Math 408, Spring 2007
Midterm Exam 2 Solutions/Comments

1. Suppose X and Y are discrete random variables with values $1, 2, 3, 4$ and joint p.m.f. given by

$$f(x, y) = \begin{cases} 1/16 & \text{if } x = y \\ 2/16 & \text{if } x < y \\ 0 & \text{if } x > y \end{cases}$$

for $x, y = 1, 2, 3, 4$.

(a) Find $P(X + Y = 6)$.

Solution. Splitting the probability into cases according the values of X and $Y(= 6 - X)$, we get

$$\begin{aligned} P(X + Y) &= f(2, 4) + f(3, 3) + f(4, 2) \\ &= \frac{2}{16} + \frac{1}{16} + 0 = \boxed{\frac{3}{16}(= 0.1875)}. \end{aligned}$$

(b) Find the conditional expectation of Y given that $X = 3$.

Solution. We first represent the joint distribution in matrix form, and compute the marginal distributions by summing the row/column entries:

$X \setminus Y$	1	2	3	4	$f_X(x)$
1	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{7}{16}$
2	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{5}{16}$
3	0	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$
4	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
$f_Y(y)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	

Next, we compute the conditional density $h(y|3)(= h(y|X = 3))$ by renormalizing the entries in the third row of the table (corresponding to $X = 3$), i.e., dividing these entries by their sum, $f_X(3) = 3/16$. This gives $h(y|3) = 0, 0, 1/3, 2/3$ for $y = 1, 2, 3, 4$, respectively. Finally, using this density $h(y|3)$ we can compute the conditional expectation requested:

$$E(Y|3) = \sum_{y=1}^{4} y h(y|3) = 3 \cdot \frac{1}{3} + 4 \cdot \frac{2}{3} = \boxed{\frac{11}{3}(= 3.66)}.$$

2. (NO CALCULATORS FOR THIS PROBLEM) Suppose that X is uniformly distributed on the interval $[0, 1]$ and that, given $X = x$, Y is uniformly distributed on the interval $[1 - x, 1]$.

   (a) Determine the joint density $f(x, y)$. (Be sure to specify the range.)

   Solution. Since X is uniformly distributed on $[0, 1]$, we have $\boxed{f_X(x) = 1,\ 0 \le x \le 1}$. Similarly, since, given $X = x$, Y is uniformly distributed on $[1 - x, 1]$, the conditional density of Y given $X = x$ is $1/(1 - (1 - x)) = 1/x$ on the interval $[1 - x, 1]$; i.e., $\boxed{h(y|x) = 1/x,\ 1 - x \le y \le 1}$ for $0 \le x \le 1$. Thus

$$f(x, y) = f_X(x)h(y|x) = \boxed{\frac{1}{x}, \quad 0 < x < 1, 1 - x < y < 1}$$

   (b) Find the probability $P(Y \ge 1/2)$. (You can leave your answer in raw form, such as $1 - 1/e$.)

   Solution. This can be computed either as an integral of the joint density over appropriate region, or as a single integral over the marginal density $f_Y(y)$ from 1/2 to 1. Since we know the joint density from part (a), the first method is the more natural one. We get (see sketch for the integration limits)

$$\begin{aligned} P(Y \ge 1/2) &= \int_{x=0}^{1/2}\int_{y=1-x}^{1} \frac{1}{x} dy dx + \int_{x=1/2}^{1}\int_{y=1/2}^{1} \frac{1}{x} dy dx \\ &= \int_{x=0}^{1/2} \frac{1 - (1 - x)}{x} dx + \int_{x=1/2}^{1} \frac{1/2}{x} dx \\ &= \frac{1}{2} + \frac{1}{2}(\ln 1 - \ln(1/2)) = \boxed{\frac{1 + \ln 2}{2} (= 0.846)}. \end{aligned}$$

   Alternate solution: We first compute the marginal density $f_Y(y)$ by integrating $f(x, y)$ with respect to x over the appropriate range, namely $1 - y \le x \le 1$:

$$f_Y(y) = \int_{x=1-y}^{1} \frac{1}{x} dx = \ln 1 - \ln(1 - y) = -\ln(1 - y), \quad 0 \le y \le 1.$$

   Integrating this density from $y = 1/2$ to $y = 1$ then gives $P(Y \ge 1/2)$:

$$\begin{aligned} P(Y \ge 1/2) &= -\int_{1/2}^{1} \ln(1 - y) dy = -\int_{u=0}^{1/2} \ln u du \quad (\text{set } u = 1 - y) \\ &= -\left[u \ln u - u\right]_0^{1/2} = -(1/2)\ln(1/2) + (1/2) = \frac{1 + \ln 2}{2}. \end{aligned}$$

   Comments: The probability in question can **not** be computed as ratios of areas. This method can only be used in cases where the joint density, $f(x, y)$, is uniform. However, here $f(x, y) = 1/x$, so is not a constant, and therefore not uniform.

