

Lab 3

Your Name

The influence of priors and data on posteriors

Remember, with Bayes' Rule, we know that: $P(\Theta|y) \propto P(\Theta)P(y|\Theta)$.

Task 1 (15 points)

In lecture, we learned that when $X \sim \text{Binomial}(\theta, n)$, and $P(\theta)$ follows a beta distribution, then $P(\theta|x, n)$ also follows a beta distribution. Said differently, the beta distribution is a conjugate prior for the binomial likelihood.

Your task is to demonstrate how the prior probability distribution of θ can influence the posterior estimate of θ , specifically when we only have a single data observation (x) from a random variate (X) that we assume is binomially distributed.

Complete the following:

1. Choose two prior distributions for θ , both being beta distributions. In other words, choose two sets of α 's and β 's to define your two prior distributions. Have one prior be "vague", and have one prior be strongly informative. (3 points)
2. For each prior, create a graphic that shows the prior distribution, the likelihood distribution, and the posterior estimate of θ , structured like the example figure below. Note that the example figure will need to be altered (e.g., the x- and y-axis limits). (7 points)
3. Write a brief interpretation of your results. No more than 4 sentences. (5 points)

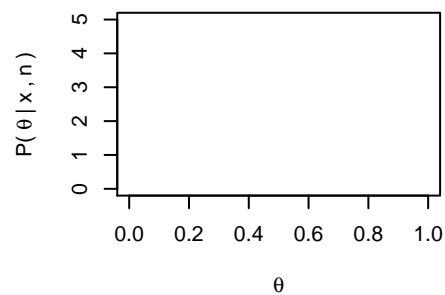
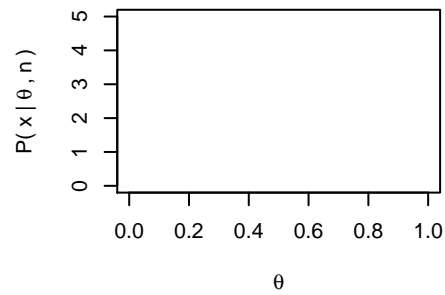
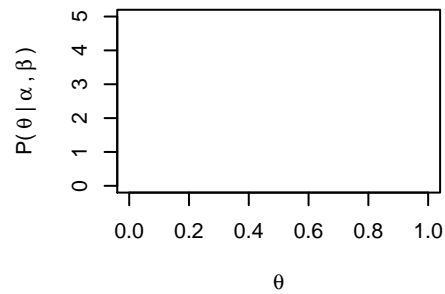
Example figure structure. You will have to populate these figures with the correct objects and syntax. And you'll need one set of these figures per prior distribution.

```
# Create a multi-panel plot with three panels across three rows, and 1 column:
par(mfrow = c(3, 1)) #c(n_rows,n_columns)

plot(NA, NA,
     xlim = c(0, 1), ylim = c(0, 5), #ylim are arbitrary here
     xlab = expression(theta), ylab = expression("P(~theta~|"~alpha~","~beta~)"))

plot(NA, NA,
     xlim = c(0, 1), ylim = c(0, 5), #ylim are arbitrary here
     xlab = expression(theta), ylab = expression("P(~x~|"~theta~","~n~)"))

plot(NA, NA,
     xlim = c(0, 1), ylim = c(0, 5), #ylim are arbitrary here
     xlab = expression(theta), ylab = expression("P(~theta~|"~x~","~n~)"))
```



I will walk you through this task, to a point. First, as in lecture, assume:

```
x = 7
n = 10
```

Here are some hints (as a reminder, don't break up the `enumerate` code block, or your document will not compile).

