

Floodrisk and race in Washington

Ailene Ettinger, ailene.ettinger@tnc.org (<mailto:ailene.ettinger@tnc.org>)

2/3/2020

Question

Are there racial inequities in floodrisk in Washington state?

Approach

Mathis has pulled together an enormous amount of work pulling together floodrisk and census data to get parcel-level race probabilities for the entire state of Washington (summarized in the README/technical document). He has already made a great story board, and figures as well.

I am helping with the statistical models and moving the writing along.

Statistical models

Data are grouped into nested levels of blocks, block groups, tracts, counties, and states.

Socioeconomic level exist only at the tract level. We have race (probability) at the parcel level.

We fit a multinomial mixed model with the following variables: -Response: Race (parcel level), multinomial distribution -Predictor: Flood Status, at the block level -Random effect: county

We fit a multinomial logit model, in which we assume that the log-odds of each response (proportion of people in each race category) follows a linear model:

$$\eta_{i,j} = \log(\pi_{i,j}/\pi_{i,J}) = \alpha_{county[i]} + \beta_{flood}x_i + \epsilon_i$$

$$\epsilon \sim N(0, \sigma_y^2)$$

$$\alpha_{county[i]} \sim N(\mu_\alpha, \sigma_\alpha^2)$$

where η is race in parcel i , $\alpha_{county[i]}$ is the county-specific intercept for race j , and x is flooded status (0,1) of parcel i .

We implement the model in Stan via the brms package.

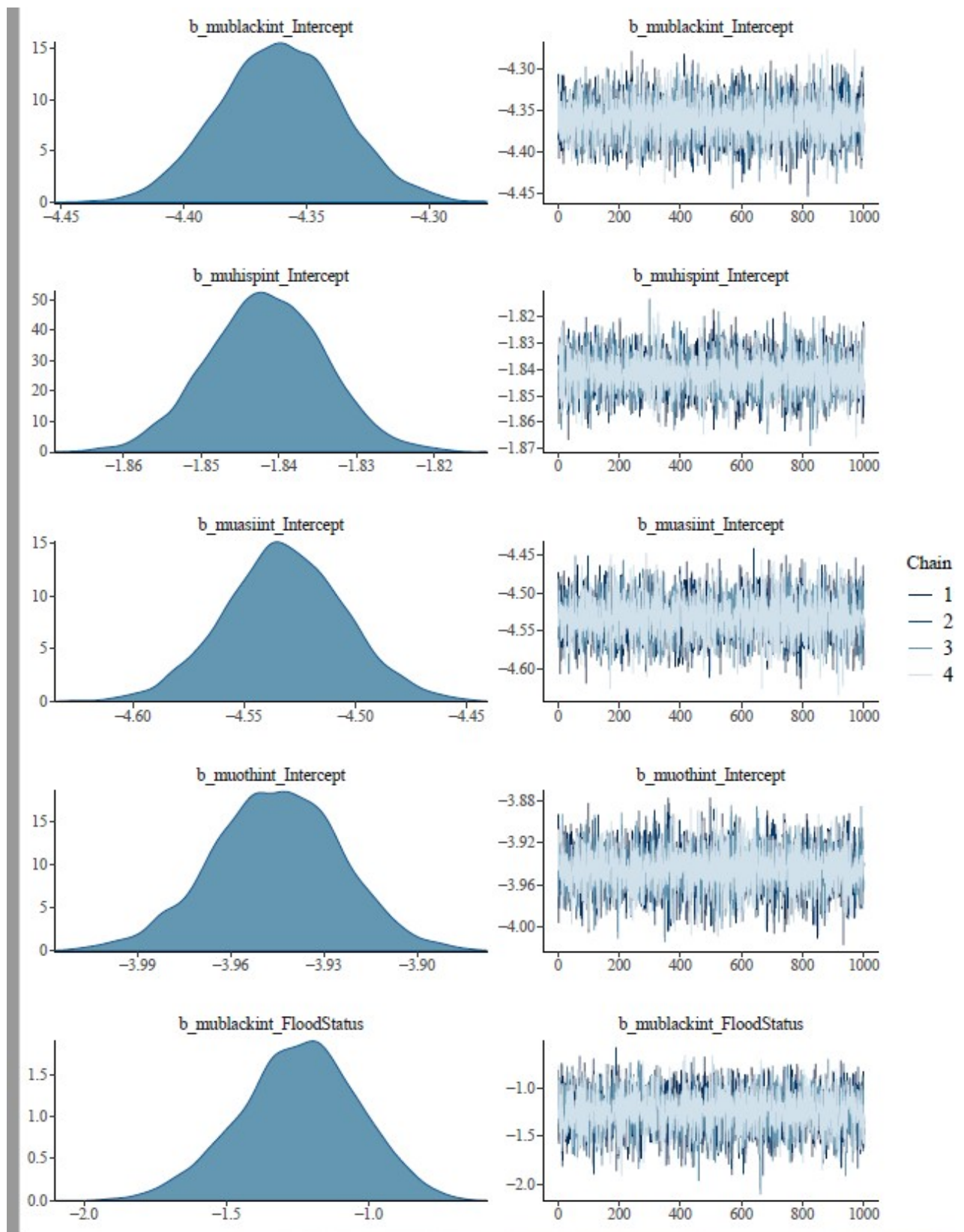
I'm not sure which model structures we ideally want:

