Lab 7, Act 2: Your Own Sampler

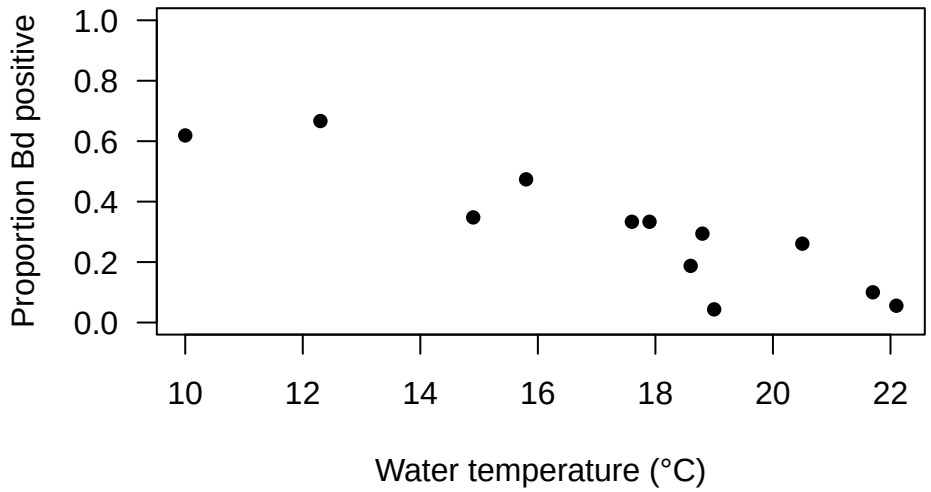
Is chytrid less common in warmer ponds?

NRES 746, name: _____ group: _____

The scenario

The pond in Act 1 was one of 12 ponds the field crew surveyed in the same summer. At each pond they swabbed a sample of frogs and recorded the water temperature. Bd grows poorly in warm water, so managers want to know whether infection is less common in warmer ponds, and if so, by how much. Here are the survey results (the Act 1 pond is P11):

Pond	Water temp. (°C)	Frogs swabbed	Bd positive
P01	10.0	21	13
P02	12.3	18	12
P03	14.9	23	8
P04	15.8	19	9
P05	17.6	18	6
P06	17.9	24	8
P07	18.6	16	3
P08	18.8	17	5
P09	19.0	23	1
P10	20.5	23	6
P11	21.7	20	2
P12	22.1	18	1



Use a logistic model: the number of positives at pond i is binomial with n_i frogs swabbed and infection probability p_i, where

$$\log \frac{p_i}{1 - p_i} = a + b\,T_i$$

and T_i is the water temperature. Give both a and b a Normal(0, 10) prior (wide enough that any signal from the data will overpower it!). You'll fit this model with your own Metropolis sampler.

How much does the chance of infection change with temperature, and how uncertain is that change?

Try Part 1 on paper first, before you open a laptop or start typing code.

Part 1: Plan it on paper (~20 min)

1a. Write the log posterior, $\log P(a, b \mid y)$, up to a constant. You can drop terms that don't depend on a or b.

1b. In plain English (not R code), describe step by step how a Metropolis sampler for (a, b) works. Act 1 is a good guide, but now there are two parameters. Cover each piece:

(i) Setup. What do you need to choose or compute before the first step?

(ii) Propose. How do you get a new candidate pair from the current one?

(iii) Accept or reject. What do you calculate, and what rule decides whether the chain moves?

(iv) Record and summarize. What do you store at each step? Once the loop ends, how do you get a 95% credible interval for b?

Checkpoint: finalize your plan (~5 min)

Look back over 1a and 1b and fix anything that no longer looks right before you start coding.

Part 2: Fit it (laptop, ~60 min)

Using AI: we recommend no AI for all of Part 2. Each task is short, and writing and debugging this code yourself is the point of the lab. R's help pages, web search, and classmates are fine. After each question, circle how you actually used AI.

What to turn in: your filled-in answers and your 2c figure as an image (upload in WebCampus). You don't need to submit your code.

The data are in `BdPondSurvey.csv` (columns `pond`, `water_temp_c`, `frogs_swabbed`, `bd_positive`), on the course website: https://kevintshoemaker.github.io/NRES-746/data/BdPondSurvey.csv. You can read it straight from that link:

```
data_url <- "https://kevintshoemaker.github.io/NRES-746/data/BdPondSurvey.csv"
bd <- read.csv(data_url)
```

Or paste the data into R directly:

```
water_temp_c  <- c(10.0, 12.3, 14.9, 15.8, 17.6, 17.9, 18.6, 18.8, 19.0, 20.5, 21.7, 22.1)
frogs_swabbed <- c(21, 18, 23, 19, 18, 24, 16, 17, 23, 23, 20, 18)
bd_positive   <- c(13, 12, 8, 9, 6, 8, 3, 5, 1, 6, 2, 1)
```

2a. *(Recommended: no AI.)* Write a function that returns the log posterior for `par = c(a, b)`. Check it: the log posterior at $(a, b) = (1, -0.1)$ minus the log posterior at $(3, -0.2)$ should be about -1.18. (Why does this check work no matter which constants you dropped?)

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2b. *(Recommended: no AI.)* Write your sampler from 1b. Run it for 10,000 steps, starting at $(a, b) = (0, 0)$, with jumps drawn from Normal distributions with standard deviation 0.2 for a and 0.05 for b. What fraction of proposals were accepted? Drop the first 1,000 steps and report a 95% credible interval for b (the 2.5% and 97.5% quantiles of the chain). Then run it again (no `set.seed()`, or a different seed) and report the interval from the second run.

Acceptance rate: _____

Run 1: _____ to _____ Run 2: _____ to _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2c. *(Recommended: no AI.)* Now **center** the temperature: use $T_i - 17$ in place of T_i (17 °C is close to the average). This changes what a means but not b. Run the sampler twice more with the same settings. Make one figure with the trace plot of b from 2b (top) and from the centered run (bottom), and report the new results. Why did centering help?

Acceptance rate: _____

Run 1: _____ to _____ Run 2: _____ to _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2d. *(Recommended: no AI.)* Fit the centered model by maximum likelihood with `optim(..., hessian = TRUE)` (as in Lab 6) and report the Wald interval (computing the SE from the Hessian matrix), $\hat{b} \pm 1.96\,\mathrm{SE}$. Compare it with your credible interval from 2c.

Wald interval for b: _____ to _____

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

2e. Where did you use AI, if anywhere, and what did you have to check or fix? If you didn't use it, what was hardest to get working?

Optional. *(Recommended: no AI.)* Every step of your centered chain gives a pair (a, b), and this allows you to compute a predicted prevalence at any temperature. Compute the prevalence at 12 °C for each step after burn-in, and report its posterior mean and 95% credible interval.

AI use (circle one): no AI / AI helped (e.g., debugging) / AI wrote the code

Reflection (~10 min)

In Act 1, the Wald and credible intervals for one pond's prevalence disagreed sharply. How closely did they agree for b here, and why might that be? What would you tell the managers about Bd and water temperature, in a sentence or two they could act on?

Also record any changes you made, new insights, and any "muddiest points" or topics you want to review.