Introduction to Bayesian Statistics

Patrick Lam

Bayesian versus Non-Bayesian

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian Approach

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian Approach

- Parameters are random variables with distributions attached to them

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian Approach

- Parameters are random variables with distributions attached to them
- Subjective notion of probability (prior) combined with data

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian Approach

- Parameters are random variables with distributions attached to them
- Subjective notion of probability (prior) combined with data
- Does not require large sample approximations

Bayesian versus Non-Bayesian

Non-Bayesian Approach:

- Parameters are fixed at their true but unknown value
- Objective notion of probability based on repeated sampling
- Large sample properties/asymptotic approximations
- Maximizing a likelihood

Bayesian Approach

- Parameters are random variables with distributions attached to them
- Subjective notion of probability (prior) combined with data
- Does not require large sample approximations
- Simulation-based approach

The Basics of Bayesian Statistics

The Basics of Bayesian Statistics

Based on Bayes' Rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ are our parameters and y is our data.

The Basics of Bayesian Statistics

Based on Bayes' Rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ are our parameters and y is our data.

We have a posterior density,

The Basics of Bayesian Statistics

Based on Bayes' Rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ are our parameters and y is our data.

We have a posterior density, sampling density (or likelihood),

The Basics of Bayesian Statistics

Based on Bayes' Rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ are our parameters and y is our data.

We have a posterior density, sampling density (or likelihood), prior density,

The Basics of Bayesian Statistics

Based on Bayes' Rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

where θ are our parameters and y is our data.

We have a posterior density, sampling density (or likelihood), prior density, and a normalizing constant (which we typically do not need to find).

Why Bayesian?

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)
- Results approximate MLE results as n increases

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)
- Results approximate MLE results as n increases
- Confidence intervals have a more intuitive meaning (we call them credible sets)

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)
- Results approximate MLE results as n increases
- Confidence intervals have a more intuitive meaning (we call them credible sets)
- Ability to find more quantities of interest (for example, $P(\theta > .3)$ or P(Obama is more left than Kerry) in ideal point estimation)

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)
- Results approximate MLE results as n increases
- Confidence intervals have a more intuitive meaning (we call them credible sets)
- Ability to find more quantities of interest (for example, $P(\theta > .3)$ or $P(\text{Obama is more left than Kerry})$ in ideal point estimation)
- Easily set up and estimate difficult models

Why Bayesian?

- Ability to incorporate prior knowledge (perhaps qualitative knowledge)
- Results approximate MLE results as n increases
- Confidence intervals have a more intuitive meaning (we call them credible sets)
- Ability to find more quantities of interest (for example, $P(\theta > .3)$ or $P(\text{Obama is more left than Kerry})$ in ideal point estimation)
- Easily set up and estimate difficult models
- Priors often help with identification

Why Not Bayesian?

Why Not Bayesian?

- It's hard

Why Not Bayesian?

- It's hard
- Computationally intensive

Why Not Bayesian?

- It's hard
- Computationally intensive
- Need defense of priors or sensitivity analyses of prior specification

Why Not Bayesian?

- It's hard
- Computationally intensive
- Need defense of priors or sensitivity analyses of prior specification
- No guarantee of Markov Chain convergence

Why Not Bayesian?

- It's hard
- Computationally intensive
- Need defense of priors or sensitivity analyses of prior specification
- No guarantee of Markov Chain convergence

Something to think about:

Is MLE/frequentist approach simply Bayesian statistics with an uninformative prior?