For extra practice, several additional review problems are printed below. Solutions to these problems can be found on the exams page of the course website. While these questions are representative of the typical scope and difficulty of individual exam questions, this review is not comprehensive, nor does it necessarily represent the total amount of time available for the exam. Additionally, while the exam will be cumulative, these problems focus primarily on content covered since the second midterm. You should use older exams, homework, and class activities to review older material.

Take-home Exam.

(1) A fair 20-sided die is rolled repeatedly until an odd number is shown. Let W be the event that the first odd number appeared on an even numbered roll (i.e if the first three rolls are 18, 2, 6, and the fourth roll is 7, then W occurs, since 4 is even).
  (a) Give at least two different ways to calculate $P(W)$, using material from our course.
  (b) Explain why it makes sense that $P(W) < 1/2$.
  (c) Create a simulation of this dice-rolling experiment in R, and use your simulation to approximate $P(W)$. *Hint: In R,* `n%%2` *returns* 0 *if* n *is even and* 1 *if* n *is odd.*

(2) Suppose X_1 and X_2 are iid Beta$(2,1)$ variables. Define two new variables Y_1 and Y_2 by

$$Y_1 = \frac{X_1}{X_2} \qquad Y_2 = X_1X_2$$

Let g be the transformation $(y_1, y_2) = g(x_1, x_2) = \left(\frac{x_1}{x_2}, x_1x_2\right)$.
  (a) Write down the formula for the joint density function of X_1 and X_2 (be sure to specify support).
  (b) Find a formula for the inverse transformation $(x_1, x_2) = g^{-1}(y_1, y_2)$.
  (c) Compute the Jacobian either of g or g^{-1}.
  (d) Use the change-of-variables formula to find a formula for the joint PDF of Y_1 and Y_2.

(3) Let $X_1, X_2, \ldots$ be iid variables, each with CDF F. For every x, define a function $R_n(x)$ to be the number of $X_1, \ldots, X_n$ that are less or equal to x.
  (a) Find the mean and variance of $R_n(x)$ (in terms of n and $F(x)$).
  (b) Show that with probability 1, $\lim_{n\to\infty} \frac{R_n(x)}{n} = F(x)$.

(4) Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ with $n \geq 2$. Let $\bar{X}_n$ be the sample mean and let $S_n^2 = \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2$ be the sample variance. Show that $\bar{X}$ and S_n^2 are independent. *Hint: Apply properties of Multivariate Normal distributions to the vector* $(\bar{X}_n, X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$.

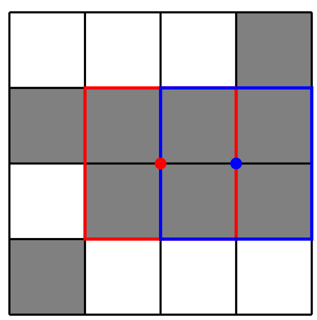
(5) Jonathan's newly-built computer will last Expo(λ) years before one of its parts fails. When that happens, he will try to fix it. With probability p, he will be able to fix it, at which point it will last an additional Expo(λ) years before one of its parts fails, independent of previous time until failure (at which point he will attempt to fix it with probability p, and so on). If at any time the computer cannot be fixed, he will build a new computer from scratch. Find the expected amount of time until Jonathan builds a new computer, along with the variance in the amount of time.

(6) A new treatment for a disease is being tested, to see whether it is better than the standard treatment. The existing treatment is effective on 50% of patients. It is believed initially that there is a 2/3 chance the new treatment is effective on 60% of patients, and a 1/3 chance that the new treatment is effective on 50% of patients. In a pilot study, the new treatment is given to 20 random patients, and is effective for 15 of them.

(a) Given this information, what is the probability that the new treatment is better than the standard treatment?

(b) A second study is done later, giving the new treatment to 20 new random patients. Given the results of the first study, what is the PMF for how many of the new patients the new treatment is effective on (Let p be the probability calculated in part (a), and express your answer in terms of p).

(7) Consider an n-by-n checkerboard divided into n^2 many 1-by-1 squares. Suppose each square is colored gray or white independently and uniformly at random. Let X denote the number of 2-by-2 inch subboards of the checkerboard that are all gray. An example of one such board is shown below (with $n = 4$). For this board, $X = 2$ with the two all-gray 2-by-2 subboards and centerpoints of these subboarded highlighted in red and blue.



(a) Suppose the vertices of the checkerboard are labeled in Cartesian coordinates, where $(0, 0)$ denotes the vertex in the lower left of the checkerboard and (n, n) denotes the vertex in the upper right of the checkerboard. Let A be the event that the vertex at $(1, 1)$ is the center of an all-gray 2-by-2 subboard, and let B be the event that the vertex at $(2, 1)$ is the center of an all-gray 2-by-2 subboard. Are A and B independent? Explain.

(b) Compute the expected value of X.

(c) What is an accurate *approximate* distribution of X? Explain why this approximation is accurate.

(d) Use your approximation in the previous part to estimate the variance in X.

(8) A researcher is studying the length of hospital stays for patients with a certain disease. Let $X \sim \text{Pois}(\lambda)$ denote the number of days a randomly selected person with the disease stays at a hospital. Data however, is only collected for patients with the disease who were admitted to the hospital (i.e. no data is available for patients with the disease who did not stay at a hospital). Therefore, averaging the length of hospital stays among this population does not estimate $E[X]$. Instead, we can assess the conditional distribution of X, given that the person stayed at least 1 day in the hospital.

(a) Find a formula for the conditional PMF of X given $X \geq 1$.

(b) Use the previous result to calculate $E[X|X \geq 1]$.

(c) Use part (a) to calculate $\text{Var}(X|X \geq 1)$.

(d) Create a plot in R for the function $f(\lambda) = E[X|X \geq 1]$ (i.e. a plot of the conditional expected value versus the expected value). Based on your plot, for what values of λ is $\lambda \approx E[X|X \geq 1]$?

(9) Consider the binomial distribution. Name at least 4 other distributions that are related to the binomial and explain how they are related to the binomial.