TEST #2
STA 5326 **Name:**____________________________________
October 29, 2009

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

Directions

- This exam is **closed book** and **closed notes**. (You will have access to a copy of the "Table of Common Distributions" given in the back of the text.)
- The different parts of a problem are sometimes unrelated. If you cannot solve part of a problem, you should still go on to look at the later parts.
- If your answer is valid only for a certain range of values, this should be stated as part of your answer. For example, if a density is zero outside of some interval, this interval should be stated explicitly.
- 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! Show enough work so that what you have done is clearly understandable.
- On problems where work is shown, **circle your answer**. (On these problems, you should give only one answer!)
- 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.)
- All the work on the exam should be your own. No "cooperation" is allowed.
- Arithmetic does **not** have to be done completely **unless numerical answers are requested**. Answers can be left as fractions or products. You do not have to evaluate binomial coefficients, factorials or large powers. Answers can be left as summations (unless there is a simple closed form such as when summing a geometric or exponential series).
- You need only pens, pencils, erasers and a calculator. (You will be supplied with scratch paper.)
- Do **not** quote homework results. If you wish to use a result from homework in a solution, you must prove this result.
- The exam has **9** pages.
- There are a total of **100** points.

Problem 1. (15%) Suppose that X is a continuous random variable with median m, that is, $P(X \leq m) = P(X \geq m) = 1/2$. Show that $E|X - a|$ is minimized when $a = m$.

This is exercise 2.18.

Problem 2. An urn initially contains two balls, one red and one green. This is a Polya urn: after a ball is drawn, it is replaced and then another ball of the same color is added to the urn. Ed decides to randomly draw balls until he finally gets a green ball. Then he will stop. Let X be the total number of balls that Ed draws from the urn.

This is very similar to Exercise $C-1$.

(a) (8%) For $k = 1, 2, 3, \ldots$, what is $P(X = k)$?

$P(X = k) = \dfrac{1}{k(k+1)}$ *just as in Exercise $C-1$.*

(b) (7%) What is EX?

$EX = \infty$ *(or undefined) just as in Exercise $C-1$. Clearly* $EX = \sum kP(X = k) = \sum 1/(k+1) = \infty$. *The mass function in part (a) was discussed (long ago) in lecture, and it was shown in class that* $EX = \infty$. *Students who have the correct answer to part (a) can just quote the lecture result for this part. No calculations need be shown.*

Problem 3. Let X have the density (pdf)

$$f(x) = cx^3 e^{-x^2/\beta^2}, \quad 0 < x < \infty, \quad \beta > 0.$$

This problem is very similar to exercise 2.22. The easiest way to do both parts is probably to substitute $u = x^2/\beta^2$, reducing the integrals to gamma functions.

(a) (7%) Find the value of the normalizing constant c. (c will depend on β.)

$c = \dfrac{2}{\beta^4}$

(b) (8%) Find EX.
[If you could not find the value of c, just leave it as c in your answer.]

$EX = \dfrac{3}{4} b\sqrt{\pi}$. *It is OK if students leave gamma functions in their answer, if they cannot remember* $\Gamma(1/2) = \sqrt{\pi}$.

Problem 4. Suppose X and Y are iid random variables with moment generating function

$$M_X(t) = M_Y(t) = \frac{1}{(1+t)(1-t)} \quad \text{for } -1 < t < 1 \text{ (and undefined or } +\infty \text{ otherwise).}$$

Find the following moment generating functions. In each case, your answer should include the range in which the mgf is well-defined and finite.

(a) (8%) Find the mgf of $3X + 5$.

The mgf is $\dfrac{e^{5t}}{(1+3t)(1-3t)}$ *for* $-1/3 < t < 1/3$.

(b) (7%) Find the mgf of $X + Y$.

The mgf is $\dfrac{1}{(1+t)^2(1-t)^2}$ *for* $-1 < t < 1$.

Problem 5. (16%) Suppose X is a random variable with moment generating function (mgf) given by

$$M(t) = e^{\beta(e^{\lambda t} - 1)} = \exp[\beta(e^{\lambda t} - 1)] \qquad \text{where } \beta > 0.$$

Use the mgf to find the mean and variance of X.

The mean is $\beta\lambda$. The variance is $\beta\lambda^2$. Note that $X \stackrel{d}{=} \lambda Z$ where $Z \sim$ Poisson(β).

mean= ____________________________ variance= ____________________________

Problem 6. Let $X \sim$ Negative Binomial(r, p) according to the textbook definition (in which the possible values are $0, 1, 2, \ldots$). See the appendix for the pmf, mean, variance, and mgf.

In the following parts, compute a **numerical** answer using an appropriate **approximation**. Briefly justify the approximation that you use.

In each part, let $\mu = EX$ and $\sigma^2 = \mathrm{Var}(X)$.

