PRACTICE PROBLEMS FOR EXAM 2

Math 3160Q – Fall 2015
Professor Hohn

Below is a list of practice questions for Exam 2. Any quiz, homework, or example problem has a chance of being on the exam. For more practice, I suggest you work through the review questions at the end of each chapter as well.

1. Let X be a continuous random variable with density function

$$f_X(t) = \begin{cases} a\,t + b & -1 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find b.

Solution: We know that $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$. Therefore

$$1 = \int_{-\infty}^{\infty} f_X(t)\,dt = \int_{-1}^{1} (at + b)\,dt = 2b.$$

Solving for b, we find $b = 1/2$.

(b) Find $\mathbb{E}(X)$. This may be in terms of a.

Solution: With $b = 1/2$,

$$\mathbb{E}[X] = \int_{-1}^{1} t(at + 1/2)\,dt = \left[\frac{1}{3}\,at^3 + \frac{1}{4}\,t^2\right]_{-1}^{1} = \frac{2}{3}\,a.$$

(c) Find $\text{Var}(X)$. This may be in terms of a.

Solution: We know that $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. We found $\mathbb{E}[X]$ in the previous part, so it remains to find $\mathbb{E}[X^2]$. We have

$$\mathbb{E}[X^2] = \int_{-1}^{1} t^2(at + 1/2)\,dt = \left[\frac{1}{4}\,at^4 + \frac{1}{6}\,t^3\right]_{-1}^{1} = \frac{1}{3}$$

Therefore,

$$\text{Var}(X) = \frac{1}{3} - \left(\frac{2}{3}\,a\right)^2 = \frac{1}{3} - \frac{4}{9}\,a^2 = \frac{1}{3}\left(1 - \frac{4}{3}\,a\right).$$

(d) Give a brief (but reasonable) explanation why any value of a where $a > b$ would not be possible.

Solution: Notice that as the value of t approaches -1 from the right, the density $f_X(t)$ approaches the value $-a + b$. Since the density must always be non-negative, it must be that $-a + b \geq 0$, which rearranges to $b \geq a$, showing that $a > b$ is not allowed (otherwise the density will output negative values).

2. Suppose $X = N(\mu, \sigma^2)$. In terms of the distribution $\Phi(x) = P(N(0,1) \le x)$ of the standard normal random variable, find the probability that X is less than $\frac{1}{2}\sigma + \mu$, or greater than $\frac{3}{2}\sigma + \mu$. That is, find $P\big((X < \frac{1}{2}\sigma + \mu) \cup (X > \frac{3}{2}\sigma + \mu)\big)$.

Solution: If A is the event that $X < \frac{1}{2}\sigma + \mu$ and B is the event $X > \frac{3}{2}\sigma + \mu$, then we want $P(A \cup B)$. Since A and B are mutually exclusive (disjoint), we have

$$P(A \cup B) = P(A) + P(B) = P(X < \frac{1}{2}\sigma + \mu) + P(X > \frac{3}{2}\sigma + \mu)$$

Using that $X \stackrel{d}{=} \sigma Z + \mu$ where $Z \stackrel{d}{=} N(0,1)$,

$$P(X < \frac{1}{2}\sigma + \mu) = P(\sigma Z + \mu < \frac{1}{2}\sigma + \mu) = P(Z < \frac{1}{2}) = \Phi(1/2)$$
$$P(X > \frac{3}{2}\sigma + \mu) = P(\sigma Z + \mu > \frac{3}{2}\sigma + \mu) = P(Z > \frac{3}{2}) = 1 - \Phi(3/2)$$

So, $P(A \cup B) = \Phi(1/2) + 1 - \Phi(3/2)$.

3. Suppose that an experiment has two outcomes 0 or 1 (such as flipping a coin). Suppose that you run n independent experiments and for the i^{th} experiment you let the random variable X_i tell you the outcome for $1 \le i \le n$. Then we can assume that for each i, that $X_i = \text{Ber}(p)$ with $p = P(X_i = 1)$ (where we will assume for this problem that p is the same for each i). Then, let $X = \sum\limits_{i=1}^{n} X_i$.

   (a) What is the state space S_X of X?

   Solution: If we add together n numbers, each of which are either 0 or 1, the possible outcomes are $0, 1, 2, ..., n$. So the state space S_X is $S_X = \{0, 1, 2, ..., n\}$.

