MATH 425
Final
Winter, 2009

Name:

- You have **120 minutes** to complete your work.
- Show all work and make it clear what your answers are.
- You are permitted three 3x5 notecards. Otherwise, books, notes, calculators and computers are not permitted on this exam.

---

problem	points	score
1	10	
2	15	
3	15	
4	5	
5	15	
6	15	
7	20	
8	10	
9	10	
10	20	
11	15	
Total	150	

1. (10 points)

(a) Provide an example of a *binomially distributed* random variable which is best modeled using a normal distribution. State any relevant assumptions. Give the values for the parameters of your example.

(b) Provide an example of a random variable that is best modeled by a *gamma distributed* random variable. State any relevant assumptions. Give the values for the parameters of your example.

a. The important point is the number of trials: a binomial random variable is well-approximated when the number of trials is greater than thirty and the probability of success is not too far from 50%. Here is the example from the second midterm

- A small college has found that on average only 1 in every 3 students accepted will actually attend. They accept 450 students for next year's class. Let X count the number of students who actually attend.

  The parameters from the normal distribution are $\mu = 150$ and $\sigma^2 = 100$.

b. Here is the example from the second midterm

- During the peak time of the Capricornids meteor shower, the average time between meteors is 4 minutes. Let X count the time (in hours) you wait until you observe 25 meteors.

  The parameters are $\alpha = 25$ (number of meteors) and $\lambda = 15$ (recall that $\frac{1}{\lambda}$ is the expected time for one event, in hours here).

2. (15 points)
A standard 52 card deck contains 4 suits (hearts, diamonds, clubs, spades) and each suit contains 13 cards. A gin hand consists of ten cards from a 52 card deck. Assume each hand of 10 cards is equally likely.

(a) Find the probability that all 10 cards are in the same suit.

(b) Find the probability that exactly 4 cards are in one suit and 3 cards in two other suits.

(c) Find the probability that the distribution of suits is 4, 3, 2, 1.

(a).

$$\frac{\binom{4}{1} \cdot \binom{13}{10}}{\binom{52}{10}} \approx 7.23 \times 10^{-8}.$$

(b).

$$\frac{\binom{4}{1} \cdot \binom{13}{4} \cdot \binom{3}{2} \cdot \binom{13}{3} \cdot \binom{13}{3}}{\binom{52}{10}} \approx 0.044.$$

(c).

$$\frac{4! \cdot \binom{13}{4} \cdot \binom{13}{3} \cdot \binom{13}{2} \cdot \binom{13}{1}}{\binom{52}{10}} \approx 0.315.$$

3. (15 points)
A doctor assumes that a patient has one of three diseases d_1, d_2 or d_3. Before any testing, he assumes an equal probability for each disease. He carries out a test that will be positive with probability 0.8 if the patient has d_1, 0.6 if the patient has disease d_2, and 0.4 if the patient has disease three.

(a) Given that the outcome of the test was positive, find the probabilities the doctor should now assign to the three possible diseases.

(b) Given that the outcome of the test was negative, find the probabilities the doctor should now assign to the three possible diseases.

Probabilities given by the problem

$$\mathbf{P}\{d_1\} = \mathbf{P}\{d_2\} = \mathbf{P}\{d_3\} = \frac{1}{3}.$$
$$\mathbf{P}(\text{positive} \mid d_1) = 0.8 \quad \mathbf{P}(\text{positive} \mid d_2) = 0.6 \quad \mathbf{P}(\text{positive} \mid d_3) = 0.4$$

We will need the probability the patient will test positive, by conditioning:

$$\begin{aligned}\mathbf{P}\{\text{positive}\} &= \mathbf{P}(\text{positive} \mid d_1) \cdot \mathbf{P}\{d_1\} + \mathbf{P}(\text{positive} \mid d_2) \cdot \mathbf{P}\{d_2\} + \mathbf{P}(\text{positive} \mid d_3) \cdot \mathbf{P}\{d_3\} \\ &= \frac{0.8 + 0.6 + 0.4}{3} = 0.6\end{aligned}$$

