

Modèles hiérarchiques bayésiens 2

Contents

Données	1
Bayesian model of the probability of reproduction according to age and ros	2

Données



Figure 1: Svalbard reindeer

The data are from my post-doc. We will analyze the probability of a reindeer to have a baby during the summer. In this system, one of the most important environmental factors is the presence of rain-on-snow events. These occur when precipitation occurs during the winter. These freeze and then form a thick layer of ice that blocks access to food resources.

However, as the Arctic continues to warm, some researchers believe that the ros will no longer have an effect. When there are sustained rain events followed by a warm period, the rain causes a release of food resources and has time to run off before freezing. We will attempt to explore these changes in the effect of ros.

Translated with www.DeepL.com/Translator (free version)

```
library(dplyr)
library(readr)
library(ggplot2)
library(tidyr)
```

```
library(cowplot)
library(lubridate)

dat <- read_csv("../donnees/SvalbardDat.csv")
dat <- dat %>% mutate(age=year-yrbirth) %>% filter(age>1)

ros <- read_csv("../donnees/ROS.csv")
```

We first transform the predictors: - *rosNorm* is the logarithm of *ros*, normalized to have a mean of 0 and a standard deviation of 1. - we create a variable *age2* for the quadratic effect of age - and finally a *period* variable that separates the study into 5 periods

Bayesian model of the probability of reproduction according to age and ros

Of course, we will also need to control for the age of the individuals. The model will therefore be a binomial model with age and ros as fixed effects. The random effects will consist of the year, the period and a slope of the ros varying with the period.

Notes:

- The model formula in **brm** follows the same syntax as **lmer** for the specification of fixed and random effects.
- Although it would be possible to add a random country effect on the **age:ros** interaction, year, ID, density and several other control variables. We omit them here to reduce the computational time of the models.

A Choose *a priori* distributions for the parameters of the model described above. Here is an example of code where only the specification of the distributions is missing. The first four lines define the *a priori* distributions for the intercept and coefficients of the three fixed effects, the next three define the distributions for the standard deviations of the random effects (**class** = "sd"), while the last one refers to the standard deviation of the individual observations (**class** = "sigma").

```
library(brms)
my_prior <- c(set_prior("", class = "Intercept"),
              set_prior("", class = "b", coef = "ages"),
              set_prior("", class = "b", coef = "age2"),
              set_prior("", class = "b", coef = "rosNorm"),
              set_prior("", class = "sd", coef = "Intercept", group = "id"),
              set_prior("", class = "sd", coef = "Intercept", group = "period"),
              set_prior("", class = "sd", coef = "rosNorm", group = "period"))
```

It is recommended to choose normal distributions in all cases. For 'sd', these distributions will be interpreted as half-normal because it is implied that these parameters are ≥ 0 . To choose the mean and standard deviation of each normal distribution, consider the interpretation of each parameter and in particular the scales of the predictors **ros**, **ages** and **age2**. In brms, the family used will be **family=bernoulli("logit")**.

As for the standard deviations of the random effects ("sd"), their distribution *a priori* can have the same width as that of the corresponding "b" coefficient.

Now draw a sample of the joint *a priori* distribution of parameters with **brm**. I suggest specifying **chains** = 1, **iter** = 1500, **warmup** = 1000 to produce a single Markov chain with 1000 run-in iterations and 500 sample iterations. Then visualize the distribution of **calf** predicted for each iteration of the *a priori* parameters.

Because of the large number of effects estimated and the fact that we are imposing only mild constraints on each distribution *a priori*, extreme or even impossible values (large positive and negative values) are to be expected; the important thing is that the density is larger within a realistic range. It may be useful to "zoom

in” on part of the `ggplot` by adding `coord_cartesian(xlim = c(..., ...), ylim = c(..., ...))` with limits in x and y .

C Now fit the model with `brm`. You can reduce the number of Markov chains to 2 to save time, but keep the default values for the number of iterations. (You can ignore the warning that the effective sample size or ESS is small). How can you evaluate the convergence of the model?

D Compare the magnitude of the ‘`rosNorm`’ coefficient to that of the random effects. What does this comparison tell you?

E Check posterior: Apply `predict` to the model to get the mean, standard deviation and 95% interval for the *hindcast* prediction. You can give a `data.frame` to the `newdata` argument to get the desired predictions (`expand.grid(rosNorm=seq(-1.6,1.6,l=30),period=unique(dat$period),...)`). Illustrate the model predictions and their credibility intervals for the different periods for a 7 year old individual.