CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

Lecture 76
Bayesian Regression, part 3

Chris A. Mack

Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

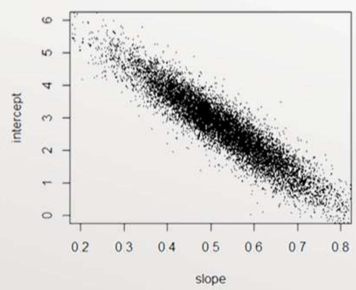
THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Bayesian Computations

- Analytical solutions of the posterior distribution are only possible for special cases (called conjugate priors)
- Usually, we need to solve Bayes' equation numerically
  - Markov Chain Monte Carlo (MCMC) sampling
  - Result is a set of points from the posterior distribution that we then summarize

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

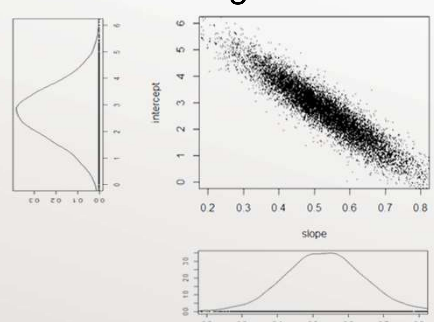
Bayesian Numerical Output



Numerical joint posterior probability distribution of the slope and intercept (straight line fit)

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Summarizing the Posterior



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Three Problems with Bayesian Regression

- Usually, Bayesian regression involves difficult computations
  - Less of an issue with modern software and computers
- What prior distributions should we use?
  - Sometimes we run through many priors to see what happens
- Is it reasonable to treat every regression parameter as a random variable?

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Frequentist + Bayesian

- Suppose we have a "constant" in our model, but we don't know the value of the constant accurately
- Frequentist approaches:
  - Let the constant "float" as a parameter, ignoring what we already know about its value
  - Fix the constant at our best estimate of its value, ignoring the uncertainty in this value
- Can we add Bayesian ideas to our frequentist regression?

One Bayesian Approach

- Take the -log of Bayes' Equation

Constant

$$-\ln(\mathrm{P}(\theta|\mathbf{y})) = -\ln(\mathrm{P}(\mathbf{y}|\theta)) - \ln(\mathrm{P}(\theta)) + \ln(\mathrm{P}(\mathbf{y}))$$

Log-likelihood

- Let's take the maximum of $\ln(\mathrm{P}(\theta|\mathbf{y}))$ (the mode) as our point estimate of the parameters
  - For normal likelihood, we want to minimize $S = \frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_\varepsilon^2} - \ln(\mathrm{P}(\theta))$

© Chris Mack, 2016 Data to Decisions 7

Hybrid Interpretation

- Let the constant (call it g) be a parameter of the fit, with a normal prior distribution with mean $\bar{g}$ and standard error s_g

$$-\ln(\mathrm{P}(\theta)) = \frac{1}{2}\left(\frac{\bar{g}-\hat{g}}{s_g}\right)^2$$

- This is the same as adding a penalty to the χ^2 by letting our prior estimate of g be a "measurement"

$$S = \left(\frac{\bar{g}-\hat{g}}{s_g}\right)^2 + \sum_{i=0}^{n}\left(\frac{y_i-\hat{y}_i}{s_y}\right)^2$$

© Chris Mack, 2016 Data to Decisions 8

Bayesian as Ridge Regression

- This "hybrid" interpretation can be applied to all model parameters:

$$S = \sum_{i=0}^{n}\left(\frac{y_i-\hat{y}_i}{s_y}\right)^2 + \sum_{k=0}^{p-1}\left(\frac{b_k-b_{k,prior}}{s_{bk,prior}}\right)^2$$

- When $b_{k,prior} = 0$, this is identical to Ridge Regression

© Chris Mack, 2016 Data to Decisions 9

Bayes Regression Summary

- Pick a prior distribution
- Calculate the likelihood function in the usual way
- Calculate the posterior distribution, usually numerically (MCMC)
- Summarize the posterior distribution (usually MAP estimate, credible interval)

© Chris Mack, 2016 Data to Decisions 10

Lecture 76: What have we learned?

- What are the major difficulties of Bayesian regression?
- Explain how frequentist and Bayesian regression concepts can merge in a "hybrid" form
- Explain how this "hybrid" form relates to ridge regression

© Chris Mack, 2016 Data to Decisions 11