Problem Set 3

Michael Fryer

2025-11-24

Collaborators: *...*

1 Entropy & Divergence

```
# Calculate Information Entropy
calc_entropy <- function(l) return(-sum(l*log(l)))
```

Suppose biased coin lands heads 60% of the time.

a) What is the entropy of this coin?

```
# 60% heads, 40% tails
p <- c(0.6, 0.4)
entropy.p <- calc_entropy(p)
```

```
## Entropy:  0.6730117
```

b) Suppose another biased coin lands tails 30% of the time. Which coin has higher entropy? Justify and explain your answer.

```
# 70% heads, 30% tails
q <- c(0.7, 0.3)
entropy.q <- calc_entropy(q)
```

```
## Entropy:  0.6108643
```

*The coin in part **a)** has a higher entropy than the coin in **b)**. This is because the variance in the probabilities in **b)** is larger than in **a)** i.e. the more unique events there are with similar probabilities of occurring, the larger the entropy will be.*

c) Suppose the first coin is the true distribution and the second coin's distribution is your approximation of the first coin. What is the KL divergence between these two distributions? Explain and justify your answer.

KL divergence is defined as:

$$D_{KL}(p, q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right)$$

```
kl <- sum(p*log(p/q))
```

```
## KL divergence:  0.02258242
```

We observe a KL divergence of 0.022. We expect this value to be small as p and q are relatively "close" together and the KL divergence is the average difference in log probabilities between the two models. This value would be much larger if instead $q = \{0.9, 0.1\}$.

2 Collider Bias and Information Criteria

Return to the textbook example in §6.3.1, which explores the relationship between age, marriage and happiness. This example makes use of the simulation model `sim_happiness()`, which is part of the rethinking package:

```r
d <- sim_happiness(seed=1515 , N_years=1000)
d.adults <- d[d$age > 17,] # only adults
d.adults$A <- (d.adults$age - 18) / (65 - 18)
d.adults$mid <- d.adults$married + 1
```

Compare the two models, `m6.9` and `m6.10`, using both PSIS and WAIC scores using `compare()`.

Fit `m6.9` *and* `m6.10`.

```r
# Model m6.9
m6.9 <- quap(
  alist(
    happiness ~ dnorm(mu, sigma),
    mu <- a[mid] + bA*A,
    a[mid] ~ dnorm(0, 1),
    bA ~ dnorm(0, 2),
    sigma ~ dexp(1)
  ),
  data=d.adults
)
# Model m6.10
m6.10 <- quap(
  alist(
    happiness ~ dnorm(mu, sigma),
    mu <- a + bA*A,
    a ~ dnorm(0, 1),
    bA ~ dnorm(0, 2),
    sigma ~ dexp(1)
  ),
  data=d.adults
)
```

Score the models.

```r
psis.compare <- compare(m6.9, m6.10, func=PSIS)
```

```r
waic.compare <- compare(m6.9, m6.10, func=WAIC)
```

Chart the models.

```
##              PSIS       SE    dPSIS      dSE    pPSIS       weight
## m6.9   2771.139 37.10321   0.0000       NA 3.823583 1.000000e+00
## m6.10 3102.217 27.77867 331.0777 34.17721 2.512896 1.280557e-72

##              WAIC       SE    dWAIC      dSE    pWAIC       weight
## m6.9   2771.107 37.08184   0.0000       NA 3.807557 1.000000e+00
## m6.10 3102.197 27.76361 331.0901 34.17543 2.503099 1.272606e-72
```

PSIS

m6.9

m6.10

2800 3000

deviance

WAIC

m6.9

m6.10

2800 3000

deviance
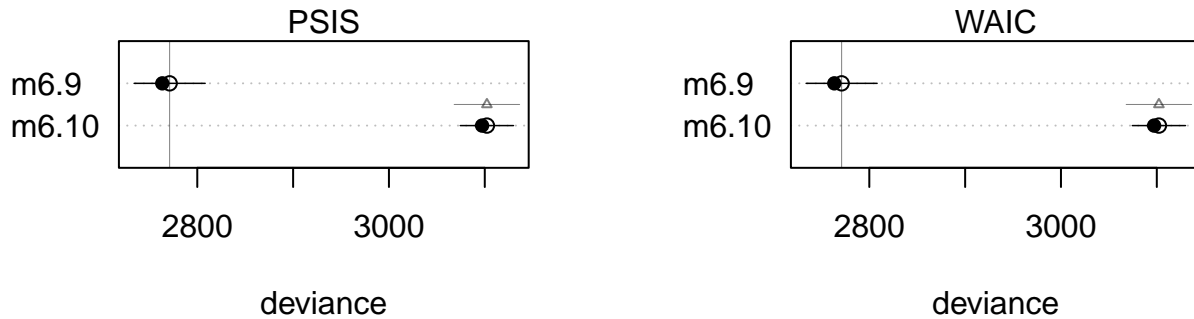
a) Which model is expected to make better predictions according to these two information criteria? Explain your answer.

