POISSON POPULATION

For this worksheet, we will deal with a population that follows the Poisson distribution. This means that we are able to count successes, but unable to count failures. Thus, the possible values (number of occurrences in a sample) are all integers. Examples of this distribution will be explored in this worksheet, as well as how the prior probabilities are updated when observing Poisson populations.

OVERVIEW OF THE BASIC IDEAS

1. First, start by opening the Discrete Bayesian Statistics Applet. Click the "Poisson" button, located on the left under the label, "Population Distribution to Pick From." Set the Uniform Prior Distribution to have a lower limit of 0 and an upper limit of 10. Also set the true population mean to be 4, if not already set to those values.

2. Look at the plot in the top left for the Prior Distribution. What type of distribution is being used?

   ____

3. From the Prior Distribution, what is the probability that the true mean is equal to 4?

   ____

4. Now, observe the plot in the upper right. This is the expected sampling distribution from a population with a Poisson distribution. What shape does this graph appear to have? What is the most likely number of occurrences to expect?

   Big Idea: The sampling distribution for a Poisson population is skewed right. The most likely value for the number of occurrences in a sample is the true population mean.

STARGAZING EXAMPLE

Suppose you went out on a clear night in the mountains to go stargazing. You want to know what the true mean number of shooting stars that you would see in a two hour span is. This would be a Poisson distribution because we can count successes, or number of occurrences, being the number of shooting stars, but can't count failures. We cannot count the number of shooting stars that didn't come in those two hours in a practical matter.

5. Being unsure of what the true mean number of shooting stars that could be seen in two hours, you decide to use a uniform prior distribution. You estimate that the true mean can be anywhere between 0 and 14. Change the appropriate values in the applet to meet these prior distribution assumptions.

6. You sit on the mountain for two hours at night, observing the stars. This time in which you are observing can be assumed to be random. During your time out, you see 6 shooting stars! Suppose the true mean number of shooting stars that can be seen in two hours of stargazing is 4. Set this value as the true population mean. Is seeing 6 shooting stars an unusual number?

7. Now, based on our sample whereas we saw 6 shooting stars, what happens to the posterior distribution? Enter the number of occurrences in the sample into the appropriate input in the applet. Describe what happens to the posterior distribution. (What value for the true mean is now seen as the most probable? At what values of the true mean did the probability decrease? Explain all you can about what changed.)

8. Using the appropriate information from the applet, complete the following statement: After observing 6 shooting stars in one two hour session, we believe that there is a 90% chance that the true mean number of shooting stars that can be seen in two hours is between _________ and __________ .

9. Based on your answer to exercise 8, would you still believe 4 shooting stars is a plausible value for the true mean number of shooting stars seen in two hours? Explain your thinking.

10. Lastly, by increasing the number of occurrences in increments of one, find the number of shooting stars we would have needed to see to not include 4 in the 90% credible interval of the possible values for the true mean. Report the number of occurrences in the sample that you found.

11. If the true mean was in fact 4 shooting stars, what is the probability of seeing the number of shooting stars you found in exercise 10? Is this higher or lower than expected? Explain your reasoning. (*Hint: Use the population sampling distribution to find the answer.*)

Big Idea: After we observe a sample over a given amount of time, we can update our prior probabilities of the possible values for the true mean number of occurrences in a given time. These updated probabilities become our posterior distribution for the true mean, in which we can use to make new claims about the true mean, such as credible intervals, and the most probable values for the true mean.

TYPING ERRORS EXAMPLE

In this example, we will look at the true number of typing errors that a person will make per page. This follows a Poisson distribution. Suppose Jack averages making 6 typing errors per page. In the exercises below, we will look at how the width of different uniform prior distributions affects the posterior distribution, as well as different observed values in the sample.

12. First, set the true population mean to 6. Also, assume that we will start by using a prior distribution for the true mean from 0 to 10. Change the appropriate values in the applet to match that assumption as well.

13. Suppose we randomly select one page that Jack has written. On this page, it is found that he made just one mistake. If the true mean number of mistakes that Jack makes per page is 6, approximate what the probability of him making one mistake on this paper was. Report that value.

    ____

14. Now update the number of occurrences in a sample to be 1 (matching our sample). Observe that posterior distribution. What is the probability that the true mean number of errors for Jack is zero? Why do you think this is?

Big Idea: If we get any sample with at least one success, then the updated probability (in the posterior distribution) that the true mean is zero is zero. This is because, if the true mean were to be zero, then it is not possible for any samples to get a single success. Thus, with at least one success, we can rule out zero as a possible value of the true mean.

15. What if we got a more likely number of errors in our sample though. Suppose we found 6 errors, the true mean. Increase the sample number of occurrences in the applet to match this data. Report the 95% credible interval for the true number of typing errors that Jack would make in one page in the context of this problem.

16. Now, what if we really had no idea of how good or bad Jack would be at typing, and thus had a much wider range of possible values for the true mean number of typing errors per page for Jack. Set the lower and upper values for the prior distribution to 0 and 13, representing this less accurate prior distribution.

17. Using the new uniform prior distribution (0 to 13) and the same data collected (6 errors in a randomly selected page), what is the new 95% credible interval for the true number of typing errors per page for Jack? How does it compare to the interval you found in exercise 15, why do you think this is?

18. Suppose, instead of a worse idea of Jack's typing accuracy, we had a better idea. To demonstrate this, set the uniform prior distribution to be from 0 to 7. Report the 95% credible interval for Jack's true mean number of typing errors after observing one sample with 6 errors. Compare this interval to the two found in exercises 15 and 17. Is this difference what you expected? Explain your reasoning.

Big Idea: With a wider uniform prior distribution, the size of the credible interval, given the same data, will be wider, in general. The opposite is also true, whereas a narrower uniform prior distribution will yield a narrower credible interval. This is because if we are more "accurate" before collecting data, we will have a more accurate credible interval, which will be narrower.

Explore more changes in the input parameters for the applet, creating hypothetical situations. This will help you to learn more about how a uniform prior distribution can be used for a Poisson population.