Homework 10 (due December 2, 2009)

Problem 1. Let $X$ and $Y$ be independent binomial random variables with parameters $(n_1, p)$ and $(n_2, p)$ respectively. Prove that $X + Y$ is a binomial random variable with parameters $(n_1 + n_2, p)$.

Solution: This problem is solved in the book, see the section on Sums of Random Variables.

Problem 2. Each day a meteorologist makes a prediction of what time the sun will rise. The difference in minutes between his prediction and the actual time the sun rises is described by a normal random variable with mean 0 and variance 36. What is the probability that in a month of 30 days his prediction is within 5 minutes of the actual sunrise on exactly 20 days?

Solution: Let $X = \text{the number of days among 30 in which his prediction is within 5 minutes}$. This is a binomial random variable where $n = 30$ and $p$ is the probability that any one day his prediction is off by less than 5 minutes. Since his error, which I will denote as the random variable $E$, is a normal random variable with mean 0 ($\mu = 0$), and variance 36 ($\sigma = 6$). Thus,

$$p = P(-5 < E < 5) = P\left(\frac{-5 - 0}{6} < \frac{E - 0}{6} < \frac{5 - 0}{6}\right)$$

$$= \Phi(5/6) - \Phi(-5/6)$$

$$= \Phi(5/6) - 1 + \Phi(5/6)$$

$$= 2\Phi(5/6) - 1 \approx 2(.7977) - 1 = .5953$$

Now since we have $p$ then,

$$P(X = 20) = \binom{30}{20}(.5953)^{20}(1 - .5953)^{10} = .1106$$

Problem 3. The weight that a bridge can support without structural damage is estimated to be 350, in units of 1000s of pounds. The weight of a single car is on average 3 (again, in 1000s of pounds) with standard deviation .3. How many cars would need to be on the bridge for the probability of structural damage to exceed .1?
Solution: Suppose that the weight of car $i$ is $X_i$, then $E[X_i] = 3$ and $Var(X_i) = .09$. We are interested in finding $n$ so that:

$$P(X_1 + X_2 + \ldots + X_n > 350) > .1$$

Since we are assuming that $n$ is large we can apply the central limit theorem and we need to subtract off $n\mu$ and divide by $\sigma\sqrt{n}$ or:

$$P(X_1 + X_2 + \ldots + X_n > 350) = P\left(\frac{X_1 + X_2 + \ldots + X_n - 3n}{.3\sqrt{n}} > \frac{350 - 3n}{.3\sqrt{n}}\right)$$

$$\approx 1 - \Phi\left(\frac{350 - 3n}{.3\sqrt{n}}\right)$$

If we want this to be greater than .1 then we have:

$$1 - \Phi\left(\frac{350 - 3n}{.3\sqrt{n}}\right) > .1$$

or

$.9 > \Phi\left(\frac{350 - 3n}{.3\sqrt{n}}\right)$

Now the first value where $\Phi(x) = .9$ is when $x = 1.29$. Thus to find the $n$ we seek we set

$$1.29 = \frac{350 - 3n}{.3\sqrt{n}}$$

or

$$0.387\sqrt{n} = 350 - 3n$$

or

$$3n + .387\sqrt{n} - 350 = 0$$

Using the quadratic formula we have that:

$$\sqrt{n} = \frac{-0.387 \pm \sqrt{0.387^2 + 4(3)(350)}}{6} = \frac{-0.387 \pm 64.8085625}{6}$$

We must take the $+$ sign since $\sqrt{n}$ cannot be negative. Thus $\sqrt{n} = 10.73$, So with 116 cars the probability will exceed .1.

**Problem 4.** The length of time to fix a certain machine is known to be an exponential random variable with mean .2 hours. If the repair man has 30 machines to fix approximate the probability that he can complete the work in under 5.5 hours. State clearly any assumptions you make.
Solution: Let \( X_i \) be the amount of time to fix machine \( i \). Then each \( X_i \) is an exponential random variable with mean \( \frac{1}{\lambda} \) which is equal to \( \frac{1}{\lambda} \) thus the variance is \( 1/\lambda^2 \) is .04 ad the \( \sigma = .2 \). Now we are interested in \( P(X_1 + X_2 + \ldots + X_{30} < 5.5) \). We can apply the Central Limit Theorem since we are dealing with the sum of identical independent random variables and \( n = 30 \) is large enough for the approximation to be valid. So,

\[
P (X_1 + X_2 + \ldots + X_{30} < 5.5) = P \left( \frac{X_1 + X_2 + \ldots + X_{30} - 30(.2)}{.2\sqrt{30}} < \frac{5.5 - (30)(.2)}{.2\sqrt{30}} \right)
\]

\[
\approx \Phi \left( -\frac{.5}{.2\sqrt{30}} \right)
\]

\[
= 1 - \Phi(.456) = 1 - .6736 = .3264
\]

Problem 5. Problem 8.2 on page 412 of the text.

