MATH 361 — FALL 2001 — TEST 1

NAME:

Do all questions. No books, notes, extra paper, calculators or computers allowed.

EXPLAIN every answer. Answers with no explanation will receive no credit.

1 (20 points)**.** A study found that 25% of the paintings at a certain art gallery are fakes, and only 75% are originals. It is known that when a particular buyer decides whether or not a painting is fake, he makes the right decision $\frac{4}{5}$ of the time and the wrong decision $\frac{1}{5}$ of the time.

If the buyer decides a painting is an original, then what is the probability it is actually a fake?

2 (20 points)**.** Let $0 < p < 1$. Two baseball teams called the Cubs and the Sox play in the World Series. In each game, the Cubs win with probability p, and the Sox win with probability $1 - p$. The first team to win 4 games will win the series.

Find

(a) P(Cubs win the series after 4 games),

(b) P(Cubs win the series after 5 games).

3 (30 points)**.** A cruise ship has 10 Americans and 12 Brazilians as passengers. For dinner, a group of 8 passengers is selected at random to dine with the captain. Find the probability that:

(a) 5 Americans and 3 Brazilians are selected;

(b) at least one American and at least one Brazilian is selected.

4 (15 points)**.** A bookshelf contains 11 books written by Austen and 20 books by Wodehouse. Randomly remove 10 books from the shelf, one by one. Find the probability that the first four books are by Austen and the fifth book is by Wodehouse.

5 (15 points). Explain in *words* the reason why

$$\binom{n}{r} = \binom{n}{n-r}.$$

Do not use formulas!

MATH 361 — FALL 2001 — TEST 1

NAME: **SOLUTIONS**

1 (20 points)**.** A study found that 25% of the paintings at a certain art gallery are fakes, and only 75% are originals. It is known that when a particular buyer decides whether or not a painting is fake, he makes the right decision $\frac{4}{5}$ of the time and the wrong decision $\frac{1}{5}$ of the time.

If the buyer decides a painting is an original, then what is the probability it is actually a fake?

Solution: First we **read** the question carefully to figure out what we are being asked for. Apparently we want to know P(painting is fake | buyer decides it is original). So we define events

F : painting is a fake,
O : painting is an original,
DF : buyer decides painting is a fake,
DO : buyer decides painting is an original.

Then we want

$$P(F|DO) = \frac{P(F \cap DO)}{P(DO)} = \frac{P(DO|F)P(F)}{P(DO|F)P(F) + P(DO|O)P(O)} = \frac{\frac{1}{5}(0.25)}{\frac{1}{5}(0.25) + \frac{4}{5}(0.75)} = \frac{5}{65} = \frac{1}{13}.$$

This can also be done using a table, as follows:

¿From the table,

$$P(F|DO) = \frac{P(F \cap DO)}{P(DO)} = \frac{\frac{1}{5}(.25)}{.65} = \frac{1}{13}.$$

Alternatively, we could argue as follows, in words. Out of 100 paintings, 75 are original and 25 are fake. The buyer decides that $\frac{4}{5}(75) = 60$ of the originals are original, and decides that $\frac{1}{5}(25) = 5$ of the fakes are original. So the buyer believes that 65 paintings are original, even though 5 of them are really fakes. Hence $P(F|DO) = 5/65 = 1/13$.

Remark. In this problem, we are not trying to find P(painting is a fake | buyer made wrong decision). That would be quite a different question.

2 (20 points)**.** Let $0 < p < 1$. Two baseball teams called the Cubs and the Sox play in the World Series. In each game, the Cubs win with probability p, and the Sox win with probability $1 - p$. The first team to win 4 games will win the series.

Find

(a) P(Cubs win the series after 4 games),

(b) P(Cubs win the series after 5 games).

Solution:

This is a Bernoulli trial situation, with "success" meaning that the Cubs win the game.

(a) The Cubs win the series after 4 games if and only if they win the first 4 games, which happens with probability p^4.

(b) Similarly, the Cubs win the series after 5 games if they win 3 and lose 1 of the first 4 games, and then win the fifth game; this happens with probability

$$\binom{4}{3} p^3(1-p)^1 \cdot p = 4p^4(1-p),$$

where the factor $\binom{4}{3}$ takes account of choosing which 3 games the Cubs win, out of the first 4 games.

Remark. The answer $\binom{5}{4}p^4(1-p)^1$ would be incorrect because it counts also the situation where the Cubs win the first 4 games and lose the 5^{th} game. But in that situation, the Cubs win the series after 4 games, not after 5 games.

Now just for fun, we also compute the following.

(c) The Cubs win the series after 6 games if they win 3 and lose 2 of the first 5 games, and then win the sixth game; this happens with probability

$$\binom{5}{3} p^3(1-p)^2 \cdot p = 10p^4(1-p)^2.$$

(d) The Cubs win the series after 7 games if they win 3 and lose 3 of the first 6 games, and then win the seventh game; this happens with probability

$$\binom{6}{3} p^3(1-p)^3 \cdot p = 20p^4(1-p)^3.$$

(e) The total probability that the Cubs win the World Series is just the sum of these four probabilities (since the events of their winning in 4,5,6 or 7 games are obviously disjoint).

3 (30 points)**.** A cruise ship has 10 Americans and 12 Brazilians as passengers. For dinner, a group of 8 passengers is selected at random to dine with the captain. Find the probability that:

(a) 5 Americans and 3 Brazilians are selected;

(b) at least one American and at least one Brazilian is selected.

Solution:

(a) This is a partitioning problem. The total number of passengers is 22. Thus there are

$$\binom{10}{5}\binom{12}{3}\Big/\binom{22}{8}$$

ways to choose a group of 8 that contains 5 Americans and 3 Brazilians.

(b) Define events

A : the group contains no Americans (so it contains only Brazilians),

B : the group contains no Brazilians (so it contains only Americans).

Then we want

$$\begin{aligned}
&P(\text{at least one American and at least one Brazilian})\\
&= P(\text{not } A \text{ and not } B)\\
&= P(A^c \cap B^c)\\
&= 1 - P(A \cup B) \qquad \text{by De Moivre's law.}
\end{aligned}$$

Notice A and B are disjoint events. Thus

$$1 - P(A \cup B) = 1 - P(A) - P(B) = 1 - \frac{\binom{12}{8}}{\binom{22}{8}} - \frac{\binom{10}{8}}{\binom{22}{8}}.$$

Alternatively, if there are k Americans and $8-k$ Brazilians, then we want $1 \le k \le 7$. Hence

$$P(\text{at least one American and at least one Brazilian}) = \sum_{k=1}^{7} \binom{10}{k}\binom{12}{8-k}\Big/\binom{22}{8}.$$

Remark. A solution like

$$\binom{10}{1}\binom{12}{1}\binom{20}{6}\Big/\binom{22}{8},$$

in which one chooses an American, then a Brazilian, and then chooses 6 others, is incorrect because it counts each group more than once, in the numerator. For example, the group (Andy, Jorge, Alice, 5 others) will be counted again as (Alice, Jorge, Andy, 5 others).

Remark. This is not a Bernoulli trial situation, since we do not "replace" the passengers after selecting them. For example, if you choose an American as the first passenger (with probability 10/22, then the chance of choosing an American as the second passenger is only 9/21.

Moral. If a problem involves the phrase "at least", then it is often a good idea to examine instead the complementary event. Exception: sometimes such a problem can be done directly, or can be done using the "union of events" formula.

4 (15 points)**.** A bookshelf contains 11 books written by Austen and 20 books by Wodehouse. Randomly remove 10 books from the shelf, one by one. Find the probability that the first four books are by Austen and the fifth book is by Wodehouse.

Solution: Here we are sampling without replacement.

First we remove four books of the 11 by Austen, which can be done in $(11)_4$ ways. Then we select a book by Wodehouse, which can be done in 20 ways. Then we select five more books (to make a total of ten), from the remaining 26 books, which is possible in $(26)_5$ ways. Finally we divide by the total number of ways of selecting 10 books from 31, in order, which is $(31)_{10}$. So the desired probability is

$$p = \frac{(11)_4 \cdot 20 \cdot (26)_5}{(31)_{10}}.$$

Alternatively, we can examine just the first 5 books:

$$p = \frac{(11)_4 \cdot 20}{(31)_5}.$$

Alternatively, we can just apply common sense to our choices of the first 5 books:

$$p = \frac{11 \cdot 10 \cdot 9 \cdot 8 \cdot 20}{31 \cdot 30 \cdot 29 \cdot 28 \cdot 27},$$

since we have 11 choices for the first book by Austen, 10 choices for the second book, and so on, and 20 choices for the book by Wodehouse.

Alternatively, we can choose 4 books by Austen, then arrange them in all possible orders, then choose a book by Wodehouse, then choose 5 more books which can be arranged in any order. This gives

$$p = \frac{\binom{11}{4} 4! \binom{20}{1} \binom{26}{5} 5!}{\binom{31}{10} 10!}.$$

It's worth checking that all these answers are the same!

Remark. The answer

$$\frac{\binom{11}{4} \binom{20}{1} \binom{26}{5}}{\binom{31}{10}}$$

is incorrect, because it does not take account of the number of ways in which the books can be arranged in different orders. That is, it does not contain the terms $4!, 5!$, and $10!$. And this really does make a difference, because $4! \cdot 5! \neq 10!$.

The following answer is incorrect for similar reasons:

$$\frac{\binom{11}{4} \binom{20}{1}}{\binom{31}{5}}.$$

5 (15 points). Explain in *words* the reason why

$$\binom{n}{r} = \binom{n}{n-r}.$$

Do not use formulas!

Solution: Choosing r objects out of n can be done in $\binom{n}{r}$ ways. But choosing r objects out of n is equivalent to specifying which $n-r$ objects you are *not* going to choose, and this can be done in $\binom{n}{n-r}$ ways. Thus

$$\binom{n}{r} = \binom{n}{n-r}.$$

MATH 361 — FALL 2001 — TEST 2

NAME:

Do all questions.
No books, notes, extra paper, calculators or computers allowed. EXPLAIN every answer.

You can use (without proof) any needed formula for the density or expectation of any distribution from class, unless the test question is clearly asking you to derive this formula.

1 (20 points)**.** Let X be a discrete random variable taking values $0, 1, 2, \ldots$.
(a) Define the density function of X.
(b) Define the expectation of X.
(c) Suppose $X \sim \text{Poisson}(\lambda)$. Evaluate EX.

2 (30=5+5+5+15 points)**.** Roll two independent fair dice. Define discrete random variables by

$$X = \text{number rolled on die \#1},$$
$$Y = \text{number rolled on die \#2}.$$

(a) Evaluate EX.
(b) Evaluate $E[2X + 4Y]$.
(c) Evaluate $E[XY]$.
(d) Let $Z = \max(X, Y)$. Find the density of Z.

3 (20 points). Consider discrete random variables X and Y, with the following joint density:

(a) Evaluate $P(X = 1)$, and $P(Y = 1)$.
(b) Are X and Y independent? Explain.

4 (20 points)**.** For each hour that a light bulb is in use, it has probability p of burning out and probability $1-p$ of not burning out.

The manufacturer wants to advertize that its bulbs will "last for 1000 hours or more before burning out". What value of p will allow the manufacturer to advertize this claim, reasonably truthfully?

Hint: let X denote the number of complete hours a bulb is in use before it burns out. Evaluate the density $P(X = x)$ of X.

5 (10 points). Twentyfive women are independently tossing fair coins, with each woman stopping when she first tosses a head. Find the probability that exactly 100 tails are tossed in total, before everyone stops.

MATH 361 — FALL 2001 — TEST 2

NAME: **SOLUTIONS**

Do all questions.
No books, notes, extra paper, calculators or computers allowed. EXPLAIN every answer.

You can use (without proof) any needed formula for the density or expectation of any distribution from class, unless the test question is clearly asking you to derive this formula.

1 (20 points)**.** Let X be a discrete random variable taking values $0, 1, 2, \ldots$.
(a) Define the density function of X.
(b) Define the expectation of X.
(c) Suppose $X \sim \text{Poisson}(\lambda)$. Evaluate EX.

Solution:

(a) (5 points) The density function is defined by

$$f_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}),$$

for all $x \in \mathbb{R}$. That is, the density function simply tells us the probability of the random variable X equalling the value x. *Note.* $P(X = x) = 0$ when $x \neq 0, 1, 2, \ldots$, because X only takes on the values $0, 1, 2, \ldots$. But this is irrelevant for part (a) of the problem.

(b) (5 points) The expectation (or mean) of a random variable is the average of the values of the random variable, weighted by how likely those values are to occur: since X takes on values $0, 1, 2, \ldots$, this means

$$EX = \sum_{x=0}^{\infty} x f_X(x) = \sum_{x=0}^{\infty} x P(X = x).$$

(c) (10 points) For the Poisson distribution we know $P(X = x) = e^{-\lambda}\lambda^x/x!$ for $x = 0, 1, 2, \ldots$, and $P(X = x) = 0$ otherwise. Hence

$$\begin{aligned}
EX &= \sum_{x=0}^{\infty} x(e^{-\lambda}\lambda^x/x!) \\
&= \sum_{x=1}^{\infty} x(e^{-\lambda}\lambda^x/x!) \qquad \text{since the } x = 0 \text{ term drops out} \\
&= e^{-\lambda}\lambda \sum_{x=1}^{\infty} \lambda^{x-1}/(x-1)! \qquad \text{since } x! = x \cdot (x-1)! \\
&= e^{-\lambda}\lambda(1 + \lambda/1! + \lambda^2/2! + \cdots) \\
&= e^{-\lambda}\lambda \cdot e^{\lambda} \\
&= \lambda.
\end{aligned}$$

2 (30=5+5+5+15 points)**.** Roll two independent fair dice. Define discrete random variables by

$$X = \text{number rolled on die \#1},$$
$$Y = \text{number rolled on die \#2}.$$

(a) Evaluate EX.
(b) Evaluate $E[2X + 4Y]$.
(c) Evaluate $E[XY]$.
(d) Let $Z = \max(X, Y)$. Find the density of Z.

Solution:

(a) The density of X is uniform, with $P(X = x) = 1/6$ for $x = 1, 2, 3, 4, 5, 6$, and $P(X = x) = 0$ otherwise. Hence

$$EX = \sum_{x=1}^{6} xP(X = x) = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6} = 3.5.$$

Similarly $EY = 3.5$. This is just common sense: the average of the numbers $1, \ldots, 6$ on the die is 3.5.

(b) $E[2X+4Y] = 2EX+4EY = 21$, where we have used linearity of expectation (Theorem 2 on page 87). *Note.* Linearity of expectation is valid whether or not X and Y are independent. It has nothing to do with independence.

(c) Because X and Y are independent, Theorem 4 on page 89 gives that

$$E[XY] = (EX)(EY) = (3.5)(3.5) = 12.25.$$

(d) [This is like Chapter 3#8.] Clearly Z only takes on values $1, 2, 3, 4, 5, 6$. For these values of z, the density is

$$\begin{aligned} f_Z(z) &= P(Z = z) \\ &= P(\max(X, Y) = z) \\ &= P(X = z, Y < z) + P(X < z, Y = z) + P(X = z, Y = z) \\ &= P(X = z)P(Y < z) + P(X < z)P(Y = z) + P(X = z)P(Y = z) \quad \text{by independence} \\ &= \frac{1}{6} \cdot \frac{z-1}{6} + \frac{z-1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} \\ &= \frac{2z-1}{36}. \end{aligned}$$

For all other values of z we have $f_Z(z) = P(Z = z) = 0$.

Alternatively, for $z = 1, 2, 3, 4, 5, 6$ the density is

$$\begin{aligned}
f_Z(z) &= P(Z = z) \\
&= P(\max(X, Y) = z) \\
&= P(\max(X, Y) \leq z) - P(\max(X, Y) < Z) \\
&= P(X \leq z, Y \leq z) - P(X < z, Y < z) \\
&= P(X \leq z)P(Y \leq z) - P(X < z)P(Y < z) \qquad \text{by independence} \\
&= \frac{z}{6} \cdot \frac{z}{6} - \frac{z-1}{6} \cdot \frac{z-1}{6} \\
&= \frac{2z-1}{36}.
\end{aligned}$$

Alternatively, one can examine the value table of $Z = \max(X, Y)$. To find the density $P(Z = z)$, we need only look to see how many entries in the table have value z. Note that every entry in the table has probability $1/36$, since $P(X = x, Y = y) = P(X = x)P(Y = y) = 1/36$, by independence of X and Y.

3 (20 points)**.** Consider discrete random variables X and Y, with the following joint density:

(a) Evaluate $P(X = 1)$, and $P(Y = 1)$.
(b) Are X and Y independent? Explain.

<u>*Solution:*</u>

(a) (8 points) Adding across the first row of the density table gives

$$P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 2) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Adding down the first column of the density table gives

$$P(Y = 1) = P(X = 1, Y = 1) + P(X = 2, Y = 1) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}.$$

(b) (12 points) X and Y are not independent, because independence would require $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x and y, whereas the table and part (a) give $P(X = 1, Y = 1) = \frac{1}{8} \neq \frac{1}{4} \cdot \frac{3}{8} = P(X = 1)P(Y = 1)$.

Remark. Knowing the *definition* of independence is crucial for this problem.

Remark. A problem of this general nature might arise in an experimental situation, where you can obtain the joint density function by measuring the values of X and Y on many repetitions of the experiment, under a wide variety of experimental conditions. Then you start wondering whether X and Y are independent, or whether they are perhaps related in some way. The joint density table can answer that question, as shown here.

4 (20 points)**.** For each hour that a light bulb is in use, it has probability p of burning out and probability $1-p$ of not burning out.

The manufacturer wants to advertize that its bulbs will "last for 1000 hours or more before burning out". What value of p will allow the manufacturer to advertize this claim, reasonably truthfully?

Hint: let X denote the number of complete hours a bulb is in use before it burns out. Evaluate the density $P(X=x)$ of X.

<u>*Solution:*</u>

To obtain $X=x$ we see that a bulb must not burn out during each of the first x hours, and then must burn out during the following hour. That is, $P(X=x)=p(1-p)^x$ for $x=0,1,2,\ldots$, which is the geometric density.

The expectation of a geometric random variable is $EX=\frac{1}{p}-1$. We want this expectation (which is the expected number of hours a bulb will be in use before burning out) to equal 1000, for that would mean the manufacturer is being reasonably truthful in claiming that its bulbs last for 1000 hours (on average) before burning out.

Solving $\frac{1}{p}-1=1000$ for p gives

$$p=1/1001\approx .001.$$

Of course, a smaller value of p would give an even longer expected lifetime for the bulb.

Alternative. Another reasonable solution would be for the manufacturer to require that half the bulbs last for 1000 hours or more. That means

$$\frac{1}{2}=P(X\geq 1000)=(1-p)^{1000},$$

where we have used the formula $P(X\geq x)=(1-p)^x$ that is valid for geometric densities. Solving the last equation for p gives

$$p=1-(1/2)^{1/1000}\approx .0007.$$

Thus the two solutions give fairly close values of p.

5 (10 points)**.** Twentyfive women are independently tossing fair coins, with each woman stopping when she first tosses a head. Find the probability that exactly 100 tails are tossed in total, before everyone stops.

Solution: Let $X_1, \ldots, X_{25}$ denote the number of tails tossed by the 25 women, respectively, before first tossing a head. Then each X_i has geometric density with $p = 1/2$, in other words, negative binomial density with parameters $\alpha_i = 1$ and $p = 1/2$. Hence by Theorem 1 on page 75, the sum

$$X = X_1 + \cdots + X_{25}$$

has negative binomial density with parameters $\alpha = 1 + \cdots + 1 = 25$ and $p = 1/2$. We are interested $P(X = 100)$.

With $x = 100$ we see from the alternate formula for the negative binomial density that

$$P(X = 100) = \binom{\alpha - 1 + x}{x} p^{\alpha}(1-p)^{x} = \binom{124}{100} (1/2)^{125}.$$

Remark. This last formula can also be derived directly, if we think instead about one woman tossing a single coin until the 25^{th} head appears: we require 25 successes (heads) and 100 failures (tails), and require the final trial to be a head; then there are $\binom{124}{100}$ ways to choose which 100 of the other 124 trials are tails. (And at the end, you just mentally divide the 125 trials into 25 groups, each of which terminates with a head; assign one group of trials to each woman.)

MATH 361 — FALL 2001 — TEST 3

NAME:

Do all questions.
No books, notes, extra paper, calculators or computers allowed. EXPLAIN every answer.

You can use (without proof) any needed formula for the density or expectation of any distribution from class, unless the test question is clearly asking you to derive this formula.

1 (20=5+5+10 points).
(a) State a formula for the variance of a random variable X, in terms of one or more expectations.
(b) Suppose X is a continuous random variable. State formulas for EX and $\mathrm{Var}[X]$ in terms of the density $f(x)$ of X.
(c) Let X denote a randomly chosen point in the interval $(-1, 1)$. Use your answer to (b) to calculate the variance of X.

2 (10 points). A certain quantity X has normal distribution with mean 50 and variance 4, that is $X \sim \text{Normal}(50, 4)$. Evaluate the probability that $52 \leq X \leq 56$.

3 (20 points)**.** Do part (a) or part (b) but NOT both.

(a) Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and define $Y = X^{1/2}$. Find the density $g(y)$ of Y. (Be sure to give the value of $g(y)$ for all $y \in \mathbb{R}$.)

(b) Let $X_1, \ldots, X_n$ be random variables with $X_i \sim \text{Gamma}(\alpha, \lambda)$, for each i, and with correlation coefficients $\rho(X_i, X_j) = \frac{1}{2}$ for each $i \neq j$. Let $X = X_1 + \cdots + X_n$. Compute $\text{Var}[X]$.

4 (20 points)**.** Consider a sequence of n independent rolls of a fair die, with the rolls represented by $X_1, \ldots, X_n$. Write $\overline{X}_n = (X_1 + \ldots + X_n)/n$ for the average of the n rolls. Estimate $P(\overline{X}_n \leq 3 \text{ or } \overline{X}_n \geq 4)$.

Hint. Start by stating the Weak Law of Large Numbers.

Note. You may use without proof the values of the mean μ and variance σ^2 for rolling a single die, if you know them. Otherwise, show how to work these values out (you need not bother to carry through the arithmetic to the end).

5 (30=5+25 points)**.** Two thirds of all transactions at a bank are "easy", lasting an average of 2 minutes each (that is, easy transactions are completed at an average rate of $\frac{1}{2}$ a transaction per minute). The other one third of transactions are "hard", lasting an average of 5 minutes each (that is, $\frac{1}{5}$ of a transaction per minute, on average).

(a) Calculate the expected duration (or length) of a transaction. (Keep your solution non-technical: just consider three typical transactions.)

(b) The bank advertizes that "more than half of all transactions last less than 2 minutes". How might the bank justify this claim?

Hint for (b). What proportion of the easy transactions last less than 2 minutes? What proportion of the hard transactions last less than 2 minutes?

You can use that $e^{-1} \approx \frac{3}{8}$ and $e^{-2/5} \approx \frac{2}{3}$.

MATH 361 — FALL 2001 — TEST 3

NAME: **SOLUTIONS**

Do all questions.
No books, notes, extra paper, calculators or computers allowed. EXPLAIN every answer.

You can use (without proof) any needed formula for the density or expectation of any distribution from class, unless the test question is clearly asking you to derive this formula.

1 (20=5+5+10 points).
(a) State a formula for the variance of a random variable X, in terms of one or more expectations.
(b) Suppose X is a continuous random variable. State formulas for EX and $\mathrm{Var}[X]$ in terms of the density $f(x)$ of X.
(c) Let X denote a randomly chosen point in the interval $(-1, 1)$. Use your answer to (b) to calculate the variance of X.

Solution:

(a) (5 points) Either $\mathrm{Var}[X] = E[(X-\mu)^2]$, where $\mu = EX$, or else $\mathrm{Var}[X] = E[X^2]-(EX)^2$.

(b) (5 points) $\mu = EX = \int_{-\infty}^{\infty} xf(x)\,dx$ and either

$$\mathrm{Var}[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

or else

$$\mathrm{Var}[X] = E[X^2] - (EX)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{\infty} xf(x)\,dx\right)^2.$$

(c) (10 points) Here the density is $f(x) = \frac{1}{2}$ for $-1 < x < 1$ and $f(x) = 0$ otherwise. (You can check that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.)

By part (b), the mean is

$$\mu = EX = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-1}^{1} x \cdot \frac{1}{2}\,dx = \frac{1}{4}x^2\bigg|_{-1}^{1} = 0,$$

and the variance is

$$\mathrm{Var}[X] = E[X^2] - (EX)^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - 0^2 = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\,dx = \frac{1}{6}x^3\bigg|_{-1}^{1} = \frac{1}{3}.$$

2 (10 points). A certain quantity X has normal distribution with mean 50 and variance 4, that is $X \sim \text{Normal}(50, 4)$. Evaluate the probability that $52 \le X \le 56$.

Solution: Here $\mu = 50, \sigma^2 = 4$ and so $\sigma = 2$. From page 125 formula (25) we know that

$$P(52 \le X \le 56) = \Phi\left(\frac{56-\mu}{\sigma}\right) - \Phi\left(\frac{52-\mu}{\sigma}\right) = \Phi(3) - \Phi(1) = .9987 - .8413 = .1574,$$

where Φ is the standard normal distribution function and the values of $\Phi(3)$ and $\Phi(1)$ come from the table on the last page of the test.

Remark.

You divide by $\sigma = 2$ here, not by $\sigma^2 = 4$. Remember that subtracting μ and then dividing by σ reduces X back to the standard normal situation. That is, if $X \sim \text{Normal}(\mu, \sigma^2)$ then $\frac{X-\mu}{\sigma} \sim \text{Normal}(0, 1)$.

The following reasoning (which is valid for any random variable X) helps you remember that the "standardized" random variable $\frac{X-\mu}{\sigma}$ has mean 0 and variance 1:

$$E[\frac{X-\mu}{\sigma}] = \frac{E[X-\mu]}{\sigma} = \frac{E[X]-\mu}{\sigma} = 0$$

and

$$\text{Var}[\frac{X-\mu}{\sigma}] = \frac{\text{Var}[X-\mu]}{\sigma^2} = \frac{\text{Var}[X]}{\sigma^2} = 1.$$

(Notice how σ^2 appears in the denominator when we pull σ out of the variance.)

3 (20 points)**.** Do part (a) or part (b) but NOT both.

(a) Suppose $X \sim \text{Gamma}(\alpha, \lambda)$ and define $Y = X^{1/2}$. Find the density $g(y)$ of Y. (Be sure to give the value of $g(y)$ for all $y \in \mathbb{R}$.)

(b) Let $X_1, \ldots, X_n$ be random variables with $X_i \sim \text{Gamma}(\alpha, \lambda)$, for each i, and with correlation coefficients $\rho(X_i, X_j) = \frac{1}{2}$ for each $i \neq j$. Let $X = X_1 + \cdots + X_n$. Compute $\text{Var}[X]$.

<u>*Solution:*</u>

(a) We have the gamma density for X:

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Method 1. The transformation $y = x^{1/2}$ is increasing on the interval $x \in (0, \infty)$ in which we are interested, and so we can apply Theorem 1 on page 119 to get $g(y)|dy| = f(x)|dx|$: in other words, since $x = y^2$ we get

$$g(y) = \left|\frac{dx}{dy}\right| f(x) = 2y f(y^2) = 2\frac{\lambda^\alpha}{\Gamma(\alpha)} y^{2\alpha-1} e^{-\lambda y^2}.$$

This formula is valid for $y > 0$, since $y = x^{1/2} > 0$. For all other y-values, the theorem says $g(y) = 0$.

Method 2. The distribution function of Y is $G(y) = P(Y \leq y)$. Since $Y = X^{1/2} \geq 0$ always, we see Y cannot be negative and so $G(y) = 0$ when $y < 0$. For $y \geq 0$ we have

$$G(y) = P(Y \leq y) = P(X^{1/2} \leq y) = P(X \leq y^2) = F(y^2).$$

Since the derivative of the distribution function is the density, we deduce that for $y > 0$ we have $g(y) = G'(y) = F'(y^2) \cdot 2y = 2yf(y^2)$, while for $y < 0$ we have $g(y) = G'(y) = 0$. Thus we end up with the same formulas for g as in Method 1.

Note. X is always nonnegative (since $P(X < 0) = 0$), and so we can indeed define Y to be the square root of X, in this problem.

(b) From class we know that the Gamma density has variance α/λ^2, and so $\text{Var}[X_i] = \alpha/\lambda^2$ for each i. Furthermore, we are given that for all $i \neq j$,

$$\frac{1}{2} = \rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}[X_i]\text{Var}[X_j]}} = \frac{\text{Cov}(X_i, X_j)}{\alpha/\lambda^2},$$

and so $\text{Cov}(X_i, X_j) = \alpha/2\lambda^2$. Hence

$$\begin{aligned}
\text{Var}[X] &= \text{Var}[X_1 + \cdots + X_n] \\
&= \text{Var}[X_1] + \cdots + \text{Var}[X_n] + \sum_{i \neq j} \text{Cov}(X_i, X_j) \\
&= n\frac{\alpha}{\lambda^2} + n(n-1)\frac{\alpha}{2\lambda^2} \qquad \text{since there are } n(n-1) \text{ pairs } (i,j) \text{ with } i \neq j, \\
&= n(n+1)\frac{\alpha}{2\lambda^2}.
\end{aligned}$$

4 (20 points)**.** Consider a sequence of n independent rolls of a fair die, with the rolls represented by $X_1, \dots, X_n$. Write $\overline{X}_n = (X_1 + \dots + X_n)/n$ for the average of the n rolls. Estimate $P(\overline{X}_n \leq 3 \text{ or } \overline{X}_n \geq 4)$.

Hint. Start by stating the Weak Law of Large Numbers.

Note. You may use without proof the values of the mean μ and variance σ^2 for rolling a single die, if you know them. Otherwise, show how to work these values out (you need not bother to carry through the arithmetic to the end).

Solution:

If X represents the outcome for one roll of a fair die, then we know the mean is

$$\mu = EX = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

and the variance is

$$\sigma^2 = \text{Var}[X] = E[X^2] - (EX)^2 = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - (3.5)^2 = \frac{35}{12}.$$

The problem asks us to estimate $P(\overline{X}_n \leq 3 \text{ or } \overline{X}_n \geq 4)$, in other words $P(|\overline{X}_n - 3.5| \geq \frac{1}{2})$. (This is the key observation — look at a number line to confirm that $|\overline{X}_n - 3.5| \geq \frac{1}{2}$ means "either $\overline{X}_n \leq 3$ or $\overline{X}_n \geq 4$".)

The Weak Law of Large Numbers (or more precisely formula (27) on page 102) says that

$$P(|\overline{X}_n - \mu| \geq \delta) \leq \frac{\sigma^2}{n\delta^2},$$

and if we simply put $\mu = 3.5, \delta = \frac{1}{2}$ and $\sigma^2 = \frac{35}{12}$, then we arrive at the estimate

$$P(|\overline{X}_n - 3.5| \geq \frac{1}{2}) \leq \frac{35}{3n}.$$

Aside. For example, with $n = 105$ we see

$$P(|\overline{X}_n - 3.5| \geq \frac{1}{2}) \leq \frac{1}{9},$$

and so it is rather unlikely (probability at most one ninth) that the average of 105 rolls will be either less than 3 or greater than 4.

5 (30=5+25 points). Two thirds of all transactions at a bank are "easy", lasting an average of 2 minutes each (that is, easy transactions are completed at an average rate of $\frac{1}{2}$ a transaction per minute). The other one third of transactions are "hard", lasting an average of 5 minutes each (that is, $\frac{1}{5}$ of a transaction per minute, on average).

(a) Calculate the expected duration (or length) of a transaction. (Keep your solution non-technical: just consider three typical transactions.)

(b) The bank advertizes that "more than half of all transactions last less than 2 minutes". How might the bank justify this claim?

Hint for (b). What proportion of the easy transactions last less than 2 minutes? What proportion of the hard transactions last less than 2 minutes?
You can use that $e^{-1} \approx \frac{3}{8}$ and $e^{-2/5} \approx \frac{2}{3}$.

Solution:

(a) In every three typical transactions, we expect two easy ones (2 minutes each, for a total of 4 minutes) and one hard one (5 minutes). This gives a total of 9 minutes for three transactions, which averages out to 3 minutes per transaction.

(b) It seems reasonable for the bank to model the duration of an easy transaction by means of a random variable $X \sim \text{Exponential}(\frac{1}{2})$, since one can think of the duration as being the waiting time for an easy transaction to finish. Similarly one can model the duration of a hard transaction with another random variable $Y \sim \text{Exponential}(\frac{1}{5})$. Then

$$
\begin{aligned}
&P(\text{transaction length} < 2)\\
&= P(\text{transaction length} < 2|\text{easy})P(\text{easy}) + P(\text{transaction length} < 2|\text{hard})P(\text{hard})\\
&= P(X < 2) \cdot \frac{2}{3} + P(Y < 2) \cdot \frac{1}{3}\\
&= (1 - e^{-\frac{1}{2}2}) \cdot \frac{2}{3} + (1 - e^{-\frac{1}{5}2}) \cdot \frac{1}{3} \qquad \text{since } P(X < a) = 1 - e^{-\lambda a} \text{ for an exponential density}\\
&= (1 - e^{-1}) \cdot \frac{2}{3} + (1 - e^{-2/5}) \cdot \frac{1}{3}\\
&\approx (1 - \frac{3}{8}) \cdot \frac{2}{3} + (1 - \frac{2}{3}) \cdot \frac{1}{3} \qquad \text{by the given information}\\
&= 19/36.
\end{aligned}
$$

Since this probablity is greater than one half, the bank's claim that more than half of all transactions last less than 2 minutes seems justified.

Problem. Explain how it is possible that the expected transaction length is 3 minutes even though more than half of all transactions take less than 2 minutes.

Aside. It would be more realistic to use Gamma densities for X and Y, such as $X \sim \text{Gamma}(3, \frac{3}{2})$ and $Y \sim \text{Gamma}(3, \frac{3}{5})$, because these Gamma densities still have $EX = 2$ and $EY = 5$, but these densities are also "peaked" somewhere near $x = 2$ and 5 respectively. (The exponential density is not peaked, being a decreasing function of x.)

Remark. It is not true that if $EX = 2$ then half of all easy transactions take less than 2 minutes. In fact we saw above that $P(X < 2) = (1 - e^{-1}) \approx \frac{5}{8}$.

MATH 361 — FALL 2001 — FINAL EXAM

NAME:

Do 10 questions. State here which 2 questions you do **not** want graded: ____________.

No books, notes, extra paper, calculators or computers allowed. EXPLAIN every answer.

You can use any needed formula for the density, expectation or variance of any distribution from class, unless the test question is clearly asking you to derive this formula.

1 (15 points)**.** In a certain town, 6% of people are sick with the flu, another 14% are infected (but not yet sick), and the other 80% are healthy. None of the sick people feel cheerful, half of the infected people feel cheerful, and three quarters of the healthy people feel cheerful.

Find the probability that a cheerful person is healthy.

2 (15 points)**.** Find the probability that in a group of n students, exactly 2 students share their birthday and the other $n-2$ students all have different birthdays.

Hint. Consider n balls and 365 boxes.

3 (15 points). Eight people stand in line, in random order. The people consist of four couples. Find the probability that no-one stands next to their partner. (You need not evaluate your answer.)

4 (15=6+3+6 points)**.** Suppose you put n pairs of socks into the laundry machine ($2n$ socks in total). The machine randomly eats $2n - k$ socks, and returns the other k socks to you, where $k \geq 2$ is fixed.

(a) Let $X_i = 1$ if the i-th pair of socks is returned, and $X_i = 0$ otherwise. Evaluate $p = P(X_i = 1)$.

(b) Evaluate EX_i.

(c) Compute the expected number of *pairs* of socks, among the k socks returned to you.

5 (15=12+3 points). Consider two boxes, with the first containing balls numbered $1, \ldots, 5$ and the second containing balls numbered $1, \ldots, 10$. One ball is drawn from each box, independently, giving numbers X and Y respectively.

(a) Find the discrete density of $Z = \max(X, Y)$.

(b) Write down an expression for EZ (you need not simplify your answer).

Hint. To help check your answer in (a), make sure the total probability adds up to 1.

6 (15=3+3+3+6 points).

(a) Define the Poisson density.

(b) Suppose that $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ are independent. Then $X_1 + X_2 \sim$ ....? (Fill in the blanks.)

(c) Explain the meaning of part (b) in terms of (for example) calls arriving at a call center.

(d) Let $X \sim \text{Binomial}(100, .02)$. Show that $P(X = 3) \approx e^{-2} 2^3 / 3!$.

7 (15=7+8 points).

(a) Let $X \sim \text{Geometric}(p)$ with $p = \frac{3}{4}$. Show $\mu = \frac{1}{3}$ and $\sigma = \frac{2}{3}$, and $P(|X - \mu| \le 2\sigma) = \frac{15}{16}$.

(b) Let $X \sim \text{Poisson}(\lambda)$ with $\lambda = \frac{1}{3}$. Evaluate μ and σ, and show $P(|X - \mu| \le 2\sigma) = \frac{4}{3}e^{-1/3}$.

Hint for (b). $\frac{1}{3} < \frac{1}{\sqrt{3}} < \frac{5}{6}$.

8 (15=6+3+6 points)**.** Buses arrive independently at the station all afternoon, at an average rate of 10 buses per hour.

(a) Let X equal the length of time (in hours) after midday until the third bus arrives. Describe the density of X (you need not write an explicit formula for this density).

(b) Find the expected time of arrival of the third bus at the station.

(c) Write down an explicit integral for the probability that the arrival time of the third bus lies within one standard deviation of the expected arrival time found in part (b). (You need not evaluate your integral.)

9 (15=3+4+8 points). Suppose $X \sim \text{Normal}(5, 1)$. Find
(a) $P(X < 4)$
(b) $E[X^2]$
(c) $E[e^{2X}]$

Hint for (c). First write this expectation as an integral.

10 (15 points)**.** Let $X \sim \text{Normal}(50, 19)$ and $Y \sim \text{Normal}(20, 9)$, and assume X and Y are independent.

Find the probability that $X \leq 3Y$.

Hint. $-3Y \sim \text{Normal}(-60, 81)$.

11 (15 points)**.** A certain baseball player has batting average $p = .2$ (meaning his chance of a hit is .2, for each at-bat).

Use the Central Limit Theorem (or a closely related formula) to show that the probability of the player getting less than or equal to 200 hits in his next 900 at-bats is about .95.

Briefly discuss any assumptions that you are making.

12 (15 points)**.** Xavier and Yolanda arrive at a cafe independently, at times X and Y minutes after the cafe opens. Assume $X, Y \sim \text{Exponential}(\frac{1}{5})$.

When Yolanda gets to the cafe, she notices that Xavier is already there. Show the probability that Xavier arrived during the first minute after the cafe opened is approximately $1/3$. (You can use that $e^{-2/5} \approx \frac{2}{3}$.)

Hint. First write down the joint density.