14.30 Statistics - Fall 2003
Exam #1 Solutions

Prepared by Eric Moos

1. True, False, or uncertain.

   (a) False. Sets can be both disjoint and exhaustive; for example, A and A^c are disjoint and exhaustive. Also, two sets A and B need not be either disjoint or exhaustive. For example, consider rolling a die; the events "roll a 3" and "roll an odd number" are neither disjoint nor exhaustive.

   (b) False. The correct relationship is

   $$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

   The simple sum $P(A) + P(B) + P(C)$ overcounts outcomes in multiple events. Subtracting out the outcomes in the pairwise intersections corrects for this overcounting, but it overcorrects. We must add back outcomes in all three events.

   (c) False. Independence has nothing to do with the "number of experiments." If it did, then we would always face existential questions like "Is flipping a coin twice one experiment or two?" Additionally, consider selecting an individual at random from the MIT student population. There is no reason that the events "the student is male" and "the student is originally from Connecticut" cannot be independent. Clearly, this is only one experiment; the events can potentially be independent. See Example 2.3 of Handout 1 as an example.

   (d) True. Given a non-zero probability event, B; two things can occur: either event A or event A^c. These are disjoint and exhaustive; hence, their probabilities sum to 1.

   (e) False. The probability that exactly one of two events occurs is

   $$P(A \cup B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$$

   (f) False. We cannot necessarily interpret $f(x)$ as a probability. If X is discrete, then the interpretation is justified. But if X is continuous, then $f(x)$ cannot be interpretted as a probability. For a continuous distribution, the probability of any particular number is 0. Recall that this is why we differentiate between probability mass functions (pmf) and probability density functions (pdf).

   (g) True. Two random variables X and Y are independent if we can separate the joint pmf/pdf into two functions: one of which is only a function of X and the other only a function of Y. Separability also requires that the domains of X and Y be independent of one another as well. Although the domain of X is unrelated to Y and vise verse, the pdf cannot be separated. There is no way to break the joint pdf up into separate functions.

2. MIT, Harvard, and BU students.

   (a) Committee forming:

      i. $\binom{130}{6}$ - There are 130 total students from which to choose six for the committee.

      ii. $\binom{30}{2}\binom{40}{2}\binom{60}{2}$ - There are $\binom{30}{2}$ ways to select the two MIT students, $\binom{40}{2}$ ways to select the two Harvard students, and $\binom{60}{2}$ ways to select the two BU students.

   (b) There are 50 women from which to choose two, and there are 80 men from which to choose the remaining four committee members. Therefore,

   $$P(\text{two women}) = \frac{\binom{50}{2}\binom{80}{4}}{\binom{130}{6}}$$

(c) Each group must have ten MIT students and twenty BU students. One group will have twenty Harvard students, while the other two groups will have only ten Harvard students. (Pity the poor group with twenty Harvard students!) How may ways are there to select the group with twenty Harvard students: $\binom{30}{10}\binom{40}{20}\binom{60}{20}$ - We must select the ten MIT students out of the thirty possible; we must select the twenty Harvard students out of the fourty possible; and we must select the twenty BU students out of the sixty possible. How many ways are there to select the first group with ten Harvard students: $\binom{20}{10}\binom{20}{10}\binom{40}{20}$ - We must select the ten MIT students out of the twenty remaining possible; we must select the ten Harvard students out of the twenty remaining; and we must select the twenty BU students out of the fourty remaining. The leftover students fill out the third group. Therefore, there are

$$\binom{30}{10}\binom{40}{20}\binom{60}{20}\binom{20}{10}\binom{20}{10}\binom{40}{20}$$

ways to split the students into the three groups.

