fixed bad notation on very last question 6/4/17 4:30pm

STA131A spring 2017 Homework 9 (last one!)
Due at the beginning of class on Wednesday June 7, 2017

format instructions DO NOT STAPLE THE 'TOP SHEET' TO YOUR HOMEWORK. TURN IN BOTH THE (UNSTAPLED) TOP SHEET AND THE HOMEWORK TOGETHER.

All homework pages (except the top sheet) must be stapled (before you come to class).

The first page ('top sheet') should contain ONLY your name, student ID, discussion section, and homework number. Use the format shown below. Do NOT staple to the rest of your homework.

The second page should ALSO contain your name, student ID, discussion section, and homework number. Use the format shown below.

You will not loose any point for not making a 'top sheet'. But if your homework goes missing you will have no way to prove you turned it in.

	Points lost if you don't follow the rule
Write your section number on the right side at the top of the first AND second pages. MAKE SURE YOU WRITE THE SECTION YOU ARE REGISTERED FOR EVEN IF YOU HAVE BEEN ATTENDING A DIFFERENT SECTION	all points (no credit)
Write your name and student ID on the left side at the top of the first AND second pages.	1
Your name should be spelled exactly the same as it is on SmartSite (no nicknames please)	all points (no credit)
Write the homework number in the middle at the top of the first AND second pages. (Any format is fine, for example, HW1, homework 1, etc.)	1
Staple all pages.	1
If your homework is on paper pulled out of a notebook, cut off all of the fringes (from the torn horizontal threads that attached the paper to the notebook).	1

For example, for homework 9 if your name is John Smith, your student ID is 123456789, and you are in section A01, then the top of your FIRST and SECOND pages should look like this

John Smith	Homework 9	A01
123456789		

Be kind to the grader.

- make sure you write your name clearly (so it is easy to read)
- write neatly

The negative binomial distribution is a generalization of the geometric distribution. Where the geometric distribution is used to model the number of coin flips to get 1 head, the negative binomial distribution is used to model the number of coin flips to get r heads.

Negative binomial has probability density function (pdf)

$$f(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r} = \frac{(y-1)!}{(y-r)!(r-1)!} p^r (1-p)^{y-r} \quad y = r, r+1, \ldots \quad 0 < p < 1$$

r is a positive integer

The negative binomial with parameter $r = 1$ is the same as the geometric distribution, which has pdf

$$f(y) = p(1-p)^{y-1} \quad y = 1, 2, \ldots, \ldots \quad 0 < p < 1$$

1. For each of the following, determine if X and Y are independent by showing certain probabilities are equal or unequal.

   (a)

X \ Y	0	1	2
0	0.1	0.2	0.3
1	0.05	0.1	0.15
2	0.03	0.03	0.04

   (b)

X \ Y	0	1	2
0	0.1	0.2	0.1
1	0.1	0.2	0.1
2	0	0.1	0.1

   (c)

X \ Y	1	2
3	0.2	0.3
4	0.1	0.4

2. The joint distribution of X and Y is

   $f_{X,Y}(x, y) = x + y \quad 0 < x < 1, \quad 0 < y < 1$

   (a) Find $P(0 < x < \frac{1}{2}, 0 < y < \frac{1}{2})$.

   (b) Find $E(Y)$. There are three different methods you may use to find this. (You only need to use one method for homework, but might want to verify your answer with a second method.)

   (1) $E(Y) = \int_0^1 y f_{X,Y}(x, y) dy$

   (2) $E(Y) = \int_0^1 y f_Y(y) dy$

   (3) $E(Y) = E_X\big[E(Y|X)\big] = \int_0^1 E(Y|X) f_X(x) dx$

(c) Find $P(X < 2Y)$.

3. The example we did in lecture on 5-24-17 was

$f_{X,Y}(x, y) = e^{-y} \quad 0 < x < y < \infty$

We found the following distributions.

The marginal pdf of X.

$f_X(x) = e^{-x} \quad x > 0$

which is Exponential(1) = Gamma(1,1)

The marginal pdf of Y.

$f_Y(y) = ye^{-y} \quad y > 0$

which is Gamma(2,1)

The pdf of Y conditional on $X = x$.

$f_{Y|X=x}(y) = e^{-(y-x)} \quad x < y < \infty$

$x < y < \infty$ is the support of Y conditional on $X = x$

x is a parameter

the parameter space for x is $(0, \infty)$
(a parameter space is the set of possible values for the parameter)

The pdf of X conditional on $Y = y$.

