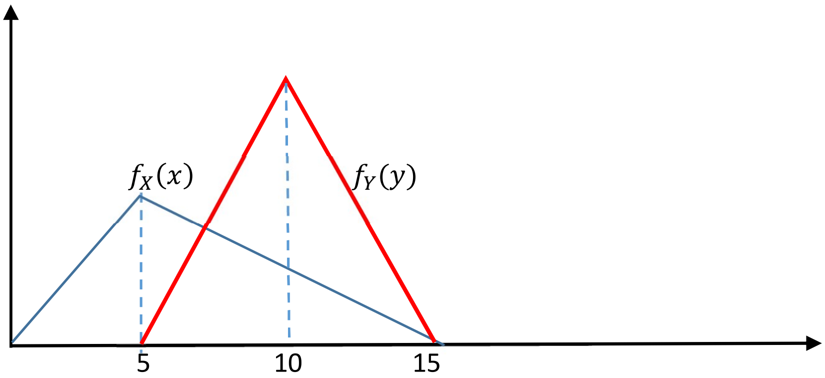
Statistics - Homework 3 (Due April 12, 2019)

1. The probability density functions of two random variables X and Y are shown in the following figure.



   (1) Derive and plot the cumulative distribution functions of X and Y, respectively.
   (2) $P(X > Y)$=?
   Hint: $P(X > Y) = \int P(X > y|y) f_Y(y)dy$
   (3) Simulate 100,000 random numbers of X and Y, respectively.

   $\{x_i, i = 1,2, \ldots, 100{,}000\}; \{y_i, i = 1,2, \ldots, 100{,}000\}$

   (4) Plot the ***relative*** histograms of the above samples.
   (5) Calculate the number of pairs $\{(x_i, y_i), i = 1,2, \ldots, 100{,}000\}$ that satisfy $x_i > y_i$.

2. Let X be a random variable with a continuous uniform density between -1 and 1, i.e., $X \sim U(-1, 1)$. Random variable Y is defined by the following transformation:

   $$Y = \frac{2}{\pi}\sqrt{1 - X^2} .$$

   (1) $Var(Y) = ?$
   (2) $P\left(Y \le \frac{1}{\pi}\right) = ?$

3. Let X be an exponential random variable with the following probability density function:

   $$f_X(x) = \lambda e^{-\lambda x}, \ \lambda > 0, \ 0 \le x < +\infty.$$

   (1) Find the coefficient of skewness and coefficient of kurtosis of X.
   (2) Generate 1000 random samples, each of size 20 ($ns = 20$), of an exponential random variable with $\lambda = 0.4$. Find the sample coefficient of skewness and sample coefficient of kurtosis of individual random samples using R. [Note: You will need to install the R package "moments" and then use its commands *skewness* and *kurtosis* for calculation of sample coefficient of skewness and sample coefficient of kurtosis.]
   (3) Repeat (2) for sample sizes of 40, 60, 80, 100, 200, 400, 600, 800 and 1000.
   (4) For each of the above sample sizes, construct the scatter plot of the sample coefficient of

skewness and sample coefficient of kurtosis. On the same plot, also show the theoretical coefficient of skewness and coefficient of kurtosis.

(5) For sample size 100, plot histograms of the sample coefficient of skewness and sample coefficient of kurtosis, respectively.

(6) For each specific sample size, calculate the mean values of the sample coefficient of skewness and sample coefficient of kurtosis, respectively.

4. Random variable X is normally distributed with an expected value of 100 and standard deviation of 30. Conduct the following simulation using R:

(1) Let V_c be the 0.975 quantile of X. Find the value of V_c.

(2) Generate 10,000 random samples of X with sample sizes ns = 20, 30, …, 90, 100, respectively.

(3) For a given random sample $(x_1, x_2, \cdots, x_{ns})$ with sample size ns, let M represent the number of *higher-than-*V_c random numbers in the random sample. Calculate the value of M for each individual random sample. [Note: The value of M will vary with sample size and random samples.]

(4) For each particular sample size (ns = 20, 30, …, 90, 100), calculate the proportion ($\hat{p}$) of the 10,000 random samples having M values equal to 0, 1, 2, 3, 4, and 5, respectively. List the values of $\hat{p}$ with respect to M and ns and plot the scatter plot of ($\hat{p}$, M) with respect to different sample sizes. [Note: the proportion is dependent on the sample size and the value of M.]

(5) Based on the Poisson distribution, calculate the probabilities ($p_{Poisson}$) that a random sample of X of size ns having M value equals to 0, 1, 2, 3, 4, and 5, respectively. List the values of $p_{Poisson}$ with respect to M and ns and plot the scatter plot of ($p_{Poisson}$, M) with respect to different sample sizes.

(6) Based on the binomial distribution, calculate the probabilities (p_{Binom}) that a random sample of X of size ns having M value equals to 0, 1, 2, 3, 4, and 5, respectively. List the values of p_{Binom} with respect to M and ns and plot the scatter plot of (p_{Binom} , M) with respect to different sample sizes.