Take Home Final Exam (STAT 5860)

Due by 11:59 pm, Apr. 25

Instructions:

1. Copy your answer to a document for each question and write your R code in the R script file. After you finish, upload both the document and the R script file to the dropbox in Elearning. If you are using R markdown then submit both compiled pdf (or word file) and .Rmd file.

2. Use `set.seed(418)` at the beginning of each problem.

3. Copy the statement below to your answer document and print your name and write the date.

   I certify here that the work on this exam is solely mine. I did not receive any assistance from others and I did not provide any assistance to others.

   PRINT YOUR NAME: DATE:

Problem 1. For the following bivariate density of X and Y,

$$f(x,y) \propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}, \quad x = 0, 1, \ldots, n, \;\; 0 \le y \le 1,$$

suppose we are interested in estimating the mean of the marginal distribution $f(x)$ of X. It can be shown that for fixed α, β, and n, the conditional densities are

$$f(x|y) \quad \text{is} \quad \text{Binomial}(n, y)$$

$$f(y|x) \quad \text{is} \quad \text{Beta}(x+\alpha, n-x+\beta).$$

1. (20 pts) Use the Gibbs sampling to generate a random sample of size 10,000 from the target bivariate density with parameter values $n = 16$, $\alpha = 2$, and $\beta = 4$. You can initialize first sample with (5, 0).
   (Hint: Use rbinom() function to generate sample from Binomial distribution and use rbeta() function to generate sample from Beta distribution.)

2. (5 pts) Now to estimate the mean of the marginal distribution $f(x)$ of X, calculate the sample mean of X from the generated random sample.

3. (5 pts) As you may have already noticed, Gibbs sampling is actually not needed in this example, since $f(x)$ can be obtained analytically from the joint density and it follows the beta-binomial distribution. However, Gibbs sampling becomes indispensable in situation where $f(x)$ cannot be obtained analytically. The theoretical mean of the beta-binomial distribution is $\frac{n\alpha}{\alpha+\beta}$. Compare the theoretical mean to the sample mean you obtained.

Problem 2. In statistics, nonlinear regression is a form of regression analysis in which data are modeled by a function which is a nonlinear combination of the model parameters. To illustrate the fitting of nonlinear regression model, we use the two-parameter nonlinear exponential regression model

$$Y_i = \beta_0 e^{\beta_1 X_i} + \epsilon_i$$

where the ϵ_i are independent normal with constant variance. To estimate the parameters β_0 and β_1, we extend the concept of least squares estimation for linear regression to nonlinear regression models. The least squares criterion here is:

$$Q = \sum_{i=1}^{n} \left(Y_i - \beta_0 e^{\beta_1 X_i} \right)^2$$

The goal for this problem is to find β_0 and β_1 that minimize Q for the given data.

1. (5 pts) A hospital administrator wished to develop a regression model for predicting the degree of long-term recovery after discharge from the hospital for severely injured patients. The predictor variable to be utilized is a number of days of hospitalization (X), and the response variable is a prognostic index for long-term recovery (Y), with large values of the index reflecting a good prognosis. Download the `hospital.txt` file from the course Elearning and load the data set to R. Draw a scatter plot.

2. (15 pts) Based on the scatter plot, we decide to use the two-parameter nonlinear exponential regression model described above. Use the Newton's method we learned in class with initial value $\beta_0 = 50$ and $\beta_1 = 0$ to find the least squares estimators that minimize Q for the given data. Write your own code without using `optim()` function.

3. (10 pts) This time use the `optim()` function with initial value $\beta_0 = 50$ and $\beta_1 = 0$ to find the least squares estimators that minimize Q for the given data.

4. (5 pts) To fit the nonlinear model in R, we use `nls()` function. Compare your parameter estimation results with `nls()` output. Below is the demonstration how to use `nls()` function.

```
> nls_fit <- nls(prog ~ beta0*exp(beta1 * days), data = hospital,
start = list(beta0 = 50, beta1 = 0))
> summary(nls_fit)
```

Problem 3. Solve the following Linear Programming problem.

$$\begin{aligned} \text{Maximize} \quad & 4x + 2y + 9z \\ \text{subject to} \quad & 2x + y + z \leq 2 \\ & -x + y - 3z \geq -3 \\ & x \geq 0, y \geq 0, z \geq 0 \end{aligned}$$

1. (15 pts) Use the `solveLP()` function in `linprog` package to solve the problem.

Problem 4. Download the `jobsubmission.txt` file from the course Elearning and load the data set to R. The data set contains a report on times (in minutes) between job submissions to a computer center. Assume that the times between job submissions follow an exponential distribution. The exponential distribution has following density function

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Here, λ is the rate parameter and the maximum likelihood estimator (MLE) of λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}.$$

1. (10 pts) Use bootstrap to estimate the standard error of the MLE of λ. Write your own code without using `boot()` function. Use 1000 bootstrap samples (or replicates).

2. (10 pts) Find the 95% confidence level percentile bootstrap confidence interval for MLE of λ. Use `quantile()` function with the MLEs you obtained from the previous problem.