Introduction to Hierarchical Models

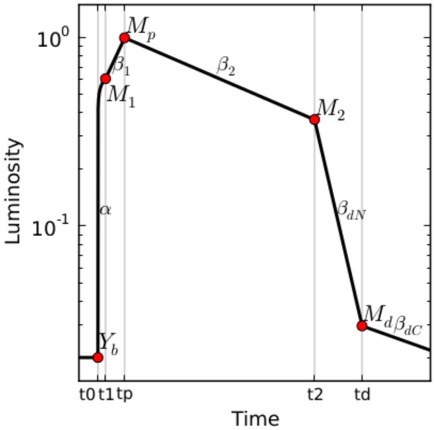
STAT 689: Statistical Computing with R and Python

April 6, 2018

Hierarchical Statistical Models

- non hierarchical models: all observations $X_1, \ldots, X_n$ share same parameter vector θ
- hierarchical models: each observation X_i has its own parameters θ_i. the parameters θ_i have their own distribution with parameters ϕ.
- best to see some hierarchical model examples
- computing for hierarchical models is generally challenging
  - number of parameters on the same order as number of observations
  - conjugate models not available
  - straightforward implementations of Gibbs / Metropolis not feasible due to parameter size

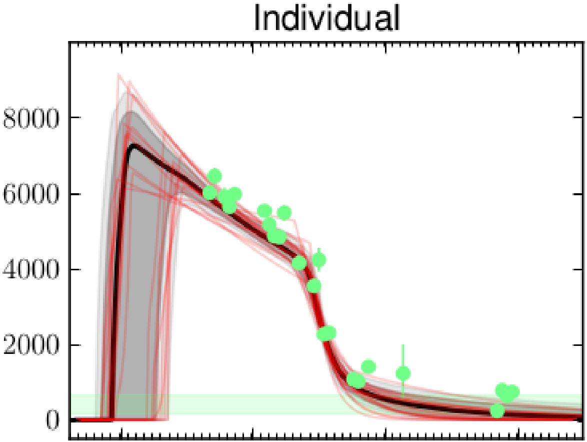
Example 1: Supernovae Model



10 Total Parameters:

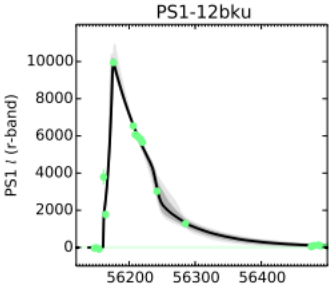
- times $(t_0, t_1, t_p, t_2, t_d)$
- magnitudes $(Y_b, M_1, M_p, M_2, M_{d\beta_{dC}})$

Example 1: Supernovae Light Curve



- construct prior and likelihood
- collect data for star (green points)
- compute posterior and plot draws (red lines)

Example 1: Supernova Light Curves



Surveys collect many supernovae (2 examples above), each with their own parameters.

- Need parameter estimates (e.g. slopes) for individual supernovae
- Need distributions of parameters across all supernove.

Example 2: 1970 Batting Averages for 18 Players

1. 1970 Batting Averages for 18 Major League Players and Transformed

i	Player	Y_i = batting average for first 45 at bats (1)	p_i = batting average for remainder of season (2)	At bats for remainder of season (3)
1	Clemente (Pitts, NL)	.400	.346	367
2	F. Robinson (Balt, AL)	.378	.298	426
3	F. Howard (Wash, AL)	.356	.276	521
4	Johnstone (Cal, AL)	.333	.222	275
5	Berry (Chi, AL)	.311	.273	418
6	Spencer (Cal, AL)	.311	.270	466
7	Kessinger (Chi, NL)	.289	.263	586
8	L. Alvarado (Bos, AL)	.267	.210	138
9	Santo (Chi, NL)	.244	.269	510
10	Swoboda (NY, NL)	.244	.230	200
11	Unser (Wash, AL)	.222	.264	277
12	Williams (Chi, AL)	.222	.256	270
13	Scott (Bos, AL)	.222	.303	435
14	Petrocelli (Bos, AL)	.222	.264	538
15	E. Rodriguez (KC, AL)	.222	.226	186
16	Campaneris (Oak, AL)	.200	.285	558
17	Munson (NY, AL)	.178	.316	408
18	Alvis (Mil, NL)	.156	.200	70

Efron and Morris JASA 1975

Example 2: Batting averages

- player i has some "true" average p_i
- average after 45 at bats is estimate of p_i

$$
\begin{aligned}
X_i &= \text{ number hits for player } i \text{ after 45 at bats} \\
X_i &\sim \mathit{Binomial}(n = 45, p_i) \\
\widehat{p}_i &= \frac{X_i}{45} \text{ MLE, could use Bayesian estimator}
\end{aligned}
$$

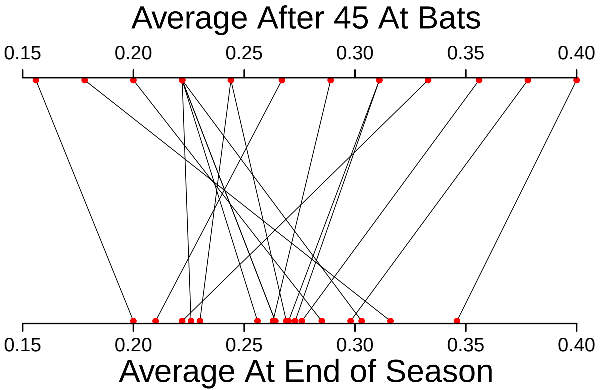
We are also interested in the distribution of p_i e.g. what is the range of typical batting averages?