3. This is a conceptually difficult problem because, in essence, we are drawing with replacement; each box we draw from has all fifteen colors in it. It is a hard because you need to turn upside down the problem to be able to use the tools learned in class. If you follow the standard step of counting all possibilities and then dividing to eliminate the double counting (because order doesn't matter), you will get 15^{10} in the numerator. However, it is not clear what goes in the denominator, because colors can be repeated (i.e., the colors are chosen with replacement). One way of approaching this problem is to translate it into a slightly different setting, one more conducive for solving. Imagine, instead of ten boxes with fifteen balls in each box, there are fifteen bins, one for each color. There is a red bin, a yellow bin, etc. So instead of drawing balls out of boxes in order to form an array of colored balls, we place ten markers in the fifteen bins, where a marker represents the selection of a colored ball. This shall be our new mechanism for choosing color arrays. For example, the following choice of balls

Color 1										
Color 2	x							x		
Color 3										
Color 4						x				
Color 5										
Color 6										
Color 7			x		x					
Color 8										x
Color 9										
Color 10							x			
Color 11		x								
Color 12									x	
Color 13										
Color 14										
Color 15				x						
	Box 1	Box 2	Box 3	Box 4	Box 5	Box 6	Box 7	Box 8	Box 9	Box 10

can be represented, equivalently, by the following arrangement of markers in color bins:

	xx		x			xx	x		x	x	x			x
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Thus, the number of ways we can choose the colored balls from the 10 boxes is the same as the number of ways we can put 10 markers in 15 bins. How many ways can the 10 markers be put in the 15 bins? Let a marker be denoted by x, and let walls of the bin be denoted by /; in order to represent fifteen bins we need only 14 walls (we just need dividers between adjacent bins). For example, the arrangement depicted above can be represented as

$$/xx//x///xx/x//x/x/x///x$$

Hence, our problem now is to calculate the number of orderings of ten x's and fourteen /'s. There are

$$\binom{24}{10} = \frac{24!}{10!14!}$$

such orderings There are 24! orderings of all 24 elements; dividing by 10! corrects for the orderings among the markers; and dividing by the 14! accounts for the orderings among the bin walls.

4. For clarity sake, let us explicity set up the problem in terms of specified events. Let $t = 0$ be today and $t = 1$ be tomorrow. Consider the following events:

$$\begin{aligned} G_t &\equiv \text{Gabriel washes at date } t \\ D_t &\equiv \text{Daniela washes at date } t \\ NW_t &\equiv \text{No one washes the dishes at date } t \end{aligned}$$

We are given the following probabilities:

$$\begin{array}{cc} P(G_0) = 5/16 & P(G_1|G_0) = 5/16 \\ P(D_0) = 5/16 & P(D_1|G_0) = 5/16 \\ P(NW_0) = 6/16 & P(NW_0|G_0) = 6/16 \end{array}$$

and

$$\begin{array}{cc} P(G_1|D_0) = 5/16 & P(G_1|NW_0) = 11/16 \\ P(D_1|D_0) = 5/16 & P(D_1|NW_0) = 5/16 \\ P(NW_1|D_0) = 6/16 & P(NW_1|NW_0) = 0 \end{array}$$

These probabilities come from the following:

$$\begin{aligned} P(G_0) &= P(G_1|G_0) = P(G_1|D_0) \\ &= P(3 \text{ or } 4 \text{ heads}) \\ &= \binom{4}{3}\left(\frac{1}{2}\right)^4 + \binom{4}{4}\left(\frac{1}{2}\right)^4 \\ &= \frac{5}{16} \\ P(D_0) &= P(D_1|G_0) = P(D_1|D_0) = P(D_1|NW_0) \\ &= P(3 \text{ or } 4 \text{ tails}) \\ &= \frac{5}{16} \\ P(NW_0) &= P(NW_0|G_0) = P(NW_1|D_0) \\ &= 1 - P(G_0) - P(D_0) = \frac{6}{16} \\ P(G_1|NW_0) &= 1 - P(D_1|NW_0) = \frac{11}{16} \end{aligned}$$

(a) The probability that Gabriela will wash tomorrow given that no one washed today is

$$P(G_1|NW_0) = \frac{11}{16}$$

(b) The probability that no one washed today given that Gabriela washes tomorrow is

$$\begin{aligned}
P(G_1) &= P(G_1|NW_0)\,P(NW_0) + P(G_1|G_0)\,P(G_0) \\
&\quad + P(G_1|D_0)\,P(D_0) \\
&= \frac{11}{16}\left(\frac{6}{16}\right) + \frac{5}{16}\left(\frac{5}{16}\right) + \frac{5}{16}\left(\frac{5}{16}\right) \\
&= \frac{116}{16^2} \\
P(NW_0|G_1) &= \frac{P(G_1|NW_0)\,P(NW_0)}{P(G_1)} \\
&= \frac{\frac{11}{16}\left(\frac{6}{16}\right)}{\frac{116}{16^2}} \\
&= \frac{66}{116} = \frac{33}{58}
\end{aligned}$$

