

EEEB UN3005/GR5005

Homework - Week 04 - Due 26 Feb 2019

Xun Zhao, xz2827

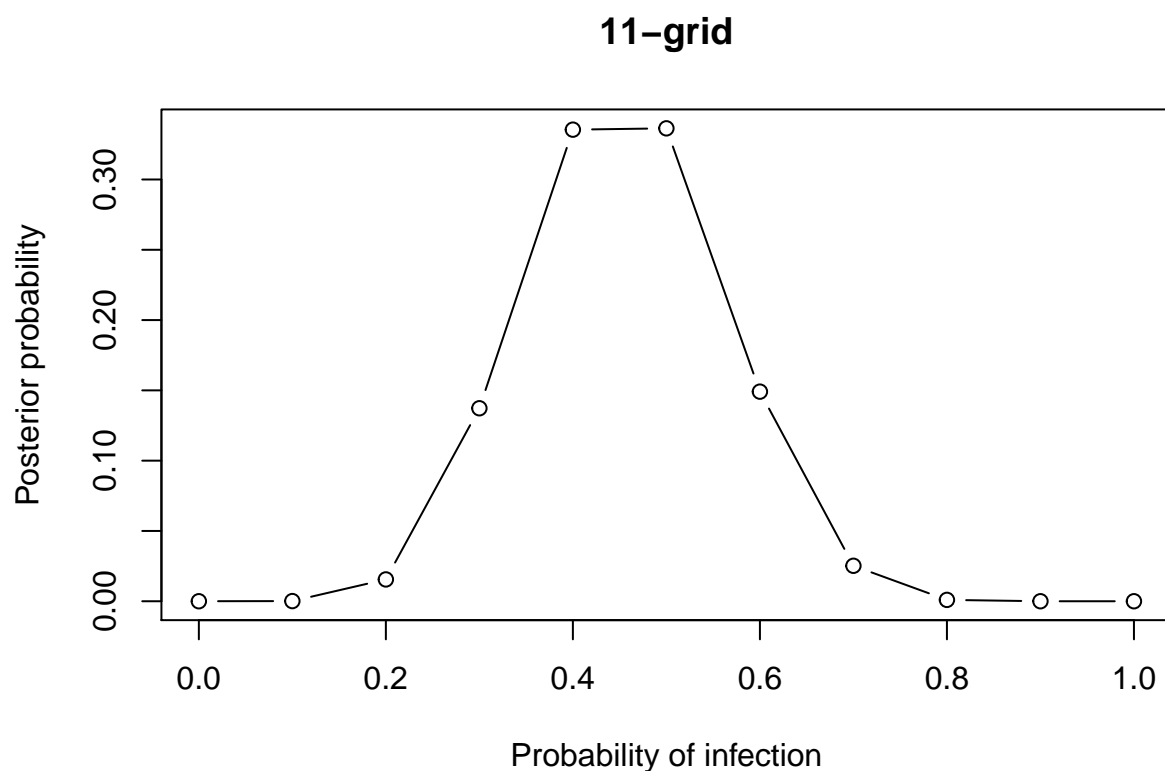
Homework Instructions: Complete this assignment by writing code in the code chunks provided. If required, provide written explanations below the relevant code chunks. Replace “USE YOUR NAME HERE” with your name in the document header. When complete, knit this document within RStudio to generate a pdf. Please review the resulting pdf to ensure that all content relevant for grading (i.e., code, code output, and written explanations) appears in the document. Rename your pdf document according to the following format: hw_week_04_firstname_lastname.pdf. Upload this final homework document to CourseWorks by 5 pm on the due date.

Problem 1 (5 points)

Imagine you’re beginning studies in a new field system, focusing on a bacterial pathogen (i.e., disease-causing organism) that infects small mammals. Your pilot sampling efforts find 9 infected animals out of 20 animals sampled. Based on this (admittedly small) sample, you’d like to estimate the probability of infection value (p) in the study population. To do so, use grid approximation with 11 grid points to construct the posterior for the probability of infection parameter. Assume a flat prior.

