Lecture 9: Wrap-up

Professor Alexander Franks

2020-12-07

Why Bayesian statistics?

- Simulation a useful strategy
- Combining knowledge
- Subjectivity / usefulness of prior
- Bias and Variance.
- Comparison to Frequentist Inference.
- Bayesian Updating.

Jobs

Jobs Companies Messages Developer Story Job Preferences

Bayesian Statistician, Research & Development

Houston Astros Baseball R&D | Houston, TX

Paid relocation

Overview Company More Jobs

About this job

Job type: **Full-time**

Experience level: **Mid-Level, Senior**

Role: **Data Scientist**

Industry: **Biomechanics, Data & Analytics, Sports**

Company size: **11–50 people**

Company type: **Private**

Technologies

r stan python mcmc bayesian

Jobs

Job description

The Houston Astros Baseball Club is accepting applications for a Bayesian Statistician to join our growing Research & Development team within Baseball Operations. We are seeking an applicant with a strong knowledge of Bayesian statistics to plan, design, and build new models, visualizations, and tools to support (and collaborate with) all facets of Baseball Operations: scouting; player development; player acquisition; video; and more. This position will work closely with a cross-functional agile team to use Bayesian methods and tools that support effective understanding of baseball data and decision making to help the Astros stay ahead of the competition.

Role Responsibilities

- Develop Bayesian models to support Baseball Operations' research in all areas of decision making including player evaluation, roster construction, in-game tactics, and more
- Implement Bayesian methods to improve the organization's understanding of baseball data
- Design Bayesian frameworks for new research methodologies and experimentation
- Communicate closely with front office, coaching staff, and scouting personnel in the gathering and application of baseball information

Jobs

Data analyst (Part-Time or Full-Time)

Center for Policing Equity - Los Angeles, CA
Full-time, Part-time

Apply Now

Data analyst (Part-Time or Full-Time) ; Location flexible

About The Center For Policing Equity

The Center for Policing Equity (CPE) is a research and action think tank that, through evidence-based approaches to social justice, conducts research and uses data to create levers for social, cultural, and policy change.

Qualifications

- Background in statistics/data science with specific experience in performing multiple regression, multilevel (hierarchical) modeling, and Bayesian inference
- Highly proficient in R and/or Python
- Ability to produce markdown notebooks with Jupyter and/or RMD/Knitr

Why Bayesian Modeling?

- Forces you to carefully and explicitly modeling your data generating process

- Prescriptive: once I have a model (including the prior distribution) in theory I know how to do inference

  - "Turn the Bayesian crank"

  - Leads to many new possible estimators

Why Bayesian Modeling?

- Forces you to carefully and explicitly modeling your data generating process
- Prescriptive: once I have a model (including the prior distribution) in theory I know how to do inference
  - "Turn the Bayesian crank"
  - Leads to many new possible estimators
- Easy to "borrow strength" and share information across observations
  - Hierarchical modeling!
- Model checking is fundamental part of the process
- Is frequentist inference still important? Yes!
  - Calibration

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
    - Includes a prior distributions for parameters
    - Could be very complex and include hiearchies of parameters

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
  - Includes a prior distributions for parameters
  - Could be very complex and include hiearchies of parameters
- 2. Compute the posterior distribution

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
    - Includes a prior distributions for parameters
    - Could be very complex and include hiearchies of parameters
- 2. Compute the posterior distribution
- 3. Summarize your results (Integrals)
    - Inference and interpretation for parameters
    - (Markov Chain) Monte Carlo

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
    - Includes a prior distributions for parameters
    - Could be very complex and include hiearchies of parameters
- 2. Compute the posterior distribution
- 3. Summarize your results
    - Inference and interpretation for parameters
    - (Markov Chain) Monte Carlo
- 4. Identify any model misfit (Posterior predictive checking)

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
  - Includes a prior distributions for parameters
  - Could be very complex and include hiearchies of parameters
- 2. Compute the posterior distribution
- 3. Summarize your results
  - Inference and interpretation for parameters
  - (Markov Chain) Monte Carlo
- 4. Identify any model misfit (Posterior predictive checking)
- 5. Refine and rebuild (go to step 1).

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
    - Includes a prior distributions for parameters
    - Could be very complex and include hiearchies of parameters
- 2. Compute the posterior distribution
- 3. Summarize your results
    - Inference and interpretation for parameters
    - (Markov Chain) Monte Carlo
- 4. Identify any model misfit (Posterior predictive checking)
- 5. Refine and rebuild (go to step 1).
- 6. Eventually... communicate results and/or make a decision!

1. Propose a Data Generating Process

1. Propose a Data Generating Process

1. Propose a Data Generating Process

- Every distribution has a "story"
  - Thinking in anologies can be very powerful
- Prior distributions can have stories too (pseudo-counts)
- How do these stories fit together?
  - Hierarchical modeling

2/3. Compute and summarizing the posterior distribution

- Easy to write down: proportional to likelihood $\times$ prior
- Hard to summarize: usual need Monte Carlo techniques
  - Estimate posterior means of interpretable parameters
  - Compute probability intervals (quantile or HPD)
  - Estimate posterior predictive distributions
- Currently, the most common approach for summarization is MCMC

Challenges in MCMC

- Modern models often have *many* parameters. Large models pose a challenge for MCMC.
- When there are thousands or more parameters
  - MCMC may take a long time to conververge to the stationary distribution
  - In Metropolis-Hastings we have many tuning parameters for the proposal distribution
  - Gibbs sampling has no tuning parameters, but does not work well for highly correlated posterior distributions
- In general, MCMC is very slow relative to optimization methods (like OLS)

