Lecture 9: Wrap-up

Professor Alexander Franks

2024-03-13

- Homework 5
- Quiz 5 - HW
- Practice Exam
- Course Evals
- OH adjustments / Study Sessions.

- Likelihood / Simplify
  - Binomial
  - Poisson (w/ exposures)
  - Normal Model
- Bayes w/ conjugate
  - Bin - Beta
  - Poisson - Gamma
  - Normal - Normal
  - Derive the posterior,
  - interpret params,
  - write posterior mean as weighted average of MLE & prior mean.

Monte Carlo

- Inversion / Rejection Sampling,
- Metropolis - Hastings*
- HMC variant.
- Diagnostics

Posterior Predictive

- Pseudo-code

- For model checking

Hierarchical Modeling

$$y_i \sim N(\theta_i, \sigma_i^2)$$

$$\theta_i \sim N(\mu, \tau^2)$$

$$P(\mu, \tau^2) \quad \text{(hyperprior)}$$

Post Means

$$E[\theta_i \mid y_i, \underline{\mu, \tau^2}]$$

$$w_i y_i + (1 - w_i)\mu$$

$\uparrow$
MLE

$$w_i = \frac{1/\sigma_i^2}{1/\sigma_i^2 + 1/\tau^2}$$

τ^2 large: close no pooling

τ^2 small: closer to

complete pooling.

$$\begin{cases} y_i \sim \text{Bin}(n_i, \theta_i) \\ \theta_i \sim \underline{\text{Beta}(a, b)} \\ \text{Prior}(a, b) \end{cases}$$

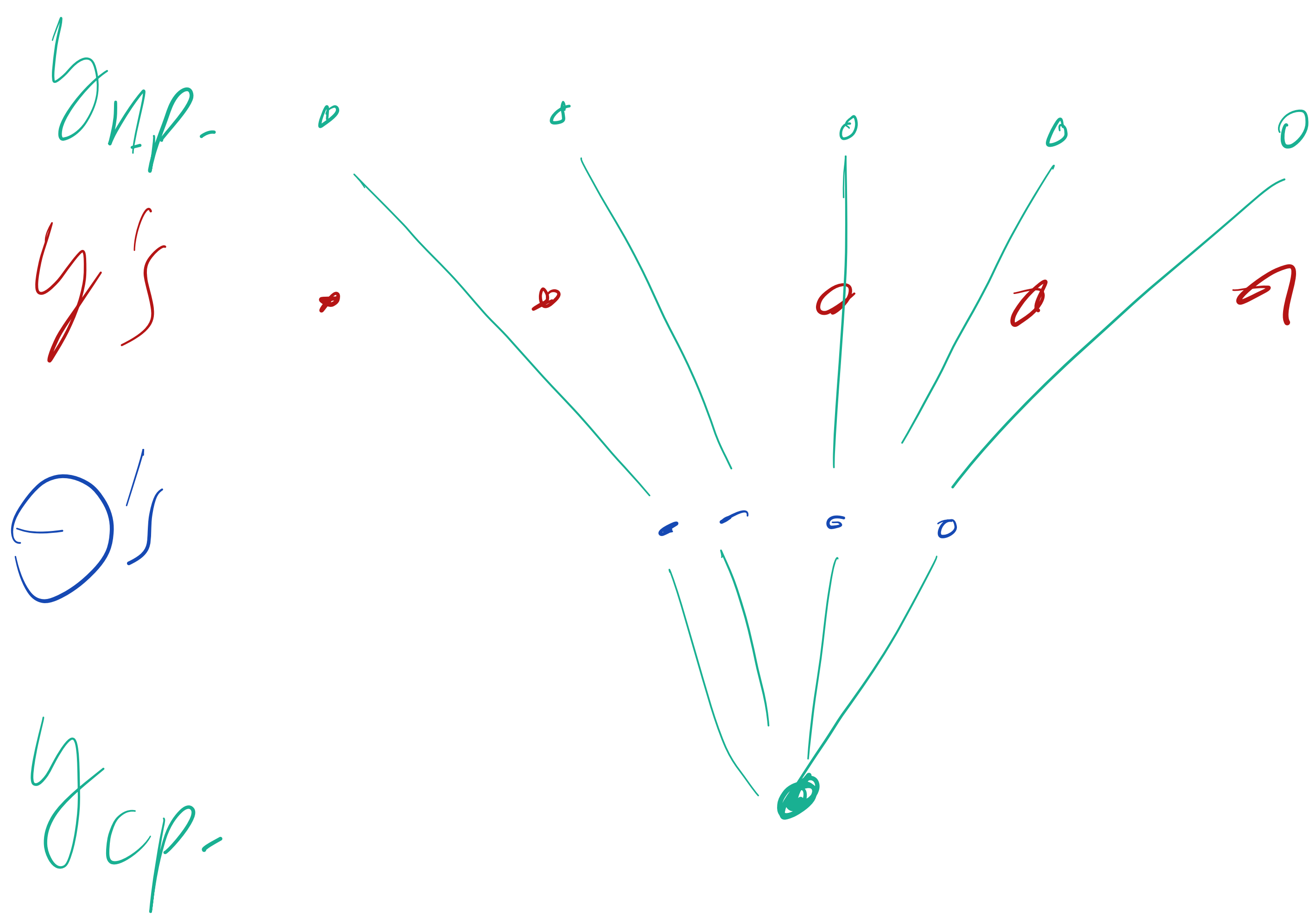
Binomial

Hierarchi

$E[\theta_i \mid a, b, y, n)$ as a weighted avg.

