Week 3 Problems

1. (Adapted from Jackman 2.3) Three polls are fielded on the weekend prior to an election, each using simple random sampling, producing the following estimates of vote intention for the incumbent candidate.

Poll	Support (%)	Sample size
A	52	350
B	49	500
C	55	600

a) Let the proportion of the electorate intending to vote for the incumbent be θ. Write out a plausible statistical model for each poll's results in which θ appears as a parameter. Specifically, specify a likelihood and prior for your model.

Let Y denote the number of supporters for the incumbent in a poll. Then we have the following likelihood and prior.

$$Y \sim \text{Binomial}(n, \theta)$$
$$\theta \sim \text{Beta}(a, b)$$

where n is the sample size for each poll.

b) Suppose your have no prior information over θ. Taking each poll separately, what is the posterior density for θ given the information in each poll. Find posterior means, standard deviations, and credible intervals using R.

With a binomial likelihood and beta prior, the posterior is a conjugate $\text{Beta}(y + a, n - y + b)$ posterior. For an uninformative prior, $a = 1$ and $b = 1$.

```
> Y <- c(350 * 0.52, 500 * 0.49, 600 * 0.55)
> n <- c(350, 500, 600)
> a <- b <- 1
> n.sims <- 10000
> draws <- matrix(NA, nrow = n.sims, ncol = 3)
> colnames(draws) <- c("A", "B", "C")
> for (i in 1:3) {
+     draws[, i] <- rbeta(n.sims, Y[i] + a, n[i] - Y[i] +
+         b)
+ }
> colMeans(draws)
        A         B         C
0.5202426 0.4899081 0.5503152

> apply(draws, MARGIN = 2, FUN = sd)
         A          B          C
0.02646951 0.02240686 0.02028511

> apply(draws, MARGIN = 2, FUN = quantile, probs = c(0.025,
+     0.975))
             A         B         C
2.5%   0.4681027 0.4463045 0.5102297
97.5%  0.5718493 0.5337669 0.5898281
```

c) Now suppose you observe each poll in order sequentially starting with A. Update your posterior with each new poll and show how your estimates change.

```
> Y <- c(350 * 0.52, 500 * 0.49, 600 * 0.55)
> n <- c(350, 500, 600)
> a <- b <- 1
> n.sims <- 10000
> draws <- matrix(NA, nrow = n.sims, ncol = 3)
> colnames(draws) <- c("A", "B", "C")
> for (i in 1:3) {
+     draws <- rbeta(n.sims, sum(Y[1:i]) + a, sum(n[1:i]) -
+         sum(Y[1:i]) + b)
+     print(c(mean = mean(draws), sd = sd(draws), quantile(draws,
+         probs = c(0.025, 0.975))))
+ }
      mean        sd      2.5%     97.5%
0.5199096 0.0261081 0.4685834 0.5715444
      mean         sd      2.5%      97.5%
0.50212930 0.01725793 0.46864962 0.53595468
      mean         sd      2.5%      97.5%
0.52208666 0.01321983 0.49620579 0.54832019
```

As we get more information, the posterior standard deviation decreases and the posterior mean gets closer to 0.5, which is where the bulk of the observations are around.

d) Now suppose you observe all the polls simultaneously such that you can pool the results together. Analytically show that your posterior is equivalent to the posterior from c).

In part c), we updated our estimates by using old posteriors as new priors. So our posteriors at each update were

$$
\begin{aligned}
\theta|y_A &\sim \text{Beta}(y_A + a, n_A - y_A + b) \\
\theta|y_A, y_B &\sim \text{Beta}(y_B + (y_A + a), n_B - y_B + (n_A - y_A + b)) \\
\theta|y_A, y_B, y_C &\sim \text{Beta}(y_C + (y_B + y_A + a), n_C - y_C + (n_B - y_B + n_A - y_A + b)) \\
&\sim \text{Beta}(y_{ABC} + a, n_{ABC} - y_{ABC} + b)
\end{aligned}
$$

which is mathematically equivalent to the posterior in the case of the polls pooled together.