Modern MCMC

- Gibbs and Metropolis samplers have a "random walk" behavior
  - Induces autocorrelation
  - Makes it difficult to explore the posterior space
- Hamiltonian Monte Carlo (HMC) is an MCMC method that borrows an idea from physics to reduce this problem

HMC

- Imagine a marble on a frictionless surface. The location of the marble is the current value of θ_t
- The negative posterior density is the "height" of the surface
- Also need to compute the gradient of the posterior ("slopes" in all directions)
- Each iteration we flick the marble with some velocity in a random direction
- Regions of high posterior density are like "wells"

HMC

- In physics the Hamiltonian is the sum of the kinetic energies of all the particles, plus the potential energy of the particles associated with the system.
    - As our proposal, we randomly sample a momentum for the marble and update its position accordingly
    - Can think of HMC as the MH algorithm with a very clever jumping/proposal rule

HMC

Try out HMC at:

https://chi-feng.github.io/mcmc-demo/app.html

- Choose "HamiltonianMC" algorithm
- Experiment by sampling from different target distributions
- Compare to the Random Walk Metropolis
- Stan uses a version of HMC

Approximate Inference

- MCMC can be very slow in high dimensional problems
- Idea: find a distribution that is easy to sample from which closely approximate $p(\theta \mid y)$
- A couple of examples
  - Laplace Approximation
  - Variational Bayes

Approximate w/ multivariate normal Distribution.

Flexible approximation to posterior

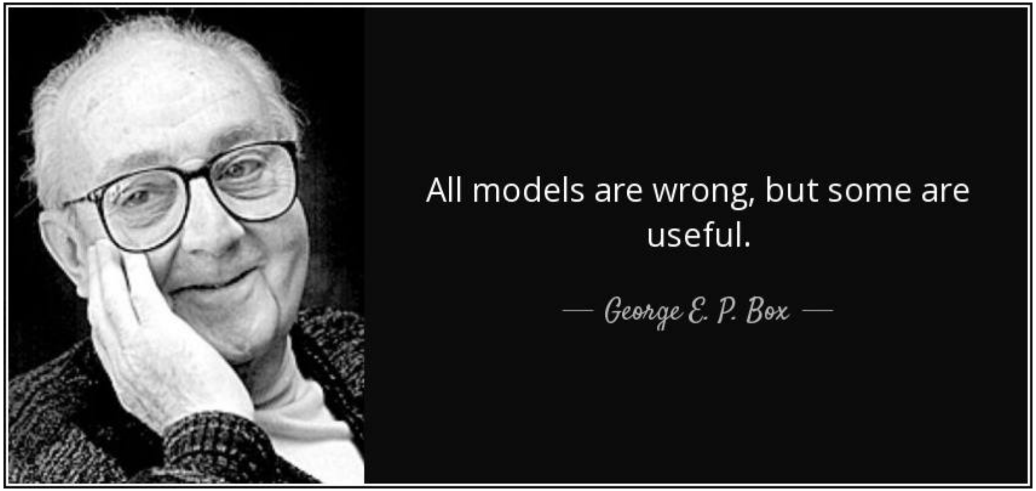
Numerical Integration

e.g. Bayesian Quadrature

3. Summarize your results

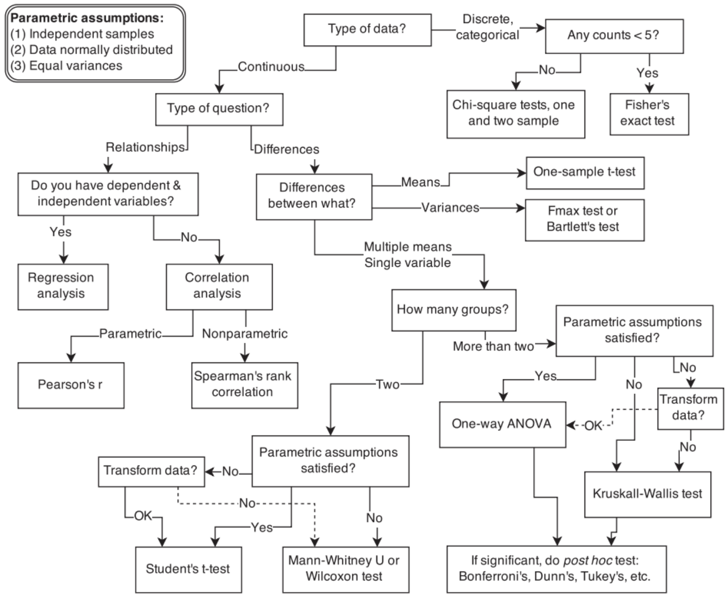
- Decision theory
  - If I have a loss function I need
- Uncertainty quantification and intervals
- What is the role of hypothesis testing?
- Predictive summaries

4. Identify model misfit



5. Refine and rebuild

Significance Testing Flowchart



Free yourself from this perspective on statistics!

Final Thoughts

- Beyond math and programming proficiency you *must* think critically
    - Sources of variation: sampling variability, measurement error, bias, signal variability, hierarchies
    - How does domain knowledge inform the DGP and my prior specification

Final Thoughts

- Beyond math and programming proficiency you *must* think critically
    - Sources of variation: sampling variability, measurement error, bias, signal variability, hierarchies
    - How does domain knowledge inform the DGP and my prior specification
- Don't constrain yourself to the basic models you've already encountered
    - Build your own "lego" masterpieces!

Final Thoughts

- Beyond math and programming proficiency you *must* think critically
    - Sources of variation: sampling variability, measurement error, bias, signal variability, hierarchies
    - How does domain knowledge inform the DGP and my prior specification
- Don't constrain yourself to the basic models you've already encountered
    - Build your own "lego" masterpieces!
- You now have the core tools necessary to become a practicing Bayesian statistician

Thanks!