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 a number greater than or equal to 11 is shown. Let W be the event that the first such number appear on an even numbered roll (i.e if the first three rolls are 3, 2, 9, and the fourth roll is 12, 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.*

Solution. (a) Several methods are available. Below are solutions using First-Step analysis and using the First Success variable.

- We can use first-step analysis. Let B be the event that the first roll is greater than or equal to 11. By LotP,

$$P(W) = P(W|B)P(B) + P(W|B^c)P(B^c) = P(W|B^c)\frac{1}{2}$$

as $P(W|B) = 0$, since the first roll is an odd numbered roll. But now note that, given B^c, if an additional odd number of rolls are required, then the total number of rolls required is even. Since future rolls are independent of past rolls, $P(W|B^c) = P(W^c) = 1 - P(W)$. Therefore,

$$P(W) = (1 - P(W))\frac{1}{2}$$

and so $P(W) = \frac{1}{3}$.

- Alternatively, let X denote the number of rolls required and note that $X \sim \text{FS}(1/2)$, which has PMF

$$P(X = k) = \frac{1}{2^{k-1}}\frac{1}{2} \qquad k \geq 1$$

Then

$$P(X \text{ is even }) = \sum_{k \text{ even}} P(X = k) = \sum_{j=1}^{\infty} \frac{1}{2^{2j-1}}\frac{1}{2} = \sum_{j=1}^{\infty} \frac{1}{4^j} = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}$$

(b) Since we can only roll a number greater than or equal to 11 in an even-number of rolls if the first roll is 10 or less (while rolling a number greater than or equal to 11 in an odd number of rolls does not have this challenge), then it makes sense that $P(W) < \frac{1}{2} < P(W^c)$.

(c) In R, we can perform this experiment by writing a function using the following code:

```
do_trials <- function(){
trials <- 0
is_big <- 0
while(is_big == 0){
is_big <- sum(sample(1:20, size = 1) >= 11)
trials <- trials + 1
}
return(trials)
}
```

We then use the `replicate` function to perform many simulations, and calculation the proportion of times the result occurred in an even number of rolls:

```
sim <- replicate(10^5, do_trials())
success <- sim %% 2
proportion <- sum(success == 0)/length(success)
proportion
## 0.3365
```

In the above simulation, the result occurred about 33% of the time.

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

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

Let g be the transformation $(y_1, y_2) = g(x_1, x_2) = \left(\frac{x_1}{x_2}, x_1 x_2\right)$.

(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.

Solution. (a) Since each is Beta$(2,1)$, they both have marginal density $f(x) = 2x$. Since X_1, X_2 are independent, then their joint density is the product of their marginals, and so their joint density is

$$4x_1x_2 \qquad 0 \le x_1 \le 1,\ 0 \le x_2 \le 1.$$

(b) Note that $y_1 y^2 = x_1^2$, so $x_1 = \sqrt{y_1 y_2}$. On the other hand, $\frac{y_2}{y_1} = x_2^2$, so $x_2 = \sqrt{\frac{y_2}{y_1}}$. Therefore,

$$(x_1, x_2) = g^{-1}(y_1, y_2) = \left(\sqrt{y_1 y_2}, \sqrt{\frac{y_2}{y_1}}\right)$$

(c) The Jacobian of g^{-1} is

$$J_{g^{-1}}(y_1, y_2) = \frac{\partial x}{\partial y} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{y_2}}{2\sqrt{y_1}} & \frac{\sqrt{y_1}}{2\sqrt{y_2}} \\ -\frac{1}{2}\frac{\sqrt{y_2}}{\sqrt{y_1^3}} & \frac{1}{2\sqrt{y_1 y_2}} \end{pmatrix}$$

which has determinant

$$\det J_{g^{-1}}(y_1, y_2) = \frac{1}{4y_1} + \frac{1}{4y_1} = \frac{1}{2y_1}$$

Since $y_1 > 0$, this is also the absolute Jacobian determinant.

(d) By the change of variables formula,

$$f_{Y_1,Y_2} = f_{X_1,X_2}(g^{-1}(y_1, y_2) = |\det J_{g^{-1}}(y_1, y_2)| = 4\sqrt{y_1 y_2}\sqrt{\frac{y_2}{y_1}}\frac{1}{2y_1} = 2\frac{y_2}{y_1}$$

which is valid for all points y_1, y_2 in the image of the unit square under g.

(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, $\lim_{n\to\infty} \frac{R_n(x)}{n} = F(x)$.

Solution. For each $1 \le j \le n$, let I_j be the indicator for the event $X_j \le x$. Then $R_n(x) = \sum_{j=1}^n I_j$ and by the fundamental bridge and linearity of expectation,

$$E[R_n(x)] = \sum_{j=1}^n E[I_j] = \sum_{j=1}^n P(X_j \le x) = nF(x)$$

Since the variables $X_1, \dots, X_n$ are independent, then the indicator variables $I_1, \dots, I_n$ are independent as well. Additionally, since each I_j is Bern(p) with $p = F(x)$, then $\text{Var}(I_j) = p(1-p) = F(x)(1-F(x))$. Therefore

$$\text{Var}(R_n(x)) = \sum_{j=1}^{n} \text{Var}(I_j) = \sum_{j=1}^{n} F(x)(1-F(x)) = nF(x)(1-F(x))$$

By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{R_n(x)}{n} = \lim_{n\to\infty} \frac{I_1 + \cdots + I_n}{n} = E[I_1] = F(x).$$

as the variables $I_1, \dots, I_n$ are iid with mean $F(x)$.