   (b) What is $\mathbb{E}[X]$?

   Solution: Since $X = \sum_{i=1}^{n} X_i$, we have

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

   Since $X_i \stackrel{d}{=} \text{Ber}(p)$ for each i, $\mathbb{E}[X_i] = p$. So

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np.$$

   We could have also realized this since if you consider flipping a coin n times where you assign 1 if it lands heads and 0 if it lands tails, then X counts the number of heads landed in those n flips. So $X \stackrel{d}{=} \text{Bin}(n, p)$, so $\mathbb{E}[X] = np$, as we discovered before.

4. Suppose that the time between customer arrivals in a store is given by an exponential random variable $X \stackrel{d}{=} \text{Exp}(\lambda)$, such that the average time between arrivals is 2 minutes. Suppose you

walk past the store and notice it's empty. What is the probability from the time you walk past the store, the store remains empty for more than 5 minutes?

Solution: We know that $\mathbb{E}[X] = \frac{1}{\lambda}$ since $X \stackrel{d}{=} \text{Exp}(\lambda)$. We are given $\mathbb{E}[X] = 2$ (min), so $\lambda = 1/2$ (min^{-1}). Suppose that we walk past the store at time t_0 (in minutes) and notice it is empty, the problem asks us to find $P(X > t_0+5 \mid X > t_0)$. However, since X is an exponential random variable, the memoryless property tells us that $P(X > t_0 + 5 \mid X > t_0) = P(X > 5)$. Now,

$$P(X > 5) = e^{-\lambda 5} = e^{-5/2}.$$

5. Let X and Y be random variables with distributions given by,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 3x & 0 \leq x < 1/3 \\ 1 & x \geq 1/3 \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & x < 0 \\ 1 - \frac{1}{2}e^{-2x} & x \geq 0 \end{cases}$$

(a) Find $P(X \leq 1/4)$, $P(Y < 0)$, and $P(Y \leq 0)$.

Solution: We have

$$P(X \leq 1/4) = F_X(1/4) = 3(1/4) = 3/4$$

and

$$P(Y \leq 0) = F_Y(0) = 1/2.$$

Also

$$P(Y < 0) = P(Y \leq 0) - P(Y = 0) = F_Y(0) - P(Y = 0)$$

and since there is a jump gap of $1/2$ at $Y = 0$, we have $P(Y = 0) = 1/2$. Now,

$$P(Y < 0) = F_Y(0) - P(Y = 0) = 1/2 - 1/2 = 0.$$

(b) Find $\mathbb{E}[X]$ and $\text{Var}(X)$.

Solution: Note that F_X is continuous (no jumps) and that we can find a density f_X

$$f_X(t) = \frac{d}{dt}F_X(t) = \begin{cases} 3 & 0 \leq t < 1/3 \\ 0 & \text{otherwise} \end{cases}$$

So, X is, in fact, a uniform random variable $X \stackrel{d}{=} \text{Unif}(0, 1/3)$. Therefore,

$$\mathbb{E}[X] = \frac{0 + 1/3}{2} = 1/6$$

and

$$\text{Var}(X) = \frac{(1/3 - 0)^2}{12} = 1/108.$$

(c) Find $\mathbb{E}[Y]$.

Solution: Because of the jump in the graph of $F_Y(t)$ at the value $t = 0$ (i.e., $F_Y(0+) - F_Y(0-) = 1/2$) we see that Y is not continuous. However, since F_Y is not a step function, Y is not a discrete random variable either. We are not out of luck though! We have

$$\begin{aligned}\mathbb{E}[Y] = \int_0^\infty (1 - F_Y(t))\,dt - \int_{-\infty}^0 F_Y(t)\,dt &= \frac{1}{2}\int_0^\infty e^{-2t}\,dt - \int_{-\infty}^0 0\,dt \\ &= \frac{1}{2}\int_0^\infty e^{-2t}\,dt = \frac{1}{4}.\end{aligned}$$

6. Let X be a continuous random variable with density given by,

$$f_X(x) = \begin{cases} ke^x & 0 < x < \ln(2) \\ 0 & \text{otherwise} \end{cases}$$

(a) Find k.

Solution: Integrating over the density,

$$\int_{-\infty}^\infty f_X(t)\,dt = k\int_0^{\ln(2)} e^t\,dt = ke^t\Big|_0^{\ln(2)} = k(2-1) = k.$$

We also know that $\int_{-\infty}^\infty f_X(t)\,dt = 1$, so $k = 1$.

