Week 9 — Homework

Caio Borges Aguida Geraldes

March 10, 2022

Previous assumptions

As it will become important to our theoretical model, it must be noted that the data in Hill and Kintigh (2009) has been collected from 1980 to 2007 by two different means: observational (1980–2000) and by interviews (1994–2007). Notice also that there are no data points for the years 1986 to 1993. Whereas the authors argue that the data for kilograms of meat and the hunter-dependent information, the number of hours for each hunting trip was not consistently recorded, as shown in Figure 1.
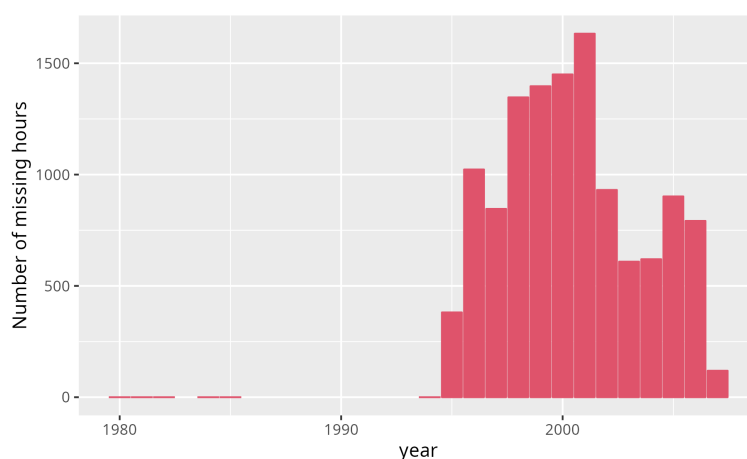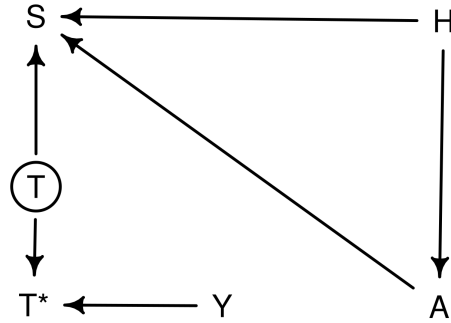

Figure 1: Count of missing data for hours expend hunting by year.

As a consequence, the scientific model must account for year, in some measure, as the variable T is unobserved, and the hours actually observed T^* are dependent on year. Hunter H, age A, and time T will affect the likelihood of scoring a game in a hunting trip. The DAG of the model assumed will be:

Exercice 1

Assuming that all the Aché individuals in the data `Achehunting` are functionally adults (as they seem to take active part in the tasks for sustaining their community), we can suppose that their age will affect their hunting success in a quadratic way: young adults will be still learning how to hunt properly, whereas older members of the community will gradually loose their ability to hunt effectively. I model the effect of age over hunting success as a quadratic linear model, where α denotes the average likelihood of any member of the community to successfully come back from a hunting trip with some game, and β the effect of z-score of age squared:

$$\text{logit}(p) = \alpha + \beta A_i^2$$

As the success of any hunting trip i, S_i, is true or false, it is modeled as a Bernoulli distribution on p_i:

$$S_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit}(p_i) = \alpha + \beta A_i^2$$

I pick very flat priors that favour values of $\beta \leq 0$, so the final model is:

$$S_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit}(p_i) = \alpha + \beta A_i^2$$
$$\alpha \sim \text{Normal}(0, 0.1)$$
$$\beta \sim \text{Normal}(0, 0.05)$$

The model is processed by `ulam` and stored in a model file `models/mA.rds`:

```
1 f.mA <- alist(
2   S ~ bernoulli(p),
3   logit(p) <- a + b*A^2,
4   a ~ normal(0, 0.1),
5   b ~ normal(0, 0.05)
6 )
7
8 mA <- ulam( flist = f.mA, dat = dat, cores = 4, chains = 4,  iter = 2000, file = "./
      models/mA")
```