$\rightarrow w_i\, y_i/n_i + (1-w_i)\dfrac{a}{a+b}$

θ_i's differ. Which is better: "no pooling" or "complete pooling"?



Pois Prior.

$$y_i \sim \text{Pois}(\lambda_i v_i)$$

$$\lambda_i \sim \text{Gamma}(a, b)$$

Prior (a, b)

$$E[\lambda_i \mid y_i, a, b]$$

$$w_i \frac{y_i}{v_i} + (1 - w_i)\frac{a}{b}$$

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

Laplace Approximation to the Posterior

- Approximate the posterior distribution using a multivariate normal distribution
- When we have a lot of i.i.d. observations, the posterior will be approximately normal
- Center the normal at the mode of the posterior
- Compute the (co)variance of the normal by computing the second derivative / hessian of the posterior at the mode

Laplace Approximation

- Approximate the posterior distribution using a normal distribution
- When we have a lot of i.i.d. observations, the posterior will be approximately normal
- Center the normal at the mode of the posterior
- Compute the (co)variance of the normal by computing the second derivative / hessian of the posterior at the mode

Laplace Approximation

- Let $\tilde{\theta}$ be the mode of the of the posterior distribution
- Use a Taylor Series approximation the log-posterior around the mode is
  - $logP(\theta \mid y) \approx logP(\tilde{\theta} \mid y) - 1/2(\theta - \tilde{\theta})H(\theta - \tilde{\theta})$
  - $H = \frac{d^2}{d\theta^2} logp(\theta \mid y)$
  - Note, linear term falls out because derivative at the mode is zero
- $p(\theta \mid y) \approx N(\tilde{\theta}, I(\theta)^{-1})$

Finding the mode of the posterior distribution

- Calculus
    - Take the log
    - Differentiate, set to zero and solve
- Computational
    - `optim` in R for one dimensional posteriors
    - `optimise` in R for multivariate p

Variational Bayes

- Find even better approximations
- Let θ be d dimensional parameter vector
- Let $\epsilon \sim MVN_d(0, \Sigma)$
- Let g_λ be a class of flexible functions parameterized by λ
  - Neural networks!
- Solve an optimization problem:

$$\arg\min_{\lambda} \text{dist}(p(\theta \mid y)), \cancel{p}(g_\lambda(\epsilon)))$$

  - Minimize the "distance" between the true posterior and the approximate one.
- Sample ϵ from a multivariate normal. $g_\lambda(\epsilon)$ will be a sample from something close to $p(\theta \mid y)$

Why Bayesian statistics?

Interpretations

- Belief / Credible vs CI.
- ↳ propogate into estimates via weighted avg.

Subject Matter Expertise

Prior → Bias / Variance

Flexibility

prior choice / Modeling.

Computation

Jobs (a quick search)

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 (a quick search)

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 model the 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"
    - Derive your own estimators

Posterior Mean.

- Models which "borrow strength" and share information across observations
    - Hierarchical modeling!
- Model checking is fundamental part of the process
- Is frequentist inference still important? Yes!
    - Calibration / Bias / Variance / MSE

A Bayesian Modeling Process (overview)

- 1. Propose a Data Generating Process
  - Includes a prior distributions for parameters
  - Could be very complex and include hierarchies of parameters
- 2. Posterior density is proportional to likelihood times prior density
- 3. Summarize the posterior density
  - Posterior means and credible intervals for parameters
  - Get these with (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

Hierarchal

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. Computing posterior summaries

- "Easy" to write down the density: proportional to likelihood $\times$ prior
- Hard to summarize: usually need Monte Carlo techniques
    - Point estimates with posterior means (or medians) of parameters
    - Compute probability intervals (quantile or HPD)
    - Compute posterior predictive distributions
- A very common approach is to approximate these summaries by generating MCMC samples

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
- In general, MCMC is very slow relative to optimization methods

Modern MCMC

- Vanilla 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 address this problem