(a) Give an upper bound for the probability that a student’s test score will exceed 85.

Solution: Here we want an upper bound for \( P(X > 85) \),

\[
P(X > 85) \leq E[X]/85 = 75/85
\]

(b) Suppose that in addition, that the professor knows that the variance of a student’s test score is equal to 25.

Solution: What can be said about the probability that a student will score between 65 and 85? Here we know,

\[
P(65 < X < 85) = P(65 - 75 < X - 75 < 85 - 75) = P(-10 < X - 75 < 10)
\]

\[
= P(|X - 75| < 10) = 1 - P(|X - 75| \geq 10)
\]

If we apply Chebyshev’s Inequality then we have that

\[
P(|X - 75| \geq 10) \leq \frac{25}{100} = .25
\]

Thus,

\[
P(65 < X < 85) = 1 - P(|X - 75| \geq 10) \leq 1 - .25 = .75
\]

(c) How many students would have to take the examination to ensure, with probability at least .9 that the class average would be within 5 of 75? Do not use the Central Limit
Theorem.

**Solution:** The average score of $n$ students is $A_n = (X_1 + X_2 + ... + X_n)/n$. The mean of $A_n = 75$ and the variance of $A_n$ is $25/n$. Now according to Chebyshev’s inequality we have that:

$$P(|A_n - 75| \geq 5) \leq \frac{25}{25n} = \frac{1}{n}.$$ 

So in order to ensure that the average is within 5 with probability of at least .9 then we need the probability that that average is more than 5 from 75 is less than .1 so we must take $1/n < .1$ or $n > 10$. So we must take at least 11 students to assure that we have the desired situation.

**Problem 6.** $X$, the average rainfall in the month of June in Minneapolis, is known to be approximately a normal random variable with mean 8 and variance 4. In Paris, the average rainfall in June is given by a normal random variable, $Y$, with mean 9 and variance 7. If it is assumed that the rainfall in Minneapolis is independent of that in Paris, what distribution ( and with what parameters) would represent the average of $X$ and $Y$?

**Solution:** Since $X$ and $Y$ are normal then the sum of $X$ and $Y$ is also normal with mean equal to the sum of the means and variance equal to the sum of the variances. Thus $X + Y$ is normal with mean 17 and variance 11. Now, the average is $1/2(X + Y)$. This makes the average normally distributed with mean $17/2$ and variance $11/4$.

**Problem 7.** Show that

$$f(x, y) = \begin{cases} \frac{1}{x}, & 0 < y < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that $f(x, y)$ is a true joint density function for two random variable $X$ and $Y$.

**Solution:** To show this we compute the integral of the density function over the region in which it is non-zero which gives:

$$\int_0^1 \int_0^x \frac{1}{x} \, dy \, dx = \int_0^1 \frac{1}{x} \int_0^x \, dy \, dx = \int_0^1 \frac{1}{x} (x - 0) \, dx = 1 - 0 = 1$$

So yes it is indeed a proper density function.

(b) Find the marginal densities $f_X(x)$ and $f_Y(y)$. 
Solution: To find the marginal density for $x$ we hold $x$ fixed and integrate over $y$. Namely,

$$f_X(x) = \int_0^x \frac{1}{y} \, dy = \left. \frac{1}{x} y \right|_0^x = 1, \quad 0 < x < 1$$

Similarly,

$$f_Y(y) = \int_y^1 \frac{1}{x} \, dx = \ln(x)|_y^1 = \ln(1) - \ln(y) = \ln \left( \frac{1}{y} \right), \quad 0 < y < 1$$

(c) Find $E[X]$ and $E[Y]$.

