Module 2 - MTH 461

Contents

I. Monte Carlo Integration 3
II. Maximum Likelihood Estimation 4
III. Least Squares Regression 5

Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the module. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- There are two ways you can write your answers, a: by handwriting (either physically or digitally), or b: by typing on a template document with file type options, which can be downloaded from the course website.
- If you had handwritten your answers/solutions on a physical paper, make sure to label it properly and please scan your document using a scanner app for convenience. Suggestions: (1) "Tiny Scanner" for Android or (2) "Scanner App" for iOS.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

R Packages:

- Below are pre-loaded general packages required for this module assignment. You can load more packages here or throughout the module if necessary.
- Note that you need to install R packages before you can use them. You can use the `install.packages()` in the R console, or go to the "Tools" tab and click "Install Packages…" in R Studio.
- Be careful on loading R packages because sometimes any two packages can have conflicting functions when calling them.

```
# pre-load packages here
library(tidyverse)
```

I. Monte Carlo Integration

Objectives: To integrate functions using the Monte Carlo method.

The power of random sampling is all over mathematics! In fact, there are classes of computational methods that use random sampling to solve challenging applied mathematics problems such as optimization, parameter inference, and integration. Such methods are typically called **Monte Carlo methods**[1].

Materials

- `monteCarlo1D.R` - This R script provides an example on how to use Monte Carlo integration of a 1D function.
- `monteCarlo2D.R` - This R script provides an example on how to use Monte Carlo integration of a 2D function.

Exercises

1. Use Monte Carlo Integration to integrate:

$$f(x) = \sqrt{\left(1 + \frac{1}{2x}\right)}$$

   from 1 to 3.

   Note: For this example skip the by hand integration and simply supply the plots.

2. Provide both a write-up of your solution by hand, and figures from your Monte Carlo Integration of

$$f(x) = \frac{1}{x^2(x^2+25)}$$

   from 1 to 5. Note: This example is challenging to integrate by hand, but involves partial fractions.

3. Another function of your choosing from the list below. But you must show both your by hand integration and your results from Monte Carlo Integration.

   a. Integrate $f(x) = x^{-1/3} + x/10$ from 0 to 1.
   b. Integrate $f(x) = 4/(1 + x^2)$ from 0 to 1.
   c. Integrate $f(x) = 2x \exp x^2$ from 0 to 1.
   d. Integrate $f(x) = \sin(x)$ from 0 to 1.
   e. Integrate $f(x) = 4\sqrt{(1 - x^2)}$ from 0 to 1.

[1] Monte Carlo methods take their name after a popular gambling destination in Monaco.

II. Maximum Likelihood Estimation

Objectives: To use the method of Maximum Likelihood Estimation to find a set of parameter that maximizes the likelihood function.

Maximum likelihood estimation is a method for figuring out model parameter values. The parameter values are selected to maximize the likelihood that the observed data was created by a defined process in the model. By using maximum likelihood estimation, we look for a set of variables that best matches the distribution of the data sample. The task is approached as an optimization or search problem.

Materials

- `mle.R` - This R script contains example R codes for performing the method of Maximum Likelihood estimation.

Exercises

1. Generate n independent uniformly distributed random numbers $(x_1, x_2, \ldots, x_n)$ on an interval from $[0, a]$ where you specify the value of a. (Pick any value of a you want, but not too crazy! Try the range $5 \leq a \leq 50$.)

   a. Explain why the likelihood of the data is just:
   $$P(x_1, x_2, ...x_n) = \left(\frac{1}{a}\right)^n.$$
   b. For the optimize function in R you need to specify whether to minimize or maximize. What is the lowest values that a can be in order for all the data points you generated to have non-zero probability? Is there an upper bound for what values a can have? (If there's no well-defined upper or lower bound, you can just run the optimize function over different possible ranges and see if you get a different estimate.)
   c. Modify the R code you received to determine the maximum likelihood choice for a from your likelihood. Plot the likelihood and show the maximum likelihood point on the curve. (You should include the figure and your code.)
   d. Increase your value of n, explain what happens to your maximum likelihood estimate. Does it get closer to the true value of a? Does this make sense? (Hint: What is the probability density function of the maximum of n uniformly distributed random variables?)

2. The data file `testDataExp.csv` contains a data set of 50 independent points sampled from an exponential distribution with unknown parameter $\lambda > 0$. Modify the R code you were given to determine the maximum likelihood estimate for λ. Remember that the probability density function of an exponential distribution is given as follows:
   $$f(x) = \lambda e^{-\lambda x}.$$

   Using what you learned in the previous problem.

   a. Explain in mathematics why the following code in R is the likelihood of your data (data) as a function of λ (lambda):

```
like <- function(lambda,data){
  return(lambda^(length(data))*exp(-lambda*sum(data)))
}
```

   b. Determine the maximum likelihood estimate for λ. For full credit you must submit:
      - R code
      - a figure showing the likelihood curve and the maximum.
      - An explanation of your reasoning.
      - Explain using Calculus what the maximum likelihood estimate should be in terms of your original data.)

III. Least Squares Regression

Objectives: To use the method of Least Squares Regression to estimate the parameters of a linear model.

A least-squares regression line, which minimizes the vertical distance between the data points and the regression line, is the line that best fits a set of data if there is a linear association between two variables.

Materials

- `lsr.R` - This R script contains example R codes for generating "synthetic" data and applying the Least Squares Regression method.

Exercises

1. Suppose that in a certain chemical process the reaction time Y (hours) is measured in response to the temperature x (degrees Fahrenheit). The data are assumed to fit a linear model and you will use R to:
   a. Generate "synthetic" data.
   b. Fit a linear model to the data.
   c. Analyze the linear model.
2. In the chamber in which the reaction takes place according to the simple linear regression model:

   $$Y = 5.00 - 0.01x + U$$

   where U is the error in the reaction time and is known to be normally distributed with mean 0 and $\sigma^2 = (0.075)^2$.

   Since in simple linear regression we assume that the independent variable (in this case temperature) is known exactly, we can also write:

   $$Y_x = 5.00 - 0.01x + U.$$

   a. Determine confidence intervals in the estimate of your values for α and β.
   b. Increase the number of points you sample, how does this impact the ability to estimate the correct parameters? Do the confidence intervals change? Give another set of estimates for α and β for your new choice.
   c. Decrease the number of points you sample (less than 30) and increase the variance in the noise U. How does this impact the ability to estimate the correct parameters? Give another set of estimates for α and β for your new choice.