THE IMPROPER PRIOR DISTRIBUTION

When performing Bayesian statistics on continuous distributions, our prior distribution does not have to proper, it can instead be improper. An improper prior distribution does not need to sum to one. Below, we will look at some examples of prior distributions that can be used and later how they differently affect the resulting posterior distribution after data collection.

1. First start by opening the Continuous Bayesian Statistics Applet for Improper Prior Distributions. Note the different inputs that can be changed in the sidebar panel.

2. Next, click the "Choose File" button under the Sample Data label. Choose the file, "busStopTimes.csv" file, which should have been saved to your computer prior to starting this worksheet. After selecting that file, how many plots show up to the right of the sidebar panel? What does each one show?

3. Set the value of k to 0, if it is not already set to that value. Then observe the plot of the Prior distribution. What type of distribution is this? How does the probability of $\theta = 3$ compare to the probability of $\theta = 12$?

4. Does the Prior distribution sum to 1 or not? (Is the area under the curve equal to 1?) Does this mean the Prior distribution is proper or improper?

5. Now, change the value of k to 1. What changes do you notice about the Prior distribution now? How are the probability of θ and θ related?

6. With k = 1, is the Prior distribution proper or improper? Why?

7. Once again, change the value of k. This time, set k equal to -1. Answer the questions asked in exercises 5 and 6 with k now equal to -1.

Big Idea: In Bayesian statistics for continuous variables, we can use improper prior distributions. This means that the sum of the probabilities, or area under the curve, does not need to equal 1. This, therefore, does not restrict us to prior distributions that are convergent. Only the posterior distribution is required to sum to 1. The prior distributions being used in this worksheet are of the type $1/\theta^k$. Thus a uniform prior is when k = 0, positive k values represent inverse relationships between θ and its probability and negative k values represent direct relationships between θ and its probability.

UNIFORM PRIOR (k = 0)

For the next sections of this worksheet, we will use an example that deals with the amount of time you have to wait for the bus after its scheduled arrival time. This is a continuous variable because the waiting time can take any value and is not restricted to just a discrete set of numbers. Our main goal is to determine the longest you will have to wait for the bus at the stop near your house.

8. If you have not already done so, upload the “busStopTimes.csv” file. Then set k to 0. What do you notice about the prior distribution? What does this say about our prior assumptions about the maximum wait time for the bus?

9. Suppose you went to wait for the bus scheduled to arrive at 10:30 every morning for 2 weeks. These weeks can be assumed to be seemingly random. At the bus stop for those 14 days, you recorded how late the bus was in minutes (recorded in busStopTimes.csv). Comment about the times seen in the distribution of the sample. (*Hint: Top-right graph*)

10. Look at the file containing the wait times for the bus.If you had to predict what the maximum wait time, θ, would be, what would your guess be? Why?

11. Based on your guess from exercise 10? Would θ = 5 minutes be a good guess for the maximum wait time? Why or why not?

12. Would 15 be a good guess for θ? Is it a better or worse guess than 5? Explain your reasoning.

Big Idea: When estimating θ, we cannot have a value less than the maximum value we already found in a sample. Thus θ = 5 would be a very bad guess, since it has a probability of 0 of being the maximum wait time. On the other hand, θ = 15 would not be a terrible guess, because it is still possible for it to be the maximum. One of the best guesses for the maximum value, θ, though, is the maximum value that you found in your sample. In this case it is 9.19, which is the most likely value for the true maximum just based on observation. Later in this worksheet, we will go over another estimate, known as the Bayes Estimate, as well as an interval estimate for θ.

13. Now observe the posterior distribution in the bottom left. Do you notice anything unusual about the posterior distribution? Is it continuous, skewed, etc.? Comment your findings below.

14. On the axis labelled θ, go to the maximum value you found in the sample of 14 wait times. What happens to the posterior distribution at this point? What are the values of posterior distribution to the left of this point?

    Big Idea: After taking a sample, the prior distribution can be updated to create the posterior distribution for the possible values of θ. As explored previously, the probability of θ less than the maximum of our sample is always zero in the posterior distribution, since we have already found a point greater than it. The rest of the posterior distribution is right skew.

BAYES ESTIMATE

Another way to estimate the maximum wait time for the bus is to use the Bayes estimate. The section below will explore how to the Bayes estimate is calculated.

15. Look at the plot for the Posterior distribution. What does the blue vertical line represent? Is the θ value where the blue line is vertical greater or less than the maximum wait time we found in our sample?

16. What is the total area under the posterior distribution? Why?

17. Based on your answer to exercise 16, what do you think the area under the posterior distribution from 0 to the Bayes estimate? What about from the Bayes estimate to infinity? What does this say about what the Bayes estimate represents? Explain your reasoning?

Big Idea: The Bayes estimate is the average of the posterior distribution. The area under the posterior distribution is equal to 1, and so the Bayes estimate is the θ value whereas the area under the curve is split to be 0.50 to the left and 0.50 to the right. This Bayes estimate differs from choosing the maximum value found in the sample and is a much more conservative estimate.

HIGHEST PROBABILITY DENSITY INTERVAL (HPD INTERVAL)

Sometimes a point estimate is not the best way to estimate θ. Instead, we can use an interval to give a range of possible values for θ. Due to the nature of Bayesian statistics, using our posterior distribution, we can add up say that there is a certain probability that the value of θ lies between two values.

18. Find the place in the applet where the HPD interval is stated. What do you notice about the lower value of the interval?

19. Next, report the 90% HPD interval. State the interval below in the context of the problem.

20. What is the 95% HPD interval for θ? Is this interval wider or narrower than the 90% HPD interval? Why?

21. Lastly, change the size of the HPD interval one last time to be a 99% HPD interval. What do you notice about the lower value of the interval compared to the lower values found in exercises 19 and 20? What are the reasons for this?