Solution:

$$E[X] = \int_0^1 x f_X(x) \, dx = \int_0^1 x \, dx = \left. \frac{x^2}{2} \right|_0^1 = \frac{1}{2}$$

$$E[Y] = \int_0^1 -y \ln(y) \, dy = -\int_0^1 y \ln(y) \, dy$$

Now to evaluate the remaining integral you can either look up the integral of $y \ln(y)$ or you can find it by integrating by parts which is a bit of a trick letting $u = \ln(y)$ and $dv = y \, dy$. Thus $du = \frac{1}{y} \, dy$ and $v = \frac{y^2}{2}$. This gives,

$$E[Y] = -\int_0^1 y \ln(y) \, dy = -\frac{y^2}{2} \ln(y)|_0^1 + \int_0^1 \frac{y^2}{2} \, dy = 0 + \frac{y^2}{4}|_0^1 = \frac{1}{4}$$

Problem 8. Let $X$ and $Y$ be continuous random variable with joint density given by:

$$f(x, y) = \begin{cases} \frac{x}{5} + cy, & 0 < x < 1, \quad 1 < y < 5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $c$.

$$\int_0^1 \int_{\frac{x}{5}}^5 cy \, dy \, dx = \int_0^1 \left[ \frac{xy}{5} + \frac{cy^2}{2} \right]|_{\frac{x}{5}}^5 \, dx = \int_0^1 \left( x + \frac{25c}{2} \right) - \left( \frac{x}{5} + \frac{c}{2} \right) \, dx$$

$$= \int_0^1 \frac{4x}{5} + 12c \, dx = \frac{2x^2}{5} + 12cx|_0^1 = \frac{2}{5} + 12c$$

Setting this equal to 1 we have that $12c = 3/5$ or $c = 1/20$.

(b) Are $X$ and $Y$ independent? Justify your answer.

One could compute both the marginal densities, but one could also observe that even though the region of definition is not a rectangle, there is no way to factor the function $f(x, y)$ into two functions of $x$ and $y$. Thus, these can’t be independent.
(c) Find \( P(X + Y > 3) \).

\[ \text{Solution:} \]

\[
P(X + Y > 3) = \int_0^1 \int_{x-x}^5 \frac{x}{5} + \frac{y}{20} \, dy \, dx = \int_0^1 \left[ \frac{xy}{5} + \frac{y^2}{40} \right]_{x-x}^5 \]

\[ = \int_0^1 \left( x + \frac{25}{40} \right) - \left( \frac{x(3-x)}{5} + \frac{(3-x)^2}{40} \right) \, dx \]

\[ = \int_0^1 \left( \frac{40x + 25}{40} \right) - \left( \frac{24x - 8x^2}{40} + \frac{9 - 6x + x^2}{40} \right) \, dx \]

\[ = \int_0^1 \left( \frac{22x + 7x^2 + 16}{40} \right) \, dx = \frac{11}{40} x^2 + \frac{7x^3}{120} + \frac{2}{5} x \bigg|_0^1 = \frac{11}{15} \]

(d) Write an expression involving integrals that represents \( P(X > 1/2 | X + Y > 3) \).

**Solution:** Using the definition of conditional probability we have that:

\[
P(X > 1/2 | X + Y > 3) = \frac{P(X > 1/2, X + Y > 3)}{P(X + Y > 3)} = \frac{\int_0^1 \int_{3-x}^5 \frac{x}{5} + \frac{y}{20} \, dy \, dx}{\frac{11}{15}}
\]

**Problem 9.** Suppose that the joint probability mass function for two discrete random variable \(X\) and \(Y\) is given by:

\[
\begin{array}{c|cccc}
  & -1 & 4 & 7 & 10 \\
  x & 0 & .09 & .12 & .06 & .03 \\
 0 & .03 & .04 & .02 & .01 \\
 4 & .12 & .16 & .08 & .04 \\
11 & .06 & .08 & .04 & .02 \\
\end{array}
\]

(a) Find the marginal mass functions for \(X\) and \(Y\).

**Solution:** To compute the marginal mass functions we add the rows to get \( p_X(x) \) and add the columns to get \( p_Y(y) \).

\[
\begin{array}{c|c|c|c|c}
  x & p_X(x) & y & p_Y(y) \\
-4 & .3 & -1 & .3 \\
 0 & .1 & 4 & .4 \\
 4 & .4 & 7 & .2 \\
11 & .2 & 10 & .1 \\
\end{array}
\]
(b) Are $X$ and $Y$ independent? Justify your answer.

**Solution:** If one checks each of the values it is easy to see that $p(x, y) = p_X(x)p_Y(y)$, thus $X$ and $Y$ are indeed independent.

(c) What is the probability $P(X + Y > 10)$?

**Solution:** To compute this probability we add up all the cases that give values greater than 10 so,

$$P(X+Y > 10) = p(4, 7)+p(4, 10)+p(11, 4)+p(11, 7)+p(11, 10) = .08+.04+.08+.04+.02 = .26$$

(d) Compute the probability $P(X = 0|Y = 10)$.