In both cross-validation approximators, a lower score is better. With that metric in mind, both WAIC and PSIS expect `m6.9` *to be a better prediction model. However, it should be noted that both the standard error (**SE**) and prediction penalty (**pPSIS** and **pWAIC**) are higher in* `m6.9`. *This means that while* `m6.9` *is a better predictor according to WAIC and PSIS, it has more uncertainty,* **SE**.

b) On the basis of the causal model, how should you interpret the parameter estimates from the model preferred by PSIS and WAIC?

*As WAIC and PSIS are cross-validation approximators, they estimate of how well a model does at **predictions**, one cannot, and should not, use them to inform any decisions about **causations** That is, WAIC and PSIS scores have no bearing on if the model is the "correct" causal model. In fact,* `m6.9` *is **not** the correct causal model,* `m6.10` *is.*

3   Tulips and Interactions

Tulips need both water and light to bloom. But do water and light work independently, or do they interact? The `tulips` dataset in the rethinking package contains experimental data on tulip blooms under different conditions.

```
data(tulips)
d <- tulips
str(d)
```

```
## 'data.frame':    27 obs. of  4 variables:
##  $ bed   : Factor w/ 3 levels "a","b","c": 1 1 1 1 1 1 1 1 1 2 ...
##  $ water : int  1 1 1 2 2 2 3 3 3 1 ...
##  $ shade : int  1 2 3 1 2 3 1 2 3 1 ...
##  $ blooms: num  0 0 111 183.5 59.2 ...
```

The dataset contains:

- `bed`: Bed identifier (grouping variable)
- `water`: Amount of water (1 = low, 2 = medium, 3 = high)
- `shade`: Amount of shade (1 = low, 2 = medium, 3 = high)
- `blooms`: Number of blooms produced

Before fitting any models, standardize the predictor variables and the outcome:

Below, I do not standardize the predictor variables and the outcome. Instead, I scale the outcome to $[0, 1]$ *and center the predictor variables such that the values are* $-1, 0, 1$. *This accomplishes a similar goal as standardizing but makes it easier to reason about.*

```
# Scale blooms to [0, 1]
d$B <- (d$blooms - min(d$blooms))
d$B <- d$B / max(d$B)
```

```
# Center water and shade st. the values are {-1,0,1}
d$W <- d$water - mean(d$water)
d$S <- d$shade - mean(d$shade)
```

a) Fit a model (`m3.1`) that predicts blooms using water and shade as independent predictors (no interaction). That is, assume the effect of water on blooms is the same regardless of shade level, and vice versa.

Use prior predictive simulation to justify your priors. Write 2-3 sentences explaining your choice of priors.

Hint: With standardized variables, you should use relatively weak priors. A slope of 2.0 would be implausibly large for most ecological relationships.

