TEST #2
STA 5326 **Name:**________________________________
November 6, 2018

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, a copy of the Table of Common Distributions from the back of our textbook, and a table of normal probabilities.
- 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). **Some problems require numerical answers.**
- 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 page. **No credit is given without work.** 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. 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).
- 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 **8** problems and **9** pages. There are a total of **100** points.

Problem 1. There are 10,000 high school students who attended summer basketball clinics last summer.

(a) (6%) Suppose that each of these 10,000 students has probability **.02** of receiving a scholarship to play college basketball. Let X be the number of these students who receive basketball scholarships. What is the exact distribution of X? (Assume the students are independent.)

The exact distribution of X is Binomial$(n = 10000, p = .02)$.

(b) (6%) What is a good approximation to the distribution of X? (Give the name of the approximate distribution and the value(s) of any parameter(s).)

The distribution of X is approximately $N(\mu = np = 200, \sigma^2 = np(1-p) = 196)$.

(c) (6%) Use this approximation to compute an approximate numerical value for $P(X > 210)$.

The normal approximation with continuity correction leads to the answer $P(X > 210) = P(X \geq 211) = P(X^* \geq 210.5) = P(Z > (210.5 - 200)/14) = P(Z > 0.75) = 1 - \Phi(0.75) = 0.2266274$ or 0.2266 using the tables. The continuity correction makes a definite improvement in this case and should be used.

In the above X^* is a normal rv with the same mean and variance as X, and $Z \sim N(0,1)$.

[**Problem 1 continued**]

(d) (6%) Suppose that each of the 10,000 students has a probability of **.0004** of eventually playing on an NBA basketball team. Let Y be the number of these students who will play on an NBA team. What is a good approximation to the distribution of Y? Give the name of the approximate distribution and the value(s) of any parameter(s). (Assume the students are independent.)

A good approximation is $Y \approx Poisson(\lambda = np = 10000 \cdot 0.0004 = 4)$.

This is similar to the example on page 31 of notes6.pdf.

(e) (6%) Use this approximation to compute an approximate numerical value for $P(Y \geq 2)$.

$P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1) = 1 - e^{-\lambda} - \lambda e^{-\lambda} = 1 - e^{-4} - 4e^{-4} = 0.9084218.$

Problem 2. (13%) Suppose the random variable X has pdf $f(x)$ and cdf $F(x)$. Define

$$h(t) = \lim_{\delta \to 0+} \frac{1}{\delta} P(t < X \leq t + \delta \,|\, X > t).$$

Derive a simple formula for $h(t)$.

This is exercise 3.25.

Problem 3. Let X_1, X_2, X_3, X_4, X_5 be iid random variables with density $f(x) = 3x^2$ for $0 < x < 1$. Define $Y = \max_{1 \le i \le 5} X_i$.

This is similar to exercise C3.

(a) (7%) Find the cumulative distribution function (cdf) of Y.

(b) (7%) Find EY.

Problem 4. (13%) Let X be a random variable with pdf $f(x)$ and cdf $F(x)$. Suppose that $f(x) = 0$ for $x < 0$ so that X is a nonnegative random variable. Show that

$$EX = \int_0^\infty (1 - F(x))\,dx, .$$

This is exercise 2.14. The easiest solution is that given in the posted solutions. This solution relies on interchanging the order of integration in a double integral. This interchange is always valid for nonnegative integrands. (It is even valid when the double integral is infinite; in that case you get the value infinity no matter which order of integration is used.)

Some students tried to give solutions using integration by parts. It is possible to give a correct solution this way, but it involves an irritating technicality which makes it actually harder than the "interchange" approach.

For any value $L \in (0, \infty)$, integration by parts leads to

$$\int_0^L x f_X(x)\,dx = x F_X(x)\Big|_0^L - \int_0^L F_X(x)\,dx \;=\; L F_X(L) - \int_0^L F_X(x)\,dx$$

which if we let $L \to \infty$ becomes

$$\int_0^\infty x f_X(x)\,dx = \infty - \infty$$

which leads to a dead end; further algebraic manipulations involve illegal or undefined operations.

