Chapter 9: Hierachical Models

Shubham Gupta

May 6, 2019

1 Introduction

- They involve multiple parameters.
- When the value of one parameter θ depends on another variable ω, the hierachical structure of these variables can be represented by a hierachical model.
- To infer these parameters, we apply the joint probability rule for the parameters.

$$P(\theta, \omega|D) \propto P(D|\theta, \omega)p(\theta, \omega).$$
$$= P(D|\theta) * P(\theta|\omega) * P(\omega).$$

- The above equation implies that value of D is dependent only on θ and independant of other variables. Similarly, the value of θ is dependant only on the value of ω and is conditionally independant of all other parameters.
- The dependencies between parameters are useful because:
  - They are meaningful for the given application
  - Because of dependencies across parameters, they can jointly inform all parameter estimates.
  - Easier convergence with smart algorithms that exploit this joint probability.

1.1 Coin flipping from a single mint

- We will use bernoulli distribution for the data and beta distribution for the prior.

$$y_i \approx dbern(\theta).$$
$$\theta \approx dbeta(a, b).$$

- We know that a and b can be represented as using mode ω and concertration κ as:

$$a = \omega(\kappa - 2) + 1.$$
$$b = (1 - \omega)(\kappa - 2) + 1.$$

- Hence, we can write θ as:

$$\theta \approx dbeta(\omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1).$$

  - The value κ controls how close the value of ω is θ.
  - Higher value of κ = Closer to value of θ

- Let us assume ω is another parameter to be estimated. Assume this to be a beta distribution: $\omega \approx beta(\omega|A_\omega, B_\omega)$

- We know the value of ω is closer to the mode of the distribution in this case i.e: $\frac{A_\omega - 1}{A_\omega + B_\omega - 2}$

- Substituting bayes rule, we get:

$$p(\theta, \omega|y) = \frac{p(y|\theta, \omega)p(\theta, \omega)}{p(y)} = \frac{p(y|\theta)p(\theta|\omega)p(\omega)}{p(y)}.$$

- We have the equations for all of the above components. We can get the posterior probability by solving the above equation.

- We can solve them using grid approximation as well as the parameters are finite.

1.2 Multiple coins from single mint

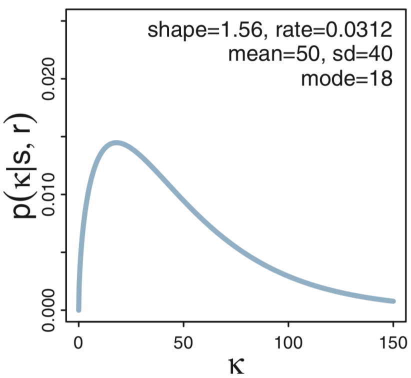
- Assume we have multiple coins for a single mint. Each coin will now have it's own parameter θ_s and we will estimate this using all the data for ω.

1.3 Real example

- For the multiple coins problem, we do not know the value for ω in advance. We will have to estimate it from the data available.

- We will assume ω follows a gamma distribution. The gamma distribution has the following formula: $gamma(\kappa|s, r)$. Here, s is the shape parameter and r is the rate parameter.

- We will use the parameters s =1.56 and r =0.0312 because these values have a boundary at 0 and infinite possible positive values.

- 

- Mean: $\mu = \frac{s}{r}$
- Mode: $\omega = \frac{s-1}{r}$
- SDev: $\sigma = \frac{\sqrt{s}}{r}$
- We can derive s and r from the above as:

$$s = \frac{\mu^2}{\sigma^2}.$$

$$r = \frac{\mu}{\sigma^2}.$$

  when the mean $\mu>0$

- It can also be written as:

$$s = 1 + \omega r.$$

$$r = \frac{\omega + \sqrt{\omega^2 + 4\sigma^2}}{2\sigma^2}.$$

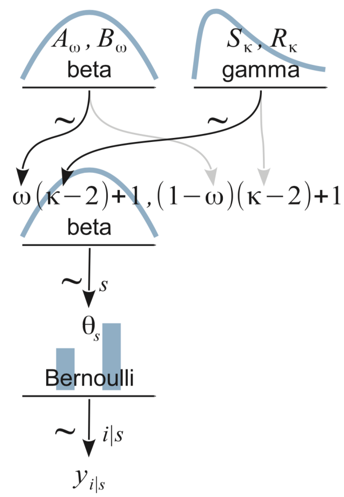


Figure 1: Hierachical Model

1.4 Theraupeutic Touch

- Relieve congestion and improve balance by manipulating "energy field" without touching the patient.
- Experiment
  - Practitioner should be able to tell if hand is near their hand without touching the hand.
  - Experimenter flips a coin. Depending on the outcome, places hand above or below practitioner hand.
  - Practitioner guesses if hand is above or below.
  - Chance performance for guessing the result is 0.5
- Questions:
  - How much did group differ from chance performance?
  - How much did each individual differ from chance performance?