The precis table shows that β is, as expected, consistently negative. The whole distribution of β is shown in Figure 2.

```
1 precis(mA)
2 #         mean          sd        5.5%        94.5%     n_eff      Rhat4
3 # a   0.1371577 0.02232976   0.1009685   0.1736372 1380.094 1.001854
4 # b  -0.2351356 0.01524043  -0.2594043  -0.2106380 1485.530 1.000031
```
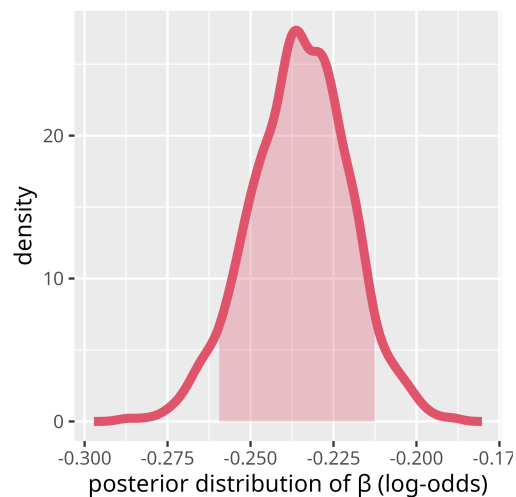

Figure 2: Posterior distribution of β with the 89% HPDI compatibility interval shown in the red area.

Using the predictive posterior distribution of the probabilities of returning with some game after a hunting trip, we can compare the model results with the data, Figure 3. The model has a lot of recall, particularly in the cases of individuals of lower age, as they seem to have not taken part on many hunting trips (hunters of age 11, 12 and 13 have only 1, 9 and 4 hunts recorded in total), so there is not enough data to see their real chances of having a successful hunt.
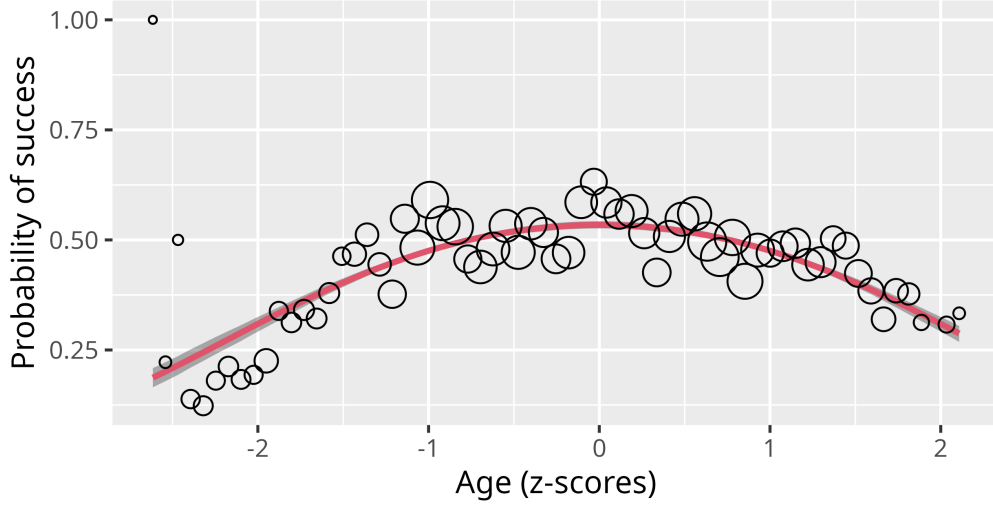
Figure 3: Posterior predictive distribution. The number of documented hunting trips per age is represented by the size of the points. The curve indicates the mean of the predictive posterior distribution with the 89% HPDI compatibility interval.

Exercice 2

Assuming that each hunter H has their own intercept, representing their physical and mental characteristics that affect their likelihood of scoring some game by each hunting trip they take, I change the previous parameter α to vary according to the hunter H, α_H. This parameter is partially pooled from the whole community of hunters, so that there is an average likelihood for any hunter to have a successful hunting trip, $\bar{\alpha}$, which is summed to an individual effect. I repeat the prior for α from the previous model as $\bar{\alpha}$ and build an non-centered distribution for the individual effects:

$$\alpha_H = \bar{\alpha} + z_{\alpha,H}\sigma_\alpha$$
$$\bar{\alpha} \sim \text{Normal}(0, 0.1)$$
$$z_{\alpha,H} \sim \text{Normal}(0, 1)$$
$$\sigma_\alpha \sim \text{Exponential}(1)$$