1. Random intercept (as shown in the equation, allowing for mean differences in floodrisk by county, but assuming the effect of floodrisk to be consistent across counties) `brms: mod <- brm (race ~ floodrisk + (1|county), data=wadat, family="multinomial", prior=c(set_prior ("normal (0, 1)"))`

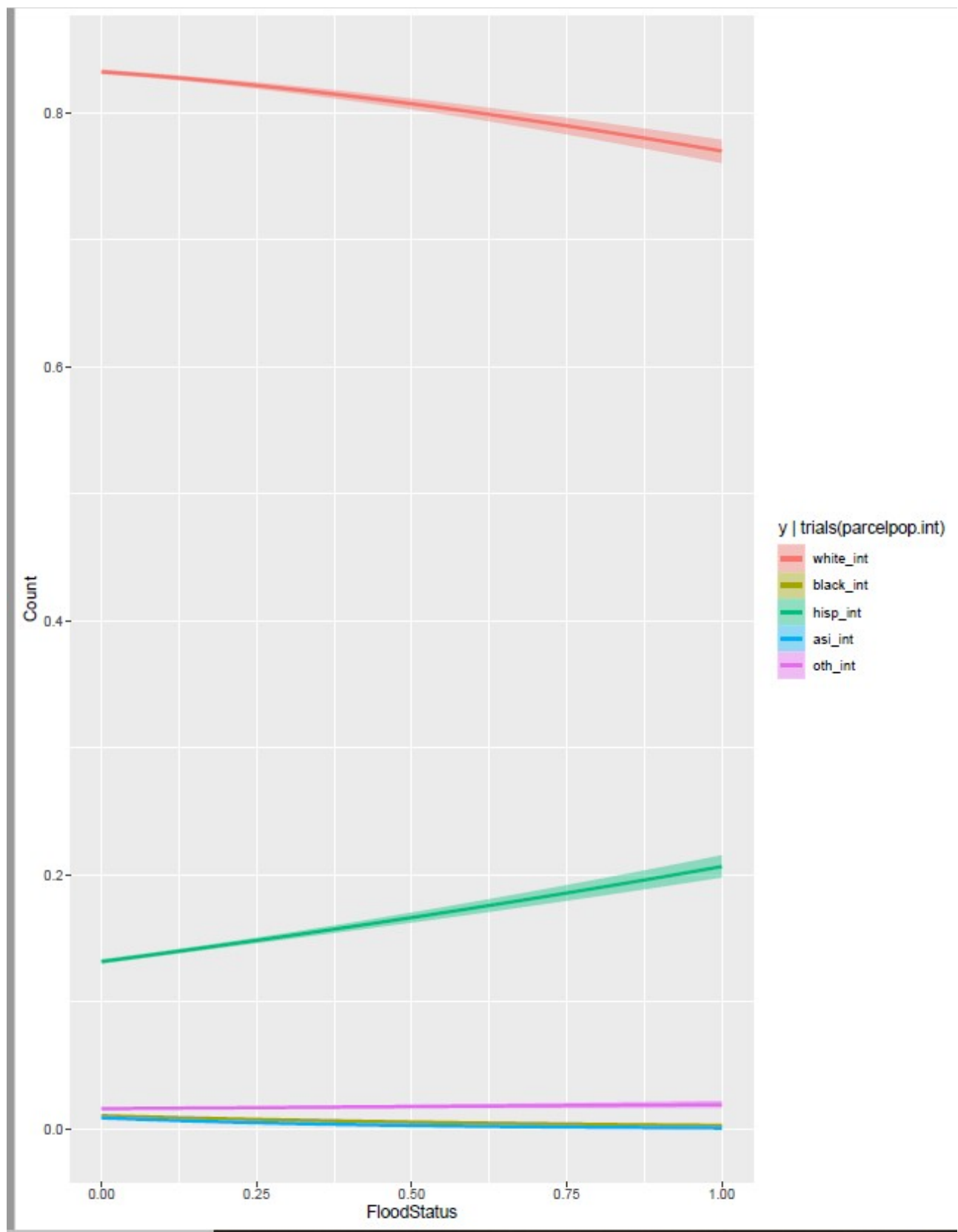
or

2. Random slopes and intercepts (allowing for mean differences in floodrisk by county and allowing the effect of floodrisk on race to vary across counties). brms: `mod <- brm (race ~ floodrisk + (floodrisk|county), data=wadat, family=multinomial(), prior=c(set_prior ("normal (0, 1)"))` ## Progress I fit a single level version of the model to a single county and for 5 counties. The multi-level version is currently running on the 5-county subset. It takes forever currently so I need to get it to run more efficiently before fitting to the full dataset (structure of the random effects, adding priors) to help it run more efficiently.

The single level model fits well



It suggests that as floodrisk increases, the probability of whites inhabiting a parcel increases and the probability of hispanics increases.



Next Steps

1. Fit multi-level model with county (using reduced dataset)
2. Fit multi-level model with county (using full dataset)
3. Outline/writing

Notes and Resources

Helpful links and code:

https://cran.r-project.org/web/packages/brms/vignettes/brms_multivariate.html (https://cran.r-project.org/web/packages/brms/vignettes/brms_multivariate.html)

<https://stackoverflow.com/questions/21082396/multinomial-logistic-multilevel-models-in-r>
(<https://stackoverflow.com/questions/21082396/multinomial-logistic-multilevel-models-in-r>)

<https://data.princeton.edu/wws509/notes/c6s2> (<https://data.princeton.edu/wws509/notes/c6s2>)