TEST #2
STA 5326 **Name:**________________________________
November 7, 2019

Please read the following directions.
DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

Directions

- The exam is closed book and closed notes. You will be supplied with scratch paper, and a copy of the Table of Common Distributions from the back of our textbook.
- During the exam, you may use ONLY what you need to write with (pens, pencils, erasers, etc) and (if you wish) an ordinary scientific calculator (TI-86 or below is fine).
- All other items (INCLUDING CELL PHONES) must be left at the front of the classroom during the exam. This includes backpacks, purses, books, notes, etc. You may keep small items (keys, coins, wallets, etc., but NOT CELL PHONEs) so long as they remain in your pockets at all times.
- Partial credit is available. (If you know part of a solution, write it down. If you know an approach to a problem, but cannot carry it out – write down this approach. If you know a useful result, write it down.)
- You must show and explain your work (including your calculations) for all the problems except those on the last two pages. **No credit is given without work or explanation!** But don't get carried away! Give enough explanation and work so that what you have done is clearly understandable.
- Make sure that the grader can easily see how you get from one step to the next. If you needed scratch paper to work something out, make sure to transfer your work to the exam.
- You should give only one answer to each problem. **Circle your answer** if there is any chance for confusion.
- Simplify your answers when it is easy to do so. But more difficult arithmetic does **not** have to be done completely (except as noted below). Answers can be left as fractions or products. You do not have to evaluate large binomial coefficients, factorials or powers. Answers can be left as summations (unless there is a simple closed form such as when summing a geometric or exponential series).
- **On this exam, problems 1(a), 1(b), and 5 require decimal answers for full credit.**
- All algebra and calculus must be done completely. (Only arithmetic can be left incomplete.)
- Do **not** quote homework results. If you wish to use a result from homework in a solution, you must prove this result.
- All the work on the exam should be your own. No "cooperation" is allowed.
- The exam has **9** problems and **11** pages. There are a total of **100** points.

Problem 1. Joe really likes Nutro Meal Replacement Bars and eats a lot of them. They are currently holding a sales promotion: each bar has probability 1/2 of having a coupon wrapped with it. (The bars are independent of each other.) **The different parts below are NOT related.**

(a) (12%) Joe is going to place an order for M Nutro Bars. He would like to have a probability of at least 0.90 of getting at least 100 coupons. What is the smallest value of M which Joe could choose? Use an appropriate approximation. (Your final answer should be an integer. You may wish to use some trial and error to find your answer.)

Solution: Let X be the number of coupons that come with an order of M Nutro Bars. Then $X \sim \text{Binomial}(n = M, p = 1/2)$. Let $X^* \sim N(\mu = M/2, \sigma^2 = M/4)$ be a normal rv with the same mean and variance as X.

Let $Z \sim N(0,1)$. From the tables we learn that $P(Z \leq 1.28) = 0.89973 \approx 0.90$ so that $P(Z \geq -1.28) \approx 0.90$. Using a normal approximation with a continuity correction, the problem becomes one of finding the smallest integer M such that $P(X \geq 100) \geq 0.90$ which implies $P(X^* \geq 99.5) = P(Z \geq (99.5 - (M/2))/\sqrt{M/4}) \geq 0.90$ which implies $\dfrac{99.5 - (M/2)}{\sqrt{M/4}} \leq -1.28$. The smallest M satisfying this may be found using a little trial and error. Alternatively, one may let $y = \sqrt{M}$ and solve the quadratic equation $\dfrac{99.5 - (y^2/2)}{y/2} = -1.28$. Then M is the value of y^2 rounded up to the next larger integer. Either way leads to $M = 218$.

Note: Using `pbinom` in R you can check that the answer 218 is correct. Without the continuity correction you get $M = 219$, which is one too large.

(b) (8%) Later Joe, hearing that Elizabeth Warren might win the next election and fearing the imminent collapse of civilization, decides to use his entire life savings to order a shipment of 40,000 Nutro Bars. What is the probability that he receives **exactly** 20,000 coupons with his shipment? Use an appropriate approximation. (To receive full credit you must give a decimal answer which does NOT use the normal tables.)

The ending of this solution is similar to the solution of exercise C5. See also the "probability of a tie" discussion on pages 25 to 27 of notes7.pdf.

Let

$$X \sim \mathit{Binomial}(n = 40000, p = 0.5) \quad \mathit{and}$$
$$X^* \sim N(\mu = np = 20000, \sigma = \sqrt{np(1-p)} = \sqrt{40000/4} = 100)\,.$$

Then

$$P(X = 20000) \approx P(20000 - 0.5 < X^* < 20000 + 0.5) = P(-0.5/100 < Z < 0.5/100)$$

where $Z \sim N(0,1)$.

$$P(-0.005 < Z < 0.005) = \int_{-0.005}^{0.005} \phi(z)\,dz \approx 0.01 \times \phi(0) = \frac{0.01}{\sqrt{2\pi}} = 0.003989423.$$

Note: The exact answer is dbinom(20000,40000,.5) = 0.003989398.

(c) (6%) Suppose that every coupon can be redeemed for a free Nutro Bar, and that each of these free bars also has probability 0.5 of coming with a coupon. Suppose that Joe buys **8** Nutro Bars and always redeems every coupon he receives for a free bar. Let X be the total number of Nutro Bars he ends up with. Give a formula for $f_X(x)$, the probability mass function (pmf) of X.

The solution uses the closure property of the negative binomial distribution on page 18 of notes6.pdf. As a special case of this, we know that adding r independent Geometric(p) rv's gives a NegBinomial(r, p) rv. This property is also mentioned in the discussion on page 14 of notes6.pdf.