(b) Let $Y = e^X$. Find the density $f_Y(y)$ of Y.

Solution: Putting the CDF of Y into terms of the CDF of X, we have

$$F_Y(t) = P(Y \le t) = P(e^X \le t) = P(X \le \ln(t)) = F_X(\ln(t)).$$

Therefore,

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \frac{d}{dt}F_X(\ln(t)) = \frac{1}{t}f_X(\ln(t)).$$

Note that

$$f_X(\ln(t)) = \begin{cases} e^{\ln(t)} & 0 < \ln(t) < \ln(2) \\ 0 & \text{otherwise} \end{cases} = \begin{cases} t & 1 < t < 2 \\ 0 & \text{otherwise} \end{cases}$$

and so

$$f_Y(t) = \frac{1}{t}f_X(\ln(t)) = \begin{cases} 1 & 1 < t < 2 \\ 0 & \text{otherwise} \end{cases}$$

(c) What type of continuous random variable is Y?

Solution: Since $f_Y(t)$ is constant on the interval $(1, 2)$ and 0 elsewhere, we have at $Y \overset{d}{=} \text{Unif}(1, 2)$.

7. Your professor (reluctantly) runs on a treadmill at the UConn gym during the winter. Suppose the time she waits in line for a treadmill is exponentially distributed, and on average, the time she waits in line for a treadmill is 7 minutes.

(a) What is the probability that the next time your professor waits for a treadmill, she will wait in line for at least 10 minutes?

Solution: Let X be the random variable which represents the amount of time your professor waits for a treadmill during her next trip to the gym. Then by assumption, X is an exponential random variable with parameter $\lambda = 1/7$ in units min^{-1} (since $\lambda = 1/\mathbb{E}[X]$ and $\mathbb{E}[X] = 7$ minutes by assumption). We have

$$P(X \geq 10) = \frac{1}{7}\int_{10}^{\infty} e^{-t/7}\, dt = e^{-10/7}.$$

(b) You notice your professor in the gym waiting in line for the treadmill. You do not know how long she has already been waiting in line. If you can, calculate the probability that she will get a treadmill in the next 5 minutes; if you can not calculate this probability, explain why not.

Solution: Suppose that your professor has already been waiting in line for some time $t_0 > 0$ when you see her. Once again, let X represent the amount of time the professor will wait in line. We want to find $P(X \leq t_0 + 5 \,|\, X > t_0)$, which by the memoryless property of the exponential random variable is equal to $P(X \leq 5) = 1 - e^{-5/7}$.

8. You are choosing between two venues to order food from which will be delivered to your house. With probability 1/3 you will choose venue A, and with probability 2/3 you will choose venue B. If you order from venue A, 15 minutes after making the call the remaining time it takes the food to arrive is exponentially distributed with average 10 minutes. If you order from venue B, 10 minutes after making the call, the time it takes the food will arrive is exponentially distributed with an average of 12 minutes. Given that you have already waited 25 minutes after calling, and the food has not arrived, what is the probability that you ordered from venue A?

Solution: Let A be the event you ordered from venue A and B be the event you ordered from venue B. Let $X \overset{d}{=} \text{Exp}(1/10)$ and $Y \overset{d}{=} \text{Exp}(1/12)$. Let T be the random variable telling us the time it takes your food to arrive after ordering. What the problem tells us is that for any time t (greater than 15min), we have $P(T > t \mid A) = P(X + 15 > t) = P(X > t - 15) = e^{-(t-15)/10}$ and $P(T > t \mid B) = P(Y + 10 > t) = P(Y > t - 10) = e^{-(t-10)/12}$. We are asked to find $P(A \mid T > 25)$. An application of Bayes' formula gives

$$P(A \mid T > 25) = \frac{P(T > 25 \mid A)P(A)}{P(T > 25 \mid A)P(A) + P(T > 25 \mid B)P(B)}$$
$$= \frac{e^{-(25-15)/10}(1/3)}{e^{-(25-15)/10}(1/3) + e^{-(25-10)/12}(2/3)} = \frac{e^{-1}}{e^{-1} + 2e^{-5/4}} = \frac{1}{1 + 2e^{-1/4}}$$

9. Distracted while listening to the latest Beyoncé album, General Xavier accidentally knocks over a large jar filled with 10,000 fair coins at Fort Knox. All the coins fall out completely randomly. Let X count the number of heads that appear when the coins fall. Then $X \overset{d}{=} \text{Bin}(10\,000, \frac{1}{2})$.