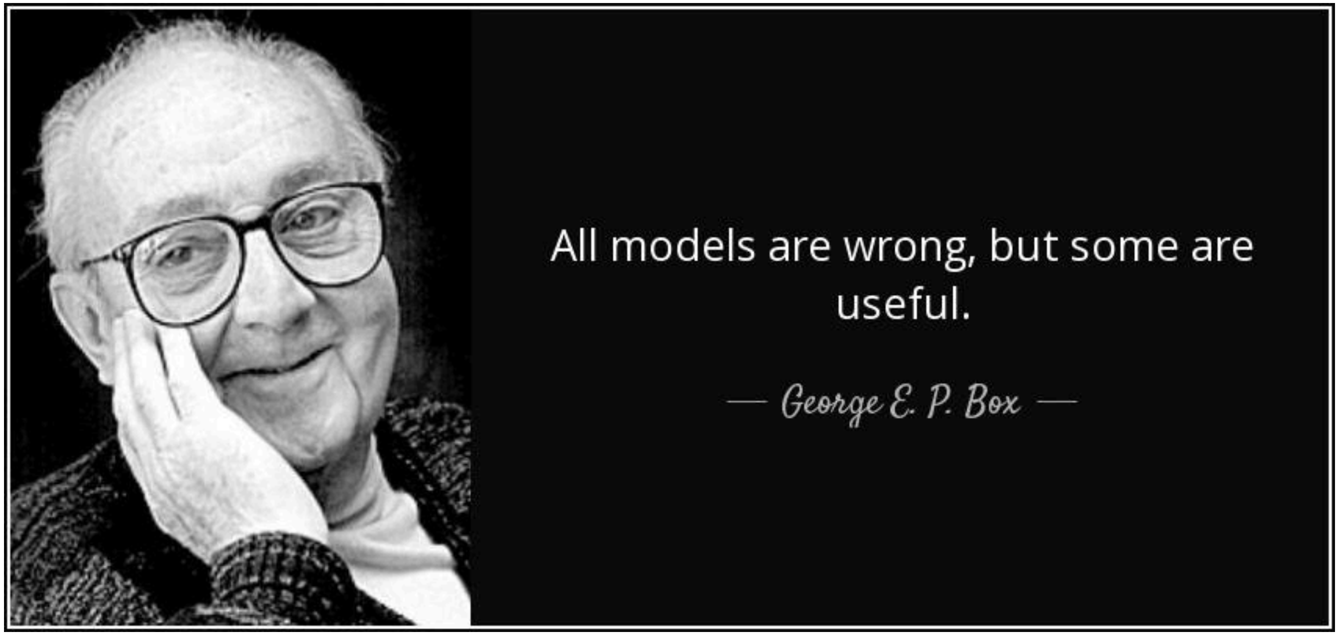
Approximate Inference

- 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

3. Summarize your results

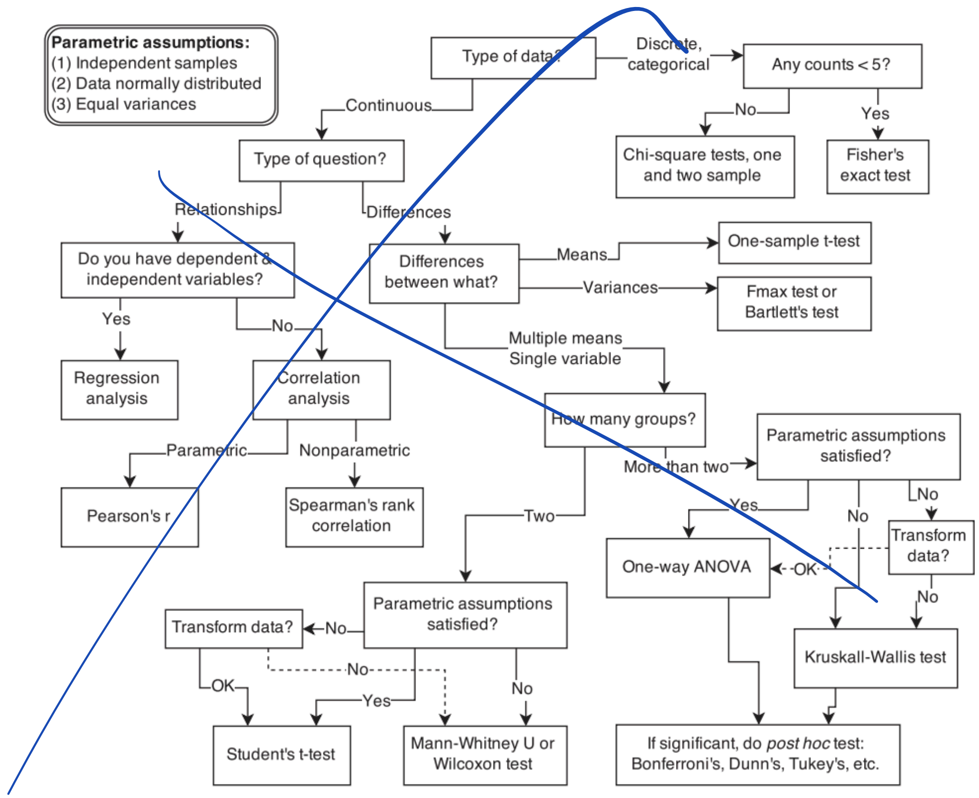
- Decision theory
  - Come up with a loss function for your problem
  - Choose the estimator/action which minimizes the posterior expected loss
- Uncertainty quantification and intervals
- What is the role of hypothesis testing?
- Predictive summaries

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 prior specifications
- 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!