(4) Let $X_1, \dots, X_n$ 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* $(\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$.

Solution. Let $\mathbf{X} = (\bar{X}_n, X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$, and note that $\mathbf{X}$ is indeed MVN, since linear combinations of its coordinates are linear combinations of the X_i, which are iid $N(\mu, \sigma^2)$, and so these linear combinations are Normally distributed.

Recall also that for MVN, if two coordinates are uncorrelated, then they are actually independent. Consider $\text{Cov}(\bar{X}, X_j - \bar{X})$ for some j. By the linearity properties of Covariance,

$$\text{Cov}(\bar{X}, X_j - \bar{X}) = \text{Cov}(\bar{X}, X_j) - \text{Cov}(\bar{X}, \bar{X}) = \text{Cov}(\bar{X}, X_j) - \text{Var}(\bar{X})$$

Recall that $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. Additionally, since X_j is independent of X_i for all $i \neq j$, then

$$\text{Cov}(\bar{X}, X_j) = \frac{1}{n}\sum_{i=1}^{n} \text{Cov}(X_i, X_j) = \frac{1}{n}\text{Cov}(X_j, X_j) = \frac{\sigma^2}{n}$$

Therefore,

$$\text{Cov}(\bar{X}, X_j - \bar{X}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

That is, $\bar{X}$ is indepdendent of each of $X_j - \bar{X}$. But now, note that

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$$

is a function of just the $X_j - \bar{X}$, and so S_n^2 is also independent of $\bar{X}$.

(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.

Solution. Let N denote the number of part failures until the computer can no longer be repaired (including the final one), and note that $N \sim \text{FS}(1-p)$. Let T_1 be the time until the first failure, let T_2 be the time until the second failure, and so on. Let T be the total times until the computer cannot be repaired, with $T = T_1 + \cdots + T_N$. By the Law of Total Expectation and linearity of conditional expectation,

$$E[T] = E[E[T|N]] = E\Big[E[T_1|N] + E[T_2|N] + \cdots + E[T_N|N]\Big] = E\Big[N\frac{1}{\lambda}\Big] = \frac{1}{\lambda(1-p)}$$

(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).

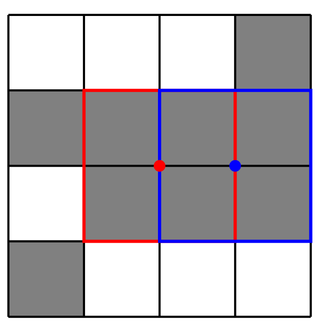
Solution. (a) Let B be the event that the new treatment is better than the standard treatment and let X be the number of people for which the new treatment was effective. By Bayes Rule,

$$P(B|X=15) = \frac{P(X=15|B)P(B)}{P(X=15)} = \frac{P(X=15|B)P(B)}{P(X=15|B)P(B)+P(X=15|B^c)P(B^c)}$$
$$= \frac{\binom{20}{15}(0.6)^{15}(0.4)^5\frac{2}{3}}{\binom{20}{15}(0.6)^{15}(0.4)^5\frac{2}{3}+\binom{20}{15}0.5^{20}\frac{1}{3}} \approx 0.9099$$

(b) Let Y be the number of new patients for whom the new treatment is effective, and let $p = P(B|X=15)$ from above. If $0 \le k \le 20$, then by LotP with extra conditioning

$$P(Y=k|X=15) = P(Y=k|X=15,B)P(B|X=15) + P(Y=k|X=15,B^c)P(B^c|X=15)$$
$$= P(Y=k|B)P(B|X=15) + P(Y=k|B^c)P(B^c|X=15)$$
$$= \binom{20}{k}0.6^k 0.4^{20-k}p + \binom{20}{k}0.5^k(1-p)$$

(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.

Solution. (a) The 2-by-2 square centered at $(1, 1)$ and the 2-by-2 square centered at $(2, 1)$ have two squares in common. If the event A occurs, then these two squares must be gray, which makes it more likely for the event B to occur. Hence, A and B are not independent.

(b) There are a total $m = (n-1)(n-1)$ interior vertices in the checkerbard. Number them 1 through m and for $1 \leq j \leq m$, let I_j be the indicator that the 2-by-2 square centered at vertex j is all gray. Since the colors of squares are independent, then $E[I_j] = P(4 \text{ squares are gray}) = \frac{1}{2^4}$. By linearity and the fundamental bridge,

$$E[X] = E\left[\sum_{j=1}^{m} I_j\right] = \sum_{j=1}^{m} E[I_j] = \frac{m}{2^4} = \frac{(n-1)^2}{2^4}.$$

(c) While the events that adjacent vertices are centers of all gray squares are somewhat dependent, this dependence is relatively weak. Moreover, for non-adjacent vertices, the events that these vertices are at the center of all gray squares are independent. Therefore, the Poisson Paradigm applies and $X \sim \text{Pois}(\lambda)$ where $\lambda = E[X] = \frac{(n-1)^2}{2^4}$ from above.