(a). The posterior probabilities for each disease if the test is positive is then obtained by Bayes theorem:

$$\begin{aligned}\mathbf{P}(d_1 \mid \text{positive}) & \quad \frac{\mathbf{P}(\text{positive} \mid d_1) \cdot \mathbf{P}\{d_1\}}{\mathbf{P}\{\text{positive}\}} \\ &= \frac{0.8}{0.6} \cdot \frac{1}{3} = \frac{4}{9}\end{aligned}$$

$$\begin{aligned}\mathbf{P}(d_2 \mid \text{positive}) & \quad \frac{\mathbf{P}(\text{positive} \mid d_2) \cdot \mathbf{P}\{d_2\}}{\mathbf{P}\{\text{positive}\}} \\ &= \frac{0.6}{0.6} \cdot \frac{1}{3} = \frac{1}{3}\end{aligned}$$

$$\begin{aligned}\mathbf{P}(d_3 \mid \text{positive}) & \quad \frac{\mathbf{P}(\text{positive} \mid d_3) \cdot \mathbf{P}\{d_3\}}{\mathbf{P}\{\text{positive}\}} \\ &= \frac{0.4}{0.6} \cdot \frac{1}{3} = \frac{2}{9}\end{aligned}$$

(b). The following probabilities are easily computed from those before (a):

$$\mathbf{P}\{\text{negative}\} = 1 - \mathbf{P}\{\text{positive}\} = 0.4$$
$$\mathbf{P}(\text{negative} \,|\, d_1) = 0.2 \quad \mathbf{P}(\text{negative} \,|\, d_2) = 0.4 \quad \mathbf{P}(\text{negative} \,|\, d_3) = 0.6$$

The posterior probabilities for each disease if the test is negative is then obtained by Bayes theorem:

$$\begin{aligned}
\mathbf{P}(d_1 \,|\, \text{negative}) & \quad \frac{\mathbf{P}(\text{negative} \,|\, d_1) \cdot \mathbf{P}\{d_1\}}{\mathbf{P}\{\text{negative}\}} \\
&= \frac{1 - 0.8}{0.4} \cdot \frac{1}{3} = \frac{4}{9} \\
\mathbf{P}(d_2 \,|\, \text{negative}) & \quad \frac{\mathbf{P}(\text{negative} \,|\, d_2) \cdot \mathbf{P}\{d_2\}}{\mathbf{P}\{\text{negative}\}} \\
&= \frac{1 - 0.6}{0.4} \cdot \frac{1}{3} = \frac{1}{3} \\
\mathbf{P}(d_3 \,|\, \text{negative}) & \quad \frac{\mathbf{P}(\text{negative} \,|\, d_3) \cdot \mathbf{P}\{d_3\}}{\mathbf{P}\{\text{negative}\}} \\
&= \frac{1 - 0.4}{0.4} \cdot \frac{1}{3} = \frac{4}{9}
\end{aligned}$$

4. (5 points)
The time T (in hours) required to repair a machine is exponentially distributed with parameter $\lambda = 2$. The repair company charges X (dollars) for a repair according to the time required using the equation

$$X = 15 + 3e^{T/2}.$$

Find the expected cost of a repair.

The cost X is a function of time T, so we can compute $E[X]$ using the known density for the exponential variable T:

$$\begin{aligned} E[X] &= \int_0^\infty (15 + 3e^{t/2}) \cdot 2e^{-2t}\, dt \\ &= \int_0^\infty 30e^{-2t} + 6e^{-3t/2}\, dt \\ &= 15 + 4 = 19 \end{aligned}$$

A quite common mistake was was to compute $E[X]$ by

$$E[X] = 15 + 3e^{E[T]/2}.$$

5. (15 points)
A shipment of 10,000 parts for manufacturing widgets has arrived at Acme. There are 50 defective parts in the shipment. Acme plans to test 100 randomly chosen parts and will reject the batch if it finds more than two defective parts. Let X be the random variable which denotes the number of defective parts in the 100 sampled.

(a) What type of distribution best models X? State the relevant parameters for this distribution. State any assumption you think relevant in deciding the best distribution.

(b) Find the probability that the batch is returned by Acme.

(a). This distribution is hypergeometric, but given the large size of the shipment compared to the sample size, and the rarity of a defective part, it is best approximated by a Poisson random variable. The probability that a part is defective is 0.005, so the parameter $\lambda = 0.5$ (the expected number of defective parts in a sample of 100).

I did except a guess of hypergeometric random variable, although there was a penalty for part (a).