(a) (9%) Suppose $r = \mathbf{10^{10}}$ and $p = \mathbf{.25}$. Find $P\left(\mu - \frac{\sigma}{2} < X < \mu + \frac{\sigma}{2}\right)$.

X is the sum of r (a very large number) iid Geometric random variables (and p is not close to 1, which you will see in the next part is important) so the CLT should give an excellent approximation. The CLT approximation is

$$P\left(\mu - \frac{\sigma}{2} < X < \mu + \frac{\sigma}{2}\right) \approx P\left(\mu - \frac{\sigma}{2} < X^* < \mu + \frac{\sigma}{2}\right) = P(-1/2 < Z < 1/2) = 0.3829$$

where $X^* \sim N(\mu, \sigma^2)$ and $Z = (X^* - \mu)/\sigma \sim N(0, 1)$.

The continuity correction will make very little difference in this case because σ is extremely large. But just to illustrate how it would be done, here are some details. You need to find the smallest and largest integers a and b in the interval $(\mu - \frac{\sigma}{2}, \mu + \frac{\sigma}{2})$. Then

$$\begin{aligned} P\left(\mu - \frac{\sigma}{2} < X < \mu + \frac{\sigma}{2}\right) &= P\left(a \le X \le b\right) \approx P\left(a - \frac{1}{2} < X^* < b + \frac{1}{2}\right) \\ &= P\left(\frac{a - \frac{1}{2} - \mu}{\sigma} < Z < \frac{b + \frac{1}{2} - \mu}{\sigma}\right). \end{aligned}$$

But $\sigma \approx 346{,}410.2$ so that $\frac{1/2}{\sigma}$ is very small and clearly makes very little difference to the answer.

[**Problem 6 continued**]

(b) (6%) Suppose $r = \mathbf{10^{10}}$ and $p = \mathbf{1 - 10^{-10}}$. Find $P\left(\mu - \frac{\sigma}{2} < X < \mu + \frac{\sigma}{2}\right)$.

In this case, r is large, but p is extremely close to 1. Plugging into the formulas for the mean and variance in the appendix, we find $\mu \approx 1$ *and* $\sigma \approx 1$. *Since* X *is nonnegative and integer-valued with* $\mu \approx \sigma \approx 1$, *it is clear that a normal approximation is very bad. But the negative binomial mgf is very close to the Poisson(1) mgf in this case (see exercise 3.15). Using a Poisson(1) to approximate the Negative Binomial gives* $P\left(\mu - \frac{\sigma}{2} < X < \mu + \frac{\sigma}{2}\right) = P(X = 1) \approx e^{-1}$, *the Poisson pmf evaluated at 1.*

The remaining problems are multiple choice. Circle the single correct response. No work is required.

Problem 7. (3%) In the neighborhood of $t = 0$, which one of the following is the moment generating function (mgf) of some distribution?

a) $\dfrac{t}{1-2t}$ **b)** $\dfrac{2t}{1+2t}$ **c)** $\dfrac{2}{1-2t}$ **d)** $\dfrac{1}{t(1-2t)}$ **e)** $\dfrac{2}{t(1-t)}$ **f)** $\dfrac{1}{1-2t}$ **g)** $\dfrac{2t}{1-t}$

The correct answer is $\frac{1}{1-2t}$ *which is the only choice with* $M(0) = 1$.

Problem 8. (3%) Suppose the possible values of a random variable X are $0, 1, 2, \ldots, 9, 10$. For this random variable, what is the set of values of t for which the moment generating function (mgf) $M(t)$ is well-defined and finite?

a) $[0, \infty)$ **b)** $(-\infty, 0]$ **c)** $(0, \infty)$ **d)** $(-\infty, 0)$ **e)** $(-\infty, \infty)$ **f)** $[0, 10]$ **g)** $(0, 10)$

h) $[10, \infty)$ **i)** $(10, \infty)$ **j)** not enough information is given to determine the answer

The answer is $(-\infty, \infty)$. *For bounded random variables, the mgf is finite for all* t.

Problem 9. (3%) Suppose X is a random variable, and y and z are positive values. Then

$$P(X > y \,|\, X > y + z)$$

is equal to which of the following?

a) 0 **b)** 1 **c)** 1/2 **d)** $\dfrac{P(y < X < y+z)}{P(X > y+z)}$ **e)** $\dfrac{P(y < X < y+z)}{P(X > y)}$

f) $\dfrac{P(X > y)}{P(y < X < y+z)}$ **g)** $\dfrac{P(X > y+z)}{P(y < X < y+z)}$ **h)** $\dfrac{P(X > y+z)}{P(X > y)}$ **i)** $\dfrac{P(X > y)}{P(X > y+z)}$

The correct answer is 1. *Note that* $X > y + z$ *implies* $X > y$ *so that* $\{X > y + z\} \subset \{X > y\}$.