(d) The variance of $\text{Pois}(\lambda)$ variable is λ, and so

$$\text{Var}(X) \approx \frac{(n-1)^2}{2^4}.$$

(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 $\lambda \approx E[X|X \geq 1]$?

Solution. (a) Using the definition of conditional probability, the conditional PMF of $X|X \geq 1$ is

$$P(X = k|X \geq 1) = \frac{P(X = k)}{P(X \geq 1)} = \frac{e^{-\lambda}\lambda^k}{k!(1 - e^{-\lambda})}$$

for $k \geq 1$. The conditional PMF is 0 for $k = 0$.

(b) We can use the conditional PMF to find the conditional expectation:

$$E[X|X \geq 1] = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!(1 - e^{-\lambda})} = \frac{e^{-\lambda}\lambda}{1 - e^{-\lambda}} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \frac{e^{-\lambda}\lambda}{1 - e^{-\lambda}} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \frac{e^{-\lambda}\lambda}{1 - e^{-\lambda}} e^{\lambda} = \frac{\lambda}{1 - e^{-\lambda}}$$

(c) Likewise, we can use LOTUS and part (a) to compute the conditional variance:

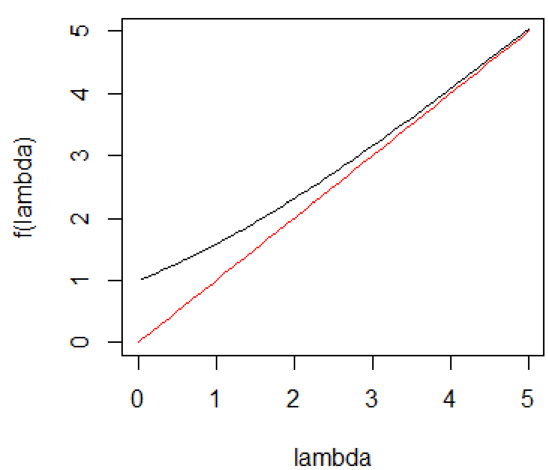
$$
\begin{aligned}
E[X^2|X \geq 1] =& \frac{e^{-\lambda}}{1-e^{-\lambda}} \sum_{k=1}^{\infty} \frac{k^2 \lambda^k}{k!} \\
=& \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\
=& \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} \sum_{k=1}^{\infty} \frac{d}{d\lambda} \frac{\lambda^k}{(k-1)!} \\
=& \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} \frac{d}{d\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
=& \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} \frac{d}{d\lambda} \lambda e^{\lambda} \\
=& \frac{e^{-\lambda}\lambda}{1-e^{-\lambda}} \left(e^{\lambda} + \lambda e^{\lambda}\right) \\
=& \frac{\lambda + \lambda^2}{1-e^{-\lambda}}
\end{aligned}
$$

Therefore,

$$
\text{Var}(X|X \geq 1) = E[X^2|X \geq 1] - (E[X|X \geq 1])^2 = \frac{\lambda + \lambda^2}{1-e^{-\lambda}} - \left(\frac{\lambda}{1-e^{-\lambda}}\right)^2
$$

(d) We will plot $f(\lambda)$ on the range $0 < \lambda < 5$, along with the function $g(\lambda) = \lambda$:

```
f <- function(lambda){
lambda/(1 - exp(-lambda))
}
lambda <- seq(0, 5, length = 100)
plot(lambda, f(lambda), type = "l", ylim = c(0,5))
lines(lambda, lambda, col = "red")
```



Based on the plot, it seems that $f(\lambda) \approx \lambda$ for $\lambda > 4$.

(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.

Solution. Many answer are possible. Here are a few:

- Bernoulli. The sum of n iid $\text{Bern}(p)$ variables is $\text{Bin}(n, p)$.
- Hypergeometric. For $w + b$ large and $p = \frac{w}{w+b}$, $\text{HGeom}(w, b, n) \approx \text{Bin}(n, p)$.
- Poisson. By the Poisson Paradigm, $\text{Bin}(n, p)$ is approximately $\text{Pois}(\lambda)$ with $\lambda = np$ with n is large and p is small.
- Normal. When n is large and p is not too close to 0 or 1, then $\text{Bin}(n, p)$ is approximately $N(np, np(1-p))$.
- Beta. The Beta distribution is the conjugate prior to the Binomial distribution.
- Multinomial. The Multinomial is the n-dimensional generalization of the Binomial distribution.
- (Tenuous). In a sequence of Bernoulli trials, the number of failures before the first success is Geometric, while the total number of success in a fixed number of trials is Binomial.
- (Tenuous). The $\text{Bin}(1, 1/2)$ variable is $\text{Dunif}(0, 1)$.