$f_{X|Y=y}(x) = \frac{1}{y} \quad 0 < x < y$

the parameter space for y is $(0, \infty)$

(a) Confirm that $f_{X,Y}(x, y) = f_{y|x}(y|x)f_x(x)$. No answer is required for this question.

(b) Confirm that $f_{X,Y}(x, y) = f_{x|y}(x|y)f_y(y)$. No answer is required for this question.

(c) We also found the expected value of X conditional on $Y = y$

$E(X|Y = y) = \frac{y}{2}$

(Note: a mistake in the notes was corrected on 5-28)

Since the distribution of X is Exponential(1), we have $E(X) = 1$.
Show that $E_Y\big[E(X|Y)\big] = E(X) = 1$. To use $E_Y\big[E(X|Y)\big]$ (the double expectation rule) we take expectation of $E(X|Y)$ with respect to Y. That is

$E_Y\big[E(X|Y)\big] = \int\limits_{y \in \mathcal{Y}} E(X|Y = y)f_Y(y)dy$

(d) Find the covariance between X and Y: i.e, find $cov(X, Y)$.

Hint: when you find $E(XY)$ it is easier if you do the integration with respect to x first. That is

$$E(XY) = \int_0^\infty \int_0^y f_{X,Y}(x, y)dxdy = \int_0^\infty \left[\int_0^y f_{X,Y}(x, y)dx \right] dy$$

where $(0, \infty)$ are the limits for y and $(0, y)$ are the limits for x.

(e) Find the correlation between X and Y.

$$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

You do not need to find $var(X)$ and $var(Y)$ from the pdfs. Use the fact that marginally both X and Y have Gamma distributions, and the variance of a Gamma variable is $\alpha\beta^2$.

4. Let X have a geometric distribution with parameter p, Y have a geometric distribution with parameter p, and X and Y be independent. Define $Z = X + Y$. Show that the distribution of Z is negative binomial with parameters $r = 2$ and p.

   Do not use the method of finding the moment generating function. Instead use the convolution rule for the sum of two independent variables (see lecture notes 5/26).

5. Let X have a Exponential distribution with parameter λ, Y have a Exponential distribution with parameter λ, and X and Y be independent. Define $Z = X + Y$. Find the distribution of Z.

   Do not use the method of finding the moment generating function. Instead use the convolution rule for the sum of two independent variables (see lecture notes 5/26).

6. Suppose that a particle counter is imperfect and independently detects each incoming particle with probability p. If the distribution of the number of incoming particles in a unit of time is Poisson with parameter λ, what is the distribution of the number of counted particles?

   Follow these steps.

   (a) Let N denote the true number of particles and X the number of counted particles. Then (from the description) we have

   $N \sim$Poisson(λ)

   Note: We define the random variable here as N (instead of, for example X, Y or Z to remind us that the N is a nonnegative integer. The notation works the same: capital N for the random variable and lower case n for a value (a number).

   $X|N = n \sim$Binomial(n, p)

   That is, the distribution of X (the number of particles counted) conditional on $N = n$ (the true number of particles) is Binomial with n trials and probability of success p.

   Notice we have a dependent support for the joint distribution of N and X because it must be $N \geq X$ (because the true number of particles must be large than or equal to the number of particles counted), so we have $X = 0, 1, \ldots, N$ and $N = X, X + 1, \ldots$.

   (b) Write down the joint density function for N and X, which has the form

   $f_{X,N}(x, n) = \mathrm{P}(X = x, N = n) = \mathrm{P}(N = n)\mathrm{P}(X = x|N = n)$

   Put in the Poisson pdf with parameter λ for $\mathrm{P}(N = n)$ and the Binomial pdf with parameters n and p for $\mathrm{P}(X = x|N = n)$.

(c) The question asks you to find the distribution of the number of counted particles, which is the marginal distribution of X. We find this marginal distribution by summing the joint pdf over the support of N

$$f_X(x) = \mathrm{P}(X = x) = \sum_{n=0}^{\infty} p(N = n)\mathrm{P}(X = x|N = n)$$

(d) Tricky step. You need to change the limits of the summation to match the values in the joint support, i.e. the values for n where $f_{X,N}(x,n) \neq 0$. (Look at (a) for the joint support.)

(e) Another tricky step. You need to manipulate the terms so that you get the pfd of a Poisson. You will also need to manipulate the limits of the summation so they correspond to the support of a Poisson. Then you will have an infinite sum of a pdf (which is 1).