(a) (3 points) What is $P(X > 5100)$? You do not need to evaluate the sum you write down.

Solution: Since $X \overset{d}{=} \text{Bin}(10\,000, 1/2)$,

$$P(X > 5100) = \sum_{k=5101}^{10,000} \binom{10000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10000-k} = \sum_{k=5101}^{10000} \binom{10000}{k} \left(\frac{1}{2}\right)^{10000}$$

(b) (5 points) Approximate $P(X > 5100)$ using a normal distribution. Leave you answer in terms of $\Phi(x)$ where $\Phi(x) = P(N(0,1) \le x)$.

Solution: Let $Z \overset{d}{=} N(0,1)$. Here $X \overset{d}{=} \text{Bin}(n,p)$ with $n = 10000$ and $p = 1/2$. So $\mu = np = 5000$ and $\sigma = \sqrt{np(1-p)} = \sqrt{2500} = 50$. We have $X \overset{d}{\approx} \sigma Z + \mu$, meaning

$$\begin{aligned} P(X > 5100) &\approx P\Big(\sigma Z + \mu > 5100\Big) = P(Z > \frac{5100 - \mu}{\sigma}) \\ &= P(Z > \frac{5100 - 5000}{50}) = P(Z > 2) \\ &= 1 - P(Z \le 2) = 1 - \Phi(2) \end{aligned}$$

(c) (5 points) Approximate $P(X > 5100)$ using a Poisson distribution. Leaving an infinite sum here is OK.

Solution: Using the Poisson approximation, we have $X \approx \text{Pois}(np) = \text{Pois}(5000)$. Therefore,

$$P(X > 5100) = 1 - P(X \le 5100) = 1 - \sum_{k=1}^{5100} \frac{5000^k}{k!} e^{-5000}$$

Where we chose to write this as a finite sum.

Note: Although n is large, p is "fixed" regardless of n and $np \not\ll n$, so a Poisson approximation should **not** be what our instincts tell us to use. Rather, with n large and $np(1-p)$ reasonably large relative to n, the normal approximation in the previous part is likely a much better approximation, much easier to calculate by hand, and should be what our instincts suggest.

10. 48000 fair dice are rolled independently. Let X count the number of sixes that appear.

(a) What type of random variable is X?

Solution: The state space of X will be $S_X = \{0, 1, 2, ..., 48000\}$ since these are all the possible times a 6 appears when rolling 48000 dice. Supposing that each of the rolls of the dice are independent, then for any $k \in S_X$, we have $P(X = k) = \binom{48000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{48000-k}$. This shows that $X = \text{Bin}(48000, 1/6)$.

(b) Write the expression for the probability that between 7500 and 8500 sixes show. That is $P(7500 \le X \le 8500)$.

Solution: Since $X = \text{Bin}(48000, 1/6)$,

$$P(7500 \le X \le 8500) = \sum_{k=7500}^{8500} \binom{48000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{48000-k}.$$

(c) The sum you wrote in part b) is ridiculous to evaluate. Instead, approximate the value by a normal distribution and evaluate in terms of the distribution $\Phi(x) = P(N(0,1) \le x)$ of a standard normal random variable.

Solution: Let $\mu = \mathbb{E}[X] = np = 48000(1/6) = 8000$ and $\sigma = \sqrt{\text{Var}(X)} = \sqrt{np(1-p)} = \sqrt{48000(1/6)(5/6)} = \sqrt{20000/3}$. Using the normal approximation, we have $X \overset{D}{\approx} \sigma Z + \mu$ where $Z = N(0,1)$. This means,

$$\begin{aligned} P(7500 \le X \le 8500) &\approx P(7500 \le \sigma Z + \mu \le 8500) \\ &= P\Big(\frac{7500-\mu}{\sigma} \le Z \le \frac{8500-\mu}{\sigma}\Big) \\ &= P\Big(\frac{-500}{\sqrt{20000/3}} \le Z \le \frac{500}{\sqrt{20000/3}}\Big) \\ &= \Phi\Big(\frac{500}{\sqrt{20000/3}}\Big) - \Phi\Big(-\frac{500}{\sqrt{20000/3}}\Big) \\ &= 2\Phi\Big(\frac{500}{\sqrt{20000/3}}\Big) - 1 \end{aligned}$$

where the very last equality used the symmetry argument $\Phi(-a) = 1 - \Phi(a)$.