(b). Let X be the number of defective parts in a sample of 100.

$$\begin{aligned}
\mathbf{P}\{X \leq 2\} &\approx e^{-0.5} + (0.5)e^{-0.5} + e^{-0.5}\frac{(0.5)^2}{2!} \\
&\approx 0.9856 \\
\mathbf{P}\{X > 2\} &= 1 - \mathbf{P}\{X \leq 2\} \approx 0.0144.
\end{aligned}$$

So, the probability of returning a shipment is about 1.4%.

6. (15 points)
A coin is chosen at random from a bin of coins whose probability p of landing heads is distributed with probability density $\frac{3}{2}(1-p^2)$. Suppose that after each coin is chosen but before it is flipped, you must predict whether it will land heads or tails. You win 1 if you are correct and lose 1 if you are incorrect.

(a) Find your expected gain if you always predict heads.

(b) Find your expected gain if you always predict tails.

(c) Find your expected gain if you are told the probability p before you make your prediction.

This problem is from Practice Test 7.15.

(a). Let X_h (X_t) denote expected winnings when betting heads (tails) and Y denote probability coin lands heads.

$$\begin{aligned} E[X_h] &= \int_0^1 E[X_h|Y=p]\frac{3}{2}(1-p^2)\,dp \\ &= \int_0^1 \frac{3}{2}(2p-1)(1-p^2)\,dp \\ &= -\frac{1}{4}. \end{aligned}$$

(b). There is no need to compute $E[X_t]$ since $E[X_t] = -E[X_h] = \frac{1}{4}$. Here are the details:

$$\begin{aligned} E[X_t] &= \int_0^1 E[X_h|Y=p]\frac{3}{2}(1-p^2)\,dp \\ &= \int_0^1 \frac{3}{2}(1-2p)(1-p^2)\,dp \\ &= \frac{1}{4}. \end{aligned}$$

(c). Always bet on heads when $0 < p < \frac{1}{2}$ and bet on tails when $\frac{1}{2} < p < 1$.

$$\begin{aligned} E[X] &= \int_0^{\frac{1}{2}} E[X_t|Y=p]\frac{3}{2}(1-p^2)\,dp + \int_{\frac{1}{2}}^1 E[X_h|Y=p]\frac{3}{2}(1-p^2)\,dp \\ &= \int_0^{\frac{1}{2}} \frac{3}{2}(1-2p)(1-p^2)\,dp + \int_{\frac{1}{2}}^1 \frac{3}{2}(2p-1)(1-p^2)\,dp \\ &= \frac{23}{64} + \frac{7}{64} = \frac{15}{32}. \end{aligned}$$

7. (20 points)
Let X and Y be two jointly distributed random variables with joint density function

$$f_{X,Y}(x,y) = \begin{cases} cxy(1-x) & \text{if } 0 < x < 1,\ 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find c.

(b) Compute the marginal density function for X and for Y.

(c) Compute the conditional density functions $f_{X|Y}$.

(d) Find the probability $\mathbf{P}(Y < \frac{1}{2} \,|\, X > \frac{1}{2})$.

(a). $f_{X,Y}$ must be a density function.

$$\begin{aligned} 1 &= \int_0^1 \int_0^1 cxy(1-x)\,dy\,dx \\ &= \frac{c}{2}\int_0^1 x - x^2\,dx \\ &= \frac{c}{12}. \end{aligned}$$

So, $c = 12$.

(b).

$$\begin{aligned} f_X(x) &= \int_0^1 12x(1-x)y\,dy \\ &= 6x(1-x). \\ f_Y(y) &= \int_0^1 12x(1-x)y\,dx \\ &= 2y. \end{aligned}$$

(c). Note that X and Y are independent:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

You can tell this even without computing (b), since you can decompose $f_{X,Y}(x,y)$ into a function of x alone and a function of y alone.