Since we are scaling and centering our data instead of standardizing it, our priors will not all be mean 0. We will predict our outcome prior will have $\mu = 0.5$ as we scaled our blooms to $[0, 1]$ and the predictor prior will have $\mu = 0$ as we centered that data. For our standard deviations, we will use $\sigma = 0.25$ for our outcome prior. This is because $\pm 2\sigma$ would put blooms at impossible values for our small world i.e. > 0 or < 1. For the same reasons, we will use $\sigma = 0.25$ for our predictor variables.

```
N <- 50
a <- rnorm(N, 0.5, 0.25)
bS <- rnorm(N, 0, 0.25)
bW <- rnorm(N, 0, 0.25)
```
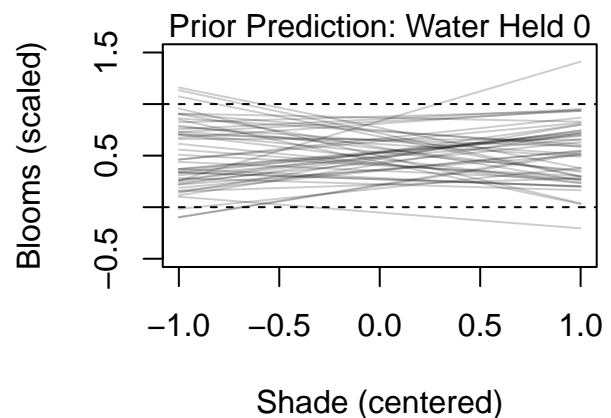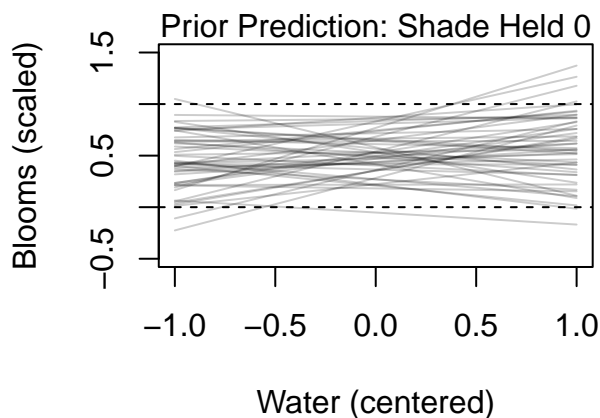


```
m3.1 <- quap(
  alist(
    B ~ dnorm(mu, sigma),
    mu <- a + bS*S + bW*W,
    a ~ dnorm(0.5, 0.25),
    bS ~ dnorm(0, 0.25),
    bW ~ dnorm(0, 0.25),
    sigma ~ exp(1)
  ),
  data=d
)
```

b) Now fit a model (`m3.2`) that includes an interaction between water and shade. This allows the effect of water on blooms to depend on the level of shade (and vice versa).

Hint: In the model formula, you can specify the interaction as $W * S$, which is shorthand for $W + S + W * S$. Or you can write out all three terms explicitly.
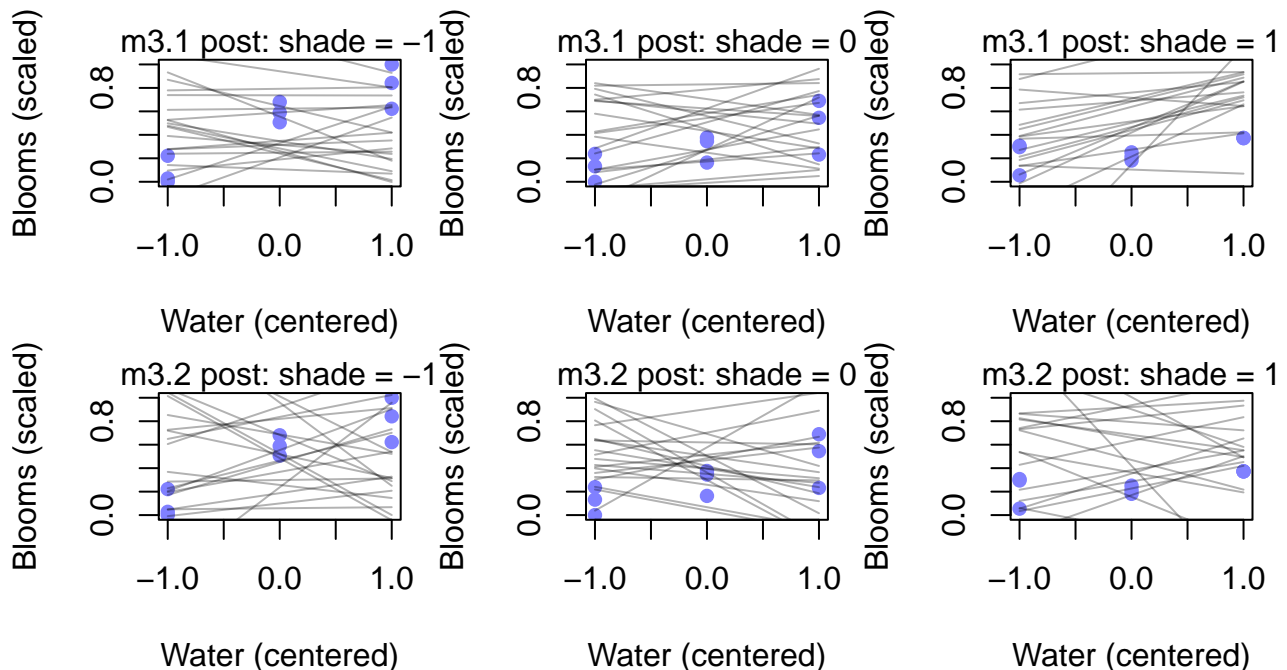
```
m3.2 <- quap(
  alist(
```

```
    B ~ dnorm(mu, sigma),
    mu <- a + bS*S + bW*W + bSW*S*W,
    a ~ dnorm(0.5, 0.25),
    bS ~ dnorm(0, 0.25),
    bW ~ dnorm(0, 0.25),
    bSW ~ dnorm(0, 0.25),
    sigma ~ exp(1)
  ),
  data=d
)
```

Plot the posterior predictions for both models (`m3.1` and `m3.2`) side by side. For each model, create a plot showing how predicted blooms change with water at different levels of shade (e.g., plot three lines for shade = -1, 0, 1).

c) Use both PSIS and WAIC to compare your two models. Which model is expected to make better predictions? Is the difference substantial?