3. (NO CALCULATORS FOR THIS PROBLEM) Let X and Y be random variables with joint density
$$f(x,y) = x - y + 1, \quad 0 \le x, y \le 1.$$

(a) Determine the conditional density of Y given $X = 1/4$. (Make sure to specify the range for this density.)

Solution. We first compute the marginal density $f_X(x)$:
$$f_X(x) = \int_{y=0}^{1} (x - y + 1)dy = \left[xy - \frac{y^2}{2} + y \right]_{y=0}^{1} = x + 0.5, \quad 0 \le x \le 1.$$
Thus, $f_X(1/4) = 3/4$. By the formula $h(y|x) = f(x,y)/f_X(x)$, it follows that
$$h(y|1/4) = \frac{f(1/4,y)}{f_X(1/4)} = \frac{1/4 - y + 1}{3/4} = \boxed{\frac{5}{3} - \frac{4}{3}y, \quad 0 \le y \le 1}.$$

(b) **Set up (but do not evaluate)** an integral (or a sum of integrals) giving the probability $P(X + Y \ge 1/2)$.

Your integral(s) should in one of the following forms (you decide which), with **concrete** numbers or functions in place of the asterisks:
$$\int_{x=*}^{x=*} \int_{y=*}^{y=*} \star \, dy \, dx, \qquad \int_{y=*}^{y=*} \int_{x=*}^{x=*} \star \, dx \, dy.$$
(The answer can be either a single expression of this type, or a sum of two such expressions. Again, leave the integral(s) unevaluated.)

Solution. [This is essentially Problem 12 from the Double Integral Problems handout.] The probability is given by the double integral over the above density function over the part of the unit square on which $x + y \ge 0.5$ (see sketch):
$$\boxed{\int_{x=0}^{0.5} \int_{y=0.5-x}^{1} (x - y + 1)dydx + \int_{x=0.5}^{1} \int_{y=0}^{1} (x - y + 1)dydx}$$

Comments: A single double integral (e.g., $\int_0^1 \int_{0.5-x}^1 \dots$) would not work here; it would overcount or undercount parts of the required region. Thus, one has to use two double integrals.

4. (NO CALCULATORS FOR THIS PROBLEM) Let X and Y be independent random variables, each having exponential distribution with mean 2. **Set up (but do not evaluate)** an integral (or a sum of integrals) giving the probability $P(|X - Y| \le 1)$. Your integral(s) should in one of the following forms (you decide which), with **concrete** numbers or functions in place of the asterisks:

$$\int_{x=*}^{x=*} \int_{y=*}^{y=*} \star \, dy \, dx, \qquad \int_{y=*}^{y=*} \int_{x=*}^{x=*} \star \, dx \, dy.$$

(The answer can be either a single expression of this type, or a sum of two such expressions. Again, leave the integral(s) unevaluated.)

Solution. [Except for the fact that the problem involves a concrete joint density $f(x, y)$, this is Problem 7 of the Double Integrals handout.] The probability is given by $\iint_R f(x, y) dx dy$, where $f(x, y)$ is the joint density of X and Y and R is the part of the range of $f(x, y)$ in which $|x - y| \le 1$. Now, by the independence and the exponential distribution of X and Y,

$$f(x, y) = f_X(x) f_Y(y) = \frac{1}{2} e^{-x/2} \cdot \frac{1}{2} e^{-y/2} = \frac{1}{4} e^{-(x+y)/2}, \quad x \ge 0, y \ge 0.$$

which $|x - y|$. The region R is the part of the first quadrant between the lines $y = x - 1$ and $y = x + 1$ (see sketch). Since the lower boundary of this region is given by two different functions depending on whether $0 \le x \le 1$ or $1 < x < \infty$, (namely, $y = 0$ in the first case and $y = x - 1$ in the second), we need to break up the integral $\iint_R f(x, y) dx dy$ into two, corresponding to the ranges $0 \le x \le 1$ and $1 < x < \infty$:

$$\boxed{\int_{x=0}^{1} \int_{y=0}^{x+1} \frac{1}{4} e^{-(x+y)/2} dy dx + \int_{x=1}^{\infty} \int_{y=x-1}^{x+1} \frac{1}{4} e^{-(x+y)/2} dy dx}$$

5. A computer generates 48 random real numbers, rounds each number to the nearest integer and then computes the average of these 48 rounded values. Assume that the numbers generated are independent of each other and that the rounding errors are distributed uniformly on the interval $[-0.5, 0.5]$. Find the approximate probability that the average of the rounded values is within 0.05 of the average of the exact numbers.

