

Problem Set 4

Michael Fryer

2025-11-27

Collaborators:

1 Log-odds and the Logistic Link

a) For each of the following, *first* state whether the probability is less than 0.5, equal to 0.5, or greater than 0.5, and explain your reasoning in one sentence. *Then* calculate the exact probability.

i) An event with odds 3:1 in favor

ii) An event with log-odds of -1.5

iii) An event with odds 1:1

b) In logistic regression, coefficients represent changes in log-odds. To interpret them, we need to understand how log-odds changes translate to odds changes.

i) Let ω denote the odds of an event, so $\log(\omega)$ is the log-odds. If the log-odds increase by Δ , show algebraically that the new odds are $\omega \cdot e^\Delta$.

ii) A logistic regression coefficient is $\beta = 0.8$. By what factor do the odds change for a one-unit increase in the predictor?

iii) If instead $\beta = -0.8$, express the change in odds both as “multiplied by” and as “divided by.” What is the relationship between your answers to **ii)** and **iii)**?

c) A common mistake is to interpret logistic regression coefficients as “The probability increases by β .” This part explores why that’s wrong. Consider a predictor with coefficient $\beta = 0.8$.

i) If the baseline probability is $p_0 = 0.10$, what is the probability after a one-unit increase in the predictor?

ii) If the baseline probability is $p_0 = 0.50$, what is the probability after the same one-unit increase?

iii) If the baseline probability is $p_0 = 0.90$, what is the probability after the same one-unit increase?

iv) A colleague claims: “Each additional unit of the predictor increases the probability by about 0.19.” Using your results from **i)-iii)**, explain in 2-3 sentences why this statement is misleading. Your explanation should reference the shape of the logistic function.

2 HMC, Interactions and Robust Priors

Recall the interaction model `m8.3`, which is a varying-slope regression model assessing the effect of ruggedness a country’s terrain on its GDP inside and outside of Africa.

```
m8.3 <- quap(
  alist(
    log_gdp_std ~ dnorm(mu, sigma),
    mu <- a[cid] + b[cid] * (rugged_std - 0.215),
    a[cid] ~ dnorm(1, 0.1),
    b[cid] ~ dnorm(0, 0.3),
    sigma ~ dexp(1)
  ),
  data = dd # See R code 9.11 to prepare dd
)
```

a) Now fit this same model using Hamiltonian Monte Carlo (HMC). The code to do this is in the book, beginning with R code 9.13. You should use the `ulam` convenience function provided by the `rethinking` package.

NOTE on Stan output: Stan generates verbose compilation messages that clutter your PDF and make grading difficult. Suppress Stan's compilation messages with `{r, ruggedness a, results="hide", message=FALSE}`.

Grading policy: Assignments with unsuppressed Stan output will receive a maximum of 50% credit.

b) Before generating any diagnostic plots, consider what you expect to see if the chains have converged properly.

i) For a well-behaved traceplot, describe in one sentence what the four chains should look like relative to each other.

ii) For a well-behaved trunkplot (trace rank plot), describe in one sentence what the histograms should look like.

Now generate traceplots and trunkplots for your model. Do your plots match your predictions? If there are any discrepancies, explain what they indicate about chain behavior.

Hint: Both `traceplot()` and `trunkplot()` include a `window` attribute that allows you to omit the first few samples that may otherwise obscure the behavior of your plots.

c) The original model uses an exponential prior `sigma ~ dexp(1)`. Now consider replacing this with a uniform prior `sigma ~ dunif(0,1)`.

i) Before fitting the model, predict: will the posterior distribution for `sigma` change substantially, change slightly, or remain essentially unchanged? Justify your prediction in 1-2 sentences, considering both the shape of the priors and what the data can tell us about `sigma`.

ii) Now fit the model with `sigma ~ dunif(0, 1)` and compare the posterior distributions. Was your prediction correct?

Hint: You may find `extract.samples()` useful.

iii) Use prior predictive simulation to visualize what each prior implies about the outcome variable. Simulate 1000 draws of `sigma` from each prior, then for each draw simulate a `log_gdp_std` value from `dnorm(1, sigma)`. Plot the resulting distributions. Do these visualizations support your prediction from part i)?

d) The slope parameter `b[cid]` controls how ruggedness relates to GDP, separately for African and non-African countries. The original model uses `b[cid] ~ dnorm(0, 0.3)`.

i) What constraint does a log-normal prior `dlnorm(0, 1)` place on the parameter values that a normal prior does not?