(d) Why do you think a normal distribution is a good choice for approximation?

Solution: Notice that n is reasonably large and $np(1-p) = 20000/3$ is also quite large relative to n. With these considerations, the normal approximation should be fairly good. Moreover, the large sum in the Poisson approximation and that $np \not\ll n$ suggest that the Poisson approximation is likely not the approximation we want to use!

11. Suppose that on average 2 people in a major city die each year from alien attack. Suppose that each attack is random and independent.

(a) If X is the number of deaths from alien attack within the next year from a randomly selected major city, what type of random variable is X?

Solution: Suppose that n is the size of the population of the major city and $p = 2/n$ is the probability that a randomly selected person drawn from that city is killed by alien attack. The total possible outcomes of X, i.e. the state space of X, is $S_X = \{0, 1, 2, ..., n\}$ and $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. This shows that $X \overset{d}{=} \text{Bin}(n, p)$.

(b) Use the Poisson approximation to approximate the probability that the next major city you visit will have at least 3 deaths due to alien attack?

Solution: The Poisson approximation says $X \overset{d}{\approx} \text{Pois}(np) = \text{Pois}(2)$ (since $p = 2/n$). We are looking for $P(X \ge 3)$. From here,

$$\begin{aligned} &P(X \ge 3) \approx P(\text{Pois}(2) \ge 3) = 1 - P(\text{Pois}(2) < 3) \\ &= 1 - \sum_{k=0}^{2} \frac{2^k}{k!} e^{-2} = 1 - \Big(1 + 2 + \frac{2^2}{2}\Big) e^{-2} = 1 - 5e^{-2}. \end{aligned}$$

(c) Why do you think a Poisson approximation is used instead of a normal approximation?

Solution: In this scenario, n is quite large compared to np (remember $np = 2$ while n is the population for an entire major city!), so the Poisson approximation seems like a good fit. Moreover, $np = 2$ stays fixed and small while n is quite large and hence $np(1-p) < np \ll n$ appears quite small with respect to n, which makes the normal approximation not as appealing.

12. Consider the following graph of the distribution $F_X(t)$ of X

defined by

$$F_X(t) = \begin{cases} 0 & t < 0 \\ .3 & 0 \le t < 1 \\ .8 & 1 \le t < 3 \\ 1 & t \ge 3 \end{cases}$$

(a) Is the random variable X discrete, continuous, or neither?

Solution: We see that the CDF is a step function, which tells us that X is a discrete random variable.

(b) What is the state space S_X of X?

Solution: The jumps in the CDF occur at 0, 1, and 3. Therefore $S_X = \{0, 1, 3\}$.

(c) What is the expected value $\mathbb{E}[X]$?

Solution: By considering the size of the gaps at each jump, we have that $P(X = 0) = .3 - 0 = .3$, $P(X = 1) = .8 - .3 = .5$, and $P(X = 3) = 1 - .8 = .2$. From here

$$\mathbb{E}[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 3 \cdot P(X = 3) = .5 + 3(.2) = 1.1.$$

13. You have a fair coin, and you want to take your professor's money. You ask the professor to play a gambling game with you. The gambling game is designed as follows: You charge the professor $\$C$ to play. You then flip the coin twice and record the number of heads that show. If 0 heads show, you pay the professor \$5. If exactly 1 head shows, you pay the professor \$2. If 2 heads show, the professor pays you \$6. Let W be the random variable representing your wealth during a play of the game.

(a) What elements are in the state space S_W of W?

Solution: If the number of heads is 0, then you have money you have made is $\$(C-5)$; if the number of heads is 1, then you have made $\$(C-2)$; if the number of heads is 2, then you have made $\$(C+6)$. Therefore, the possible outcomes of W are

$$S_W = \{C-5, C-2, C+6\}.$$

(b) What is the least amount of money $\$C$ you should charge your professor so that on average you don't lose money?