If you buy a single bar and get a coupon, then you get another bar, and you keep getting bars until you finally get a bar without a coupon. That is, you keep conducting trials (getting bars) until you get one with no coupon. Getting "no coupon" has probability 1/2 on each trial. Therefore, when you buy one bar, the total number of bars you get has a Geometric($p = 1/2$) distribution. If you buy $r = 8$ bars, then this process gets repeated with each of the $r = 8$ bars, and the number of bars you end up with is the sum of $r = 8$ independent Geometric($p = 1/2$) rv's, which has a NegBinom($r = 8, p = 1/2$) distribution. This is the definition of the negative Binomial discussed in lecture which the text calls the 'alternate' definition. The pmf is given in the appendix:

$$\begin{aligned} f_X(x) &= \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, r+2, \ldots \\ &= \binom{x-1}{7} (1/2)^x, \quad x = 8, 9, 10, \ldots \end{aligned}$$

Problem 2. (12%) Prove the "two-way" rule for expectations (also known as "The Law of the Unconscious Statistician"): If $Y = g(X)$ and $Eg(X)$ exists, then

$$\int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx\,.$$

Assume that g is differentiable and strictly increasing.

This is exercise 2.21.

Problem 3. Let X have the density (pdf)

$$f(x) = \frac{2}{\beta^2}\, x e^{-x^2/\beta^2}, \quad 0 < x < \infty \quad (\text{where } \beta > 0).$$

This is similar to exercise 2.22. As in that problem, you can do the integrals either using integration by parts or by using a substitution like $u = x^2$ or $u = x^2/\beta^2$.

(a) (10%) Find EX.

Answer: $b\sqrt{\pi}/2$

(b) (10%) Find $\mathrm{Var}(X)$.

Answer: $b^2\left(1-\dfrac{\pi}{4}\right)$

Problem 4. (12%) A *truncated* discrete distribution is one in which a particular class cannot be observed and is eliminated from the sample space. In particular, if X has range 0, 1, 2, ... and the 0 class cannot be observed, the 0-*truncated* random variable X_T has mass function (pmf)

$$P(X_T = x) = \frac{P(X = x)}{P(X > 0)}, \quad x = 1, 2, 3, \ldots$$

Find the pmf, mean, and variance of the 0-*truncated* random variable starting from $X \sim \text{Poisson}(\lambda)$. (Circle these quantities in your solution.)

This is exercise 3.13(a).

Problem 5. (12%) There are 10,000 motorcyclists in a city. The probability that an individual motorcyclist will suffer a serious or fatal accident in a one year period depends on their age as described in the table below. The table also lists the number of motorcyclists in each age category.

Group	# in Group	Prob. of Serious or Fatal Accident
Age $<$ 21 years	2,000	8×10^{-4}
Age between 21 and 50	5,000	2×10^{-4}
Age $>$ 50 years	3,000	4×10^{-4}

Assuming the motorcyclists are independent, what is the probability of exactly 5 serious or fatal accidents in the next year? (Use an appropriate approximation.)

This is similar to exercise C2.

Answer:
Let $X = (\#$ *of serious or fatal accidents*$)$.
Then Poisson$(\lambda = 3.8)$ *and* $P(X = 5) = \dfrac{\lambda^5 e^{-\lambda}}{5!} = 0.1477127$

In the remaining problems you will receive full credit just for stating the correct answer. You are not required to show work or give explanation.

Problem 6. (5%) Let $X_1, X_2, X_3, \ldots, X_n$ be iid exponential random variables with mean $1/\lambda$. For $t > 0$, what is the value of $P(X_1 + X_2 + \cdots + X_n > t)$? (For full credit, your answer should be a **summation** in which the terms are given explicitly.)

This question is based on the discussion on pages 15-16 of notes7.pdf. The desired sum is given on page 16; just replace r by n in the range of summation:

$$\sum_{i=0}^{n-1} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

Problem 7. (5%) Let $X \sim \text{Poisson}(\lambda = 13)$. What is the moment generating function (mgf) of $Y = \dfrac{X-5}{19}$. (You may use the appendix.)

Answer: $M_Y(t) = e^{-(5t/19)} \exp(13(e^{t/19} - 1))$

Use the appendix for the Poisson mgf. Then use the scaling properties on page 10 of notes5.pdf. Another relevant example is on page 24 of notes5.pdf.

Problem 8. (4%) Let $X_1, X_2, X_3, \ldots, X_n$ be iid exponential random variables with mean β. Define $Y = \min X_i$. For $t > 0$, what is $h_Y(t)$, the hazard function of Y? (Give a simple expression.)

This question is based on the discussion on page 6 of notes7.pdf which gives the answer $h_Y(t) = \dfrac{n}{\beta}$.

This may be derived by combining the two facts:
The minimum of n *iid exponential rv's with mean* β *is an exponential rv with mean* β/n.
The hazard function $h(t)$ *of an exponential rv with arbitrary mean* β *is* $h(t) = 1/\beta$ *for* $t > 0$.

Problem 9. (4%) Suppose $X_1, X_2, X_3, \ldots$ are iid with mean μ and variance $\sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Which **one** of the following responses is a correct statement of the **Strong** Law of Large Numbers (**S**LLN)?

See page 17 of notes7.pdf.

a) $\lim_{n\to\infty} P\left(\bar{X}_n = \mu\right) = 1$

b) $\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| < \varepsilon\right) = 1$ for all $\varepsilon > 0$

c) $\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| = 0\right) = 1$

d) $P\left(\lim_{n\to\infty} |\bar{X}_n - \mu| > \varepsilon\right) = 1$ for all $\varepsilon > 0$

e)$\star$ $P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$

f) $\lim_{n\to\infty} P\left(|\bar{X}_n - \mu| > \varepsilon\right) = 1$ for all $\varepsilon > 0$