STAT 491 - Week 15: Hierarchical Modeling

Hierarchical Modeling

Hierarchical models are mathematical descriptions involving multiple parameters such that the credible values of some parameters meaningfully depend on the values of other parameters. In other words, we can *borrow strength* across variables for improved estimation.

- The textbook discusses coins minted at the same factory and how a factory may generate head-biased or tail-biased coins.

- We will extend the free throw shooting idea from earlier by using free throw shooting for NBA players.

player	position	FTA	FTM
Aldridge	F	55	42
Durant	F	103	92
Favors	F	23	11
Green	F	67	46
Horford	F	29	22
James	F	162	113
Jordan	F	56	22
Love	F	75	63
Olynyk	F	30	22
Beal	G	61	50
Curry	G	114	103
Harden	G	115	101
Irving	G	84	76
Leonard	G	102	95
Oladipo	G	6	6
Parker	G	14	14
Paul	G	33	29
Wall	G	93	78

- Let ω be a parameter that describes free throw ability for players based on their position and let θ be a player specific parameter for free throw shooting ability.

-Then we are interested in the joint posterior distribution, $p(\theta, \omega|\mathcal{D})$. The joint posterior can be factored as:

$$\begin{aligned} p(\theta, \omega|\mathcal{D}) &\propto p(\mathcal{D}|\theta, \omega) \times p(\theta, \omega) \\ &= \end{aligned}$$

- The implication of the refactoring of this equation is that the data depend only on the value of θ (the player specific free throw shooting percentage). Moreover, θ depends on ω. Any model that can be factored into a chain of dependencies like this is a *hierarchical model.*

- Practically, the implication of this model is that the estimated free throw shooting percentage of a player depends on their actual shots *and* the results from the other players in the same group.

- Hierarchical models are also common in educational statistics settings where students are grouped into schools/classrooms.

Single Player Model

First consider modeling the free throw shooting for a single player.

Let $y_i = 1$ if the i^{th} free throw is made and 0 otherwise.

- Now consider a re-parameterization of the beta distribution, where instead of using a and b, the distribution is expressed in terms of the mode ω and a concentration parameter κ. Formally $\omega = (a-1)\ (a+b-2)$ and $\kappa = a+b$.

Now we can rewrite the prior as:

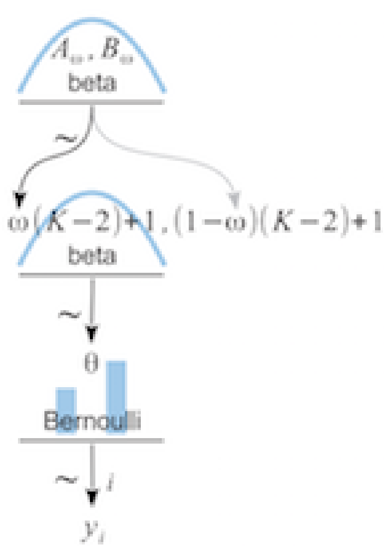
$$\theta \sim Beta(a, b) = Beta(\omega(\kappa-2)+1, (1-\omega)(\kappa-2)+1).$$

- The implication of the previous equation is that the value of θ now depends on ω and κ. The parameter θ will be near ω and the concentration around ω depends on κ.

- Up to this point, nothing has really changed. We could go ahead and specify values for ω and κ in the same fashion as we have previously with a and b. However, rather than directly choosing ω we are going to consider it to be a parameter and place a prior distribution on it. Specifically let $\omega \sim Beta(a_\omega, b_\omega)$. For now let κ be a constant.

-Again, we are interested in the joint posterior distribution, $p(\theta, \omega|\mathcal{D})$.

- This hierarchical model can also be visualized



Multiple Players from Same Position

- Now consider estimating the free throw shooting percentages for S players form the same position. Let θ_s denote the shooting percentage for the s^{th} player.

- The notation will change slightly so that $y_{i|s}$ is the i^{th} observation from the s^{th} player, so the sampling model is now $y_{i|s}|\theta_s \sim Bernoulli(\theta_s)$.

- For each θ_s, we specify $\theta_s \sim beta(\omega(\kappa - 2) + 1, (1 - \omega)(\kappa - 2) + 1)$. It may help to think of these as "latent samples" from this distribution.

- This results in a set of $S + 1$ parameters that we are interested in learning about.

$$p(\theta_1, \ldots, \theta_S, \omega|y) = \frac{p(y|\theta_1, \ldots, \theta_S, \omega) \times p(\theta_1, \ldots, \theta_S|\omega) \times p(\omega)}{p(y)}$$

- Now we are also going to place a prior on κ, the concentration parameter. We have the constraint that $\kappa - 2$ must be non-negative. So consider $\kappa \sim Gamma_+(S_\kappa, R_\kappa)$, where $Gamma_+()$ is a shifted gamma distribution. So we have

$$p(\theta_1, \ldots, \theta_S, \kappa, \omega|y) = \frac{p(y|\theta_1, \ldots, \theta_S, \kappa, \omega) \times p(\theta_1, \ldots, \theta_S|\kappa, \omega) \times p(\omega) \times p(\kappa)}{p(y)}$$

- This model can also be visualized:

