Statistical Rethinking Winter 2020 – Homework Week 4

Alessandro Gentilini `.@gmail` (just auditing)

January 9, 2021

1. Consider three fictional Polynesian islands. On each there is a Royal Ornithologist charged by the king with surveying the birb population. They have each found the following proportions of 5 important birb species:

	Birb A	Birb B	Birb C	Birb D	Birb E
Island 1	0.2	0.2	0.2	0.2	0.2
Island 2	0.8	0.1	0.05	0.025	0.025
Island 3	0.05	0.15	0.7	0.05	0.05

Notice that each row sums to 1, all the birbs. This problem has two parts. It is not computationally complicated. But it is conceptually tricky. First, compute the entropy of each island's birb distribution. Interpret these entropy values. Second, use each island's birb distribution to predict the other two. This means to compute the K-L Divergence of each island from the others, treating each island as if it were a statistical model of the other islands. You should end up with 6 different K-L Divergence values. Which island predicts the others best? Why?

```r
d <- matrix(c(.2,.2,.2,.2,.2,
.8,.1,.05,.025,.025,
.05,.15,.7,.05,.05)
,nrow=3,ncol=5,byrow=TRUE)
d
```

```
##      [,1] [,2] [,3]  [,4]  [,5]
## [1,] 0.20 0.20 0.20 0.200 0.200
## [2,] 0.80 0.10 0.05 0.025 0.025
## [3,] 0.05 0.15 0.70 0.050 0.050
```

```r
H <- function (p)
{
    return(-sum(p*log(p)))
}
```

Each rows has the proportions and that's mean that they can be thought as frequencies and so probabilities.

```r
H(d[1,])
```

```
## [1] 1.609438
```

```r
H(d[2,])
```

```
## [1] 0.7430039
```

```r
H(d[3,])
```

```
## [1] 0.9836003
```

Since in Island 1 the five birds are equally distributed we have that the entropy for the first island is the highest because we are quite uncertain about what bird we will see when we arrive for the first time on the Island 1. Island 2 has the lowest entropy because I would bet the first bird I will see is a Birb A since its proportion is quite high, namely 80%.

```
D_KL <- function(p,q)
{
    return(sum(p*(log(p)-log(q))))
}
```

Prior to compute the divergences I would say that Island 1 will be the best predictor because it has the five species equally distributed and so no surprise will appears when we will see a lot of Birb A on Island 2 or a lot of Birb C on island 3. On the other hand Island 2 will perform bad with respect to Island 3 because Birb A are rare on Island 3 and very common on Island 2. Island 3 will perform bad with respect to Island 2 because Birb C are very common on Island 3 and rare on Island 2.

Use Island 1 to predict 2 and 3:

```
D_KL(d[2,],d[1,])
```

```
## [1] 0.866434
```

```
D_KL(d[3,],d[1,])
```

```
## [1] 0.6258376
```

Use Island 2 to predict 1 and 3:

```
D_KL(d[1,],d[2,])
```

```
## [1] 0.9704061
```

```
D_KL(d[3,],d[2,])
```

```
## [1] 1.838845
```

Use Island 3 to predict 1 and 2:

```
D_KL(d[1,],d[3,])
```

```
## [1] 0.6387604
```

```
D_KL(d[2,],d[3,])
```

```
## [1] 2.010914
```

Actually both the divergences associated with Island 1 are the lowest ones.