5. Graphs.

(a) The cdf graphed in 5.1 can be written as follows:

$$F(x) = \begin{cases} 0 & x \in (-\infty, 1) \\ 0.1 & x \in [1, 2) \\ 0.2 & x \in [2, 3) \\ 0.3 & x \in [3, 4) \\ 0.3 + 0.1(x-4) & x \in [4, 9) \\ 0.8 & x \in [9, 10) \\ 0.8 + 0.2(x-10) & x \in [10, 11) \\ 1 & x \in [11, \infty) \end{cases}$$

where the slopes of the two non-flat areas are calculated by

$$\begin{aligned}
\frac{0.8-0.3}{9-4} &= \frac{0.5}{5} = 0.1 \\
\frac{1-0.8}{11-10} &= \frac{0.2}{1} = 0.2
\end{aligned}$$

and the equations are just lines. This cdf has the following pmf/pdf:

$$f(x) = \begin{cases} 0.1 & x \in \{1, 2, 3\} \\ 0.1 & x \in [4, 9) \\ 0.2 & x \in [10, 11) \\ 0 & elsewhere \end{cases}$$

See the attachment for the graph of the pmf/pdf.

(b) In order to calculate A, we can integrate under the pdf and set the area equal to 70%.

$$\begin{aligned}
0.70 &= \int_0^A 1.05x^{0.5}dx \\
&= 1.05\left(\frac{2}{3}\right)x^{1.5}\Big|_0^A \\
&= 0.70\left(A^{1.5}\right) \\
A^{1.5} &= 1 \\
A &= 1
\end{aligned}$$

In order to calculate B, we need only evaluate the pdf at A.

$$\begin{aligned} B &= f(A) \\ &= 1.05\,(1)^{0.5} \\ &= 1.05 \end{aligned}$$

C is the simplest to calculate as it is merely all the probability not under the $1.05x^{0.5}$ curve. Hence,

$$\begin{aligned} C &= 1 - 0.70 \\ &= 0.30 \end{aligned}$$

D is the height of a rectangle with area 0.30 and width 0.5

$$\begin{aligned} 0.30 &= 0.5D \\ D &= 0.6 \end{aligned}$$

6. Bivariate calculations.

   (a) Since the exponential function is always positive, the pdf satisfies the positivity requirement. Therefore, we need only check that the pdf integrates to 1 over the entire sample space. Notice that X and Y are independent: $f(x,y) = f_X(x)\, f_Y(y) = e^{-x}e^{-y}$.

$$\begin{aligned} \int_0^\infty \int_0^\infty e^{-x}e^{-y}dydx &= \int_0^\infty e^{-x}dx \int_0^\infty e^{-y}dy \\ &= \left(-e^{-x}\big|_0^\infty\right)\left(-e^{-x}\big|_0^\infty\right) \\ &= (-0+1)(-0+1) \\ &= 1 \end{aligned}$$

   Therefore, this is a well-defined pdf.

   (b) The joint cdf is found as follows:

$$\begin{aligned} F(x,y) &= \int_0^x \int_0^y e^{-u}e^{-v}dudv = \int_0^x e^{-v}dv \int_0^y e^{-u}du \\ &= \left(-e^{-v}\big|_0^x\right)\left(-e^{-u}\big|_0^y\right) \\ &= \left(1-e^{-y}\right)\left(1-e^{-x}\right) \end{aligned}$$

   Using this, we can calculate the following probability easily:

$$\begin{aligned} P(X<1, Y<1) &= F(1,1) \\ &= \left(1-e^{-1}\right)\left(1-e^{-1}\right) \end{aligned}$$

