Bayesian Hierarchical Modeling in JAGS

Marcel Niklaus

10. Juli 2015

Not content of this talk

- Bayesian Hypothesis test
- Model comparison

Why Bayes

- We are actually interested in the probability of a model given data
- p-values are based on probability of (unobserved) data given model (conceptually hard)

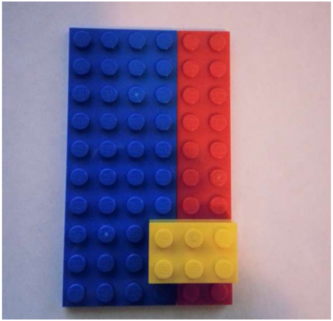
Why Bayes

- We are actually interested in the probability of a model given data
- p-values are based on probability of (unobserved) data given model (conceptually hard)
- Surprise exercise; think of a situation you are interested in data given model!

General Principles of Bayesian Analysis

- Uncertainty (of parameter estimate) is quantified by probability (distributions)
- Observed data is used to update the **prior** information to yield **posterior** information
- Bayesian workflow: set prior beliefs $\Rightarrow$ get data $\Rightarrow$ summarize posterior beliefs

Formal Bayes Theorem



- Bayes rule is rooted in conditional probability and is uncontroversial.
- $P(Red|Yellow) = P(Yellow|Red) * P(Red)/P(Yellow)$
- $P(Red|Yellow) = (4/20) * (20/60)/(6/60)$
- $2/3 = 1/5 * 1/3 * 10$

Formal Bayes Theorem (only slide with many formulas)

- Bayes Theorem tells us how to rationally revise prior beliefs in light of the data to yield posterior beliefs.
- $P(R|Y) = P(Y|R) * \frac{P(R)}{P(Y)}$
- Yellow = data, red = hypothesis θ
- $p(\theta|data) = \frac{p(data|\theta) * p(\theta)}{p(data)}$
- $posterior = \frac{likelihood * prior}{marginal\ \ likelihood}$
- $p(\theta|D) \propto p(D|\theta) * p(\theta)$
- Posterior is the likelihood weighted by the prior

Let's play a game: Rock Paper Scissor

`http://87.106.45.173:3838/felix/BayesLessons/BayesianLesson1.Rmd`

- 6 Trials of Rock Paper Scissor
- Let's ignore the ties!

Prior

Pick your Prior: Who is going to win?

- θ: the rate with which the female wins
- $\theta = 1$: female always wins
- $\theta = 0$: male always wins
- $\theta = 0.5$: equal odds
- Uncertainty (of parameter estimate) is quantified by probability (distributions)
- We assume our beliefs can be represented by a beta distribution
- A beta distribution is constraint to lie within 0-1 (perfect for proportion)

Pick your Prior: Who is going to win?

- $\theta \sim Beta(1, 1)$: Uninformed prior: Do you think (before seeing the data) that it is equally likely that $\theta = 0.01$ and $\theta = 0.5$?
- $\theta \sim Beta(3, 3)$: Slightly informed prior

We are now ready to play: Go!

Likelihood

The likelihood is the workhorse of Bayesian inference. It represents the data part.

- What's the likelihood we observe the data (y wins given n trials) given the parameter θ
- Probability density of the data, considered as a function of θ
- The likelihood of a hypothesis (θ), $L(H|data)$, conditions on the data as if they are fixed while allowing the hypotheses (our parameter θ to vary
- Binomial likelihood: $L(p|n, y) = \binom{n}{y} p^y (1-p)^{n-y}$
- `http://shiny.stat.calpoly.edu/MLE_Binomial/`

After playing

- The posterior distribution summarizes our state of uncertainty about the true value of θ after having observed the game.