One consequence of the independence is that

$$f_{X|Y}(x,y) = f_X(x) \quad \text{and} \quad f_{Y|X}(x,y) = f_X(x).$$

Here is the computation

$$\begin{aligned} f_{X|Y}(x,y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{12xy(1-x)}{2y} = 6x(1-x). \end{aligned}$$

(d) Since X and Y are independent

$$\begin{aligned} \mathbf{P}(Y < \frac{1}{2} \,|\, X > \frac{1}{2}) &= \mathbf{P}\left\{Y < \frac{1}{2}\right\} \\ &= \int_0^{\frac{1}{2}} 2y\,dy \\ &= \frac{1}{4}. \end{aligned}$$

An alternative is the long using the *definition of condition probability*:

$$\begin{aligned} \mathbf{P}(Y < \frac{1}{2} \,|\, X > \frac{1}{2}) &= \frac{\mathbf{P}\left\{Y < \frac{1}{2}, X > \frac{1}{2}\right\}}{\mathbf{P}\left\{X > \frac{1}{2}\right\}} \\ &= \frac{\int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^1 12xy(1-x)\,dx\,dy}{\int_{\frac{1}{2}}^1 6x(1-x)\,dx} \end{aligned}$$

8. (10 points)
Let X and Y be independent normal random variables with mean 0 and standard deviation 1.

(a) Find the joint density of U and V where

$$U = X \qquad V = \frac{X}{Y}.$$

(b) Find the marginal density of V.

I had intended that U and V were defined by

$$U = X \qquad V = \frac{Y}{X},$$

which makes the algebra slightly simpler, but the problem is otherwise identical.

The random variables X and Y are independent and standardly distributed random variables, so their joint density is

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$

(a). From the definition of U and V we can compute the inverse transformation from uv-coordinates back to xy-coordinates by

$$x = u \qquad y = \frac{u}{v}.$$

The Jacobean of this transformation is given by

$$J(u,v) = \begin{pmatrix} 1 & 0 \\ \frac{1}{v^2} & -\frac{u}{v^2} \end{pmatrix} = -\frac{u}{v^2}.$$

We can now compute the joint density of U and V:

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(x,y) \cdot \left|J(u,v)\right| \\ &= \frac{u}{2\pi v^2} \cdot e^{-u^2(v^2+1)/2v^2}. \end{aligned}$$

(b). We need to integrate $f_{U,V}(u,v)$ over u. This is an easy integration with the substitution

$$t = u^2(v^2+1)/2v^2 \qquad dt = u\frac{v^2+1}{v^2}du.$$

Here is the the calculation:

$$
\begin{aligned}
\int_{-\infty}^{\infty} f_{U,V}(u,v)\,du &= \int_{-\infty}^{\infty} \frac{u}{2\pi v^2} \cdot e^{-u^2(v^2+1)/2v^2}\,du \\
&= \frac{1}{2\pi v^2} \int_{-\infty}^{\infty} \frac{v^2}{v^2+1} \cdot e^{-t}\,dt \\
&= \frac{1}{2\pi(v^2+1)} \int_{-\infty}^{\infty} e^{-t}\,dt \\
&= \frac{1}{2\pi(v^2+1)} \cdot \left[e^{-u^2(v^2+1)/2v^2} \Big|_{u=-\infty}^{\infty} \right] \\
&= \frac{1}{2\pi(v^2+1)} \cdot \cdot \left[e^{-u^2(v^2+1)/2v^2} \Big|_{u=0}^{\infty} \right] \\
&= \frac{1}{\pi(v^2+1)}
\end{aligned}
$$

The density $f_V(v)$ is the *Cauchy density* which is mentioned briefly in Chapter 5.6.3.

9. (10 points)
Suppose that X_i $(i = 1, 2, 3)$ are independent Poisson random variables with respective means λ_i $(i = 1, 2, 3)$. Let $U = X_1 + X_2$ and $V = X_2 + X_3$.

(a) Find $Cov(U, V)$.

(b) Find $Var(U + V)$.

The mean and variances for X_i $(i = 1, 2, 3)$ are

$$E[X_1] = \lambda_1 \quad E[X_2] = \lambda_2 \quad E[X_3] = \lambda_3$$
$$Var(X_1) = \lambda_1 \quad Var(X_2) = \lambda_2 \quad Var(X_3) = \lambda_3$$

Since X_i and X_j are independent when $i \neq j$, the variances add:

$$Var(X_i + X_j) = Var(X_i) + Var(X_j) = \lambda_i + \lambda_j \qquad \text{when } i \neq j.$$

(a).

$$\begin{aligned} Cov(U + V) &= Cov(X_1 + X_2, X_2 + X_3) \\ &= Cov(X_2, X_2) \\ &= Var(X_2) = \lambda_2. \end{aligned}$$

(b).

$$\begin{aligned} Var(U + V) &= Var(U) + Var(V) + 2Cov(U, V) \\ &= (\lambda_1 + \lambda_2) + (\lambda_2 + \lambda_3) + 2\lambda_2 \\ &= \lambda_1 + 4\lambda_2 + \lambda_3. \end{aligned}$$

10. (20 points)
Suppose X and Y are chosen uniformly and independently on the interval $(0,1)$. Compute the following probabilities.