2 Shrinkage

- Estimates of low-level params are pulled together than they would if they were higher-level params. This pulling is called **shrinkage**.
- It occurs because:
  - Subset of data is directly dependant on the low-level parameter.

– The higher-level params that depend on the low-level params.

- Shrikage occurs because of hierachical models, not bayesian estimation.
- Intuitively, shrinkage occurs because data from all individuals influence the higher-order params, and these params in-turn influence the estimates for each individual.

3 Extending the hierachy

- We can model problems as hierachical models of multiple levels.
- Baseball players
  - They bat. Sometimes they get a hit.
  - Different positions for each player. Categorize by player positions.
  - Hence, we can estimate abilities for each player AND each position.

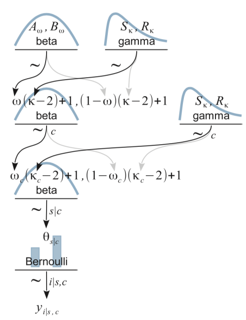


Figure 2: Baseball Hierachical model

  - Each player is denoted by s.
  - Number of oppourtunities to bat: $N_{s|c}$.
  - Number of hits: $z_{s|c}$
  - Primary position of a player: c_s

4 NCSU Hierachical models

- Presentation is available here: https://www4.stat.ncsu.edu/ reich/ABA/notes/Hier.pdf.
- Hierachical models are similar to divide and conquer problems.
- They are simple to implement because of MCMC.
- There are 3 main layers bayesian modelling:

- Data Layer: $[Y|\theta, \alpha]$ is the likelihood of the data Y.
- Process Layer: $[\theta|\alpha]$ is the model for parameters θ that define latent data generation process.
- Prior Layer: α define the prior for the hyperparameters.

4.1 Data Layer

- $S_t \implies$ susceptable individuals
- $I_t \implies$ infected individuals at time t.
- Y_t is the number of observed cases at time t.
- Data layer models our ability to process I_t.
- NO false positives and false negative probability of p.

4.2 Process Layer

- Scientific understanding of the disease is used to model how it will spread.
- We will use the Reed-Forest model

$$I_{t+1} \sim Binomial[S_t, 1 - (1-q)^{I_t}].$$

$$S_{t+1} = S_t - I_{t+1}.$$

- This model assumes that all the infected individuals at time t are removed before time $t+1$
- q is probability of non infected person coming in contact with infected person and getting the disease.

4.3 Prior Layer

- The process layer expresses disease dynamics up to a few unknown parameters.
- These unknown parameters are the priors
- Prior Layer:

$$I_t \sim Poisson(\lambda_1).$$

$$S_t \sim Poisson(\lambda_2).$$

$$p, q \sim beta(a, b).$$

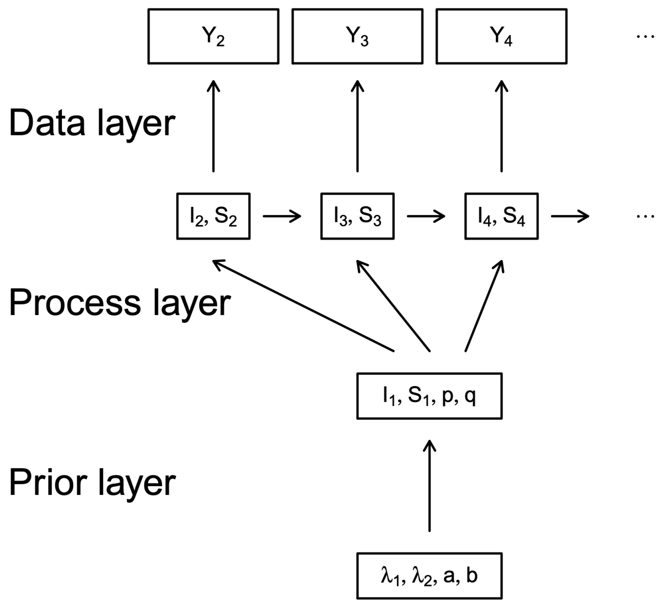


Figure 3: DAG

4.4 Hierachical models and MCMC

- MCMC is efficient for hierachical models with even larger number of parameters.
- Only consider "connected" nodes when we update each parameter.

$$1.[\theta_i|.].$$
$$2.[\mu|.].$$
$$3.[\sigma^2|.].$$
$$4.[\tau^2|.].$$

- Each of the above updates is drawn from a 1-D normal or inverse gamma distribution.
- Didn't really undestand what is happening here. I'll come back to this after a exploring few more simple examples.

5 Oxford lecture

- Lecture PDF is available here: http://www.stats.ox.ac.uk/ filippi/Teaching/HM2016/Lecture.pdf
-