ABE6933 SML HW2

Directions

Please submit **ONE PDF** file including all your reports (answer + code + figures + comments; must be easily readable and have file size under a few megabytes) and **ONE R code script**. The R script is supplementary material to ensure that your code runs correctly. If you are using RMarkdown, please also include your `.Rmd` file. If using Python, please submit the Python notebook in lieu of the R script.

Place these two (or three) files in a folder, make a zip or rar archive, and submit the archive electronically via Dropbox file request at `tinyurl.com/nbliznyuk-submit-files` (on the landing page, enter your name so that we know it is you and email so that you get a confirmation).

Please **submit only ONE solution** on behalf of the entire work group, **NOT separate/individual solutions** by different group members. You can have multiple submissions, in which case only the most recent will be graded.

Deadline: 27-Sep-2022, 10:00 PM EST.

Practice/Optional Problems (do not submit)

1. ISLR ch. 3: 5,8,9,13
2. Complete the R lab in Section 3.6 of ISLR.
3. Vector calculus (of several variables); the (linear) least squares problem: solve the problem in the posted pdf, file name "vector.calculus.review.SML.FA16.pdf" [DO NOT SUBMIT: the pdf file contains the solution; please solve on your own without consulting the solution; then check the solution. References in the pdf file like 6.1, 6.2, etc, refer to the subproblems of this problem.]

Required Problems (for submission)

ISLR ch. 3: 4, 10 (except 10-h)

Required Typed Problems

Typed Problem 1.

Let $Y_1, ..., Y_n$ be iid rvs with $E(Y_i) = a$ and $E(Y_i^2) = b$, so that $Var(Y_i) = b - a^2$.

Define $T = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, where $\bar{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ is the sample mean.

1.1. (Optional) Use the properties/calculus of expectations to find $E(T)$. If you are not able to find $E(T)$, you can use use $E(T) = (n-1)Var(Y_i)$ in subsequent subproblems.

1.2. Suppose we estimate the population variance $Var(Y_i)$ by cT for some constant $c > 0$. What value of c results in an unbiased estimator of the population variance? (The answer you should get is $c = 1/(n-1)$.) Let $T_1 = cT$ be this unbiased estimator.

1.3. Let $Y_1, ..., Y_n$ be iid $Normal(\mu, \sigma^2)$, where μ and σ^2 are the population mean and variance, respectively. One can show that $T_2 = T/n$ is the MLE for σ^2; you can take this fact for granted.

Use R to examine the small-sample properties of T_1 and T_2 as follows:

(a) Generate the data as follows:

```
m=1000; n=4; # n is the sample size; m is the # of replications
set.seed(0);
M = matrix(rnorm(m*n),nrow=m); # default parameters in rnorm are mean=0, sd=1;
# M is an m-by-n matrix with replications of the experiment stored in rows
```

(b) For each row of M, evaluate and store values of T_1 and T_2, in separate vectors. (Optional): you can do this without loops using apply() function

(c) Plot histograms of T_1 and T_2.

(d) "Monte Carlo integration" is estimation of population moments of a rv X by the corresponding sample moments whenever one can simulate iid variates $X_1, X_2, \ldots$ from the sampling distribution of X. I.e., using the law of large numbers (and another result known as the continuous mapping theorem) $\bar{X}_n \to E(X)$ and $S_n^2 \to Var(X)$ as $n \to \infty$, where $\bar{X}_n$ and S_n^2 are the sample mean and the sample variance, respectively. Use "Monte Carlo integration" to estimate bias, variance and MSE of the two estimators. Specifically, you can estimate $E(T_1)$ and $E(T_2)$ using the respective sample means, and (population) variances of T_1 and T_2 using the sample variances of T_1 and T_2.

Briefly discuss your findings in (c) and (d).

1.4. Suppose we are now interested in the population standard deviation, i.e., $\sigma = \sqrt{\sigma^2}$. Explain/argue whether $\sqrt{T_1}$ is unbiased for estimation of σ, and why. Feel free to extend the simulation study in 1.3 to reinforce your answer.