

Homework 5

Juliette J. Rubin

March 12, 2019

Question 1 - Personal project!

<https://discourse.mc-stan.org/t/poisson-binomial-distribution-any-existing-stan-implementation/4220/6> I found it a little difficult to ascertain exactly what kinds of data can be used with these different models, but I think this one could be used in assessing my behavioral bat attack data, since this data is coded in a binomial fashion (caught/escaped) but there is often overdispersion, since some moth treatments are caught about 50% of the time, while others are only rarely caught so there are a lot of zeros.

Question 2 - Philosophy!

I very much enjoyed reading the Borjes excerpt. It seems to me that modern statistics has several reasons for invoking this garden of forking paths metaphor. First, I believe a pejorative way to use it would be to indicate that a modern statistician can get lost in the garden of forking paths - that is, one can keep “going left,” trying to stick to a particular pattern of modeling choices, but can still end up completely lost and off-course. This would be someone who gets lost in the labyrinth that is modern big data and flexible modeling techniques. Second, and more uplifting, the garden of forking paths can refer to the many ways of perceiving reality, such that none of the models you build are more correct than others, they are just representing different realities at a given time. I would imagine this would make some scientists (most policy-makers?) frustrated that there is no “correct” or “real” path, but only the current path/model that you have chosen. This can make Bayesian modeling feel slippery and ambiguous, but also full of possibility since there are new and interesting ways to consider and analyze data at every turn. The difficulty here is that there would be no reproducibility, or at least, low likelihood of this phenomenon, which until now the community has held very tightly to as an indication of good science (and when you consider this anti-vaxx thing, it does seem important that we still have this test). I wonder if one were to take the exact path as laid out by someone who previously has traveled the garden of forking paths, whether they would find themselves at the same resolution? Most likely not, since running the same Bayesian model with exactly the same parameters and conditions ends up with slightly different values with each stochastic run...

Question 3 - Planets!

```
## [1] 0.2307692
```

Question 4 - Pandas!

```
A_twins<-0.1
A_single<-0.9
B_twins<-0.2
B_single<-0.8
prior<-0.5

#likelihood that next birth is twins
post_nextA<- (A_twins*A_twins*prior)/(sum(A_twins*A_single)/2)
post_nextA
```

```

## [1] 0.1111111
#0.11 is prob of 2 twins in row if A
post_nextB <- (B_twins*B_twins*prior)/(sum(B_twins*B_single)/2)
post_nextB

## [1] 0.25
#0.25 prob of 2 twins in row if B

#likelihood that it's species A, given twins born
post_twins<-(A_twins*prior)/(sum(A_twins,B_twins)/2)
post_twins

## [1] 0.3333333
#0.33

#probability of species A, given twins then singleton born
post_A <- (A_twins*A_single*prior)/(sum((A_twins*A_single),(B_twins*B_single))/2)
post_A

## [1] 0.36
#0.36

#probability of species A, given test
#prob correct A result
prob_A<-0.8
#Prob incorrect A result (actually B)
prob_Ax<-0.2
#Prob correct B result
prob_B<-0.65
#Prob incorrect B result (actually A)
prob_Bx<-0.35

#likelihood that it is correctly identified as species A
post_id<-(prob_A*prior)/(sum(prob_A,prob_Bx)/2)
post_id

## [1] 0.6956522
#0.695

#likelihood it's species A given test and observed data
postA_real <- (post_id*post_A*prior)/(sum(((1-post_id)*post_id),((1-post_A)*post_A))/2)
postA_real

## [1] 0.5664405
#0.566 probability that it's really panda species A

```