I include a non-centered parametrization also for β, so that $\beta = z_\beta\sigma_\beta$. The complete model used is as follows:

$$S_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit}(p_i) = \alpha_{H_i} + \beta A_i^2$$
$$\alpha_j = \bar{\alpha} + z_{\alpha,j}\sigma_\alpha \quad \text{for } j = 1 \ldots 147$$
$$z_{\alpha,j} \sim \text{Normal}(0, 1)$$
$$\bar{\alpha} \sim \text{Normal}(0, 0.1)$$
$$\sigma_\alpha \sim \text{Exponential}(1)$$
$$\beta = z_\beta \sigma_\beta$$
$$z_\beta \sim \text{Normal}(0, 1)$$
$$\sigma_\beta \sim \text{Exponential}(1)$$

The model is processed by `ulam` and stored in a model file `models/mAI.rds`:

```
f.mAI <- alist(
  S ~ bernoulli(p),
  logit(p) <- abar + zA[I]*sA + zB*sB*A^2,
  vector[Nh]:zA ~ normal(0, 1),
  zB ~ normal(0, 1),
  abar ~ normal(0, 0.1),
  sA ~ exponential(1),
  sB ~ exponential(1),
  gq> vector[Nh]:a <<- abar + zA*sA,
  gq> real[1]:b <<- zB*sB
)

mAI <- ulam( flist = f.mAI, dat = dat, cores = 4, chains = 4,  iter = 2000, file = ".
    /models/mAI")
```

The precis table for the fitted model:

```
precis(mAI, pars=c("abar", "b", "sA", "sB"))
#              mean         sd        5.5%       94.5%      n_eff      Rhat4
# abar   0.01441468 0.05768174 -0.07571847  0.1057518  898.4824 1.001905
# b     -0.28926879 0.02240940 -0.32476412 -0.2537649 3728.3862 1.000386
# sA     0.57322113 0.05804444  0.48758282  0.6734943  805.6828 1.000326
# sB     0.66486412 0.48710636  0.18359077  1.6266105 2285.0854 1.003606
```

Comparing to the previous model, the effect of age is overall lower and has more variation, so that stronger and weaker effects are possible. The effect is still negative, as expected. See Figure 4. The effect of the individual α_H still has a lot of variation depending on how many trips were documented for that particular individual. See Figure 5.

The variation of α and β is described by the parameters σ_α and σ_β. The mean variation of β is slightly higher than that of α, and it is itself subject to much more uncertainty, having a very long tail, see Figure 6. This indicates that much of the
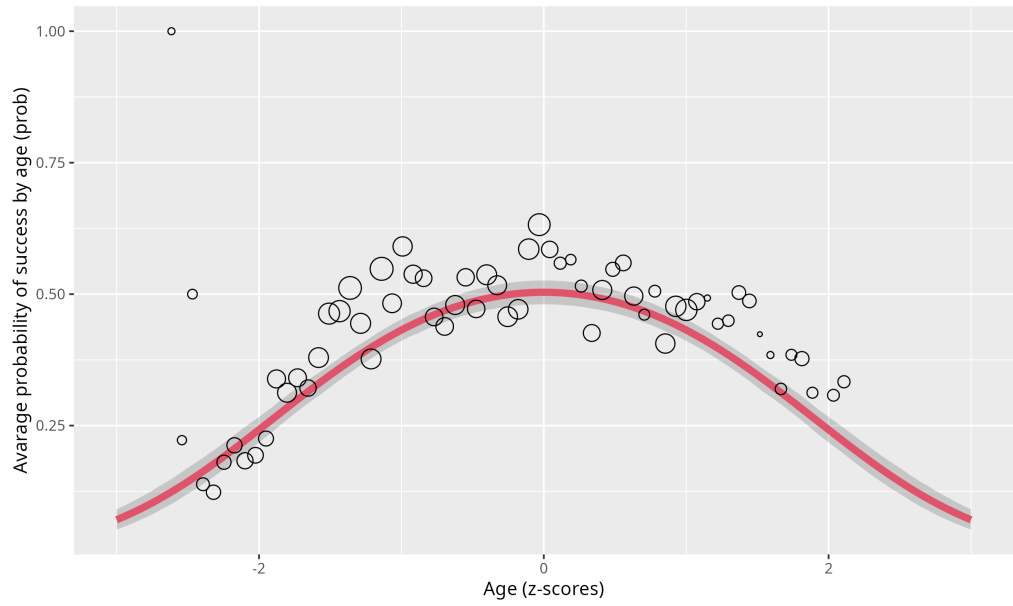
Figure 4: Posterior predictive distribution, marginalized over individuals (using $\bar{\alpha}$ in the link function). The number of documented hunting trips per age is represented by the size of the points. The curve indicates the mean of the predictive posterior distribution with the 89% HPDI compatibility interval.
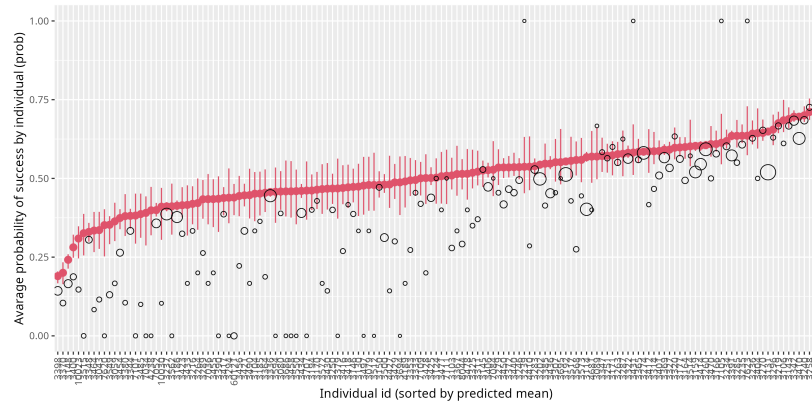