Markov chain Monte Carlo

- Posterior distribution = Prior distribution $*$ Likelihood density
- Analytical calculation of posterior only possible for simple models (high-dimensionality integration problem)
- Draw samples from posterior and summarize the distribution of those samples (it works!)
- Metropolis-Hastings algorithm and the Gibbs sampler
- JAGS WinBugs and STAN can do this for you
- Algorithm to approximate an unknown distribution

JAGS

- $\leftarrow$ is equal to: $y \leftarrow a + b$
- $\sim$ is distributed as..
- Binomial likelihood: $y \sim dbin(\theta, nAttempts)$
- Normal likelihood: $y \sim dnorm(mean, precision)$
- $precision = 1/Var$
- Loop: for (i in 1:3){bla[i]=1+i}
- bla = 2,3,4, bla[2] = 3

Soccer Shootout

Let's estimate soccer players' ability to score a penalty in a world cup!

- Open Soccer.R in your R console
- Set your working directory (line 8)
- Jags model: ShootoutAbility.txt
- What is their ability θ, [0-1]?
- $Y\theta$ Prior: remember the game

Convergence: Has the MCMC Gibbs sampler converged on posterior (after starting from random values)

- Trace plot: "tight horizontal band"
- Autocorrelation plot: "Chains should forget previous visits with time": drop off quickly
- Gelman-Rubin-Brooks diagnostic: Between and within chain variance: 1 indicated convergence (F-value); less than 1.2. See gelman.diag [CODA]
- Geweke test of non-stationarity; Heidelberger-Welch test etc.

Soccer Shootout: Afrika vs. Europa and America

- Jags model: ShootoutAbilityDifference.txt
- theta[Cindex[i]] can either be theta[1], or theta[2]

Hierarchical Modeling

- With nested data (e.g. data for participants is organized on more than 1 level), a multilevel model is appropriate.
- The classic: students grouped in classes, which nest in school districts, which in turn nest in states.
- All Bayesian Models are hierarchical because every parameter has a prior.

Easy IQ example: IQ measurements

- Scenario: We measure some IQs
- $y \sim N(\mu, \tau)$
- $\mu \sim N(100, 0.004)$
- $\tau \sim G(0.001, 0.001)$

)tau $\sim G(0.001, 0.001)$

Linear Mixed Models

- Linear Regression model
- Mixed because it includes coefficients that vary over group (random: participants, items) and some that don't (fixed, treatment group).
- Random Effects: zero mean restriction $N(0, \omega)$)
- Random intercepts model: random intercept, fixed slope
- Random intercepts and slope model: random intercept, random slope

In Bayesian Statistics, all parameters are random, i.e. drawn from an overarching distribution!

Easy IQ example: IQ regressed on math grade

- $y \sim N(\mu, \tau)$
- $\mu <- \ x0 + x1 * math$

Easy IQ example: IQ regressed on math grade; Random Intercept

- $y \sim N(\mu[i], \tau)$
- $\mu[i] < -x0[G] + x1 * math$
- for all groups j: $x0[j] \sim dnorm(100, 0.0044)$
- $x1 \sim dnorm(0, 0.001)$

Easy IQ example: IQ regressed on math grade; Random Intercept and Slope

- $y \sim N(\mu, \tau)$
- $mu < -x0[G] + x1[G] * math$
- for all groups j: $x0[j] \sim dnorm(100, 0.0044)$
- for all groups j: $x1[j] \sim dnorm(0, 0.01)$

Simple linear regression

Let's estimate soccer player's ability to score a penalty in a world cup!

- Open Exam_ple.R in your R console
- Set your working directory (line 7)
- Jags model: ExamSimple.txt

Random intercept model

- b0[school[i]]
- for all schools: $b0[j] \sim dnorm(school.b0, school.tau)$
- $school.b0 \sim dnorm(0, .0001)$

Random intercept and slope model

- $b0[school[i]] + b1[school[i]] * lrt[i]$
- for all schools: $b0[j] \sim dnorm(school.b0, school.tau)$
- for all schools: $b1[j] \sim dnorm(school.b1, school.tau.b1)$
- school.b0 $\sim$ dnorm(0,.0001)
- school.b1 $\sim$ dnorm(0,.0001)

Multi-Level Models

Covariates that capture the way groups vary are included

- Level 1: Regression for London Reading Test
- Level 2: Group Level with covariates: Regression for entry score

- Ein