Problem Set #5 Fall 2017
Due Wednesday, October 18, 2017.

1. A certain market has both an express checkout line and a regular checkout line. Let $X$ denote the number of customers in line at the express checkout at a particular time of day, and $Y$ be the number of customers in line at the regular checkout at the same time. The joint pmf of $X$ and $Y$ is given by

\[
p(x,y) = \begin{array}{ccc}
0 & 1 & 2 \\
0 & 0.10 & 0.04 & 0.02 \\
1 & 0.08 & 0.20 & 0.06 \\
2 & 0.06 & 0.14 & 0.30 \\
\end{array}
\]

(a) Find the distribution, expected value, standard deviation of $X$?
(b) What is the probability $P(X+Y \leq 2)$ and the conditional probability $P(X+Y \leq 2|X \leq 1)$?
(c) Find $P(X \leq 1, Y \leq 1)$? Are $X$ and $Y$ independent?
(d) Find the covariance of $X$ and $Y$. Are $X$ and $Y$ independent?

2. Caroline and Sarah have agreed to meet between 5:30pm and 6:30pm for dinner at Witherspoon street. Let $X$ be Caroline’s arrival time (in the unit of minutes) and $Y$ be Sarah’s arrival time (in the unit of minutes), relative to 6:00pm. Assume that their arrival times are independent.

(a) What is the joint pdf of $X$ and $Y$, if $X$ and $Y$ are normally distributed as $N(0,10^2)$.
   If they can wait each other for 10 minutes, what is the probability that they actually meet? What is $corr(X+Y,Y)$?
(b) What is the joint pdf of $X$ and $Y$, if $X$ and $Y$ are uniformly distributed on $[-30,30]$.
   If they can wait each other for 10 minutes, what is the probability that they actually meet? Hint: A good starting point is to draw a picture.

3. Suppose that a population has three genotypes with $P(aa) = 0.16$, $P(aA) = 0.48$ and $P(AA) = 0.36$. Sample $n$ people at random with replacement. Let $X$ be the number of people with genotype aa and $Y$ be the genotype with number of people with genotype aA.

(a) What are the marginal distributions of $X$ and $Y$?
(b) What is the probability that $P(X = 5, Y = 3)$, if $n = 10$?
(c) What is the covariance between $X$ and $Y$, if $n = 1$?
4. Suppose that the SAT math scores follow the normal curve very well with mean 514 and standard deviation 117 (based on 1.7 million seniors who took the SAT at any time during their high school years through June 2012).

(a) Find, approximately, the score obtained by someone who ranked 10th percentile from the top on the SAT math test.

(b) How many percent of them scored above 650?

(c) Pick 100 person at random, what is the probability that exactly 15% of them scored above 650? (Use the normal approximation with continuity correction)

(d) Suppose that the mean and standard of the SAT scores are unknown. But, John Doe and Jane Doe took the test. John got 640 and his percentile ranking was 84% and Jane got 750 with percentile ranking 95%. Approximately, what are the average and the standard deviation of the SAT math scores.

5. Simulate 40000 random numbers from Gamma distribution with \( \alpha = 3 \) and \( \lambda = 1/30 \) [in R, use \( \text{rgamma}(40000, 3, 1/30) \)]. This is regarded as the adjusted gross income (AGI, in thousand dollars) of Princeton residents. Fix this data set.

(a) Report the histogram, mean and standard deviation of the Princeton AGIs. How many percents of families have AGI at least $100K?

(b) Next we use simulations to check the square-root law and the central limit theorem.

   i. Compute the standard deviation \( s_n \) for these 1000 sample averages. Plot \( (n, \sqrt{n}s_n) \) for \( n = 16, 25, 100, 400 \). Is it about a constant?

   ii. Draw the histograms for these 1000 sample averages and check their normality distribution using QQ-plot for 100 and 400.

   iii. Compute the 1000 sample percentages with family AGI at least $100K. Draw the histograms of these percentages and check the normality using qqnorm for \( n = 100 \) and 400.

Try to avoid using loops if you can. Matrix operations in R are faster and take less computer memories. See Problem 6(c) in Homework #4.