We can improve on the above. We can replace $F_X(x)$ in the integration by parts by any function whose derivative is $f_X(x)$. If we use $F_X(x) - 1$, this leads to

$$\begin{aligned}
&\int_0^L x f_X(x)\,dx \\
&= x(F_X(x) - 1)\Big|_0^L - \int_0^L (F_X(x) - 1)\,dx \;=\; L(F_X(L) - 1) - \int_0^L (F_X(x) - 1)\,dx \\
&= -L(1 - F_X(L)) + \int_0^L (1 - F_X(x))\,dx
\end{aligned}$$

which if we let $L \to \infty$ becomes

$$\int_0^\infty x f_X(x)\,dx = -\lim_{L\to\infty} L(1 - F_X(L)) + \int_0^\infty (1 - F_X(x))\,dx$$

which gives the desired result so long as $\lim_{L\to\infty} L(1 - F_X(L)) = 0$ whenever EX is finite. It is a little irritating but not too difficult to prove this, but it does have to be proved. Here is a proof:

$$\textit{If } \int_0^\infty x f_X(x)\,dx < \infty, \textit{ then } \lim_{L\to\infty} \int_L^\infty x f_X(x)\,dx = 0.$$

Therefore, if we take the limit as $L \to \infty$ in the following:

$$0 \le L(1 - F_X(L)) = \int_L^\infty L f_X(x)\,dx \le \int_L^\infty x f_X(x)\,dx\,,$$

we obtain $\lim_{L\to\infty} L(1 - F_X(L)) = 0$ as desired.

Problem 5. (13%) In the Land of Nod the years are independent of each other with each year having probability p of drought. A family in Nod decides to plant one fruit tree at the beginning of each year. A new fruit tree needs its first 3 years to be without drought, otherwise it dies. What is the probability of the event that the first 4 trees the family plants all die and the 5th tree is the first to survive?

This is a disguised version of exercise 3.3. The solution and answer are the same.

Problem 6. Suppose the random variable X has density

$$f_X(x) = \frac{4}{\pi(x+1)(x^2+1)} \quad \text{for } 0 < x < \infty.$$

(a) (3%) Is EX finite? Answer "Yes" or "No" and justify your answer.
(You are NOT required to calculate EX.)

The answer is "Yes". See page 5 of notes4.pdf. Use a similar "eyeballing" strategy here. We want to know if

$$\int_0^\infty x f_X(x)\,dx = \frac{4}{\pi}\int_0^\infty \frac{x}{(x+1)(x^2+1)}\,dx \qquad (\dagger)$$

is finite. When x is large, $\frac{x}{(x+1)(x^2+1)} \approx \frac{x}{x^3} = \frac{1}{x^2}$ *and* $\int_c^\infty \frac{1}{x^2}\,dx$ *is finite for any $c > 0$. Therefore, the integral ($\dagger$) is also finite.*

(b) (3%) Is EX^2 finite? Answer "Yes" or "No" and justify your answer.
(You are NOT required to calculate EX^2.)

Answer: No. The argument is similar to the above. We want to know if

$$\int_0^\infty x^2 f_X(x)\,dx = \frac{4}{\pi}\int_0^\infty \frac{x^2}{(x+1)(x^2+1)}\,dx$$

is finite. When x is large

$$\frac{x^2}{(x+1)(x^2+1)} \approx \frac{1}{x}$$

and $\int_c^\infty \frac{1}{x}\,dx = \infty$ *for any $c > 0$. Therefore, the integral of interest is also ∞.*

The problems on this page require no work. You will receive full credit just for stating the correct answer.

Problem 7. (3%) Let $S_n = \sum_{i=1}^{n} X_i$ where $X_1, X_2, \ldots, X_n$ are iid with a distribution F which has a finite mean and variance. (The variance is positive.) Give an example of a distribution F for which S_n is NOT approximately normal even when $n = 10^9$. (State the name of the distribution and the value(s) of any parameter(s).)

The example given in lecture used the distribution F which is Poisson($\lambda = 10^{-9}$). If a student states Poisson($\lambda = 10^{-9}$), they should get full credit. Of course, any other really really small value of λ is also OK (so long as $10^9\lambda$ is not large).

With this F, using the closure property of the Poisson distribution, the distribution of S_n with $n = 10^9$ is Poisson($\lambda = 1$) which is definitely **not** *approximately normal.*

Many similar examples could be given. For example, F could be the distribution of Bernoulli($p = 10^{-9}$) or Gamma($\alpha = 10^{-9}, \beta$).

Note: This question is based on discussion in lecture. We were discussing how large n has to be in the Central Limit Theorem for the normal approximation to be good. This example was written on the board and is NOT stated anywhere in the posted lecture notes or homework exercises.

Problem 8. Suppose you observe a Poisson process with rate λ starting at time zero. Let T_4 be the time of the fourth arrival in this Poisson process.

(a) (4%) What is the distribution of T_4? (State the name of the distribution and the value(s) of any parameter(s).)

See notes7.pdf, page 15. The answer is Gamma($\alpha = 4, \beta = 1/\lambda$).

(b) (4%) Suppose $t > 0$. Give a formula for $P(T_4 > t)$ which expresses this probability as a sum of simple terms.

See notes7.pdf, page 16.