e) Suppose the electorate is composed of 3,000 voters who must all cast a vote. Using simulation and your posterior from d), predict the probability that the incumbent will win the election given majority rule.

```
> y <- sum(c(350 * 0.52, 500 * 0.49, 600 * 0.55))
> n <- sum(c(350, 500, 600))
> a <- b <- 1
> n.sims <- 10000
> post.draws <- rbeta(n.sims, y + a, n - y + b)
> post.votes.dist <- sapply(post.draws, FUN = rbinom, n = 1,
+     size = 3000)
> mean(post.votes.dist > 3000 * 0.5)

[1] 0.9141
```

f) Now suppose that the electoral rule dictates that a candidate can only win with a supermajority (55%). What is the probability that the incumbent will win the election given this new rule?

```
> mean(post.votes.dist > 3000 * 0.55)
[1] 0.037
```

2. A prior is considered flat or uninformative if it is proportional to a constant.

$$p(\theta) \propto c$$

For a model with a Binomial(n, θ) likelihood,

a) show that a Beta(1,1) prior on θ is uninformative.

The PDF of a Beta(1,1) prior on θ is

$$
\begin{aligned}
p(\theta) &= \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)}\theta^{1-1}(1-\theta)^{1-1} \\
&= 1
\end{aligned}
$$

b) show that if we transform the parameter such that we have a new parameter $\Theta = \exp(\theta)$, a Beta(1,1) prior on θ is no longer uninformative on Θ.

Let $\Theta = \exp(\theta)$. To find the prior distribution of Θ given a Beta(1,1) prior on θ, we need to use the change of variables formula.

$$
\begin{aligned}
\theta &= \log \Theta \\
\frac{d}{d\Theta}\log \Theta &= \frac{1}{\Theta} \\
p(\theta) &= 1 \\
p(\Theta) &= 1\left|\frac{1}{\Theta}\right| \\
&= \frac{1}{\Theta}
\end{aligned}
$$

We can see that the prior distribution of Θ under a Beta(1,1) prior on θ depends on Θ and so is not uninformative.

3. (From Jackman Example 2.9) Geddes (1990) examined the relationship between foreign threat and social revolution in Latin America using data that generated the following 2-by-2 table.

	Revolution	No revolution
Defeated and invaded or lost territory	1	7
Not defeated for 20 years	2	74

Each observation is a 20-year period for each Latin American country, with any given country contributing multiple observations. Propose a model to assess the hypothesis that suffering a military loss within a 20-year period increases the chances of revolution within a country. What is the probability that military defeat is positively associated with revolution?

We can model the problem as two separate binomial distributions. Let θ_0 denote the probability of revolution conditional on no military defeats and θ_1 denote the probability of revolution given a military defeat. We are interested in the following quantity: $P(\theta_1 > \theta_0)$

If we assume that θ_1 and θ_0 are independent a priori and using an uninformative Beta prior, we have the following model:

$$
\begin{aligned}
Y_0 &\sim \text{Binomial}(76, \theta_0) \\
Y_1 &\sim \text{Binomial}(8, \theta_1) \\
\theta_0 &\sim \text{Beta}(1, 1) \\
\theta_1 &\sim \text{Beta}(1, 1)
\end{aligned}
$$

Our posteriors are then

$$
\begin{aligned}
\theta_0 | Y_0 &\sim \text{Beta}(3, 75) \\
\theta_1 | y_1 &\sim \text{Beta}(2, 8)
\end{aligned}
$$

The probability that military defeat is positively associated with revolution is $P(\theta_1 > \theta_0)$, a quantity which we can find via simulation of the posteriors.

```
> n.sims <- 10000
> theta0.post <- rbeta(n.sims, 3, 75)
> theta1.post <- rbeta(n.sims, 2, 8)
> mean(theta1.post > theta0.post)

[1] 0.9479
```