Terenin and Draper., 2017, Jaynes, 2003].

1. States of uncertainty are represented by real numbers.

Conditional Probability

Bayesian statistics emphasizes conditional probability, following from Jaynes' desiderata [Terenin and Draper., 2017, Jaynes, 2003].

1. States of uncertainty are represented by real numbers.
2. Qualitative correspondence with common sense (details omitted).

Conditional Probability

Bayesian statistics emphasizes conditional probability, following from Jaynes' desiderata [Terenin and Draper., 2017, Jaynes, 2003].

1. States of uncertainty are represented by real numbers.
2. Qualitative correspondence with common sense (details omitted).
3. Consistency with true-false logic (details omitted).

Conditional Probability, densities

Some notation:

- y is observed data, and
- θ is parameters to be learned by conditioning on the information contained in y.

Conditional Probability, densities

Some notation:

- y is observed data, and
- θ is parameters to be learned by conditioning on the information contained in y.
- $p(\theta|y)$ is the posterior of θ conditioned on y.

Bayes' Rule

Bayes' rule allows one to reduce uncertainty about θ using the data y.

$$p(\theta|y) =$$

Bayes' Rule

Bayes' rule allows one to reduce uncertainty about θ using the data y.

$$\begin{aligned} p(\theta|y) &= \frac{p(\theta)p(y|\theta)}{p(y)} \\ &\propto p(\theta)p(y|\theta) \\ &= \textit{prior} \cdot \textit{model} \end{aligned}$$

Fun Example, baseball

Suppose you are interested the average of baseball players' batting averages, θ. The simplest model might be

$$p(y|\theta) = \binom{K}{y} \theta^y (1-\theta)^{K-y}.$$

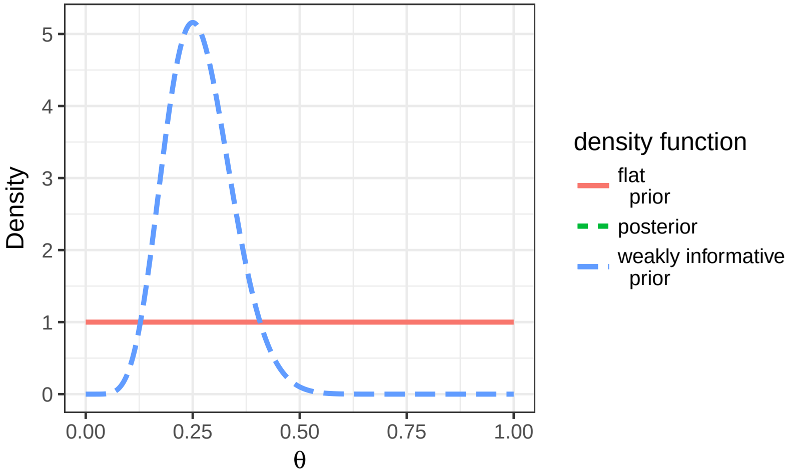
Next a prior, $p(\theta)$.

Some baseball data

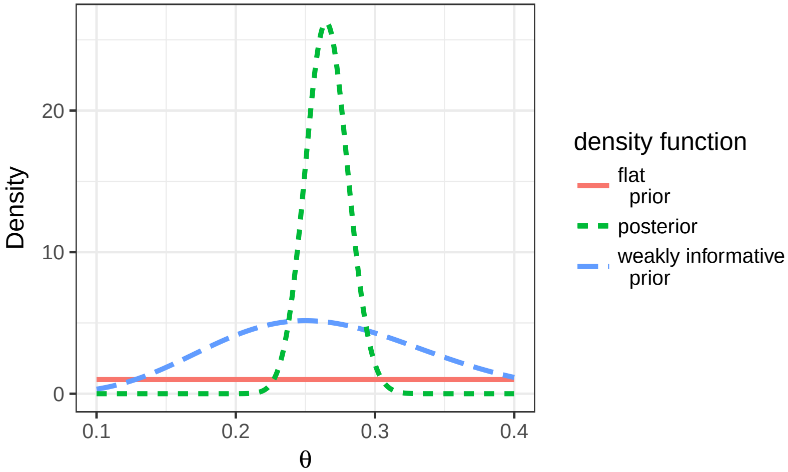
Choosing a prior before looking at (current) data

if some baseball knowledge **then**
  prior $\leftarrow$ weakly informative (utilizes prior information)
else
  prior $\leftarrow$ flat (maximizes entropy)
end if

Fun Example, baseball



Fun Example, baseball



Less-identifiable Posterior

Let $\theta = (\mu_{sp}, \delta_{sp}, \sigma_{sp}, \beta, \sigma)'$, $K = \#$ predictors, $P = \#$ sibling pairs, and $N = \#$ observations.

$$p(\theta|y) \propto \left(1 + \frac{\mu_{sb}}{3}\right)^{-2} \prod_{k=1}^{K} \left(1 + \frac{\beta_k}{3}\right)^{-2}$$
$$\cdot \exp\left(-\sigma_{sb}\right)\sigma^{-P/2} \exp\left(\frac{-\sum_{p=1}^{P} \delta_{sb,p}^2}{2\sigma_{sb}^2}\right)$$
$$\cdot \exp\left(-\sigma\right)\sigma^{-N/2} \exp\left(\frac{-\sum_{n=1}^{N} (y_n - X_n\beta)^2}{2\sigma^2}\right);$$

outline

Approximations

All of statistics is based on approximations.

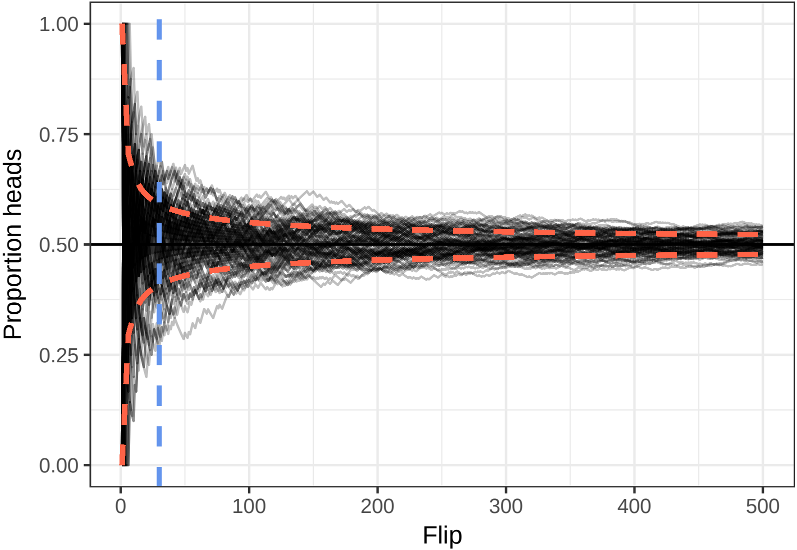
- Frequentist: Central Limit Theorem; $N \rightarrow \infty$

Approximations

All of statistics is based on approximations.

- Frequentist: Central Limit Theorem; $N \to \infty$
- Bayesian: Markov Chain Monte Carlo approximates $\mathbb{E}f$

CLT, $N \to \infty$



Bayesian Summary Statistics

Calculate posterior summary statistics via choice of f.

$$\mathbb{E}f = \int f(\theta)p(\theta|y)d\theta$$

Convergence Almost Surely

$$\mathbb{P}\left(\lim_{N\to\infty} N^{-1}\sum_{n=1}^{N} f(\theta_n) = \mathbb{E}f(\theta)\right) = 1$$