Solution: We want to find some C such that $\mathbb{E}[W] \geq 0$, the least amount C we should charge would be chosen so that $\mathbb{E}[W] = 0$. We have

$$\begin{aligned}
\mathbb{E}[W] &= (C-5)P(W=C-5) + (C-2)P(W=C-2) + (C+6)P(W=C+6) \\
&= (C-5)P(0 \text{ heads}) + (C-2)P(1 \text{ heads}) + (C+6)P(2 \text{ heads}) \\
&= (C-5)\left(\frac{1}{2}\right)^2 + (C-2)\binom{2}{1}\left(\frac{1}{2}\right)^2 + (C+6)\left(\frac{1}{2}\right)^2 \\
&= C - \frac{5}{4} - 1 + \frac{6}{4} = C - \frac{3}{4}.
\end{aligned}$$

So, the least amount will be $C = \frac{3}{4}$ dollars (i.e., 75 cents).

14. Suppose that X is a normal random variable with mean 75. Suppose that you know $\text{Var}(\frac{1}{2}X + 42) = 25$. Calculate $P(X < 60)$. You can leave your answer in terms of Φ, the CDF of a standard normal.

Solution: We know that for any random variable X, we have $\text{Var}(aX+b) = a^2\,\text{Var}(X)$. So, if $\text{Var}(\frac{1}{2}X + 42) = 25$, then $\frac{1}{4}\text{Var}(X) = 25$ which implies that $\text{Var}(X) = 100$. This shows that $X \stackrel{d}{=} N(75, 100)$. Now, letting Z be a standard normal,

$$\begin{aligned}
P(X < 60) &= P(\sqrt{100}\,Z + 75 < 60) \\
&= P\Big(Z < \frac{60-75}{10}\Big) \\
&= P(Z < -1.5) \\
&= \Phi(-1.5) \\
&= 1 - \Phi(1.5).
\end{aligned}$$

15. (a) UConn's women's basketball team takes about 2500 shots each year with a probability of scoring on a given shot being 4/5. Use the Central Limit Theorem (normal approximation) to approximate the probability that next year, of the 2500 shots taken by the team, they score on at least 2015 shots?

Solution: Let Z represent a standard normal random variable. Let X be the number of shots made by the team next year. Then, it is reasonable that X is a binomial random variable with parameters $n = 2500$ and $p = 4/5$. Let $\mu = np = 2000$ and $\sigma = \sqrt{np(1-p)} = 20$. Therefore, using the normal approximation of a binomial random variable,

$$P(X \geq 2015) \approx P(\sigma\,Z + \mu \geq 2015) = P\Big(Z \geq \frac{2015-\mu}{\sigma}\Big) = P\Big(Z \geq \frac{15}{20}\Big) = 1 - \Phi\Big(\frac{15}{20}\Big)$$

(b) Every year, each UConn undergraduate student participates in a lottery for a chance to win a dinner with Geno. Each year, an average of 3 students win this lottery. Using a Poisson approximation, approximate the probability that at most 2 students will win dinner with Geno next year.

Solution: Let X be the number of students who win dinner with Geno. Let n be the undergraduate population at UConn and p be the probability an individual undergraduate wins a dinner with Geno. The assumption of the problem is that $np = 3$. In reality, it is reasonable to assume X is a binomial random variable with parameters n and p. Using the Poisson approximation of a binomial random variable, we have

$$\begin{aligned} P(X \leq 2) \approx P(\text{Pois}(3) \leq 2) &= P(\text{Pois}(3) = 0) + P(\text{Pois}(3) = 1) + P(\text{Pois}(3) = 2) \\ &= e^{-3} + 3e^{-3} + \frac{3^2}{2} e^{-3} \\ &= \frac{17}{2} e^{-3}. \end{aligned}$$

(c) Give a brief, but reasonable explanation as to why a normal approximation is a reasonable choice for the first part, whereas a Poisson approximation is a reasonable choice for the second.

Solution: In part (a), we notice that $np(1-p) = 400$ is reasonably large compared to $n = 2500$, which is our first indication that a normal approximation is reasonable here. In contrast, in part (b), we have that $np = 3 \ll n$ where n is the (large!) student population at UConn. This indicates that a Poisson process is a prudent choice. Although based on gut feeling rather than theory, it is worth noting that in part (a) we are asking for the probability of a large range of values for X, which would make an approximation by a discrete random variable (like a Poisson process) in which we need to sum all terms much less tractable; in part (b), there is a very limited range of values of X we need to consider, which is easy to sum, and suggests that an approximation by a continuous random variable might not be judicious (like a normal approximation).