Lab3

Jun Rao

9/12/2019

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 $X \sim Binomial(\theta, n)$, 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)

```
theta = seq(0,1,length.out = 10000)
alph_1 = 2
alph_2 = 10
beta_1 = 2
beta_2 = 10

# Prior Distribution
prior_vague = dbeta(theta, alph_1, beta_1)
prior_informative = dbeta(theta, alph_2, beta_2)

#Now, we will calculate log-prior probability
log.prior_vague = dbeta(theta, alph_1, beta_1, log = TRUE)
log.prior_informative = dbeta(theta, alph_2, beta_2,log = TRUE)

#Define x and n
x = 7
n = 10
#For each value of ϑ sequence, calculate the log-likelihood of observing x =
7, given that parameter value
log.likelihood = dbinom(x,n,theta,log=TRUE)
```

```
#Postirior
log_post_vague = log.likelihood + log.prior_vague
log_post_informative = log.likelihood + log.prior_informative
```

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)

```
# Create a multi-panel plot with three panels across three rows, and 2 column
:
par(mfrow = c(3, 2)) #c(n_rows,n_columns)

plot(exp(log.prior_vague)~theta, type ='l', col = 'blue',
     #xlim = c(0, 1), ylim = c(0, 5), #ylim are arbitrary here
     xlab=expression(""~theta~""),
     ylab=expression("P("~theta~" | "~alpha~", "~beta~")")
     )


plot(exp(log.prior_informative)~theta, type ='l', col = 'red',
     #xlim = c(0, 1), ylim = c(0, 20), #ylim are arbitrary here
     xlab=expression(""~theta~""),
     ylab=expression("P("~theta~" | "~alpha~", "~beta~")")
     )

plot(exp(log.likelihood)~theta, type='l', col='blue',
     xlab=expression(""~theta~""),
     ylab=expression("P(X | "~theta~", n)"))

plot(exp(log.likelihood)~theta, type='l', col = 'red',
     xlab=expression(""~theta~""),
     ylab=expression("P(X | "~theta~", n)"))

plot(exp(log_post_vague)~theta, type='l',col='blue',
     xlab = expression(theta),
     ylab = expression("P("~theta~"|"~x~","~n~")"))

plot(exp(log_post_informative)~theta, type='l',col = 'red',
     xlab = expression(theta),
     ylab = expression("P("~theta~"|"~x~","~n~")"))
```
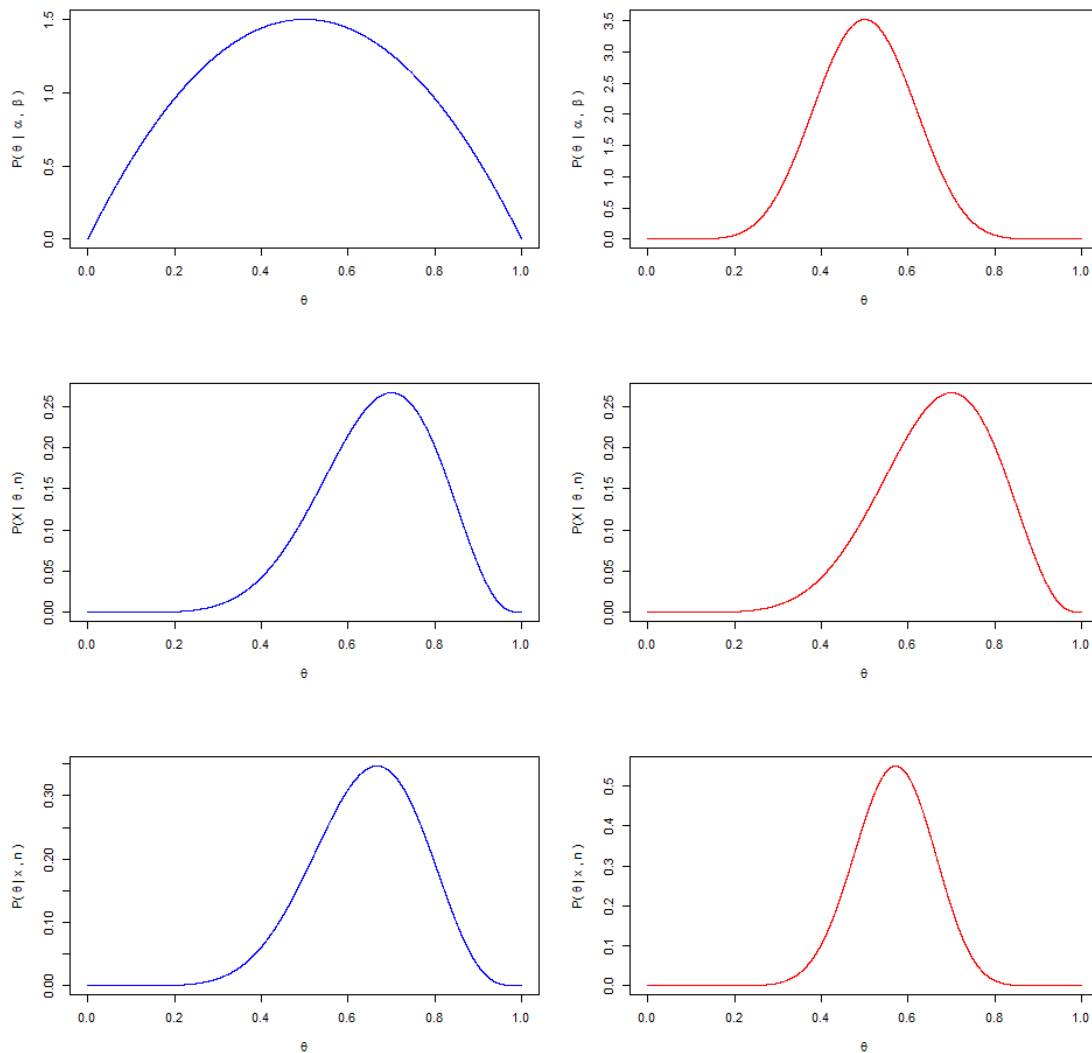
3.Write a brief interpretation of your results. No more than 4 sentences. (5 points)

