Homework Set #5 – Due 11:59pm, Tuesday, February 25.
MATH 445 WINTER 2025

Note:

- This homework assignment is based on the Chapter 7 material.
- You need to submit two files (a PDF file containing your answers and an R script file containing your R code).
- Your answers must be presented in the same order as the problems given in this assignment, including all graphs and `R` outputs, if applicable (i.e., the graphs and outputs must be in the same order and must not be relegated to the back of your assignment).
- **Do not** present your answer in the R script file only. Your answers must be presented in the PDF file.
- Use your judgment to include the minimal amount of `R` code in your answers as necessary.
- All the `R` code must be well documented and submitted as a separate R script file (a file with the ".R" extension), even if you have already presented it in the PDF file.

The grading scheme is as follows:

- This homework assignment will be graded out of 30 points.
- Three out of the four problems will be checked carefully. Each of the three problems will be graded out of 10 points.
- For the problem that will not be graded, two points will be automatically deducted for each missing part.
- If the R code is missing or incomplete, at most 5 points will be deducted.
- If the answers are not presented in the right order, at most 5 points will be deducted.

1. Power calculations: Sometimes you cannot obtain large enough observations in your sample because of the cost, time, or both. This situation is very typical in biology and psychology. For example, it might be too expensive or time consuming to find patients with certain rare medical conditions. Let us suppose that we have $m = 10$ observations available in your sample. Even if you have such a small sample, as long as the true value of σ^2 is far enough from the value in the null hypothesis, the test has enough power to reject the null hypothesis.

   (a) By referring to HW3 #5 or HW4 # 1, calculate the power of the test at $\alpha = 0.05$ by varying the true value of $\sigma^2 = r$ of the Gamma$(r, 1)$ distribution, assuming that $m = 10$, $H_0\colon \sigma_0^2 = 1$, and $H_1\colon \sigma_0^2 > 1$. Specifically, calculate the power of the test for $\sigma^2 = 1, \ldots, 6$ with an increment of 0.5 using $B = 10,000$.

   (b) Plot the results using the `plot()` function with the σ^2 values on the x-axis and the power of the test on the y-axis. Then, perform a linear interpolation using the `lines()` function. After that, add a horizontal line in the same plot at $y = 0.8$.

   (c) By examining (b), estimate (by eye) the value of σ^2 that achieves the power of the test $= 0.8$.

2. Recall that the cdf of Exp(θ) is given by $F(x) = 1 - e^{-x/\theta}$, $x > 0$.

   (a) Calculate the first quartile Q_1 (the number which satisfies $F(Q_1) = 0.25$) of Exp(θ). Note that Q_1 is the same as the 25-th percentile.

   (b) Based on (a), suggest a reasonable estimator of θ, denoted by $\hat{\theta}_{0.25}$, for Exp(θ) using the sample first quartile ($\hat{Q}_1$) calculated from a random sample of size m.

   (c) Describe how to estimate Var$(\hat{\theta}_{0.25})$ using B samples of Exp(θ) of size m.

   (d) Generate $B = 10,000$ Exp(θ) samples with $\theta = 3$ and $m = 10$, and report Var$(\hat{\theta}_{0.25})$ in your answer.

   (e) Note the theoretical value (using the formula for the large-sample asymptotic variance) of the $100p$-th percentile is given by $[p(1-p)]/[m\{f(x_p)\}^2]$, where x_p is the $100p$-th percentile. Noting that $f(x) = \theta^{-1}e^{-x/\theta}$, $x > 0$ is the pdf of Exp(θ), compute the theoretical value of Var$(\hat{\theta}_{0.25})$.

   (f) Compare the values you obtained in (d) and (e), and comment. Especially, if you see any noticeable differences between these two values, explain why.

   (g) For $X_i \sim$ Exp(θ), because $E[X_i] = \theta$, the sample mean $\bar{X}$ is another reasonable estimator of θ. Let $\hat{\theta}_{\text{mean}} = \bar{X}$. Calculate the theoretical value of bias$(\hat{\theta}_{\text{mean}})$ and Var$(\hat{\theta}_{\text{mean}})$ with $\theta = 3$ and $m = 10$.

(h) Based on (d) and (g), decide which parameter estimator is more efficient.

3. The distance (in kilometers) from an earthquake event to the nearest populated place can be found in `earthquakes.R`.

   Assuming that the these distances are exponentially distributed, compute 95% confidence intervals for rate $\lambda = 1/\theta$ using nonparametric percentile bootstrap, semi-parametric percentile bootstrap (using $\mathrm{Exp}(\bar{x})$ for data generation), and studentized bootstrap (bootstrap-t). Because the parameter θ is estimated by $\bar{x}$ (the sample mean), the rate parameter λ is estimated by $1/\bar{x}$.

4. Let $\{x_1, x_2, \ldots, x_m\}$ be a random sample. Let $\theta = \mu$ (the population mean) be the parameter of interest. It is estimated by the sample mean, i.e., $\hat{\theta} = \bar{x}$.

   (a) Show that the jackknife estimated bias of $\hat{\theta}$ is zero.

   (b) Show that the jackknife estimated variance of $\hat{\theta}$ is s^2/m, where

$$s^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2.$$