STAT 5701 Homework 2 – Fall 2015

This homework is due on Wednesday, October 21 at 11:59pm. Submit your solutions in a pdf document on Moodle. Include your R code (which must be commented and properly indented) in the pdf file. Please name the pdf file `<your last name>-HW1.pdf`. Please also submit one text file with your R code, which must be commented and properly indented.

1. Suppose we can see a realization of $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$. A $100(1-\alpha)\%$ random confidence interval for $\sigma^2$ is
$$\left( \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \right), \tag{1}$$
where $\chi^2_{u,\nu}$ is the $100{*}u$th percentile of the Chi-squared distribution with $\nu$ degrees of freedom, $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$, and $\bar{X} = n^{-1}\sum_{i=1}^n X_i$.

   (a) Perform a simulation study with 10000 replications to make a 99% *score* approximate confidence interval for the coverage probability of the random confidence interval for $\sigma^2$ defined in (1) when $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$. Set $\mu = 68, \sigma = 3$ and perform the study for $n = 10$ and $\alpha = 0.05$. Comment on the agreement between $1 - \alpha$ and the corresponding 99% *score* approximate confidence interval for the coverage probability.

   (b) Perform a simulation study with 10000 replications to make a 99% *score* approximate confidence interval for the coverage probability of the random confidence interval for $\sigma^2$ defined in (1) when $X_1, \ldots, X_n$ are iid Exp(1). Perform the study for each $(n, \alpha) \in \{10, 50, 500\} \times \{0.01, 0.05\}$. Comment on the agreement between $1 - \alpha$ and the corresponding 99% *score* approximate confidence interval for the coverage probability.

   (c) Perform a simulation study with 10000 replications to make a 99% *score* approximate confidence interval for the coverage probability of the random confidence interval for $\sigma^2$ defined in (1) when $(X_1, \ldots, X_n)'$ have the $n$-variate Normal distribution with mean vector $\mu 1_n$ and covariance matrix $\Sigma \in \mathbb{S}^n_0$. The $(i,j)$th entry of $\Sigma$ is $\Sigma_{ij} = \sigma^2 \rho^{|i-j|}$. Set $\mu = 68, \sigma = 3, \rho = 0.7$ and perform the study for each $(n, \alpha) \in \{10, 50, 500\} \times \{0.01, 0.05\}$. Comment on the agreement between $1-\alpha$ and the corresponding 99% *score* approximate confidence interval for the coverage probability.

2. In the two independent samples t-test model, we have access to a realization of $X_1, \ldots, X_{n_1}$ iid $N(\mu_1, \sigma^2)$ and access to a realization of the independent sequence $Y_1, \ldots, Y_{n_2}$ iid $N(\mu_2, \sigma^2)$. Let $\bar{X} = n_1^{-1}\sum_{i=1}^{n_1} X_i$, $\bar{Y} = n_2^{-1}\sum_{i=1}^{n_2} Y_i$, $S_1^2 = (n_1 - 1)^{-1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$, and $S_2^2 = (n_2 - 1)^{-1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$. To test
$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$
we use the random test statistic
$$T = \frac{\bar{X} - \bar{Y} - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$
where
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

When $H_0$ is true, $T \sim \text{Tdist}(n_1 + n_2 - 2)$. The random $p$-value is

$$2\texttt{pt}(-|T|, n_1 + n_2 - 2),$$

where $\texttt{pt}(\cdot, \nu)$ is the cdf of the t-distribution with $\nu$ degrees of freedom.

(a) Perform a simulation study to estimate the distribution of the random $p$-value when $n_1 = n_2 = 25$, $\mu_1 = \mu_2 = 68$, and $\sigma = 3$. What is the actual distribution of the random $p$-value in this case? Explain.

(b) Fix $n_1 = n_2 = 50$, $\mu_1 = 68$, $\sigma = 3$, and significance level $\alpha = 0.05$. Produce a plot of a simulated estimate of the power curve of this test formed by increasing $\mu_2$ (from 68). Select the sequence of values for $\mu_2$ so that the simulated estimate of the power increases from roughly 5% to roughly 99%.

(c) Fix $\mu_1 = 68$, $\mu_2 = 68.5$, $\sigma = 3$, and significance level $\alpha = 0.05$. Use simulation to find a value of $n_1 = n_2$ such that the simulated estimate of the power is roughly 95%.