7. Covariance is defined as $cov(X,Y) = E\Big\{\big[X - E(X)\big]\big[Y - E(Y)\big]\Big\}$. Note that the outer expectation is being taken with respect to the joint distribution of X and Y. Viewing $\big[X - E(X)\big]\big[Y - E(Y)\big]$ as a function of X and Y, we find the expectation of this function for two continuous variables as

$$E\Big\{\big[X - E(X)\big]\big[Y - E(Y)\big]\Big\} = \int_{y\in\mathcal{Y}} \int_{x\in\mathcal{X}} \big[X - E(X)\big]\big[Y - E(Y)\big] f_{X,Y}(x,y)dxdy$$

assuming here X and Y have independent support.

However, for computation we usually use
$cov(X,Y) = E(XY) - E(X)E(Y)$.

Show that $cov(X,Y) = E(XY) - E(X)E(Y)$.

Note: you do NOT need to use any integrals.

8. If two random variables X and Y are independent, then $cov(X,Y) = 0$. However the reverse is not necessarily true. That is, if $cov(X,Y) = 0$, then X and Y may be either independent or dependent. However, if X and Y have a joint normal distribution, then if $cov(X,Y) = 0$ then X and Y are independent. This is also true if X and Y are Bernoulli. That is, if $cov(X,Y) = 0$ then X and Y are independent.

Show that if X and Y are both Bernoulli variables and $cov(X,Y) = 0$, then X and Y are independent. Hints

(a) If Z is a Bernoulli variable, then $E(X) = \mathrm{P}(X = 1)$

(b) If W and Z are both Bernoulli variables, then WZ is also a Bernoulli variable, so $E(WZ) = \mathrm{P}(WZ = 1)$.

(c) If W and Z are both Bernoulli variables, then $\mathrm{P}(XY = 1) = \mathrm{P}(X = 1, Y = 1)$

(d) Start with setting $cov(X,Y) = 0$, which means
$0 = E(XY) - E(X)E(Y) = \cdots$
$=$ (some function of probabilities of X and Y that show independence)

9. If X and Y are both Bernoulli variables, then X and Y are independent if $\mathrm{P}(Y = 1|X = 0) = \mathrm{P}(Y = 1|X = 1)$. If X and Y are not independent, then $\mathrm{P}(Y = 1|X = 1) - \mathrm{P}(Y = 1|X = 0)$ is a measure of the dependency. Since this difference is a measure of dependency and both X and Y are Bernoulli, and two

Bernoulli variables are dependent if and only if $cov(X, Y) = 0$, then there must be a relationship between the covariance and $\mathrm{P}(Y = 1|X = 1) - \mathrm{P}(Y = 1|X = 0)$. Show that

$$\mathrm{P}(Y = 1|X = 1) - \mathrm{P}(Y = 1|X = 0) = \frac{cov(X, Y)}{var(X)}$$

Hint: In question 8b you started with expected values and changed them to probabilities. In this question you start with the probabilities $\mathrm{P}(Y = 1|X = 0)$ $\mathrm{P}(Y = 1|X = 1)$ and change them both to expected values.

10. X is a random variable with a mean 2 and variance of 3, and Y is a random variable with mean of 4 and variance of 5, and the covariance between X and Y is 3. Define $W = 2X - 3Y + 4$.

    (a) Find the $E(W)$.

    (b) Find the $var(W)$.

    (c) Find $cov(X, 2Y)$

    (d) Find $cov(4X, 2Y)$

11. Two Bernoulli variables have a joint distribution.

X \ Y	1	2
3	0.2	0.3
4	0.1	0.4

    (a) Find the covariance between X and Y.

    (b) Find the correlation between X and Y.

12. All of the rules for working with the joint distribution of two random variables extend to joint distributions of more than two random variables. Suppose we have a joint distribution of continuous variables X, Y, and Z and the support is

$$0 \leq x \leq 1$$
$$0 \leq y \leq 2$$
$$0 \leq z \leq 3$$

For each of the following, show the steps you would take to find the distribution. This should have the form

pdf we want = some function of other pdf

any integrations necessary to find the necessary pdfs.
OR

pdf we want = some integration of other pdfs

For example, to find $f(x|y, z)$

$$f(z|x, y) = \frac{f_{x,y,z}(x, y, z)}{f_{x,y}(x, y)}$$

$$f_{x,y}(x,y) = \int_0^3 f_{x,y,z}(x,y,z)dz$$

Now you try

(a) $f_z(z)$

(b) $f_{x,y|z}(x, y|z)$

(c) $f_{x|z}(x|z)$