---

2. Recall the marriage, age, and happiness collider bias example from Chapter 6. Run models m6.9 and m6.10 again. Compare these two models using WAIC (or LOO, they will produce identical results). Which model is expected to make better predictions? Which model provides the correct causal inference about the influence of age on happiness? Can you explain why the answers to these two questions disagree?

```
## R code 6.21
library(rethinking)
d <- sim_happiness( seed=1977 , N_years=1000 )



## R code 6.22
```

```
d2 <- d[ d$age>17 , ] # only adults
d2$A <- ( d2$age - 18 ) / ( 65 - 18 )

## R code 6.23
d2$mid <- d2$married + 1
m6.9_M <- quap(
    alist(
        happiness ~ dnorm( mu , sigma ),
        mu <- a[mid] + bA*A,
        a[mid] ~ dnorm( 0 , 1 ),
        bA ~ dnorm( 0 , 2 ),
        sigma ~ dexp(1)
    ) , data=d2 )
m6.9_M
```

```
##
## Quadratic approximate posterior distribution
##
## Formula:
## happiness ~ dnorm(mu, sigma)
## mu <- a[mid] + bA * A
## a[mid] ~ dnorm(0, 1)
## bA ~ dnorm(0, 2)
## sigma ~ dexp(1)
##
## Posterior means:
##       a[1]       a[2]          bA      sigma
## -0.2350877  1.2585517 -0.7490274  0.9897080
##
## Log-likelihood: -1353.08
```

```
## R code 6.24
m6.10_no_M <- quap(
    alist(
        happiness ~ dnorm( mu , sigma ),
        mu <- a + bA*A,
        a ~ dnorm( 0 , 1 ),
        bA ~ dnorm( 0 , 2 ),
        sigma ~ dexp(1)
    ) , data=d2 )
m6.10_no_M
```

```
##
## Quadratic approximate posterior distribution
##
## Formula:
## happiness ~ dnorm(mu, sigma)
## mu <- a + bA * A
## a ~ dnorm(0, 1)
## bA ~ dnorm(0, 2)
## sigma ~ dexp(1)
##
## Posterior means:
##             a            bA         sigma
##  1.649248e-07 -2.728620e-07  1.213188e+00
```

```
##
## Log-likelihood: -1548.31
```

```r
WAIC(m6.9_M)
```

```
##       WAIC      lppd  penalty  std_err
## 1 2713.971 -1353.247 3.738532 37.54465
```

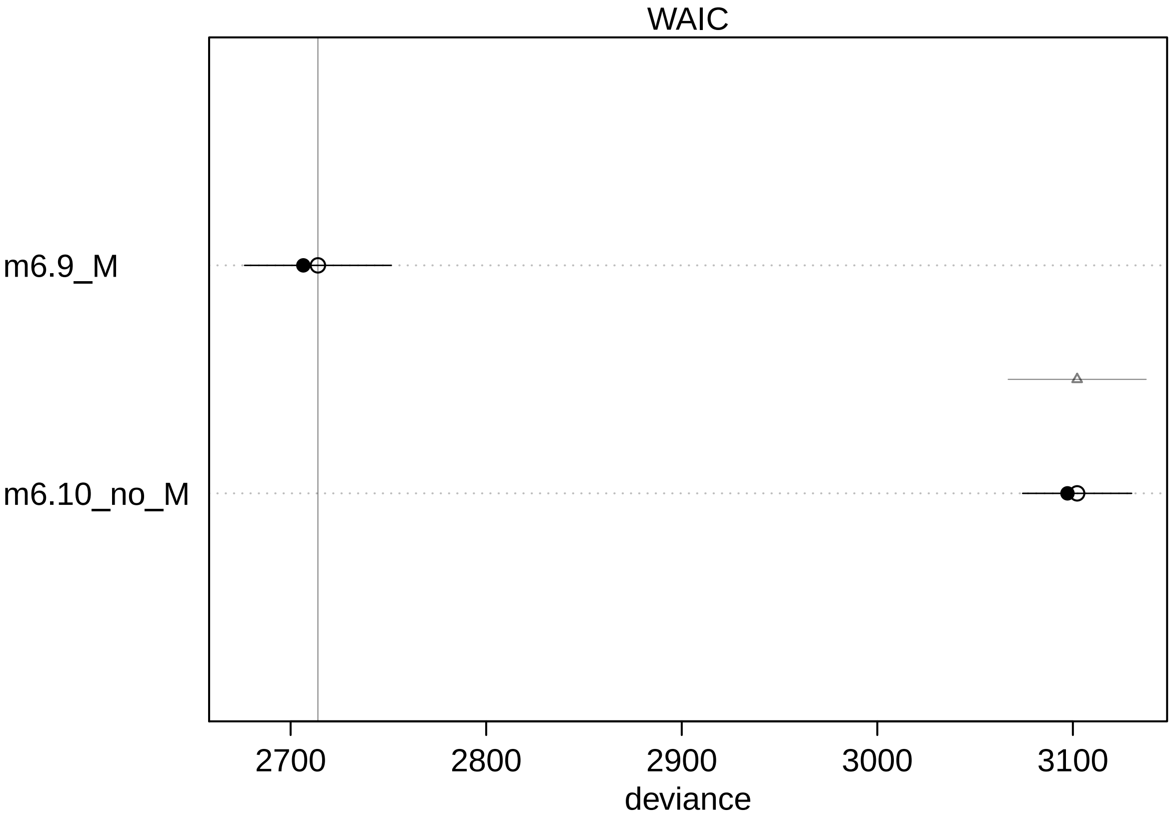
```r
WAIC(m6.10_no_M)
```

```
##       WAIC      lppd  penalty  std_err
## 1 3101.906 -1548.612 2.340445 27.74379
```

```r
compare(m6.9_M,m6.10_no_M,func=WAIC)
```

```
##                WAIC       SE    dWAIC      dSE    pWAIC       weight
## m6.9_M     2714.256 37.51051   0.0000       NA 3.861508 1.000000e+00
## m6.10_no_M 3101.883 27.68399 387.6271 35.34372 2.328445 6.727359e-85
```

```r
plot(compare(m6.9_M,m6.10_no_M,func=WAIC))
```



