18.05 Problem Set 4, Spring 2018

(due outside 4-174 Monday, Mar. 5 at 9:30 AM)

Problem 1. (10 pts.) **Joint distributions.** Suppose X and Y have joint pdf $f(x,y) = c(xy^2 + x^2y)$ on $[0,1] \times [0,1]$.

(a) Find c and the joint cdf $F(x,y)$.

(b) Find the marginal cdf F_X and F_Y and the marginal pdf f_X and f_Y.

(c) Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

(d) Find the covariance and correlation of X and Y.

(e) Are X and Y independent?

Problem 2. (10 pts.) **Independence.** Suppose X and Y are random variables with the following joint probability mass function.

$X\backslash Y$	1	2	3
1	1/15	3/15	1/15
2	1/10	3/10	1/10
3	1/30	1/30	1/10

(a) How do you know this is a valid probability mass function?

(b) Are X and Y independent?

Problem 3. (20 pts.) **Group work.** Ruthi, Efrat, Ani and Elan all have 18.05 exams to grade. Some exams are easier to grade than others, so the length of time it will take each to grade their set is random. Suppose that for each of them time in hours follows an exponential(1/3) distribution. Suppose also that the times for each of them are independent.

(a) Compute the mean and standard deviation of the time it takes one TA to finish grading.

(b) What is the probability it will take Ruthi more than 3 hours to finish?

(c) What is the probability it will take Ruthi and Efrat both more than 3 hours?

(d) Let T be the time it takes the fastest grader to finish. What is the pdf for T?

(Hint: It is not a hard computation to find $P(T > x)$.)

(e) Simulate this setting in R. Use your simulation to give estimates for the questions in parts **(a)** and **(b)**.

(Give your estimates and include a printed or handwritten version of your code.)

(f) Use your simulation to generate a density histogram for T. Use 10000 samples and a bin width of 0.1. Superimpose on it a plot of the pdf of T from part **(d)**.

Hint: The `pmin` *function in R is useful here.*

(Print out your histogram and include a printed or handwritten version of your code.)

(g) The random variable T is the minimum time it takes all the graders to finish. You should have found that it follows an exponential distribution. What about the maximum time it takes to finish? Using either an exact calculation or by simulation in R say whether or not the maximum is plausibly an exponential distribution.

As usual, if you use R turn in your code and a printout of your plots.

Problem 4. (10 pts.) **Opinion Polling.** In naming a theoretical dorm in the location of old Bexley Hall, suppose that 60% of MIT undergrads want to continue calling it Bexley, 20% support naming it Epsilon Eridani, and the rest are split among Kobol, Bajor, and Kashyyyk. A poll asks 400 random people which name they support.

(a) Use the central limit theorem to estimate the probability that less than 50% of those polled want to call it Bexley.

(b) Use the central limit theorem to estimate the probability more than 25% of those polled prefer Kobol, Bajor, or Kashyyyk.

Problem 5. (10 pts.) **Experimental Error and Mistakes.** A scientist had n independent measurements of a physical quantity Q. It was known that the measurement error followed a uniform distribution uniform$(-1, 1)$. That is, the measurements followed a uniform distribution uniform$(Q - 1, Q + 1)$.

(a) To estimate Q the scientist took the average of all n measurements. Assuming n is relatively large, estimate the probability that the average was off by more than 0.1 from the true value of Q. (Your answer will need to depend on n. Give it in the form of a probability of the standard normal distribution Z that could be easily computed using R once a value for n is known.)

(b) Suppose $X_1, X_2, \ldots$ are independent uniform$(-1, 1)$ random variables. Suppose also that X is the sum of the first n and Y is the sum of X_{n-11} to X_{2n-12}. Compute $\mathrm{Cov}(X, Y)$ and $\mathrm{Cor}(X, Y)$. You should assume that $n \geq 12$.

Hints: Use the linearity rules for covariance. What is the size of the overlap?