Solution. Let $X_1, \dots, X_{48}$ denote the 48 rounding errors, and $\overline{X} = (1/48) \sum_{i=1}^{48} X_i$ their average. We need to compute $P(|\overline{X}| \le 0.05)$. Since a rounding error is uniformly distributed on $[-0.5, 0.5]$, its mean is $\mu = 0$ and its variance is $\sigma = \int_{-0.5}^{0.5} x^2 dx = [x^3/3]_{-0.5}^{0.5} = 1/12$. By the Central Limit Theorem, $\overline{X}$ has approximate distribution $N(\mu, \sigma^2/n) = N(0, (1/12)/48) = N(0, 1/24^2)$. Thus $24\overline{X}$ is approximately standard normal, so

$$\begin{aligned} P(|\overline{X}| \le 0.05) &\approx P(24 \cdot (-0.05) \le 24\overline{X} \le 24 \cdot 0.05) \\ &= \Phi(1.2) - \Phi(-1.2) = \boxed{2\Phi(1.2) - 1 = 0.7698}. \end{aligned}$$

6. An insurance company insures two types of cars, economy cars and luxury cars. The damage claim resulting from an accident involving an economy car has normal $N(7,1)$ distribution, the claim from a luxury car accident has normal $N(20,6)$ distribution.

   Suppose the company receives three claims from economy car accidents and one claim from a luxury car accident. Assuming that these four claims are mutually independent, what is the probability that the total claim amount from the three economy car accidents exceeds the claim amount from the luxury car accident?

   Solution. [This is a problem of the same type as 5.3-10, worked out in class, which involved the weights of three-pound and one-pound grocery bags.] Let X_1, X_2, X_3 denote the claim amounts from the three economy cars, and X_4 the claim from the luxury car. Then we need to compute $P(X_1 + X_2 + X_3 > X_4)$, which is the same as $P(X_1 + X_2 + X_3 - X_4 > 0)$. Now, since the X_i's are independent and normal with distribution $N(7,1)$ (for $i = 1,2,3$) and $N(20,6)$ for $i = 4$, the linear combination $X = X_1 + X_2 + X_3 - X_4$ has normal distribution with parameters $\mu = 7 + 7 + 7 - 20 = 1$ and $\sigma^2 = 1 + 1 + 1 + 6 = 9$. Thus, the probability we want is

$$\begin{aligned} P(X > 0) &= P\left(\frac{X-1}{\sqrt{9}} > \frac{0-1}{\sqrt{9}}\right) \\ &= P(Z > -0.33) = 1 - P(Z \leq -0.33) = P(Z \leq 0.33) \approx \boxed{0.6293} \end{aligned}$$

   Comments: As pointed out in class in connection with the grocery bag problem, one has to work with three individual variables representing the economy car damages (X_1, X_2, X_3 in the above notation), not with three times a single such variable (i.e., of the form $3X_1$). The latter would mean that the variables are "in lockstep", contrary to the independence assumption, and would result in a greater variance, namely $3^2 + 6 = 15$ instead of $3 + 6 = 9$, and lead to $\Phi(1/\sqrt{15}) = 0.6026$ as answer.

7. Let X_1, X_2, X_3, X_4 be a random sample of size 4 from a normal distribution with mean 2 and variance 10, and let $\overline{X}$ be the sample mean.

   (a) Determine a such that $P(\overline{X} \leq a) = 0.90$.

   Solution. The sample mean $\overline{X}$ is normal with mean $\mu = 2$ and variance $\sigma^2/n = 10/4 = 2.5$, and standard deviation $\sqrt{2.5} = 1.58$, so

$$0.90 = P(\overline{X} \leq a) = P\left(\frac{\overline{X} - 2}{1.58} < \frac{a-2}{1.58}\right) = \Phi\left(\frac{a-2}{1.58}\right).$$

   From the normal table, we get $(a - 2)/1.58 = 1.28$, so $a = \boxed{4.02}$.

(b) Determine b such that $P\left(\sum_{i=1}^{4}(X_i-\overline{X})^2\le b\right)=0.90$.

Solution. Standardizing the expression inside $P(\dots)$ by dividing by $\sigma^2 = 10$, the given probabilitity can be written as

$$P\left(\sum_{i=1}^{4}\left(\frac{X_i-\overline{X}}{\sqrt{10}}\right)^2\le\frac{b}{10}\right).$$

Since $\sum_{i=1}^{4}\left(\frac{X_i-\overline{X}}{\sqrt{10}}\right)^2$ has $\chi^2(3)$ distribution (note the degree is $n-1=3$, not n, since the normalization is with the sample mean $\overline{X}$), we need to choose b such that $b/10$ is equal to the 90-th percentile of this distribution, i.e., $b/10=\chi_{0.10}(3)$. From the table, we get $\chi_{0.10}(3)=6.251$. Thus $b=10\cdot 6.251=\boxed{62.51}$.

Comments: The most common error was to multiply through with $n-1(=3)$, resulting in $b=20.8$ as answer. Multiplication by $n-1$ is needed when working with S^2, in order to cancel the factor $1/(n-1)$ in the definition of S^2 (namely, $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$). However, here we are dealing with a sum of the form $\sum_{i=1}^{n}(X_i-\overline{X})^2$, so the only normalization needed is division by σ^2. See problem 5.3-6(b) for a similar type of calculation.