`m6.9_M` is the model that make better predictions.

`m6.10_no_M` is the model that provides the correct causal inference about the influence of age on happiness.

The answers to these two questions disagree because the statistical association between age and happiness in married couples (the older the sadder) it is not a causal association so it gives good prediction (and so `m6.9_M` has lower WAIC) but it does not give hints about the causal relationship between age and happiness.

---

3. Reconsider the urban fox analysis from last week's homework. Use WAIC or LOO based model comparison on five different models, each using weight as the outcome, and containing these sets of predictor variables:

```
(1) avgfood + groupsize + area
(2) avgfood + groupsize
(3) groupsize + area
(4) avgfood
(5) area
```

Can you explain the relative differences in WAIC scores, using the fox DAG from last week's homework? Be sure to pay attention to the standard error of the score differences (dSE).

```r
data(foxes)
foxes$A <- standardize(foxes$area)
foxes$W <- standardize(foxes$weight)
foxes$F <- standardize(foxes$avgfood)
foxes$G <- standardize(foxes$groupsize)
```

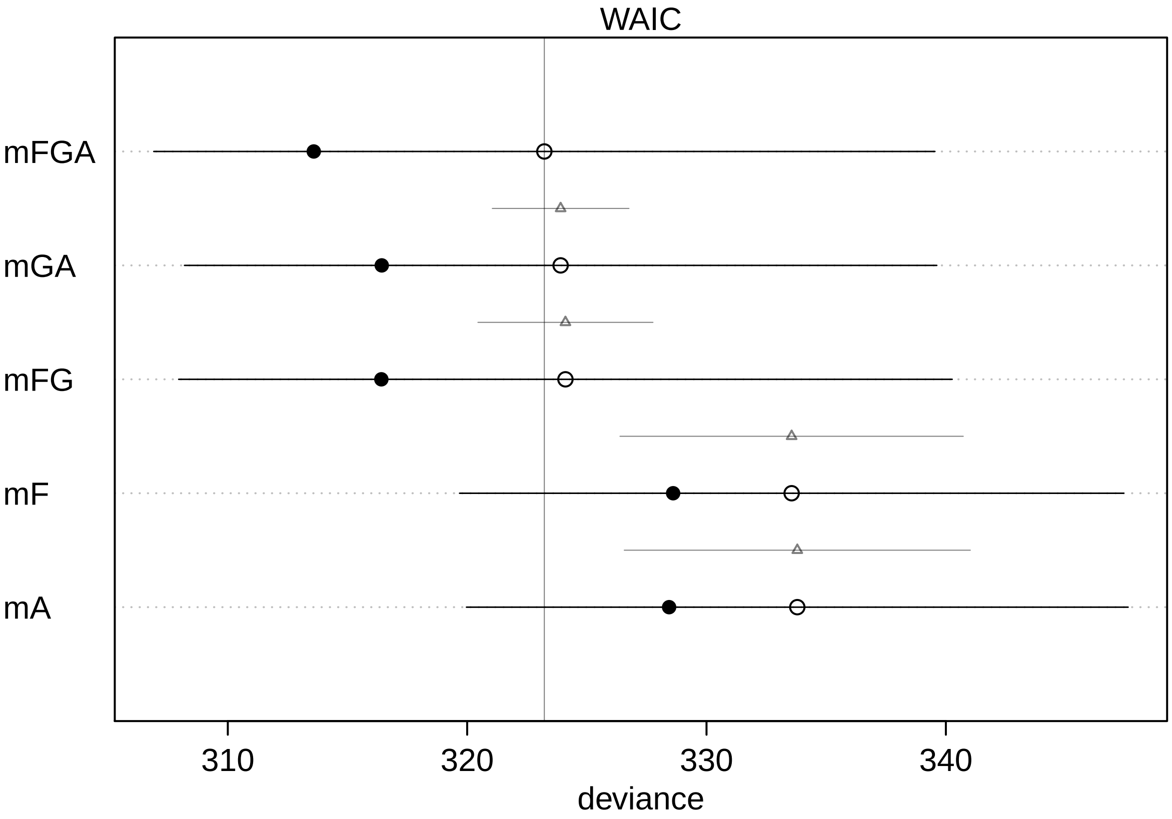
```r
mFGA <- quap(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- a + b_A*A+b_F*F+b_G*G,
        a ~ dnorm( 0 , .2 ),
        b_A ~ dnorm( 0 , .5 ),
        b_F ~ dnorm( 0 , .5 ),
        b_G ~ dnorm( 0 , .5 ),
        sigma ~ dexp( 1 )
    ), data=foxes )

mFG <- quap(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- a + b_F*F+b_G*G,
        a ~ dnorm( 0 , .2 ),
        b_F ~ dnorm( 0 , .5 ),
        b_G ~ dnorm( 0 , .5 ),
        sigma ~ dexp( 1 )
    ), data=foxes )

mGA <- quap(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- a + b_A*A+b_G*G,
        a ~ dnorm( 0 , .2 ),
        b_A ~ dnorm( 0 , .5 ),
        b_G ~ dnorm( 0 , .5 ),
        sigma ~ dexp( 1 )
    ), data=foxes )

