

540 Exam 2

Nate Wiecha

Analysis of bear surveys

Introduction

Our data is a summary of surveys of black bear sightings from amateur naturalists, who conduct surveys around the U.S. and upload their results to iNaturalist. The data contain one row for each location, and include the number of sightings, the number of surveys at that location, data on the characteristics of the location, and the ecological region, which can be one of four broad areas of the United States. The locations are randomly sampled, 50 km² areas from across the U.S. I recoded the grassland, forest, and cropland variables to be equal to 1 if the location is at least 1/3 that type and 0 otherwise to simplify the model, since otherwise the three variables would be (more) collinear and have a small scale, which might cause issues with the model estimation.

Our objective is to determine whether black bears have adapted differently to the different ecological regions. If the determinants of bear presence differ between ecological regions, that would be evidence of local adaptation. The independent variables are the proportion of each location that is forest, grassland, or cropland, the annual average temperature and precipitation of the location, the location's human population, and whether the location includes protected lands, as well as the location's ecological region.

Models and hypotheses

A hierarchical model is the most natural for this problem, since our hypothesis is that coefficients for some of the regressors differ by ecological region. A hierarchical model with ecological region as a process layer will allow the regressors to differ by ecological region while taking advantage of the strengths of MCMC.

To model the bear sightings, a binomial distribution would be most appropriate since our outcome is the number of "successes" (bear sightings) out of a fixed number of surveys. However, the data have a large number of zeros, so I used a zero-inflated binomial distribution for the response. I used the usual logit link function to fit the Bayesian GLM.

To test the hypothesis of local adaptation, I first used DIC and WAIC to test if a hierarchical model was actually appropriate and fit the data better than a non-hierarchical model with no local adaptation. As discussed below a hierarchical model was more appropriate than a nonhierarchical model. To determine what local adaptation we have evidence for, I performed pairwise comparisons, by the four regions, of the slope estimates using the MCMC samples from the posterior distributions of the slopes.

The fully hierarchical model is below, i.e., with both the binary and positive portions hierarchical. The other models tested were a model with a non-hierarchical binary portion of the model and a hierarchical positive portion, and with both nonhierarchical binary and positive portions. I display the fully hierarchical model below since the other two are special cases of this one. For simplicity X indicates the design matrix which includes an intercept.

Data layer:

$$Y_i = Y_{bin,i}Y_{pos,i} \quad Y_{bin,i} \in \{0, 1\}, \quad Y_{pos,i} \in \{0, 1, 2, \dots, N_i\}$$

For observation i .

Process layer:

$$Y_{bin,i} \sim \text{Binomial}(1, q_i), \quad \text{logit}(q_i) = \mathbf{X}_i \boldsymbol{\delta}_k$$

$$Y_{pos,i} \sim \text{Binomial}(N_i, p_i), \quad \text{logit}(p_i) = \mathbf{X}_i \boldsymbol{\gamma}_k$$

For observation i and the observation's ecological region k .

Prior layer:

$$\boldsymbol{\delta}_k \sim \text{MVN}(\boldsymbol{\lambda}, \tau_a \mathbf{I})$$

$$\boldsymbol{\gamma}_k \sim \text{MVN}(\boldsymbol{\beta}, \tau_e \mathbf{I})$$

For ecological region k .

Hyperprior layer:

$$\boldsymbol{\lambda} \sim \text{MVN}(\mathbf{0}, 0.001\mathbf{I})$$

$$\boldsymbol{\beta} \sim \text{MVN}(\mathbf{0}, 0.001\mathbf{I})$$

$$\tau_a, \tau_e \sim \text{InvGamma}(0.1, 0.1)$$

Computation

Sampling was performed using JAGS in R. 10,000 samples were used for burn-in and two chains of 20,000 samples each were used as the posterior sample. Convergence was assessed graphically, with the effective sample size, and with the Geweke diagnostic.

For the preferred model (see below), graphics showed that all chains except for the coefficient for the cropland variable had converged reasonably well. The effective sample size, however, showed that only several of the chains for the coefficients had fully converged. The Geweke diagnostic showed that all the coefficients of interest except for the cropland variables had converged.

Since the time required to even run a relatively small number of samples from the three models under consideration was very high, I decided to omit the cropland variable from further analysis and consider the other coefficients' chains as having converged reasonably well on the basis of two out of three diagnostics. With more time this could perhaps be improved with more MCMC samples, and the cropland variable was not too worrying since there were not as many observations that had cropland compared to forest and grassland.

Model Comparisons

The three models under comparison were the fully hierarchical model, where both the binary and positive portions' linear predictors have slopes that can vary by ecological region; a partially hierarchical model, where only the positive portion's linear predictor slopes can vary by ecological region; and a non-

hierarchical model, where all slopes are constant across ecological region. Models were compared using DIC and WAIC. The results are displayed in Table 1.

Table 1. Model comparisons

Model	WAIC	DIC
Fully hierarchical	15,886	1,040
Partially hierarchical	5,133	1,040
Non-hierarchical	6,977	1,134

DIC has no preference between the fully and partially hierarchical models, and WAIC prefers the partially hierarchical model, so I chose the partially hierarchical model to test the hypothesis that black bears exhibit some local adaptation.

Fit was assessed with posterior predictive checks. Since the minimum of the sample will always be zero, minimum and range were not useful, so I checked the maximum and mean. The maximum and mean both showed no evidence of lack of fit (figure 1). The Bayesian p-value for the maximum was 0.28, and for the mean was 0.23.

