**Fall 2006 Probability Practice Problems**

Here are some problems for extra practice on expectations and variance. Note that these are not meant to be indicative of the difficulty of an exam problem!

**Problem 1.** A couple decides to have children until they have both a boy and a girl. What is the expected number of children that they’ll end up with? Assume that each child is equally likely to be a boy or a girl and genders are mutually independent.

**Solution.** There are many ways to solve this problem. We’ll do it from first principles.

Suppose that a couple has children until they have both a boy and a girl. A tree diagram for this experiment is shown below.

Let the random variable $R$ be the number of children the couple has. From the definition of expectation, we have:

$$E[R] = \sum_{w \in S} R(w) \cdot \Pr(w)$$

$$= \left(2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \ldots\right) + \left(2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \ldots\right)$$

$$= 2 \left(2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \ldots\right).$$

The only difficulty is evaluating the sum. We can use the general formula

$$1 + 2r + 3r^2 + 4r^3 + \ldots = \frac{1}{(1-r)^2}$$
which is obtained by differentiating the formula for the sum of an infinite geometric series. Setting \( r = 1/2 \) gives:

\[
1 + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + \ldots = 4
\]

We have to tweak this a little to get the sum we’re interested in. Subtracting 1 from each side and then dividing both sides by 2 does the trick:

\[
2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + \ldots = \frac{4 - 1}{2} = \frac{3}{2}
\]

So from (1) we have

\[
E[R] = 2 \left( \frac{3}{2} \right) = 3.
\]

A much simpler approach uses the fact that the “mean time to failure” is \( 1/p \) where \( p \) is the probability of failure in one step. If we consider having a child of opposite sex to the first a “failure” of that child, then the mean time to failure is the expected number of children after the first until the couple has both a boy and a girl. But the probability of a failure at the \( k \)th child after the first is \( 1/2 \) for all \( k \geq 1 \). So the expected number of children after the first is \( 1/(1/2) = 2 \), and the expected number of children including the first is \( 1+2 =3 \).

**Problem 2.** A classroom has sixteen desks arranged as shown below.

If there is a girl in front, behind, to the left, or to the right of a boy, then the two of them flirt. One student may be in multiple flirting couples; for example, a student in a corner of the classroom can flirt with up to two others, while a student in the center can flirt with as many as four others. Suppose that desks are occupied by boys and girls with equal probability and mutually independently. What is the expected number of flirting couples?
**Solution.** First, let’s count the number of pairs of adjacent desks. There are three in each row and three in each column. Since there are four rows and four columns, there are $3 \cdot 4 + 3 \cdot 4 = 24$ pairs of adjacent desks.

Number these pairs of adjacent desks from 1 to 24. Let $F_i$ be an indicator for the event that occupants of the desks in the $i$-th pair are flirting. The probability we want is then:

$$E \left[ \sum_{i=1}^{24} F_i \right] = \sum_{i=1}^{24} E \left[ F_i \right] = \sum_{i=1}^{24} \Pr \left( F_i = 1 \right)$$

The first step uses linearity of expectation, and the second uses the fact that the expectation of an indicator is equal to the probability that it is 1.

The occupants of adjacent desks are flirting if the first holds a girl and the second a boy or vice versa. Each of these events happens with probability $1/2 \cdot 1/2 = 1/4$, and so the probability that the occupants flirt is

$$\Pr \left( F_i = 1 \right) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$  

Plugging this into the previous expression gives:

$$E \left[ \sum_{i=1}^{24} F_i \right] = \sum_{i=1}^{24} \Pr \left( F_i = 1 \right) = 24 \cdot \frac{1}{2} = 12$$

**Problem 3.** Here are seven propositions:

$$x_1 \lor x_3 \lor \neg x_7$$
$$\neg x_5 \lor x_6 \lor x_7$$
$$x_2 \lor \neg x_4 \lor x_6$$
$$\neg x_4 \lor x_5 \lor \neg x_7$$
$$x_3 \lor \neg x_5 \lor \neg x_8$$
$$x_9 \lor \neg x_8 \lor x_2$$
$$\neg x_3 \lor x_9 \lor x_4$$

Note that:

1. Each proposition is the OR of three terms of the form $x_i$ or the form $\neg x_i$.
2. The variables in the three terms in each proposition are all different.
Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

(a) What is the probability that a single proposition is true? **Solution.** Each proposition is true unless all three of its terms are false. Thus, each proposition is true with probability:

$$1 - \left(\frac{1}{2}\right)^3 = \frac{7}{8}$$

(b) What is the expected number of true propositions?

**Solution.** Let $T_i$ be an indicator for the event that the $i$-th proposition is true. Then the number of true propositions is $T_1 + \ldots + T_7$ and the expected number is:

$$E [T_1 + \ldots + T_7] = E [T_1] + \ldots + E [T_7]$$

$= 7/8 + \ldots + 7/8$

$= 49/8 = 6\frac{1}{8}$

(c) Use your answer to prove that there exists an assignment to the variables that makes all of the propositions true.

**Solution.** A random variable can not always be less than its expectation, so there must be some assignment such that:

$$T_1 + \ldots T_7 \geq 6\frac{1}{8}$$

This implies that $T_1 + \ldots T_7 = 7$ for at least one outcome. This outcome is an assignment to the variables such that all of the propositions are true.

**Problem 4.** The hat-check staff has had a long day, and at the end of the party they decide to return people’s hats at random. Suppose that $n$ people have their hats returned at random. We previously showed that the expected number of people who get their own hat back is 1, irrespective of the total number of people. In this problem we will calculate the variance in the number of people who get their hat back.

