Bayesian Data Analysis, class 4a

Andrew Gelman

Chapter 4: Large-sample inference and frequency properties of Bayesian inference

Discussion of homework due beginning of Class 3b

- Theory problem
- Computing problem
- Applied problem

Theory problem

- Prove that a posterior density is proper by bounding the integral over the range $(\alpha, \beta) \in (-\infty, \infty) \times (-\infty, \infty)$
- Recall that $p(\theta) \propto 1/\theta$ has an infinite integral as $\theta \to \infty$

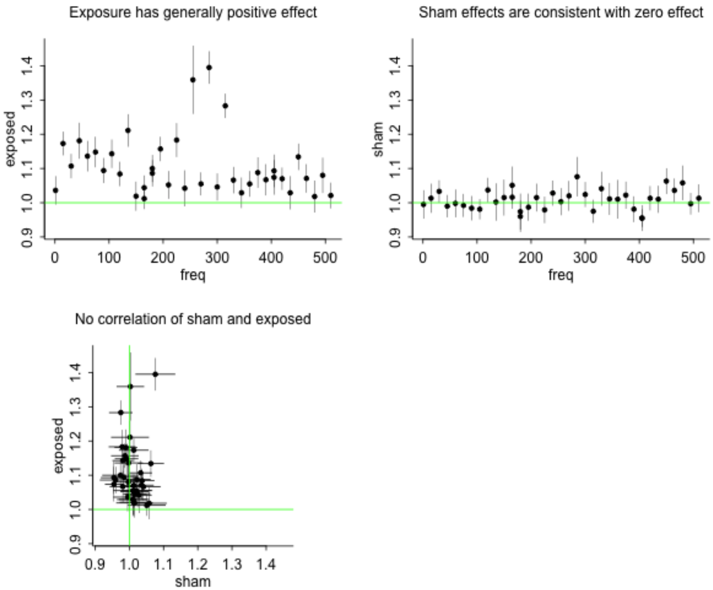
Computing problem

- Compute posterior density on a discrete grid:
  - Compute unnormalized log-density on the grid
  - Rescale and exponentiate
  - Normalize to sum to 1
- Make contour plot
- Sum over rows or columns to get marginal posterior density
- Loop 1000 times:
  - Sample from marginal, then conditional
  - Add noise to get random draw in grid box
- Graph the simulations

Applied problem

- Paired experiment on chicken brains
- How does the effect depend on frequency? Is there evidence from the data that the effect is not constant across frequencies?
- Should we use the sham data in the estimates?

Basic (Bayesian?) data analysis



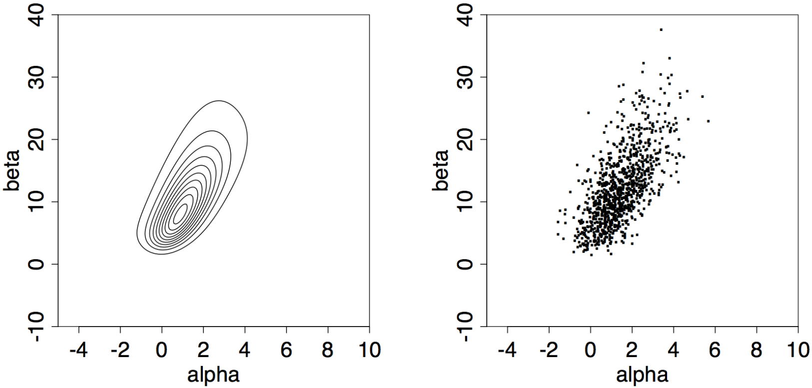
That sham treatment

- What is the role of the "sham" treatment? Why is it performed at all?
- Consider two different summaries of treatment effect: (a) mean ratios for exposed treatments vs. controls (1.036 at 1 Hz, 1.173 at 15 Hz, etc.), or (b) mean ratio for exposed treatments vs. controls, divided by mean ratio for sham treatments vs. controls (1.036/0.995 at 1 Hz, 1.173/1.013 at 15 Hz, etc.). Which of these two is a better estimate of the treatment effects? Use the data to address this question.
- Bayesian analysis

4. Large-sample inference and frequency properties

- Normal approximations to the posterior distribution
- Large-sample theory
- Counterexamples to the theorems
- Frequency evaluations of Bayesian inferences

Example posterior distribution



Contours at $0.05, 0.15, \ldots, 0.95$

4.1. Normal approximation to the posterior distribution

- Approximate log-posterior as quadratic based on mode and curvature:

$$\log p(\theta|y) = \log p(\hat{\theta}|y) - \frac{1}{2}(\theta-\hat{\theta})^T \left[\frac{d^2}{d\theta^2}\log p(\theta|y)\right]_{\theta=\hat{\theta}}(\theta-\hat{\theta}) + \ldots,$$

- Normal distribution centered at $\hat{\theta}$ with inverse variance $\left[\frac{d^2}{d\theta^2}\log p(\theta|y)\right]_{\theta=\hat{\theta}}$
- Posterior density and 2-dimensional contour plots:
  - For the normal model, the log posterior density (relative to the mode) is $-1/2$ times a χ^2_2 random variable
  - `qchisq(.95,2)` $= 5.99$
  - 95% of posterior mass has density above $e^{-5.99/2} = 0.05$ relative to the mode
  - Thus, take contours out to 0.05

4.2. Large-sample theory

- Data $y_1, \ldots, y_n \sim f(y)$, modeled as $p(y|\theta)$
- In the limit $n \to \infty$:
  - If θ is discrete and finite, $p(\theta|y) \to$ point mass at the true θ (or the model closest to f)
  - If θ is continuous on a compact set, $p(\theta|y) \to$ a point mass at the true θ (or the model closest to f)
  - Under some conditions, $p(\theta|y)$ approaches a normal distribution

4.3. Counterexamples to the theorems

- Unidentified parameters (e.g., $y = \theta_1 + \theta_2$)
- Model changing with sample size
- Unbounded likelihoods, for example this mixture model:

$$p(y) = \prod_{i=1}^{n} \left(\frac{1}{2} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2\sigma_1^2}(y_i - \mu_1)^2} + \frac{1}{2} \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2\sigma_2^2}(y_i - \mu_2)^2} \right)$$

  Blows up when $\sigma_1 \rightarrow 0$ and $\mu_1 = y_i$ for any i

- Improper posteriors
- Constrained priors
- Boundary estimates
- Tails

4.4. Frequency evaluations of Bayesian inferences

- Efficiency of point estimation
- Coverage of posterior intervals

4.5. Bayesian interpretations of other statistical methods

- Maximum likelihood
- Unbiased estimates
- Confidence intervals
- Hypothesis testing
- Multiple comparisons
- Classical (non-model-based) nonparametric methods

Summary of Chapter 4

- Normal approximation to the posterior distribution . . .
- . . . and its limitations
- Connections to non-Bayesian ideas