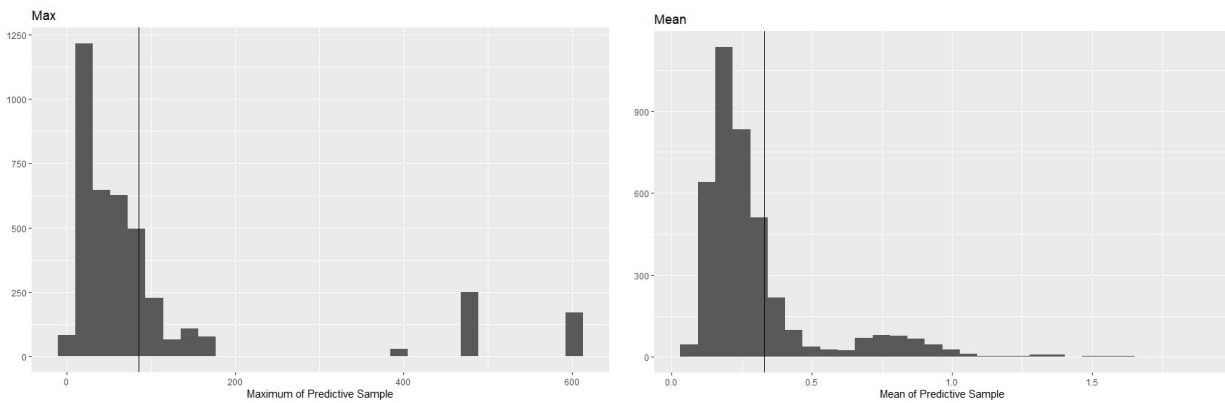


Figure 1. Posterior Predictive Checks

Results

To check for local adaptation by ecological region, I checked, for each independent variable, whether some ecological regions had different coefficients than other ecological regions. I performed this check pairwise by region. Since inference on the actual regression estimates is of less interest than the differences between ecological regions, I do not summarize the posterior distributions for all 28 estimated slopes displayed below aside from the pairwise comparisons.

The estimated regression coefficients (using posterior mean) for the positive part of the model for each ecological region are displayed in Table 2. Estimates that differ with at least 95% probability from any other estimates in the same row are bolded. There is evidence of local adaptation with respect to several variables and ecological regions.

Table 2. Linear predictor slope estimates

	Marine West Coast Forest	Mediterranean California	North American Deserts	Northwestern Forested Mountains
Forest	-0.120	-0.115	-0.122	-0.121
Grassland	-0.118	-0.275	-0.289	-0.462
Cropland (no convergence)	-7.059	-7.059	-0.758	-7.061
Temperature	-0.004	-0.101	-0.051	0.022
Precipitation	0.000	0.003	-0.001	-0.001
Human population	0.001	0.004	-0.007	-0.007
Protected	-0.079	-0.275	-0.283	-0.503

In sum, this analysis finds the following evidence of local adaptation of black bears:

- Temperature is associated with decreased expected black bear sightings in Mediterranean California; slightly decreased (less Mediterranean California) expected bear sightings in Marine West Coast Forest; and greater expected bear sightings in Northwestern Forested Mountains
- Precipitation is associated with no change in expected black bear sightings in Marine West Coast Forest; increased expected bear sightings in Mediterranean California; and decreased expected bear sightings in Northwestern Forested Mountains
- Human population is associated with increased expected black bear sightings in Mediterranean California and decreased expected bear sightings in Northwestern Forested Mountains

Discussion

This analysis used hierarchical modeling to find evidence for local adaptation of black bears to different ecological regions in the United States. The effects of temperature, precipitation, and human population on bear populations (indicated by frequency of bear sightings) differed by ecological region, indicating that bears' habitat preferences are different depending on the local ecology.

There are several limitations to the analysis. The data were randomly sampled locations rather than the full data, and did not include the spatial location. The full dataset with spatial information included could have provided more complete data, and could allow accounting for spatial covariance between observations. Another limitation is that this model did not achieve full, unambiguous convergence in a reasonable run time. With (possibly greatly) more time available it might be feasible to achieve full convergence of all parameters' posterior samples.

```

#####
#                               Hierarchical Positive Model           #
#                               Non-Hierarchical Binary Model         #
#####

hier2_model_string <- textConnection("model{
#Data layer
for(i in 1:n){
  Y_bin[i] ~ dbinom(q[i], 1)
  Y_pos[i] ~ dbinom(pi[i], N[i])

#Process
logit(q[i]) <- a + inprod(X[i,], delta[]) #nonrandom slope/intercept
logit(pi[i]) <- b[ecoregions[i]] + inprod(X[i,], gamma[,ecoregions[i]]) #gamma is random slope

# Likelihood
like[i] <- ifelse(Y[i]==0, q[i] + (1-q[i]) * dbin(Y[i], pi[i], N[i]), (1-q[i])*dbin(Y[i],pi[i], N[i]))
}

#Prior layer
for(j in 1:p){
  delta[j] ~ dnorm(0, 0.001)
  for(k in 1:n_eco){
    gamma[j,k] ~ dnorm(beta[j], taue) #random effect for coefficient j, in eco k
  }
}

for(k in 1:n_eco){
  b[k] ~ dnorm(0, 0.001)
}

a ~ dnorm(0, 0.001)

#Hyperprior layer
for(u in 1:p){
  beta[u] ~ dnorm(0, 0.001)
}
taue ~ dgamma(0.1, 0.1)

# Posterior Predictive Checks
for(i in 1:n){
  Y2_bin[i] ~ dbinom(q[i], 1)
  Y2_pos[i] ~ dbinom(pi[i], N[i])
  Y2[i] <- Y2_bin[i] * Y2_pos[i]
}

D[1] <- mean(Y[] - Y2[])
D[2] <- max(Y2[])
D[3] <- mean(Y2[])
}
")

```