Figure 5: Posterior predictive distribution, marginalized over age ($A = 0$ in the link function). The number of documented hunting trips per age is represented by the size of the points. The curve indicates the mean of the predictive posterior distribution with the 50% HPDI compatibility interval.

variation is explained by the effect of the characteristics of the individuals on the chance of a hunting trip being successful.
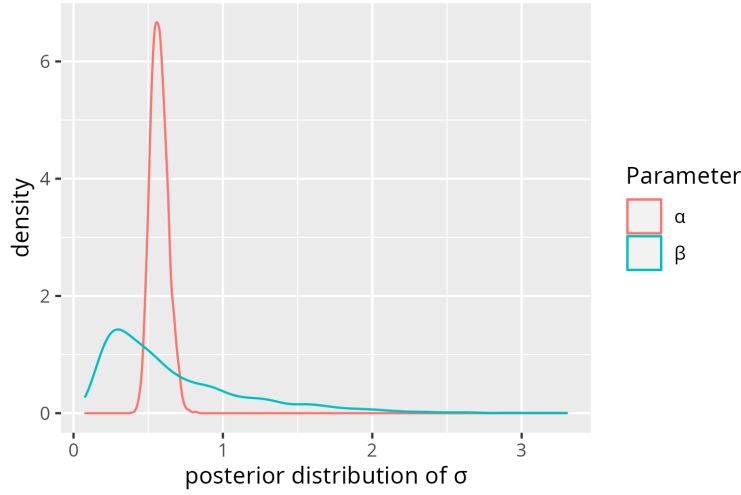

Figure 6: Posterior distribution of σ_α and σ_β.

Exercice 3

Complete cases   To account for effect of the time taken for each trip on success on the complete cases, I assume that longer trips will give more opportunities for *some* game to be scored, so I include a linear effect γ over the time T. In this model, the time taken for each trip is assumed to be independent of the hunters characteristics, believing that most of the variation is due to weather, luck and patterns of geographic distribution of the fauna being hunted. To account for measurement errors of T, I include a submodel for T_i, that is a Normal distribution with mean on the average time taken for the recorded trips $\bar{T}$ and standard deviation σ_T.

$$T_i \sim \text{Normal}(\bar{T}, \sigma_T)$$
$$\sigma_T \sim \text{Exponential}(1)$$

The full model is as follows:

7

$$S_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit}(p_i) = \alpha_{H_i} + \beta A_i^2 + \gamma T_i$$

$$T_i \sim \text{Normal}(\bar{T}, \sigma_T)$$
$$\sigma_T \sim \text{Exponential}(1)$$

$$\alpha_j = \bar{\alpha} + z_{\alpha,j}\sigma_\alpha \qquad \text{for } j = 1 \ldots 147$$
$$z_{\alpha,j} \sim \text{Normal}(0, 1)$$
$$\bar{\alpha} \sim \text{Normal}(0, 0.1)$$
$$\beta = z_\beta \sigma_\beta$$
$$z_\beta \sim \text{Normal}(0, 1)$$
$$\gamma \sim \text{Normal}(0, 1)$$
$$\sigma_\alpha, \sigma_\beta \sim \text{Exponential}(1)$$

The model is processed by `ulam` and stored in a model file `models/mAITc.rds`:

```
f.mAITc <- alist(
  S ~ bernoulli(p),
  logit(p) <- abar + zA[I]*sA + zB*sB*A^2 + g*T,
  vector[Nh]:zA ~ normal(0, 1),
  abar ~ normal(0, 0.1),
  zB ~ normal(0, 1),
  g ~ normal(0,1),
  T ~ normal(Tbar, sigma_T),
  sigma_T ~ exponential(1),
  sA ~ exponential(1),
  sB ~ exponential(1),
  gq> vector[Nh]:a <<- abar + zA*sA,
  gq> real[1]:b <<- zB*sB
)

mAITc <- ulam(flist = f.mAITc, data=dat.cmp, chains = 4, cores = 4, iter = 2000, file
      = "./models/mAITc")
```

Incomplete cases To account for effect of the time taken for each trip on success on the incomplete cases, the assumptions are the same as in the previous model. The difference is in the submodel for T_i, that now has the work of imputing values for T. We must not use the value for $\bar{T}$ directly in the normal distribution of T_i, as the real

mean is unknown, so we derive a mean μ_T from another normal distribution with mean on the complete cases $\bar{T}$:

$$T_i \sim \text{Normal}(\mu_T, \sigma_T)$$
$$\mu_T \sim \text{Normal}(\bar{T}, 1)$$
$$\sigma_T \sim \text{Exponential}(1)$$

The full model is as follows:

$$S_i \sim \text{Bernoulli}(p_i)$$
$$\text{logit}(p_i) = \alpha_{H_i} + \beta A_i^2 + \gamma T_i$$

$$T_i \sim \text{Normal}(\mu_T, \sigma_T)$$
$$\mu_T \sim \text{Normal}(\bar{T}, 1)$$
$$\sigma_T \sim \text{Exponential}(1)$$

$$\alpha_j = \bar{\alpha} + z_{\alpha,j}\sigma_\alpha \qquad \text{for } j = 1\dots147$$
$$z_{\alpha,j} \sim \text{Normal}(0, 1)$$
$$\bar{\alpha} \sim \text{Normal}(0, 0.1)$$
$$\beta = z_\beta \sigma_\beta$$
$$z_\beta \sim \text{Normal}(0, 1)$$
$$\gamma \sim \text{Normal}(0, 1)$$
$$\sigma_\alpha, \sigma_\beta \sim \text{Exponential}(1)$$

The model is processed by `ulam` and stored in a model file `models/mAITi.rds`:

```
1  f.mAITi <- alist(
2    S ~ bernoulli(p),
3    logit(p) <- abar + zA[I]*sA + zB*sB*A^2 + g*T,
4    vector[Nh]:zA ~ normal(0, 1),
5    abar ~ normal(0, 0.1),
6    zB ~ normal(0, 1),
7    g ~ normal(0,1),
8    T ~ normal(mu_T, sigma_T),
9    mu_T ~ normal(Tbar, 1),
10   sigma_T ~ exponential(1),
11   sA ~ exponential(1),
12   sB ~ exponential(1),
```

```
13   gq> vector[Nh]:a <<- abar + zA*sA,
14   gq> real[1]:b <<- zB*sB
15 )
16
17 mAITi <- ulam(flist = f.mAITi, data=dat.inc, chains = 4, cores = 4, iter = 2000, file
        = "./models/mAITi")
```

References

HILL, Kim; KINTIGH, Keith. Can Anthropologists Distinguish Good and Poor Hunters? Implications for Hunting Hypotheses, Sharing Conventions, and Cultural Transmission. *Current Anthropology*, v. 50, n. 3, p. 369–378, 2009. DOI: 10.1086/597981.