Create the comparison plot using `plot(compare(m3.1, m3.2))` and interpret what you see. Write 3-4 sentences explaining:

- Which model performs better
- How large the difference is relative to the standard error
- What this tells you about whether water and shade interact in affecting tulip blooms

Score the models.

```
psis.compare <- compare(m3.1, m3.2, func=PSIS)
```

```
waic.compare <- compare(m3.1, m3.2, func=WAIC)
```

Chart the models.

```
##            PSIS         SE       dPSIS          dSE       pPSIS   weight
## m3.1 104.3408 0.06198899 0.00000000           NA 0.008895423 0.509854
## m3.2 104.4196 0.07427911 0.07884246 0.02103329 0.011417252 0.490146
```

```
##            WAIC          SE        dWAIC           dSE         pWAIC      weight
## m3.1 104.3406 0.06080206 0.00000000           NA 0.008799849 0.5098477
## m3.2 104.4194 0.07286196 0.07879181 0.02103741 0.011296355 0.4901523
```
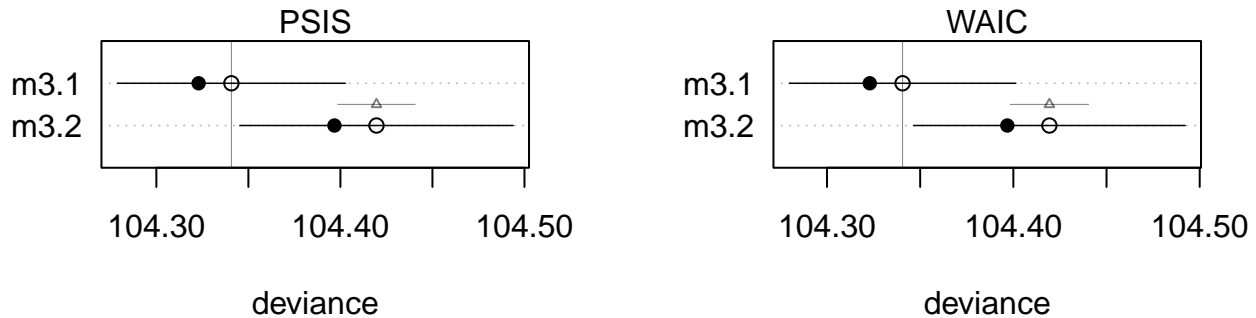

When comparing our two models using PSIS and WAIC we see similar results. With each method, `m3.1`
is expected to perform better. When comparing the difference, **dPSIS** *and* **dWAIC***, we observe that it is*
about half of the standard error. Additionally, the **dSE** *for* $m3.2$ *is 0.02. This implies that there is very*
little difference between these two models from a predictions perspective. According to these comparisons
alone, there isn't strong support for including the interaction effect. If there was, we would expect to see $m3.2$
outperforming $m3.1$*.*

d) Examine the pointwise PSIS diagnostics for your preferred model. Are there any problematic observations
(high Pareto k values)? If so, which observations are influential and why might they be outliers?