Mean Squared Error

Much of statistics revolves around minimizing

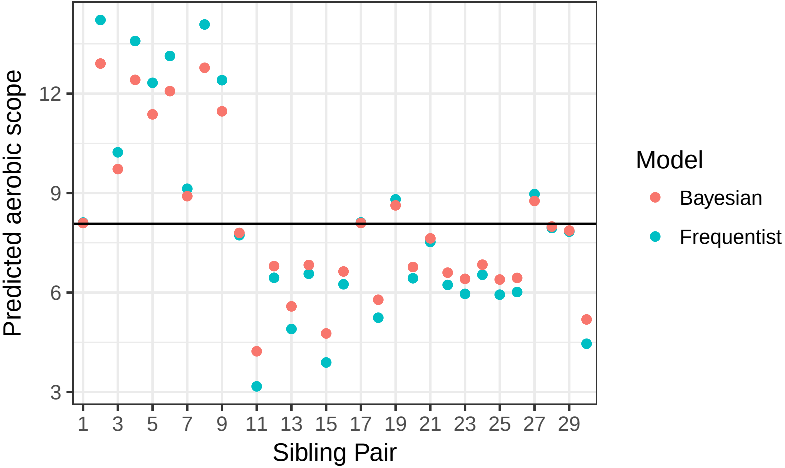
$$\begin{aligned}\mathbb{E}(\hat{\theta} - \theta)^2 &= \mathbb{E}(\theta - \mathbb{E}\hat{\theta})^2 + (\mathbb{E}\hat{\theta} - \theta)^2 \\ &= \mathbb{V}\hat{\theta} + \mathit{Bias}^2(\hat{\theta})\end{aligned}$$

outline

Salmonids

Dr. Banet designed an experiment to directly test the impact of maternal stress on the next generation of Pacific salmonids. Stress was simulated via cortisol baths that were randomly applied to sibling pairs at fertilization.

Salmonids, results



Salmonids, model

Frequentist:

$$AerobicScope_n \sim \mathcal{N}(SiblingPair_{j[n]} + X\beta, \sigma^2)$$

Salmonids, model

Frequentist:

$$AerobicScope_n \sim \mathcal{N}(SiblingPair_{j[n]} + X\beta, \sigma^2)$$

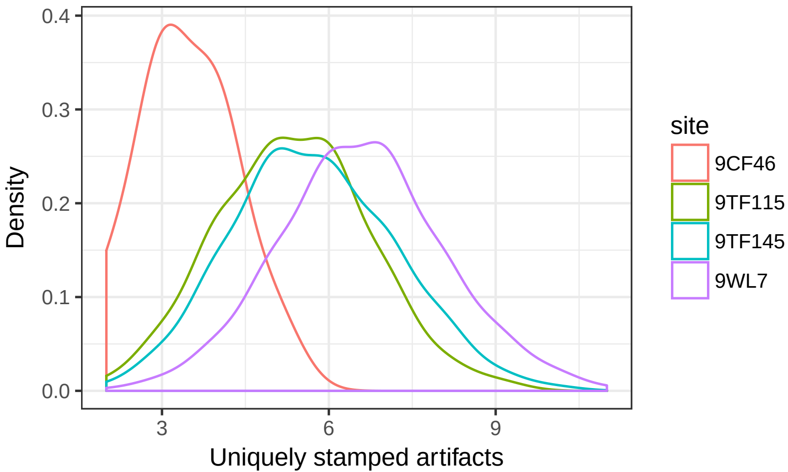
Bayesian:

$$AerobicScope_n \sim \mathcal{N}(SiblingPair_{j[n]} + X\beta, \sigma^2)$$
$$SiblingPair_j \sim \mathcal{N}(\mu_{sp}, \sigma_{sp}^2)$$

Archaeological Types

(Soon to be Dr.) Hensler surveyed multiple sites in middle South, Georgia to estimate the number of artifact classes that population groups at these sites produced. Such measures of *richness* provide evidence for interpretations of changing frequency of interaction with groups living on coastal Georgia.

Archaeological Types, results



Archaeological Types, model

Frequentist: ?
Bayesian:

$$
\begin{aligned}
x_n &:= d_{k_n} \text{ for } n = 1, \ldots, N \\
k_n &\sim \mathit{Multinomial}(1, \boldsymbol{\pi}) \\
\boldsymbol{\pi} &\sim \mathit{Dirichlet}(\mathbf{1})
\end{aligned}
$$

outline

references I

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*, volume e. CRC press Boca Raton, FL, 2014.

Edwin T Jaynes. Probability theory: The logic of science. 2003.

Alex Terenin and D. Draper. Cox's theorem and the jaynesian interpretation of probability. *arXiv:1507.06597*, 2017.