Let $X_i = 1$ if the $i$th person gets his or her own hat back and 0 otherwise. Let $S_n := \sum_{i=1}^{n} X_i$, so $S_n$ is the total number of people who get their own hat back. Show that

(a) $E [X_i^2] = 1/n$.

**Solution.** $X_i = 1$ with probability $1/n$ and 0 otherwise. Thus $X_i^2 = 1$ with probability $1/n$ and 0 otherwise. So $E [X_i^2] = 1/n$.

(b) $E [X_iX_j] = 1/n(n-1)$ for $i \neq j$.

**Solution.** The probability that $X_i$ and $X_j$ are both 1 is $1/n \cdot 1/(n-1) = 1/n(n-1)$. Thus $X_iX_j = 1$ with probability $1/n(n-1)$, and is zero otherwise. So $E [X_iX_j] = 1/n(n-1)$. 
(c) $E[S_n^2] = 2$. Hint: Use (a) and (b).

Solution.

\[
E[S_n^2] = \sum_i E[X_i^2] + \sum_{i \neq j} E[X_iX_j]
\]
\[
= n \cdot \frac{1}{n} + n(n - 1) \cdot \frac{1}{n(n - 1)}
\]
\[
= 2.
\]

(d) $\text{Var}[S_n] = 1$.

Solution.

\[
\text{Var}[S_n] = E[S_n^2] - E^2[S_n]
\]
\[
= 2 - (n(1/n))^2
\]
\[
= 2 - 1
\]
\[
= 1.
\]

(e) Explain why you cannot use the variance of sums formula to calculate $\text{Var}[S_n]$.

Solution. The indicator random variables, $X_i$, are not even pairwise independent. This can be seen by comparing the marginal and conditional probability of a particular person, Alice, getting her hat back. The marginal probability, unconditioned on any other events, is $1/n$ as we’ve computed before. However, if compute this probability conditioned on the event that a second person, Bob, got his hat back, we find that the probability of Alice getting her hat back is $1/(n-1)$.

(f) Using Chebyshev’s Inequality, show that $\Pr(S_n \geq 11) \leq .01$ for any $n \geq 11$.

Solution.

\[
\Pr(S_n \geq 11) = \Pr(S_n - E[S_n] \geq 11 - E[S_n])
\]
\[
= \Pr(S_n - E[S_n] \geq 10)
\]
\[
\leq \frac{\text{Var}[S_n]}{10^2} = .01
\]

Note that the $X_i$’s are Bernoulli variables but are not independent, so $S_n$ does not have a binomial distribution and the binomial estimates from Lecture Notes do not apply.

Problem 5. There are about 250,000,000 people in the United States who might use a phone. Assume that each person is on the phone during each minute mutually independently with probability $p = 0.01$.

(To keep the problem simple, we are putting aside the fact that people are on the phone more often at certain times of day and on certain days of the year, and that if someone is on the phone, it is hardly independent of whether they were on the phone the previous minute.)
(a) What is the expected number of people on the phone at a given moment? Solution. Let $I_i$ be an indicator for the event that the $i$-th person is one the phone. The number of people on the phone is then:

$$\sum_{i=1}^{25,000,000} I_i.$$ 

The expectation of this sum is $250,000,000 \cdot 0.01 = 2,500,00$ by linearity of expectation.

(b) Suppose that we construct a phone network whose capacity is a mere one percent above the expectation. Upper bound the probability that the network is overloaded in a given minute. (Use the approximation formula given in the notes. You may need to evaluate this expression in a clever way because of the size of numbers involved. For example, you could first evaluate the logarithm of the given expression.) Solution. The network is overloaded if the fraction of people calling is greater than $1.01 \cdot 0.01 = 0.0101$. In complementary terms, the network is overloaded if the fraction of people not calling is less than $1 - 0.0101 = 0.9899$. Define the following variables:

$$n := 250,000,000 \quad \text{people in the US}$$
$$p := 0.99 \quad \text{prob. not on phone}$$
$$\alpha := 0.9899 \quad \text{min. allowable fraction not on phone}$$

In these terms, the solution to the problem is $F_{n,p}(\alpha n)$. We can upper bound this approximately using the formula from the notes:

$$F_{n,p}(\alpha n) \approx \frac{1 - \alpha}{1 - \alpha/p} \cdot \frac{2^n H(\alpha)}{\sqrt{2\pi \alpha(1 - \alpha)n}} \cdot p^n (1 - p)^{(1 - \alpha)n}.$$ 

By first evaluating the logarithm of this expression, we find that this is about $e^{-120}$.

(c) What is the expected number of minutes (approximately) until the system is over-loaded for the first time?

Solution. Applying the “expected time to failure” formula with probability $p = e^{-120}$ gives $1/p = e^{120}$.

Problem 6. You flip a fair coin 100 times. Upper bound the probability that you get 75 or more heads using Markov’s bound, Chebyshev’s bound and Chernoff bounds respectively.

Solution. The expected number of heads is 50, so 75 heads is $75/50 = 3/2$ times the expected value of the number of heads. Thus, Markov’s gives $2/3$.

Let $X_i$ be 1 if the $i$-th flip is a heads, and 0 otherwise. The variance of a $X_i$ is $1/2 - 1/4 = 1/4$. Let $S = \sum_{i=1}^{100} X_i$. The variance of $S$ is $100/4 = 25$, and the standard deviation is 5. So Chebyshev’s theorem says that $Pr[S \geq 75] = Pr[|S - E[S]| \geq 25] \leq \frac{25}{25} = \frac{1}{25}$.

Chernoff’s gives $Pr[S \geq 3/2E[S]] \leq e^{-(3/2ln3/2-1/2)50}$, which is about $e^{-5.4}$.