In the right side (informative), prior clearly influences the posterior graph. The vague prior does not have any high concentration of density. Liklihood is not influenced by the priors.

Task 2 (20 points)

In lecture, we learned that when

(1) $Y \sim Normal(\theta, \sigma^2)$,

(2) σ^2 is known,

(3) $\theta \sim Normal(\mu_0, \tau_0^2)$,

and

(4) we have one data observation, y, then the posterior:

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

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:

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

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

In this task, we will show how the 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.

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 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 function.

```r
#Based on the # of observation in normal distribuation to generate data
data_generation <- function(num_observation){
        return(rnorm(num_observation, mean=mu_true, sd=sigma))
}

# Create a sequence of values for theta
theta = seq(-15,15,length.out = 10000)

#The prior
```

```r
log.prior = dnorm(theta,mu_0,tau_0,log = TRUE)

#Likelihood function
log.likelihood <-function(input_data, input_theta,input_sigma = sigma){
        result = 0;
        for(data in input_data){
                result = result + dnorm(data,input_theta,input_sigma,log = TRUE)
        }
        return(result)
}

#Calculate the postrior
data.1 <- data_generation(n_compare[1])
data.2 <- data_generation(n_compare[2])
log.likelihood.1 <- log.likelihood(data.1,theta)
log.likelihood.2 <- log.likelihood(data.2,theta)
postrior.1 <-log.prior + log.likelihood.1
postrior.2 <-log.prior + log.likelihood.2


# Create a multi-panel plot with three panels across three rows, and 2 column
:
par(mfrow = c(3, 2)) #c(n_rows,n_columns)

plot(exp(log.prior)~theta, type ='l', col = 'blue',
     xlab=expression(""~theta~""),
     ylab=expression("P("~theta~" | "~mu[0]~", "~tau[0]~"^2)")
     )
plot(exp(log.prior)~theta, type ='l', col = 'red',
     xlab=expression(""~theta~""),
     ylab=expression("P("~theta~" | "~mu[0]~", "~tau[0]~"^2)")
     )

plot(exp(log.likelihood.1)~theta, type='l', col='blue',
     xlab=expression(""~theta~""),
     ylab=expression("P(y | "~theta~", "~sigma~"^2)")
     )

plot(exp(log.likelihood.2)~theta, type='l', col = 'red',
     xlab=expression(""~theta~""),
     ylab=expression("P(y | "~theta~", "~sigma~"^2)")
     )

plot(exp(postrior.1)~theta, type='l',col='blue',
     xlab = expression(theta),
     ylab=expression("P(y | "~theta~", "~sigma~"^2)")
     )
```

```
plot(exp(postrior.2)~theta, type='l',col = 'red',
     xlab = expression(theta),
     ylab=expression("P(y | "~theta~", "~sigma~"^2)")
     )
```
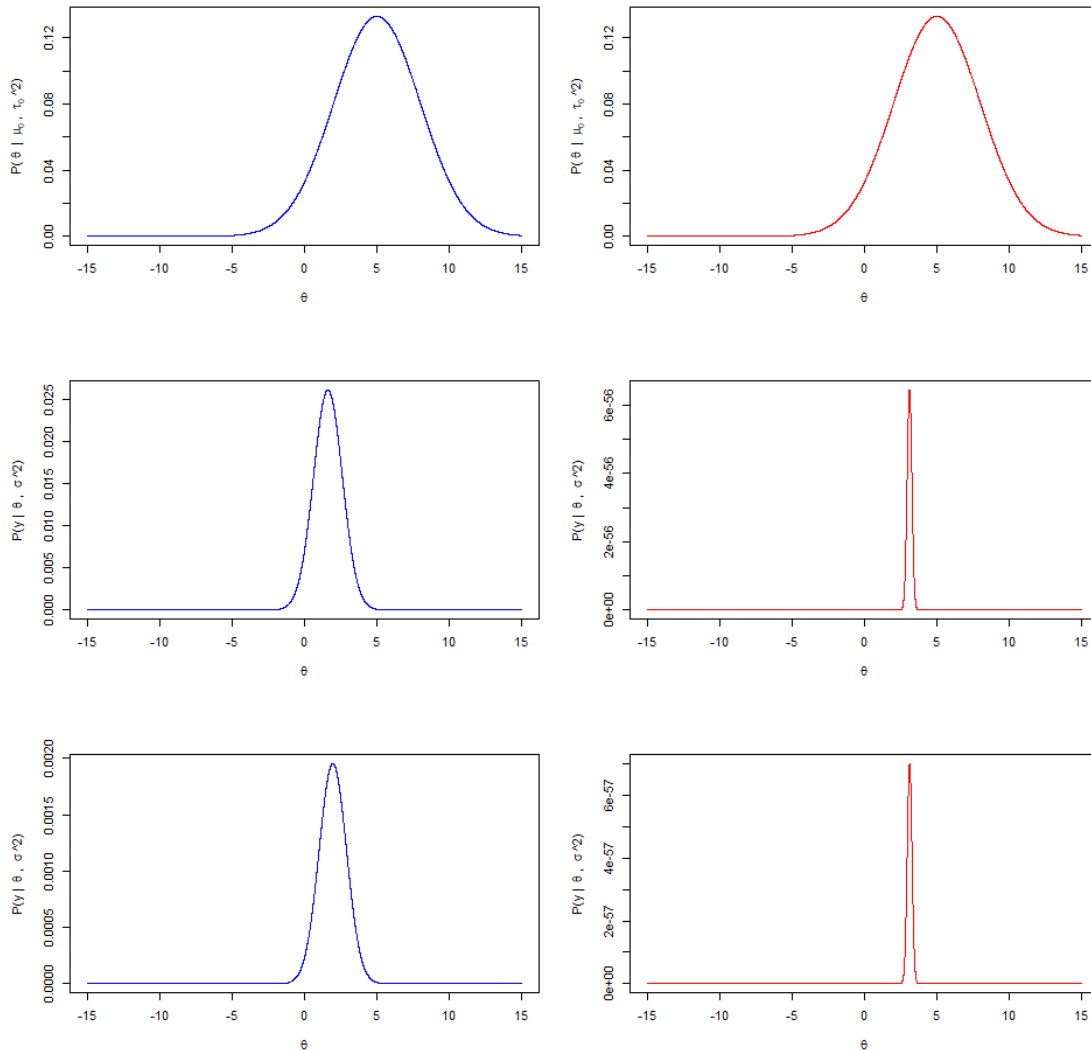

In the blue part, we don't have enogh data (the number of observation is 2) to get accurate likehood value. Whereas, we have data in red part more than blue part. Hence, we can conclude that the more data we have, the more likelihood we get. Then, we can get the accurate posterior.