Hint: Use `PSIS(model, pointwise=TRUE)` and examine the k values.

```
psis <- PSIS(m3.1, pointwise=TRUE)
plot(psis$k, psis$penalty,
     xlab="PSIS Pareto k", ylab="PSIS Penalty",
     col=rangi2, lwd=2
)
```
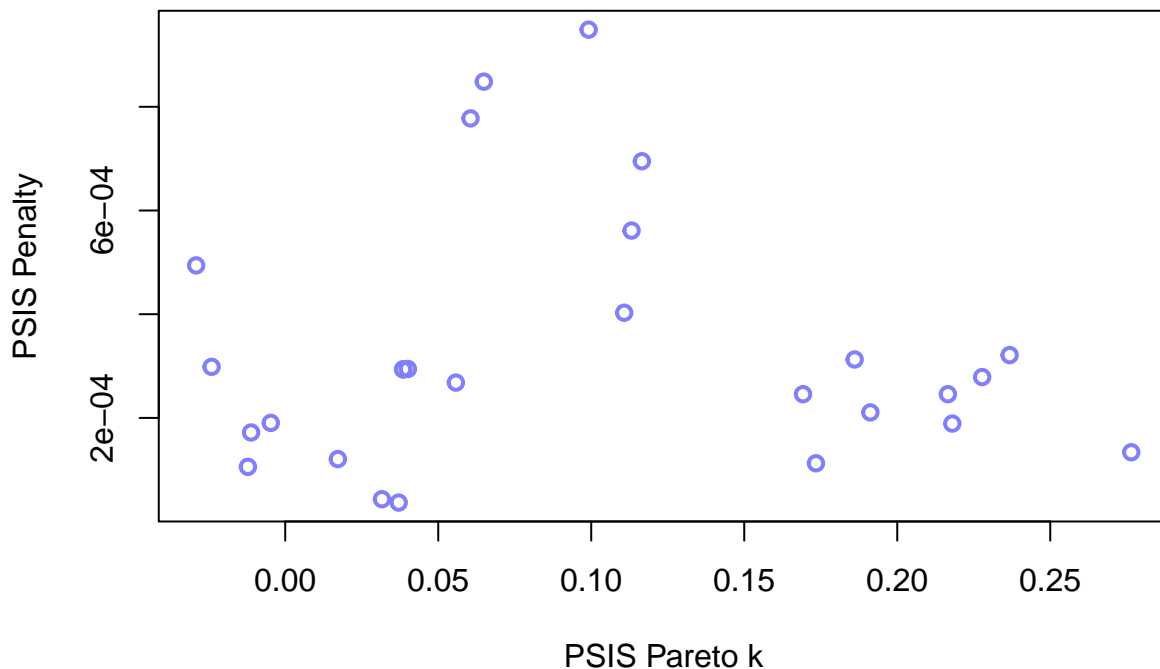


```
## Max Pareto k-value: 0.276
```

When there are Pareto k values above 0.5, their importance weight can be unreliable. This is an indicator

of a potential outlier. The highest Pareto k value we observe is 0.294 which is well below the 0.5 warning threshold.

e) Based on your analysis, write 2-3 sentences answering: Do water and light work independently to affect tulip blooms, or do they interact? What does this mean biologically?

*It is very likely that water and light interact to affect tulip bloom. This is evidenced by the shifing slopes in **b)** for $m3.2$ when holding shade. This would not be able to happen if there was not interaction. A possible biological explaination for such an interaction is that tulips need both light and water to produce blooms. This is, of course, obvious. Plants need light and nutrients to grow.*

4 Reflection

Look back at Problems 2 and 3. Both involve interactions, but of different kinds:

- In Problem 2, you examined collider bias, where conditioning on a variable (marriage status) creates a spurious association between age and happiness
- In Problem 3, you examined multiplicative interactions, where the effect of one variable (water) depends on the level of another (shade)

Write 3-4 sentences explaining: What is the difference between these two types of "interactions"? How are they similar and how are they different from a causal modeling perspective?

5 AI Declaration

Please declare your collaborators in the class and how you used AI (if at all) to complete this assignment. If you used AI, include the prompts you used and explain what you learned from its responses that you didn't understand initially.

6 Appendix

```r
# Plots a standardized predictor x
plot_prior_prediction <- function(x, ylim, xlab, ylab, title="Prior Prediction") {
  plot(NULL, xlim=range(x), ylim=ylim, xlab=xlab, ylab=ylab)
  mtext(title)
  # Graph all the lines
  for (i in 1:N) {
    curve(a[i] + b[i]*x,
      from=min(x), to=max(x),
      col=col.alpha("black", 0.2), add=TRUE
    )
  }
  # Bounds of impossibilities
  abline(h=0, lty=2)
  abline(h=1, lty=2)
}


# Plot the posterior predictions for blooms
plot_post_prediction_blooms <- function(m, m.name) {
  for (s in -1:1) {
    i <- which(d$S==s)
    plot(d$W[i], d$B[i],
      xlim=c(-1, 1), ylim=c(0, 1),
      xlab="Water (centered)", ylab="Blooms (scaled)",
      pch=16, col=rangi2
    )
    mtext(sprintf("%s post: shade = %d", m.name, s))
    post <- link(m, data=data.frame(S=s, W=-1:1))
    mu <- post$mu
    for (j in 1:20) lines(-1:1, mu[j,], col=col.alpha("black", 0.3))
  }
}
```