Introduction to Bayesian analysis

Data

Already used for maximum likelihood exercises, the thermal_range.csv dataset represents the result of an experiment to determine the effect of temperature (*temp*) on the number of eggs (*num_eggs*) produced by a species of mosquito. Three replicates were measured for temperature values between 10 and 32 degrees Celsius.

```
library(brms)

therm <- read.csv("../donnees/thermal_range.csv")
head(therm)
```

```
##   temp num_eggs
## 1   10        1
## 2   10        1
## 3   10        2
## 4   12        4
## 5   12        4
## 6   12        6
```

Bayesian estimation of the thermal optimum model

Let's remember the model used previously for this dataset. The average number of eggs produced is given by a Gaussian curve:

$$N = N_o \exp\left(-\frac{(T - T_o)^2}{\sigma_T^2}\right)$$

In this equation, T_o is the optimum temperature, N_o is the number of eggs produced at this optimum and σ_T represents the tolerance around the optimum (the higher σ_T is, the slower N decreases around the optimum).

a) It is possible to estimate the parameters of a non-linear model like this one in *brms*. For example:

```
brm(bf(num_eggs ~ No * exp(-(temp-To)^2/sigmaT^2), No + To + sigmaT ~ 1, nl = TRUE),
    data = therm)
```

Note:

- We need to enclose the formula in a `bf` function and specify the argument `nl = TRUE` (for non-linear).
- After the non-linear formula of the model, we need to add a term describing the parameters. Here, `No + To + sigmaT ~ 1` only means that we estimate a fixed effect for each parameter. If one of the parameters varied according to a group variable, we could write for example `No ~ (1|group), To + sigmaT ~ 1`.

Since we are going to use a negative binomial distribution with a logarithmic relationship to represent the mean of the response (`family = negbinomial` in *brms*), we need to modify the formula above to represent the logarithm of the mean number of eggs N. Rewrite the `bf` function by applying this transformation.

b) Choose appropriate prior distributions for three parameters in the equation obtained above. In the `set_prior` statement, the parameter name is specified with `nlpar` for a non-linear model. For example, `set_prior("normal(0, 1)", nlpar = "To")` assigns a standard normal distribution to the parameter `To`.

Note: Don't forget to specify the lower bound for `sigmaT`.

Also add a prior distribution for the θ parameter of the negative binomial distribution with `set_prior("gamma(2, 0.1)", class = "shape")`. You can visualize this distribution in R with `plot(density(rgamma(1E5, 2, 0.1))`. Since the variance of the negative binomial distribution is $\mu + \mu^2/\theta$, where μ is the mean, we want to avoid values of θ too close to zero. With the specified parameters, θ is small for values close to 0 and greater than 50 (with a θ so large, the negative binomial distribution almost matches that of Poisson).

c) Fit the non-linear model with `brm`, using the formula and prior distributions specified in the previous parts, with a negative binomial distribution of the response. Visualize the shape of the estimated N vs. T function with `marginal_effects`. Determine the mean value and the 95% credibility interval for the posterior distribution of each parameter.

d) Compare the results in (c) with the maximum likelihood estimates and confidence intervals obtained in lab 3, reproduced in the table below.

Parameter	Estimate	Interval
N_o	123.2	(104.2, 147.2)
T_o	23.9	(23.4, 24.5)
$sigma_T$	6.82	(6.33, 7.42)
k	0.103	(0.059, 0.186)

Note: The parameter k corresponds to $1/\theta$ for the negative binomial distribution.

e) Check the posterior prediction intervals with `pp_check(..., type = "intervals")`. Do the observations appear to be consistent with the fitted model?

f) As we will see next week, the *Stan* program that *brms* uses produces a sample of the joint posterior distribution of the model parameters. By default, this sample includes 4000 parameter sets. The `posterior_epred` function of *brms* calculates the mean prediction according to each of these parameter sets for a new dataset given by the `newdata` argument, like the `predict` function in the case of regression models.

Use the `posterior_epred` function to calculate the ratio of mean egg production at 25 degrees C compared to 20 degrees C, and a 95% credibility interval for this ratio.