STP598: Computational Statistics

Yiannis Kamarianakis

E-mail: yiannis76@asu.edu; Office: WXLR 531

October 4, 2017

Assignment 3 Part 1 (20 points)

- Conduct a Monte Carlo study to estimate the coverage probabilities of the 4 bootstrap confidence intervals that were presented in class. Sample from a normal population and check the empirical coverage rates for the sample mean.
- Find the proportion of times that the intervals miss on the left and the proportion of times that the confidence intervals miss on the right.

Assignment 3 Part 2 (20 points)

- The $t-test$ is robust to mild departures from normality. Use Monte Carlo simulation to investigate whether the empirical $Type-I$ error rate of the t-test is approximately equal to the nominal significance level α, when the sampled population is non-normal.
- $\alpha = 0.99$ if your ASUID is an odd number and $\alpha = 0.90$ if your ASUID is an even number.
- Discuss the simulation results for the cases where the sampled population is (i) $\chi^2(1)$, (ii) $Uniform(0, 2)$, and (iii) Exponential(rate=1).
- In each case, test $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$

Assignment 3 Part 2 (20 points)

- The $t-test$ is robust to mild departures from normality. Use Monte Carlo simulation to investigate whether the empirical $Type-I$ error rate of the t-test is approximately equal to the nominal significance level α, when the sampled population is non-normal.
- $\alpha = 0.99$ if your ASUID is an odd number and $\alpha = 0.90$ if your ASUID is an even number.
- Discuss the simulation results for the cases where the sampled population is (i) $\chi^2(1)$, (ii) $Uniform(0, 2)$, and (iii) Exponential(rate=1).
- In each case, test $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$

Assignment 3 Part 3 (20 points)

- Tests for association based on Pearson product moment correlation and Spearman's rank correlation coefficient are implemented in *cor.test*. Show empirically that the nonparametric Spearman's test is less powerful when the sampled distribution is bivariate normal.
- Find an example of an alternative (a bivariate distribution (X, Y) such that X and Y are dependent) such that the nonparametric test has better empirical power against this alternative.

Assignment 3 Part 4 (10 points)

- Compute a Monte Carlo estimate $\hat{\theta}$ of

$$\theta = \int_0^{0.5} e^{-x} dx \tag{1}$$

by sampling from $Uniform(0, 0.5)$, and estimate the variance of $\hat{\theta}$.

Assignment 3 Part 5 (10 points)

- Generate a random sample of size 1000 from a normal location mixture. The components of the mixture have $N(0, 1)$ and $N(0, 3)$ distributions with mixing probabilities p_1 and $p_2 = 1 - p_1$. Graph the histogram of the sample with density superimposed, for $P_1 = 0.75$.
- Repeat with different values of p_1 and observe whether the empirical distribution of the mixture appears to be bimodal. Make a conjecture about the values of p_1 that produce bimodal mixtures.

Assignment 3 Part 6 (10 points)

- Simulate from a continuous Exponential-Gamma mixture. Suppose that the rate parameter Λ has Gamma(r, β) distribution and Y has Exp(Λ) distribution. Generate 1000 random observations from this mixture with $r = 4$ and $\beta = 2$ and plot the histogram.

Assignment 3 Part 7 (10 points)

- Generate 500 random observation from the 3-dimensional multivariate normal distribution having mean vector $\mu = (0, 1, 2)$ and covariance matrix $\Sigma = \begin{bmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & -0.5 \\ 0.5 & -0.5 & 1 \end{bmatrix}$ using the Choleski factorization method. Use the *R pairs* plot to graph an array of scatter plots for eaach pair of variables. For each pair of variables, check visually that the location and correlation approximately agree with the theoretical parameters of the corresponding bivariate normal distribution.

Assignment 3

- E-mail your responses in a **single** pdf file by Wednesday Oct. 25 midnight.
- Use the following file name: LASTNAME_FIRSTNAME_ASUID_ASSIGNMENTNUMBER
- Prepare your pdfs carefully; each week some of you will present their work.
- Include the script containing the `R` commands you used; use the same file name as before.