Big Idea: The Highest Probability Density (HPD) intervals are used to give a range of most probable values for θ. If we are given a 90% HPD interval, that means that there is a 90% chance that the true value of θ lies between the two end points of the interval. Thus, to capture more area, an increase in the percentage value of the HPD interval is associated with an increase in width of the interval. Also, the lower end point of the interval will always be the same, which is the maximum value seen in the sample for reasons that were explained in previous big ideas.

INVERSELY RELATED PRIOR DISTRIBUTION (k > 0)

In this next section, we will explore what happens to the posterior distribution and the estimates of θ when k is greater than zero.

22. Start by Increasing the value of k to be greater than 0. Set k = 1. What do you instantly notice about the prior distribution? What happens as θ increases? Is it a linear relationship, does it have concavity? What is the equation of the prior distribution? Explain your findings below.

23. What does this prior distribution say about our prior assumptions of the possible values of θ? Are lower values or larger values more likely?

Big Idea: When k is greater than 0, θ and its probability are inversely related. This means that as θ increases, the prior probability of that θ decreases. This prior distribution weights smaller values of θ more heavily. We have a stronger belief that θ is a low value than a high one.

24. With k = 1, we are more heavily weighting lower values of θ in this prior distribution compared to the exercises above. Do you believe that the Bayes estimate will be higher, lower, or the same as when k was equal to 0? Explain your reasoning.

25. Also, do you believe that the interval will be wider or narrower for $k = 1$ as compared to when $k = 0$? Why?

26. Let's discover if your predictions from the previous exercises are correct. Report the values of the Bayes estimates for $k = 0$ and $k = 1$. Which is higher, were you correct in your prediction? In hindsight, can you explain why this is?

27. Also, report both the 90% HPD intervals for $k = 0$ and $k = -1$. What do you notice about the lower end points? Which interval is wider? Is this what you predicted above? Can you explain why one interval is wider than the other?

Big Idea: When $k > 0$, the prior distribution more heavily weights the lower values of θ. Thus, the posterior distribution will have a slightly heavier weight on the lower values as well. This causes the Bayes estimate and HPD interval to also be pulled to the left, toward lower θ values, since the probabilities are higher for those lower θ values.

28. What if we used a prior distribution whereas $k = 2$? What does this say about our prior assumptions of the possible θ values? Are we more or less heavily weighting low θ values compared to when $k = 1$?

29. Predict whether the Bayes estimate would be less or greater than what it was when k equalled 1. Also, will the HPD interval be wider or narrower for $k = 2$ compared to $k = 1$? Write down your predictions and your reasoning below.

30. Change the value of k to 2. Explore whether or not your predictions were correct for exercise 29.

31. Now, observe the posterior distribution as you increase k from 0 to 2. What happens to the posterior distribution? Explain in terms of concavity and the more probable values of θ.

    Big Idea: As k increases, we weight the lower values of θ. Plus, the Bayes estimate of θ decreases and the HPD interval decreases with an increase in k due to the higher probability of smaller θ values in the posterior distribution.

DIRECTLY RELATED PRIOR DISTRIBUTION (k < 0)

Lastly, we will explore the effects on the posterior distribution if we use an improper prior distribution using a k value less than 0.

32. Start by changing k to a value less than 0. Set k = -1. Comment about the prior distribution below. Is it increasing? Is it concave? What is the equation of the prior distribution?

33. What type of assumptions are being made about θ prior to data collection? Are lower or higher values weighted more heavily? What does this mean in the context of the bus stop wait times example?

34. What about when k = -2? What is the equation of the prior distribution? Are the lower or higher values of θ more heavily weighted?

35. Predict whether a prior with k = -1 or k = -2 will have a wider HPD interval. Also which value of k will have a lower Bayes estimate? Explain your reasoning.

36. Report the Bayes estimates of θ and the HPD intervals for both k = -1 and k = -2. Were your predictions in exercise 35 correct? Observe the plot of the posterior distribution to explain why this Bayes estimates and the HPD intervals are different. As k increases,

what happens to the concavity of the posterior distribution and the probabilities of higher θ values?

Big Idea: When $k < 0$, we are weighting higher values of θ more than the lower values. Thus the posterior distribution has slightly higher probabilities at higher values of θ than when $k = 0$. As k decreases, the Bayes estimate will increase as well as the width of the HPD interval due to the weighting of higher θ values. As shown in both this section and the previous one, k is inversely related to both the Bayes estimate and the width of the Highest Probability Density interval.

THE BIG IDEAS IN REVIEW

One major idea in this worksheet is that the prior distribution that you do not need to have a proper prior distribution. This means that we can use an improper prior distribution whereas the prior probability does not need to sum to 1. In this case we used prior distributions that were related to $1/\theta^k$. Above, you discovered some of the relationships between k, the Bayes estimate, and the Highest Probability Density interval. This is the major idea to gained from this worksheet. Thus, to review this idea, fill out the table below for the bus stop example and notice the relationships between the variables that you report.

Bayes Estimates for Various k Values

k	Bayes Estimate	Upper Bound of 95% HPD interval
-3		
-2		
-1		
-0.5		
0		
0.5		
1		
2		
3		

Lastly, if you are given a population distribution in which the lowest value is not equal to zero, you can still use the same ideas learned in this applet. Instead of using the exact estimate values found in the applet, you can simply add the lower value of the population to your answers. This adjustment is what makes this set of prior and posterior distributions so useful. I encourage you to try more examples and discover more above how the improper prior distribution can be useful to estimate θ.