Solution to Exercise 4

Department of Methodology and Statistics
Utrecht University

Bayesian Statistics, Week X

A: Specifying the logistic regression model.

```
# Bayesian Statistics - Exercise 4 Model file
# fitting the model to the raw data, and conducting a posterior predictive check

model{

 # likelihood of the data
for (i in 1:N) {
depr[i] ~ dbern(prob[i])  # equivalent: depr[i] ~ dbin(prob[i], 1)
prob[i] <- exp(a + b1 * gender[i]) / (1 + exp(a + b1 * gender[i]))
# logit(prob[i]) <- (a + b1 * gender[i]) # equivalent formulation, gives equal results.
# prob[i] <- ilogit((a + b1* gender[i])) # should be equivalent, but wrong results!
}

 # prior distributions for the intercept and regression coefficient(s)
a ~ dnorm(0, 0.01)
b1 ~ dnorm(0, 0.01)

}
```

B: Specifying the odds ratio.

```
# Bayesian Statistics - Exercise 4 Model file
# fitting the model to the raw data, and conducting a posterior predictive check

# Bayesian Statistics - Exercise 4 Model file
# logistic regression analysis, and predicting the outcome of two new patients

model{

 # likelihood of the data
for (i in 1:N) {
depr[i] ~ dbern(prob[i])  # equivalent: depr[i] ~ dbin(prob[i], 1)
prob[i] <- exp(a + b1 * gender[i]) / (1 + exp(a + b1 * gender[i]))
# logit(prob[i]) <- (a + b1 * gender[i]) # equivalent formulation, gives equal results.
# prob[i] <- ilogit((a + b1* gender[i])) # should be equivalent, but wrong results!
}

 # prior distributions for the intercept and regression coefficient(s)
a ~ dnorm(0, 0.01)
b1 ~ dnorm(0, 0.01)

 # calculate other quantities of interest
odds.ratio <- exp(b1)

}
```

B: Specifying the odds ratio.

Advantage of specifying the odds ratio within the model:

if you calculate the odds ratio based on the posterior parameter mean of b_1, it is only a point estimate. By doing it within the model, the uncertainty about b_1 is taken into account and you get a distribution, including the 95% CCIs, for the odds ratio.

In this case, you could also have obtained the 95% CCI of the odds ratio by using the 95% CCI of b_1, but that logic does *not* apply for other quantities of interest that are functions of more than one parameter (such as the probabilities, depending on both a and b_1)! Why is that?
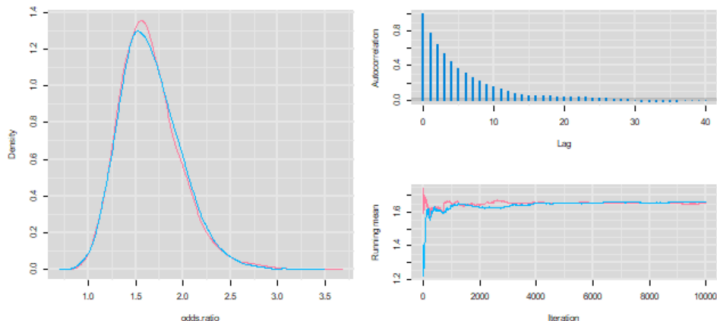
C: Assessing convergence.

Without centering (see Figure), the convergence is not good.
Center your variables before continuing (see solutions
exercise 3 for an example of centering in the model file).
Plots for odds.ratio

Diagnostics for odds.ratio

D: Interpreting results.

Women more likely to suffer from major depressive disorder
than men: their *odds* are 1.67 times as high. The 95% CCI
of the OR does not contain 1 so we are quite certain that
women have a higher risk of depression.

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 2
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

               Mean      SD  Naive SE  Time-series SE
a           -1.7592  0.08849  0.0006257       0.0008094
b1           0.4932  0.19559  0.0013830       0.0018013
odds.ratio   1.6694  0.33171  0.0023455       0.0030855

2. Quantiles for each variable:

              2.5%     25%     50%     75%   97.5%
a           -1.9355 -1.8186 -1.7585 -1.6988  -1.588
b1           0.1122  0.3611  0.4909  0.6227   0.885
odds.ratio   1.1188  1.4349  1.6339  1.8639   2.423
```

E: Model comparison using the DIC.

Advantage: the DIC can also be used to compare non-nested, qualitatively different models (on same data).

The DIC often differs only a little between models with and without an extra predictor, especially for barely-useful predictors. In those cases, the DIC often points to a different conclusion than the 95% CCIs, if you interpret it strictly as 'smaller DIC = better model'. And sometimes the DICs are (almost) equal: so do you keep the predictor because it does no harm, or lose it because it offers no improvement?

"The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed."