STAT 532 Assignment 1 (Fall 2015)

Due Wednesday, Sept. 3

For these exercises, use the marbles-in-the-bucket situation described in class. Please provide nicely organized answers, preferably using what you learned from Jim last year.

1. (1 pt) What base R function can you use to obtain $Pr(X = x|\theta)$ for specific values of x and θ?

2. (1 pt) Write your own R function that outputs the probability of drawing ONE gold ball for a specified value of θ. The first input should be θ.

3. (1 pt) Now, starting with your function from 2. and using the `apply()` function create the 2-way table discussed in class (columns are x values and rows are θ values). Provide a nicely formatted table here.

4. (2 pts) Sum across columns and rows of your table. Briefly discuss the results (see next question for more discussion).

5. (4 pts) Using this context, clearly explain to someone what a likelihood function is and how it differs from a probability mass function. What notation can help make this distinction?

6. (2 pts) What rule do we often use for estimating θ? (hint: What is the maximum likelihood principle?)

7. (3 pts) How does an estimator differ from an estimate? Explain the distinction to a non-statistician in your own words (do not simply provide textbook definitions).

8. (4 pts) What are the possible values the maximum likelihood **estimator** can take on? Briefly justify your answer.

9. (3 pts) Provide an appropriate display of the sampling distribution of the ML estimator.

10. (5 pts) Now, suppose you actually observed 1 gold ball in your sample.

    (a) What is the maximum likelihood **estimate**?

    (b) Plot the likelihood function. Think carefully about what type of plot is most appropriate. I suspect you may default to using a histogram...is it most appropriate? Why or why not?

    (c) Plot the normalized likelihood function.

11. (3 pts) Consider a simple linear regression problem. You can use `lm()` or `glm()` to get the same estimates.

    (a) What method is being used in each and what distributional assumption are needed for each?

    (b) Briefly compare how confidence intervals are obtained for the two methods.

12. (2 pts) Write down some form of Bayes Theorem (whatever first pops into your head and helps you remember it). Then, rewrite it with probability mass function notation to match the bucket problem.

13. (3 pts) How do we obtain a posterior distribution (a probability distribution) starting from a likelihood function for the bucket case. (Do not use the "proportional to" shortcut and be sure to write out all parts explicitly).

14. (6 pts) Now we will construct a new table in R with the following columns in order: θ, Prior probabilities, Value of the likelihood, product of the prior and likelihood, value of the posterior distribution. Construct this table for the following situations.

    (a) Suppose the filler of the bucket rolled a pair of dice and put in as many gold marbles as appeared on the two faces showing.

    (b) Suppose the filler rolled one die for the number of gold marbles and the remaining marbles were blue.

    (c) Suppose the filler flipped a fair coin for each of the 12 marbles and if it was a head put in a gold marble, and if it was tails put in a blue marble.

15. (3 pts) Plot and compare the posterior distributions for the different priors – do they make intuitive sense? Briefly explain.

16. (2 pts) Compare the posterior distributions to the normalized likelihood function. Do any match? Why or why not?

17. (4 pts) How would the situation be complicated if there were no information about the method used to fill the bucket with the two colors of marbles? Draw a probability distribution that might be useful to capture the following situations and briefly explain your choice. You can do this by hand (neatly) or using R if you would like. Think deeply about how these differ from the setting in 14.

    (a) You know the filler has a strong preference for the color blue.

    (b) You have played this game with many different fillers and all have chosen close to equal composition between blue and gold marbles.

    (c) You have played this game with many different fillers and found that all put far more of one color than the other (though you haven't noticed a preference for the dominant color in general).

    (d) You have no knowledge about how the filler might undertake their job and you're not willing to speculate.