**Solution:** To compute this we recall the definition of conditional probability to give:

$$P(X = 0|Y = 10) = \frac{P(X = 0, Y = 10)}{P(Y = 10)} = \frac{.01}{.1} = .1$$

(e) Find the expected value $E[|X − Y|]$.

**Solution:** To find the expected value we computer we add up the function with its associated probability over both $x$ and $y$. Namely,

$$E[|X − Y|] = \sum_{x} \sum_{y} |x − y|p(x, y)$$

or

$$E[|X − Y|] = 3p(-4, 1) + 8p(-4, 4) + 11p(-4, 7) + 14p(-4, 10) + 1p(0, -1) + 4p(0, 4) + 7p(0, 7) + 10p(0, 10) + 5p(4, -1) + 0p(4, 4) + 3p(4, 7) + 6p(4, 10) + 12p(11, -1) + 7p(11, 4) + 4p(11, 7) + 1p(11, 1)$$

(f) Find the cumulative density for $X$. Express the function with a piecewise definition and sketch a graph.
Solution: The cumulative density function can be written as a piecewise function by:

\[ F_X(x) = \begin{cases} 
0 & \text{ if } x < -4 \\
.3 & \text{ if } -4 \leq x < 0 \\
.4 & \text{ if } 0 \leq x < 4 \\
.8 & \text{ if } 4 \leq x < 11 \\
1 & \text{ if } 11 \leq x 
\end{cases} \]
**Bonus Problem 1.** Let \( X \) be the number of 1’s and \( Y \) be the number of 2’s that occur in \( n \) rolls of a fair die. Compute \( \text{Cov}(X,Y) \).

Hint: First define
\[
X_i = \begin{cases} 
1 & \text{if roll } i \text{ is a 1} \\
0 & \text{otherwise} 
\end{cases}
\]
\[
Y_i = \begin{cases} 
1 & \text{if roll } i \text{ is a 2} \\
0 & \text{otherwise} 
\end{cases}
\]

and write \( X \) and \( Y \) in terms of the \( X_i \)’s. Clearly,
\[
X = \sum_{i=1}^{n} X_i, \quad Y = \sum_{j=1}^{n} Y_j.
\]

Consequently, due to the properties of covariance we have that:
\[
\text{Cov}(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i,Y_j).
\]

Now we compute \( \text{Cov}(X_i,Y_j) \). First, we assume that the results of any particular roll are independent of any other roll, and because of this \( X_i \) and \( Y_j \) are independent if \( i \neq j \). Consequently \( \text{Cov}(X_i,Y_j) = 0 \) when \( i \neq j \). In the case when \( i = j \) we first compute \( E[X_i] = P\{ \text{roll } i \text{ is a 1} \} = \frac{1}{6} \). Similarly \( E[Y_j] = \frac{1}{6} \). Now consider \( X_iY_i \). In order for this product to be nonzero, roll \( i \) would have to be both a 1 and a 2, which is of course not possible, and thus \( E[X_iY_i] = 0 \) and thus \( \text{Cov}(X_i,Y_i) = E[X_iY_i] - E[X_i]E[Y_i] = 0 - \frac{1}{36} = -1/36 \).

Consequently,
\[
\text{Cov}(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i,Y_j) = \sum_{i=1}^{n} \text{Cov}(X_i,Y_i) = -\frac{n}{36}.
\]

**Bonus Problem 2.** Give a combinatorial proof of the identity:
\[
\binom{n}{k} = \sum_{i=k}^{n} \binom{i-1}{k-1}, \quad n \geq k
\]

Hint: Consider the set of number 1 through \( n \), How many subsets of size \( k \) have \( i \) as their highest numbered member?

**Bonus Problem 3.** Prove that:
\[
0 = \sum_{i=1}^{n} (-1)^i \binom{n}{i}
\]
Hint: Use the Binomial Theorem.

**Bonus Problem 4.** Suppose that $P(E|F) = P(E)$. Use the definition of conditional probability to show that $P(F|E) = P(F)$.

**Bonus Problem 5.** Jeff’s factory produces 3 meter beams. Suppose that the amount of weight that their beams can hold before being damaged is a exponentially distributed random variable with mean 500 pounds. The factory sends a beam periodically to a testing center to have its strength tested. The testing center rates the beams on a scale of 1 to 5. A score of 1 is given to beams that support more than 700 pounds before failure, a score of 2 is given to beams that support between 550 and 700 pounds before failure, a score of 3 is given to beams that can support between 450 and 550 pounds before failure and a score of 4 is given if it can support less than 450 pounds before fail. What is the expected score of a randomly chosen beam from Jeff’s factory? What is the standard deviation of the score?