(a) $\mathbf{P}\{X+Y<\frac{1}{2}\}$.

(b) $\mathbf{P}\{|X-Y|<\frac{1}{2}\}$.

(c) $\mathbf{P}\{\max(X,Y)<\frac{1}{2}\}$.

(d) $\mathbf{P}\{\min(X,Y)<\frac{1}{2}\}$.

Since X and Y are independent

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

(a).

$$\mathbf{P}\{X+Y<\frac{1}{2}\}=\int_0^1\int_0^{\frac{1}{2}-x} dy\,dx=\frac{1}{8}$$

(b). I broke this integral into two cases: $\{X<\frac{1}{2}\}$ and $\{X\geq\frac{1}{2}\}$.

$$\mathbf{P}\{|X-Y|<\frac{1}{2}\}=\int_0^{\frac{1}{2}}\int_0^{x+\frac{1}{2}} dy\,dx+\int_{\frac{1}{2}}^1\int_{x-\frac{1}{2}}^1 dy\,dx=\frac{3}{4}$$

(c). If $\max(X,Y)<\frac{1}{2}$ then both $X<\frac{1}{2}$ and $Y<\frac{1}{2}$.

$$\mathbf{P}\{\max(X,Y)<\frac{1}{2}\}=\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} dy\,dx=\frac{1}{4}$$

(d). This might be easiest by computing $\mathbf{P}\{\min(X,Y)>\frac{1}{2}\}$, which holds when both $X>\frac{1}{2}$ and $Y>\frac{1}{2}$:

$$\mathbf{P}\{\min(X,Y)<\frac{1}{2}\}=1-\int_{\frac{1}{2}}^1\int_{\frac{1}{2}}^1 dy\,dx=\frac{3}{4}$$

Alternatively, you could break the probability into two cases: $\{X<\frac{1}{2}\}$ and $\{X\geq\frac{1}{2}\}$. In the first case, $\{0<Y<1\}$ and in the second case $\{Y<\frac{1}{2}\}$. This requires computing two integrals

$$\mathbf{P}\{\min(X,Y)<\frac{1}{2}\}=\int_0^{\frac{1}{2}}\int_0^1 dy\,dx+\int_{\frac{1}{2}}^1\int_0^{\frac{1}{2}} dy\,dx=\frac{3}{4}$$

11. (15 points)
The servicing of a machine requires two separate steps, with the time needed for the first step being an exponential random variable with mean 0.3 hours and the time for the second step being an independent exponential random variable with mean 0.4 hours.

(a) Compute the mean and variance for the time to service one machine.

(b) Let X be the random variable denoting the time to service 100 machines. What type of distribution best models X? State the relevant parameters for this distribution. State any assumption you think relevant in deciding the best distribution.

(c) Approximate the probability that 100 machines can be serviced in under 65 hours.

(a). Let X_i be the time to service ith machine.

$$E[X_i] = 0.3 + 0.4 = 0.7 \qquad Var(X_i) = (0.3)^2 + (0.4)^2 = (0.5)^2.$$

(b). X is best modeled by a normal distribution with parameters

$$E[X] = 100 \cdot E[X_i] = 70 \qquad Var(X) = 100 \cdot Var(X_i) = 25.$$

X is the sum of 100 independent and identically distributed random variables $X_1, \ldots, X_{100}$ where each X_i is the time for the repair of the i machine.

A few people said that X is best modeled by a gamma distribution, but the time to repair one machine X_i is NOT an exponentially distributed (or gamma distributed) random variable. It is the sum of two exponential random variables with *different parameters.*

(c).

$$\begin{aligned}
\mathbf{P}\{X \leq 65\} &= \mathbf{P}\{\frac{X-70}{5} \leq \frac{65-70}{5}\} \\
&= \Phi(-1) \\
&= 1 - \Phi(1) = 1 - 0.8413 = 0.1587
\end{aligned}$$