**Problem Set 2** (two pages)

Statistics 24510-30040 (W19)

Due Tuesday, January 22 at the beginning of class.

Requirements Provide detailed derivations. Select only the relevant part of the output to be inserted. Attach your code or output as an appendix if necessary. Discussions allowed, the assignment should be devised and written by yourself completely.

## Problem assignments (Related to 2.3, 3.6, and various parts in Rice on asymptotic confidence intervals)

1. (*Change of variables*, *bivariate application*)

   Suppose $X$ and $Y$ are independent continuous random variables.

   (a) Derive the bivariate density of $S = X + Y$ and $R = \frac{X}{X+Y}$.

   (b) Furthermore, suppose $X$ and $Y$ are independent exponential random variables with density $f(x) = \lambda e^{-\lambda x}, \ x \geq 0$.

   i. Find the bivariate density of $S = X + Y$ and $R = \frac{X}{X+Y}$, based on part (a).

   ii. From the formula in part (i) above, find the density of $S = X + Y$. What is the name (and parameters) of the density?

   iii. From the formula in part (i), find the density of $R = \frac{X}{X+Y}$. What is the name (and parameters) of the density?

   (c) Furthermore, suppose $X \sim Gamma(\alpha_1, \lambda)$ and $Y \sim Gamma(\alpha_2, \lambda)$ are independent Gamma random variables with density $f(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha), \ x \geq 0$.

   i. Find the bivariate density of $S = X + Y$ and $R = \frac{X}{X+Y}$, based on part (a).

   ii. From the formula in part (i) above, find the density of $S = X + Y$. What is the name (and parameters) of the density?

   iii. From the formula in part (i), find the density of $R = \frac{X}{X+Y}$. What is the name (and parameters) of the density?

2. (*Binomial confidence intervals*) $X \sim binom(n, p)$, $\hat{p} = X/n$ is the MLE of $p$.
   Here let $z_{\alpha/2}$ be the right quantile of standard normal satisfying $P(N(0,1) > z_{\alpha/2}) = \alpha/2$.

   (a) Use the *arcsine* variance stabilizing transformation $g(p) = 2 \arcsin \sqrt{p}$ to derive a $(1-\alpha)100\%$ confidence interval for $p$.

   (b) The endpoints of 95% Wald asymptotic confidence interval (with plug-in $\hat{p}$ for $p$) has the form $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$. Find the smallest integer $n_o$ such that for any sample size $n \geq n_o$, The length of the confidence interval is $\leq 0.5$, for any value of $\hat{p} \in [0, 1]$.

3. (*R simulations of binomial asymptotic confidence intervals*)

   (a) Generate $k = 100$ draws from $Binom(n, p)$ for $n = 30, p = 0.1$ (using the `rbinom` command).
   For each draw, obtaining MLE $\hat{p}$, using test level $\alpha = 0.05$ to compute

   - Wald confidence interval (with endpoints $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ )
   - Wilson confidence interval (with endpoints $\left(z_{\alpha/2}^2 + 2n\hat{p} \pm z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4n\hat{p}(1-\hat{p})}\ \right)/2(n + z_{\alpha/2}^2)$ )
   - The arcsine confidence interval (R command for arcsine function is `asin`) you just derived.

   In your simulation,

i. What is the (theoretically) expected proportion of the $k$ confidence intervals that should contain the true parameter $p = 0.1$?

ii. For each type of confidence interval, find the proportion of the $k$ confidence intervals actually contain the true parameter $p = 0.1$.

iii. Compare and comment on the goodness of each of the three confidence intervals.

(b) Repeat the simulation in part (a) with increased sample size $n = 150$.
Do any of your previous conclusions for $n = 30$ change? Why?

4. (*Bivariate normal exercises*)

An early criminologist believed he could summarize human characteristics for purposes of identification from a few simple measurements. Let $X$ = height, $Y$ = length of left arm, $W$ = width of skull, and $Z$ = length of ear. Measurements are in centimeters. Suppose $(X, Y)$ and $(W, Z)$ are bivariate normal vectors,

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} 140 \\ 70 \end{bmatrix}, \begin{bmatrix} 100 & 25 \\ 25 & 25 \end{bmatrix} \right), \qquad \begin{bmatrix} W \\ Z \end{bmatrix} \sim N\left( \begin{bmatrix} 15 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \right)$$

Suppose that $(X, Y)$ and $(W, Z)$ are independent. The criminologist proposed two summary measures,

$$M_1 = X + Y + W + Z, \qquad M_2 = X - 2Y + W - 3Z$$

(a) Find $E(M_1), Var(M_1)$.

(b) Find the bivariate normal distribution of $(M_1, M_2)$.

(c) Find $P(M_1 \leq 250)$.

5. (*Correlation of bivariate normal*)

Suppose $(X, Y)$ has a bivariate normal distribution with expectations $(3, 1)$, variance $(9, 16)$, and correlation $\rho$. Let $W_a = 12 + aX + Y$ and $V = 19 + X + 2Y$.

(a) For $\rho = 1/3$, find $a \in \mathbb{R}$ such that $W_a$ and $V$ are independent.

(b) Is there $\rho^* \in (-1, 1)$ such that no $a \in \mathbb{R}$ can make $W_a$ and $V$ independent?
If yes, find all such $\rho^*$. If no, explain why not.

(c) Fix $a = 1$. Find $\rho$ such that $W_a$ and $V$ are independent.

(d) Is there $a^* \in \mathbb{R}$ such that no $\rho \in (-1, 1)$ can make $W_{a^*}$ and $V$ independent?
If yes, find all such $a^*$. If no, explain why not.