(d) Suppose that we perform this test with $X_1, \ldots, X_{n_1}$ iid $N(\mu_1, \sigma_1^2)$ independent of $Y_1, \ldots, Y_{n_2}$ iid $N(\mu_2, \sigma_2^2)$. We will see what happens when $\sigma_1 \neq \sigma_2$. Perform a simulation study to estimate the distribution of the random $p$-value in the following cases:

   i. $n_1 = n_2 = 100$, $\mu_1 = \mu_2 = 68$, $\sigma_1 = 3$ and $\sigma_2 = 6$
   ii. $n_1 = 20$, $n_2 = 100$, $\mu_1 = \mu_2 = 68$, $\sigma_1 = 3$ and $\sigma_2 = 6$
   iii. $n_1 = 100$, $n_2 = 20$, $\mu_1 = \mu_2 = 68$, $\sigma_1 = 3$ and $\sigma_2 = 6$.

   Since $H_0$ is true in all cases, we hope that the random $p$-value is approximately Unif$(0, 1)$. Create a QQ-plot for each case that compares the data percentiles (of the realizations of the random $p$-value) to the corresponding percentiles of Unif$(0, 1)$. Comment on the agreement. Is the 0.01 data quantile approximately equal to 0.01? Is the 0.05 data quantile approximately equal to 0.05?

(e) Suppose that $n_1 = n_2$, and $(X_1, Y_1), \ldots, (X_{n_1}, Y_{n_1})$ are iid $N_2((\mu_1, \mu_2)', \Sigma)$, where $\Sigma \in \mathbb{S}_0^2$. The two-independent samples t-test requires that $\Sigma = \sigma^2 I$ so we will see what happens when $\Sigma \neq \sigma^2 I$. In particular, suppose that $\Sigma$ has $(i, j)$th entry $\Sigma_{ij} = \sigma^2\{1(i = j) + \rho 1(i \neq j)\}$. Perform a simulation study to estimate the distribution of the random $p$-value in the following cases:

   i. $n_1 = 20$, $\mu_1 = \mu_2 = 68$, $\sigma = 3$ and $\rho = 0.01$.
   ii. $n_1 = 20$, $\mu_1 = \mu_2 = 68$, $\sigma = 3$ and $\rho = 0.1$.

   Since $H_0$ is true in both cases, we hope that the random $p$-value is approximately Unif$(0, 1)$. Create a QQ-plot for each case that compares the data percentiles (of the realizations of the random $p$-value) to the corresponding percentiles of Unif$(0, 1)$. Comment on the agreement. Is the 0.01 data quantile approximately equal to 0.01? Is the 0.05 data quantile approximately equal to 0.05?

3. Suppose that we can see a realization of $X_1, \ldots, X_n$ iid Exp$(\mu)$ where $\mu$ is the unknown distribution mean. The density for Exp$(\mu)$ evaluated at $x$ is

$$f(x; \mu) = \frac{1}{\mu} e^{-x/\mu} 1(x > 0).$$

(a) Show that $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ is the maximum likelihood estimator of $\mu$.

(b) Consider the competing *shrinkage* estimator $\hat{a}\bar{X}$, where

$$\hat{a} = \underset{a \in \mathbb{R}_+}{\arg\min}\, E\{(a\bar{X} - \mu)^2\}$$

and $\mathbb{R}_+ = \{b \in \mathbb{R} : b > 0\}$. Derive a closed-form expression for $\hat{a}$. Is the estimator $\hat{a}\bar{X}$ unbiased for $\mu$?

(c) Report closed-form expressions for $E\{(\bar{X} - \mu)^2\}$ and $E\{(\hat{a}\bar{X} - \mu)^2\}$. Which estimator has the smaller mean-squared error?

(d) For each $(n, \mu) \in \{5, 10, 50\} \times \{0.5, 1, 10\}$, perform a simulation study that compares the performance of $\bar{X}$ and $\hat{a}\bar{X}$ as estimators of $\mu$: report simulated estimates of $E\{(\bar{X} - \mu)^2\}$ and $E\{(\hat{a}\bar{X} - \mu)^2\}$. Compare these simulated estimates to their formulas derived in part 3c.