Why Not Fit Models Separately?

Fit separate model to each batter

- fit separate Binomial model for each batter
- $\widehat{p}_i = \frac{X_i}{45}$

Result: Overdispersed estimate of population distribution of p_i



Qualitative Reasoning

Problem:

- if we estimate population distribution from $\widehat{p}_i$, we infer more great hitters and more terrible hitters than there actually are
- if we estimate population distribution of plateau duration from individual SN fits, we infer more long duration plateaus and more short duration plateaus than there actually are

Solution:

- inferences can be improved by "shrinking" estimates towards the center of the population distribution
  - The average of player averages after 45 at bats is .265.
  - Clemente has .400 average after 45 at bats. Shrink Clemente estimate **down** towards .265.
  - Alvis has 0.156 average after 45 at bats. Shrink Alvis estimate **up** towards 0.265.

Hierarchical Models

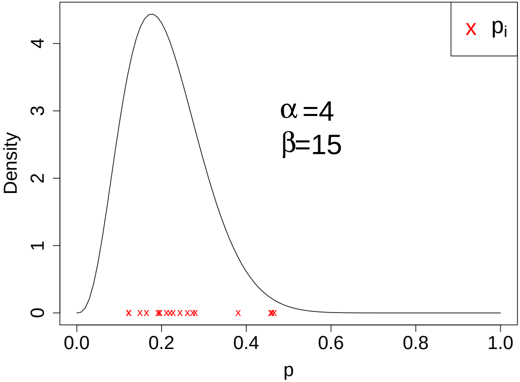
Problem is clear, but solution details murky:

- how much to shrink?
- how to construct confidence / credible intervals for observation level parameters?
- how do we estimate, quantify uncertainty in population level parameters?

Bayesian Hierarchical models are a formal method for answering these questions.

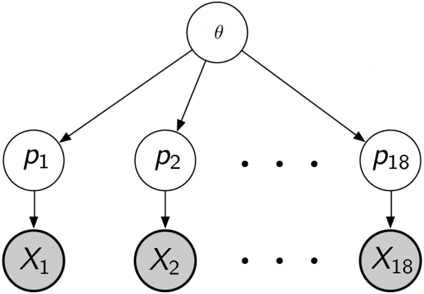
Example: Batting Averages

- $p_i \sim Beta(\alpha, \beta)$
- $\theta = (\alpha, \beta)$ is unknown.
- Player i has true (unobserved) average p_i.
- We observe $X_i \sim Binomial(45, p_i)$ hits for player i.



Goal: Use $(X_1, \ldots, X_n)$ to infer $(p_1, \ldots, p_n)$ and $\theta = (\alpha, \beta)$.

Example: Batting Averages



The missing arrows imply conditional independence.

- X_1 and θ are <u>not</u> independent.
- X_1 and X_2 are <u>not</u> independent.
- X_1 is independent of X_2 and θ <u>given</u> p_1.

These Graphical models can assist in thinking about and constructing the likelihood and prior.

Example: Batting Averages

The posterior is

$$\begin{aligned}\pi(\theta, \vec{p}|\vec{x}) &\propto f(\vec{x}|\theta, \vec{p})\pi(\theta, \vec{p}) \\ &\propto f(\vec{x}|\vec{p})\pi(\vec{p}|\theta)\pi(\theta) \\ &\propto \left(\prod_{i=1}^{n} f(x_i|p_i)\pi(p_i|\theta)\right)\pi(\theta)\end{aligned}$$

where

$$\begin{aligned}f(x_i|p_i) &= \binom{45}{x_i} p_i^{x_i}(1-p_i)^{45-x_i} \\ \pi(p_i|\theta) &= \frac{1}{B(\alpha, \beta)} p_i^{\alpha-1}(1-p_i)^{\beta-1} \\ \pi(\theta) &= \text{ prior on p distribution shape}\end{aligned}$$

References

- Coagulation example on course website.
- Overview of Hierarchical Models by Tom Loredo: `http://www.stat.tamu.edu/~jlong/astrostat/ASTRO-WG4-HBMIntro.pdf`
- Batting Averages
  - Efron and Morris `http://www.medicine.mcgill.ca/epidemiology/hanley/bios602/MultilevelData/EfronMorrisJASA1975.pdf`
  - Batting Average Study by Lawrence Brown `https://projecteuclid.org/download/pdfview_1/euclid.aoas/1206367815`
- Bayesian Data Analysis by Gelman: Chapters 5, 11.6, and 15.