Before fitting any models, use `extract.prior()` on a model with `b[cid] ~ dlnorm(0, 1)` to draw 50 regression lines from the prior predictive distribution. Plot these lines for both African and non-African countries. What do you observe about the implied relationships?

ii) Given what you know about the ruggedness-GDP relationship from the original model (positive slope for African countries, negative slope for non-African countries), predict what will happen to the fitted slopes if you use `b[cid] ~ dlnorm(0, 1)`.

iii) Fit the model and check your prediction. Which prior is more appropriate for this scientific question, and why?

3 Binomial Regression

We started the course sampling marbles from a bucket to estimate its contents, followed by tossing a globe to estimate the proportion of its surface covered in water. Each made use of the binomial distribution, which was ideal to introduce the fundamentals of Bayesian inference. Nevertheless, Binomial regression – which is any type of GLM using a binomial mean-variance relationship – introduces complications that we needed to postpone until now.

Return to the prosocial chimpanzee experiment in section §11.1 of the textbook. In this experiment, chimpanzees chose between two levers: one that delivered food to both themselves and a partner (prosocial option), and one that delivered food only to themselves. The outcome variable `pulled_left` records whether the actor pulled the left lever (`=1`) or not (`=0`). The `treatment` variable encodes the experimental condition (which combines whether the prosocial option was on the left and whether a partner was present).

Consider the HMC model with individual actor intercepts and treatment effects:

```
m11.4 <- ulam(  
  alist(  
    pulled_left ~ dbinom(1, p),  
    logit(p) <- a[actor] + b[treatment] ,  
    a[actor] ~ dnorm(0, 1.5),  
    b[treatment] ~ dnorm(0, 0.5)  
  ),  
  data = dat_list, chains=4, log_lik=TRUE # See sec 11.1 to prepare dat_list  
)
```

NOTE on Stan output: Stan generates verbose compilation messages that clutter your PDF and make grading difficult. Suppress Stan’s compilation messages with `{r, ruggedness a, results="hide", message=FALSE}`.

Grading policy: Assignments with unsuppressed Stan output will receive a maximum of 50% credit.

a) Before fitting any models, consider the prior `a[actor] ~ dnorm(0, 1.5)`.

i) What is the implied prior probability that a chimpanzee pulls the left lever when all treatment effects are zero? (That is, what does `a[actor] = 0` imply about left-lever pulling probability?)

ii) The prior has standard deviation 1.5 on the log-odds scale. Compute the probability of pulling left when `a[actor]` is one standard deviation above and below the mean (i.e., at `+1.5` and `-1.5`). What range of baseline probabilities does this prior consider plausible?

iii) Is this prior regularizing, weakly informative, or strongly informative? Justify in one sentence.

iv) Use `extract.prior()` to draw 1000 samples from `a[actor] ~ dnorm(0, 1.5)`, transform to the probability scale with `inv_logit()`, and plot the distribution. Does this match your calculations from i)-ii)?

b) Now fit both `ulam()` (HMC) and `quap()` (Laplace approximation) versions of this model. Compare the posterior distributions for the actor coefficients. What differences do you observe, and for which actor(s) are the differences most pronounced?

c) Actor 2 pulled the left lever on every trial.

i) Explain why this creates an asymmetric posterior distribution for `a[2]` that `quap` cannot capture.

- ii) What would happen to the posterior for `a[2]` as the number of trials increases, assuming Actor 2 continues to always pull left?
- d) Now change the prior on the actor intercepts to `dnorm(0, 10)` and fit both models again.
- i) Before fitting, predict: on the probability scale, what kind of chimpanzee behavior does `dnorm(0, 10)` encode? (Hint: compute `logistic(10)` and `logistic(-10)`.)
- ii) How will this prior change the posterior for Actor 2 under HMC versus quap?
- iii) Fit the models and check your predictions.

4 AI Declaration

5 Appendix