Statistical Computing: Homework 3

Due on April 3 (Monday) 23:30

1. In statistics, Jarque-Bera (JB) test is popularly used to check the normality of the underlying distribution for a given data set $\{X_i : i = 1, ..., n\}$. The JB test statistic is defined as

$$JB = \frac{n}{6}\left(\text{skewness}^2 + \frac{1}{4}(\text{kurtosis} - 3)^2\right),$$

where

$$\text{skewness} = \frac{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^3}{s^3},$$

$$\text{kurtosis} = \frac{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^4}{s^4},$$

in which $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$.

Under normality, JB test statistic has an asymptotic χ^2_ν distribution with df $\nu = 2$. Therefore, at the significant level α, we reject normality if $JB > \chi^2_2(1-\alpha)$. ($\chi^2_2(\cdot)$ is the quantile function)

Use the Monte Carlo method to assess the power function of the JB test (for testing normality H_0) against the following two specific alternatives (H_1):

- Alternative Model I: skew-normal distribution (hw1 Problem #4) with the skew parameter θ:

$$f(x;\theta) = 2\,\phi(x)\,\Phi(\theta x), \quad x \in R, \quad \theta > 0.$$

- Alternative Model II: t(ν) with df $\nu = 1/\theta,\ 0 < \theta \leq 1$.

The values of θ, in either case, characterize the level of the distribution deviated from normality. In particular, the model with $\theta = 0$ reduces to the standard normal distribution (i.e., H_0) for both alternative cases.

Goal: Sketch the power function $\pi(\theta)$ at a grid $\mathcal{G}$ on $\theta \in [0, 1]$ (having at least 5 grid points) for Alternative Model I or II (at your own choice). In particular $\pi(0)$ corresponds to the type I error for testing normality (H_0). Since the testing power also depends on the sample size n used to perform the JB test, you need to evaluate the power function for the following two scenarios:

(a) data with size $n = 20$ (small sample scenario);

(b) data with size $n = 100$ (large data scenario).

Provide the Monte Carlo error for approximating $\pi(\theta)$ (use large enough Monte Carlo replicates m to ensure that the MC s.e. of $\hat{\pi}(\theta_i)$ is less than 0.03).

Your simulation may take the following flow:

- Set the sample size n; set the number of MC replicates m; and fixed the type I error at $\alpha = 0.05$.
- For each $\theta_i \in \mathcal{G}$,

  (1) do $j = 1 : m$,

    (a) generate iid random variable from the alternative model with the given θ_i;

    (b) compute the JB test statistic and obtain the testing result (reject/not-reject);

  end do

  (2) compute the rejection rate among m Monte Carlo replicates as an estimate $\hat{\pi}(\theta_i)$ for $\pi(\theta_i)$
- Plot $\hat{\pi}(\theta_i)$ v.s. θ_i (line plot) to sketch the power function $\pi(\theta)$ on $\theta \in [0, 1]$.