Plot your posterior distribution. (Using base R will likely be easier here, since it won’t require you to organize your grid and posterior vectors into a data frame.) Based on visual inspection of the plot, which values of the probability of infection parameter have the most posterior support?

```
Bayes = function(grid.len){  
  grid = seq(0, 1, length.out = grid.len)  
  prior = rep(1, grid.len)  
  likelihood = dbinom(9, size = 20, prob = grid)  
  unstd.post = likelihood * prior  
  std.post = unstd.post / sum(unstd.post)  
  plot(grid, std.post, type = 'b',  
        xlab = 'Probability of infection',  
        ylab = 'Posterior probability',  
        main = paste(c(as.character(grid.len), '-grid'), collapse = ''))  
}  
Bayes(11)
```

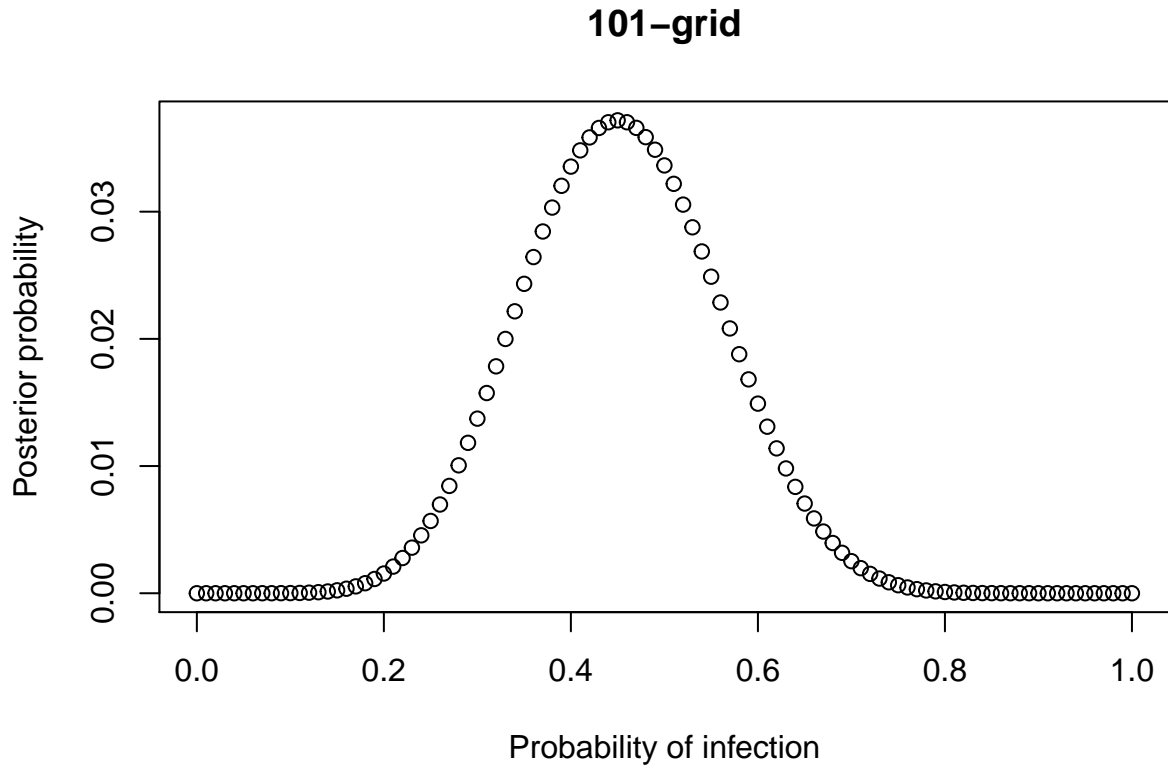


Problem 2 (2 points)

Repeat the previous problem using grid approximation with 101 points. Assume a flat prior, and again plot the resulting posterior distribution.

How has the posterior distribution changed as a result of increasing the number of points used in your grid approximation?

`Bayes(101)`



Answer:

- (1) The curve is smoother than 11-grid.
- (2) The maximum position is more accurate when $\Delta x = |x_{i-1} - x_i|$ is small.
- (3) The curve distributes like a Gaussian distribution with symmetrical shape.

Problem 3 (3 points)

Thus far you have been assuming a flat prior for the probability of infection parameter in your computations (i.e., equal prior probability across all potential parameter values). Do you think this is a realistic assumption? What, if anything, do you know about the probability of infection parameter that would allow you to formulate a sensible prior? Where might you gather information with which to inform your choice of prior?

(1)

In my opinion, it is not realistic, because the bacteria (assuming it is pathogenic) may not infect the host for 100% or 0% probability. As the bacteria was discovered by someone, there might be some pre-experiment about the infection. That can be used as the prior to provide more information about it.

(2)

For example, if we want to estimate the probability of infection of the bacteria to human, we might search for some published data about the experiment of this bacteria infecting monkey, or other primates. In addition, if the bacteria belongs to a bacteria family that all the family members are pathogenic, we can use the data from other bacteria species as the prior.

More generally, we can use the average infection probability of all known bacteria to formulate a prior.