1. Create a sequence of possible values for θ .
2. Before you create your final figures, you can test how different shape parameters (α and β) affect the prior distribution of θ , $P(\theta|\alpha, \beta)$. It may help to know that the mean of a beta distribution is $E[Y] = \frac{\alpha}{\alpha+\beta}$. Wikipedia also gives some easy examples to follow.
3. Choose your two sets of beta distribution parameters.
4. For each set of beta parameters:
 - (a) For each value of your θ sequence, calculate the log-prior probability, given your prior beta parameters. Store these log-prior probabilities in an object. In the PDF function, you can use the option `log = TRUE` to output the density value on the natural-log scale. **NOTE:** When you plot the PDF, transform back to standard density scale with the `exp()` function.
 - (b) For each value of your θ sequence, calculate the log-likelihood of observing $x = 7$, given that parameter value. Store these likelihood values in an object. Again use the option `log = TRUE`, but back-calculate to the original density scale when you plot.
 - (c) Because we know that the posterior of θ , given our data, is a beta distribution, calculate the posterior distribution of θ .
 - (d) Plot the three distributions, similar to my example of a three-panel plot, above.
5. Write your interpretation of how the strength of prior influences your posterior.

Task 2 (20 points)

In lecture, we learned that when (1) $Y \sim \text{Normal}(\theta, \sigma^2)$, (2) σ^2 is known, (3) $\theta \sim \text{Normal}(\mu_0, \tau_0^2)$, and (4) we have one data observation, y , then the posterior:

$$\theta|y, \sigma^2 \sim \text{Normal}\left(\frac{\frac{\mu_0}{\tau_0^2} + \frac{y}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2} + \frac{1}{\sigma^2}}\right).$$

This is with a single data point, but in most cases we have several data points. It can be shown that, when we have multiple data observations, such that y is now a vector of n data observations, then the posterior becomes:

$$\theta|y, \sigma^2 \sim \text{Normal}\left(\frac{\frac{\mu_0}{\tau_0^2} + \frac{\bar{y}n}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}\right).$$

Here \bar{y} is the average of our data observations, and again, n is the number of data observations.

In this task, we will show how the *number of data points* can influence our likelihood, and therefore can influence our posterior, even with a fixed prior. First, assume:

```
# Prior parameters on unknown mean (\theta):
# Assume we know that the mean is "around 5"
mu_0 = 5.0
tau_0_sq = 9.0
tau_0 = sqrt(tau_0_sq)

# Known sigma_sq
sigma_sq = 2.0
sigma = sqrt(sigma_sq)

# True mean of the random variate Y
# This is what we need to estimate, but
# to simulate data, we have to assign it here
mu_true = 3.0

# Compare these two values of "n"
# "n" is equal to the number of data observations
n_compare = c(2, 75)
```

Complete the following:

1. Simulate data from the “true” distribution of Y . You’ll have two data sets, each with different numbers of data observations. See the code above for the number of data observations you should use in your comparison.
2. For each data set, create a graphic that shows the prior distribution, the likelihood distribution, and the posterior estimate of θ , as in Task 1. (15 points, including the creation of your function, see below)
3. Write a brief interpretation of your results. Specifically, comment on how the number of data observations affects the posterior. No more than 4 sentences. (5 points)

Here are some hints:

1. Create a sequence of values for θ that could possibly result from the normal prior assumption (with the parameters of this normal prior defined in the code, above). For instance, -100 is not a likely value of θ ...
2. For each value of θ in your sequence, calculate the prior density, on the log scale, and store these densities in an object. Remember that the prior distribution will not change between the two data sets.
3. Create a function to calculate the total likelihood of your data, given your parameters, $P(y|\theta, \sigma^2)$. The inputs should be your data vector and your parameters (θ, σ^2 , but remember σ^2 is fixed). The function should then output the log-likelihood of your data. Remember that on the log-scale, the total likelihood of your data is the *sum of the likelihoods of each data point*.
4. For each of your two data sets:
 - (a) For each value of θ in your sequence, use the function you created to calculate the total log-likelihood of your data set.
 - (b) Now, you will have two vectors: (1) one vector that contains the log-prior density for each value of θ in your sequence, and (2) one vector that contains the total log-likelihood of your data set for each value of θ in your sequence. Therefore, these two vectors should be of the same size.
 - (c) You can now calculate the log-posterior distribution of θ given your data. To do this, for each value of θ in your sequence, you will *add* the log-prior and the log-likelihood. Thus, you will have a value of your posterior density for each value of θ in your sequence.
 - (d) Create the three panel plot, as in Task 1. Remember to convert to the regular density scale using the `exp()` function.
5. Write your interpretation.