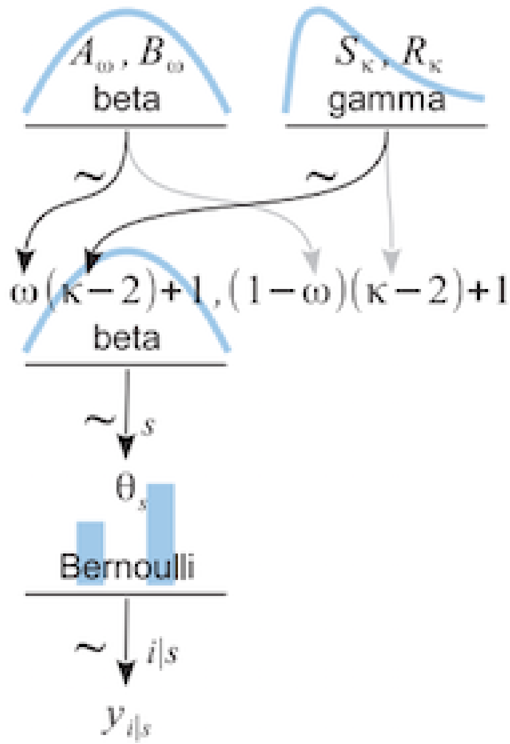


Figure 1: hierarchical model with for multiple players

```
# Data
guards <- playoff.free.throws %>% filter(position == 'G')
Nsubj <- nrow(guards)
dataList = list(z=guards$FTM, N=guards$FTA,Nsubj = Nsubj)

# Model - see page 250 in text
modelString <- 'model {
  for ( s in 1:Nsubj ) {
    z[s] ~ dbin( theta[s], N[s] )
    theta[s] ~ dbeta( omega*(kappa-2)+1, (1-omega)*(kappa-2)+1)
  }
  omega ~ dbeta(1, 1)
  kappa <- kappaMinusTwo + 2
  kappaMinusTwo ~ dgamma(.01, .01)
}'
writeLines( modelString, con='HierModel.txt')


# RUN JAGS
jags.hier <- jags.model( file = "HierModel.txt", data = dataList, n.chains = 3, n.adapt = 100000)
```

JAGS Code for Hierarchical Model

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 9
##    Unobserved stochastic nodes: 11
##    Total graph size: 40
##
## Initializing model
```

```
update(jags.hier, 10000)

num.mcmc <- 5000
codaSamples <- coda.samples( jags.hier, variable.names = c('omega', 'kappa','theta'), n.iter = num.mcmc)
#par(mfrow=c(3,4))
#traceplot(codaSamples)
guards <- guards %>% mutate(FT.PCT  = FTM / FTA)
HPDinterval(codaSamples)[[1]][2,]
```

```
##     lower     upper
## 0.8550327 0.9621387
```

```
mean(codaSamples[[1]][,2])
```

```
## [1] 0.9035692
```

	player	position	FTA	FTM	FT.PCT	post.mean	lower	upper
theta[1]	Beal	G	61	50	0.820	0.851	0.776	0.913
theta[2]	Curry	G	114	103	0.904	0.897	0.848	0.943
theta[3]	Harden	G	115	101	0.878	0.881	0.831	0.928
theta[4]	Irving	G	84	76	0.905	0.896	0.842	0.945
theta[5]	Leonard	G	102	95	0.931	0.914	0.870	0.962
theta[6]	Oladipo	G	6	6	1.000	0.898	0.813	0.986
theta[7]	Parker	G	14	14	1.000	0.911	0.838	0.987
theta[8]	Paul	G	33	29	0.879	0.882	0.811	0.949
theta[9]	Wall	G	93	78	0.839	0.857	0.795	0.909

Shrinkage in Hierarchical Models

- One principle in hierarchical models is that lower level parameters are pulled closer together based on the higher level parameters. This phenomenon is known as shrinkage, as estimates are pulled closer together. This is also commonly referred to as borrowing strength.

- Shrinkage results because the lower level parameters are influenced by:

  1. the

  2.

Multiple Players from Multiple Positions

- Now the hierarchy is extended to another level, where each group will have distinct hierarchical parameters.

- Essentially, we are modeling that guards and forwards will have different free throw shooting parameters ω and κ that influence overall group level averages.

- This model can also be visualized:

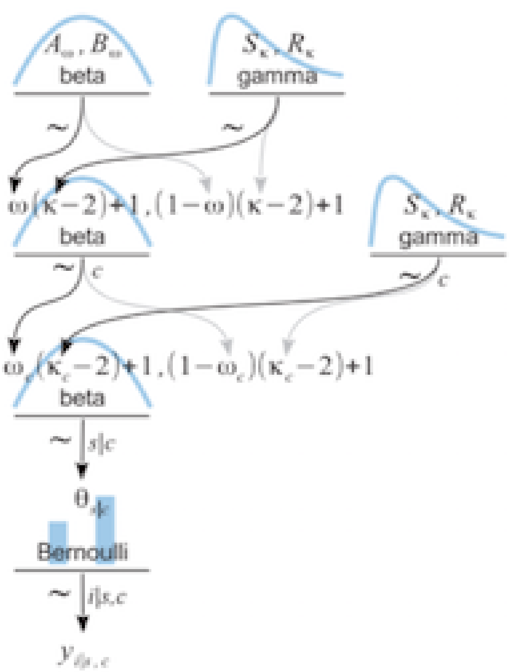


Figure 2: hierarchical model with for multiple players from multiple groups