mF <- quap(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- a + b_F*F,
        a ~ dnorm( 0 , .2 ),
        b_F ~ dnorm( 0 , .5 ),
        sigma ~ dexp( 1 )
    ), data=foxes )
```

```
mA <- quap(
    alist(
        W ~ dnorm( mu , sigma ),
        mu <- a + b_A*A,
        a ~ dnorm( 0 , .2 ),
        b_A ~ dnorm( 0 , .5 ),
        sigma ~ dexp( 1 )
    ), data=foxes )
```

```
result = compare(mFGA,mFG,mGA,mF,mA,func=WAIC)
result
```

```
##          WAIC       SE      dWAIC      dSE    pWAIC      weight
## mFGA 323.2240 16.31874  0.0000000       NA 4.816574 0.422742735
## mGA  323.9054 15.71619  0.6814363 2.856346 3.736868 0.300679727
## mFG  324.1058 16.15889  0.8817601 3.657606 3.845488 0.272022213
## mF   333.5564 13.87511 10.3324365 7.172261 2.475594 0.002412215
## mA   333.7930 13.82256 10.5690125 7.230663 2.677808 0.002143109
```

```
plot(result)
```



```
result@dSE
```

```
##          mFGA      mFG      mGA        mF        mA
## mFGA       NA 3.657606 2.856346 7.1722613 7.2306625
## mFG  3.657606       NA 5.796206 6.7168697 6.9529765
## mGA  2.856346 5.796206       NA 6.4245077 6.4747077
## mF   7.172261 6.716870 6.424508        NA 0.8402555
## mA   7.230663 6.952976 6.474708 0.8402555        NA
```

mGA vs mFGA, applying formula “R code 7.29” (at page 228):

```
zscore<-2.6
0.6814363 +c(-1,1)*2.856346*zscore
```

```
## [1] -6.745063  8.107936
```

So, no, to me these models are not very easy to distinguish by expected out-of-sample accuracy because the above interval of the difference include zero.

mFG vs mFGA, applying formula “R code 7.29” (at page 228):

```
0.8817601 +c(-1,1)*3.657606*zscore
```

```
## [1] -8.628016 10.391536
```

So, no, to me these models are not very easy to distinguish by expected out-of-sample accuracy because the above interval of the difference include zero.

mA vs mFGA, applying formula “R code 7.29” (at page 228):

```
10.5690125 +c(-1,1)*7.230663*zscore
```

```
## [1] -8.230711 29.368736
```