   (c) Because X and Y are independent,

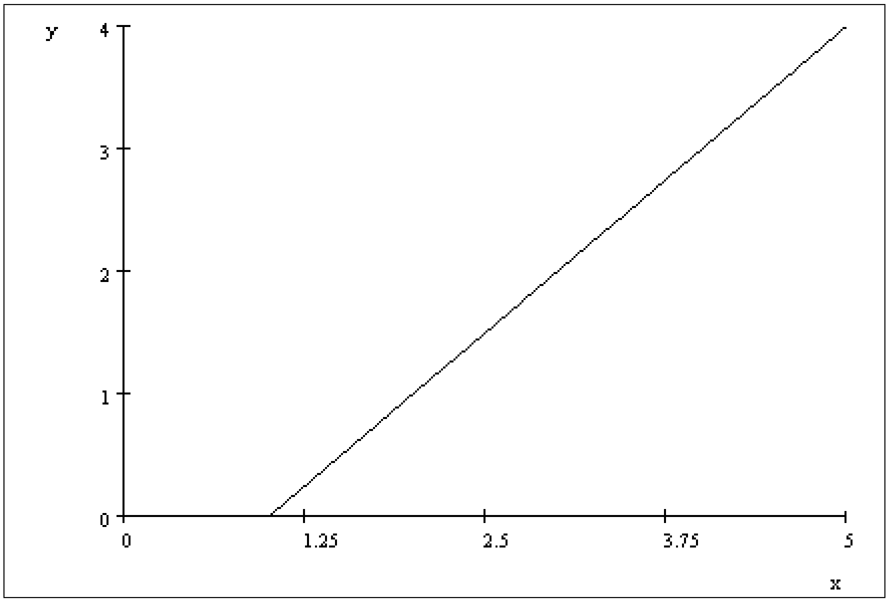
$$f(x) = f_X(x) = e^{-x}$$

   and

$$f(x|y) = f(x) = e^{-x}$$

   Our calculation in part 1 guarantees that the marginal (and, therefore, the conditional) distribution integrates to 1.

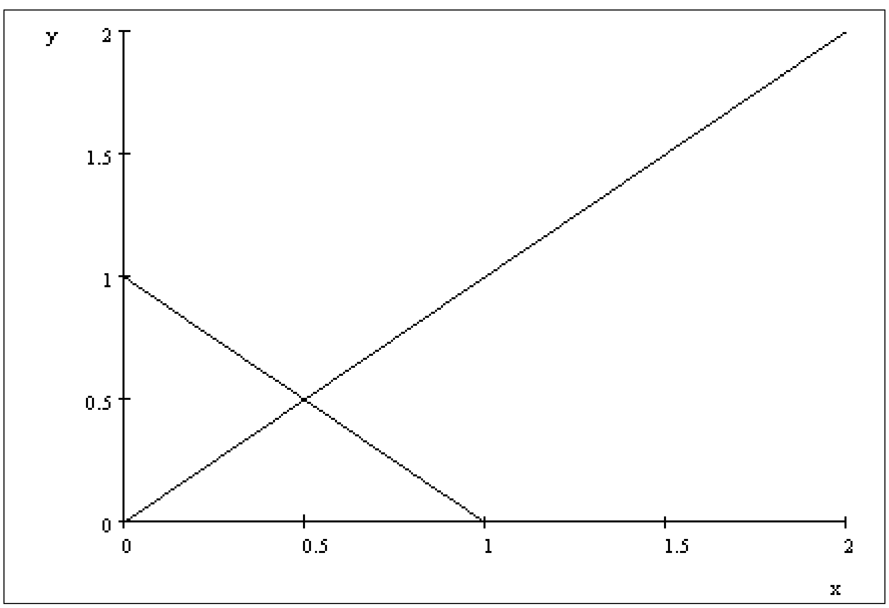
   (d) In order to set up the limits of integration, it is helpful to graph the regions of integration. Consider the equation $X - Y = 1$.



Therefore,

$$\begin{aligned} P(X - Y > 1) &= P(Y < X - 1) \\ &= \int_1^\infty \int_0^{x-1} e^{-x-y} dy dx \end{aligned}$$

(Incidentally, the probability turns out to be $\frac{1}{2}e^{-1}$.) The next probability is a bit tougher. Consider the graphs of $X + Y = 1$ and $X = Y$.



Therefore,

$$P(X + Y > 1, X > Y) = \int_{0.5}^{1} \int_{1-x}^{x} e^{-x-y} dy dx + \int_1^\infty \int_0^x e^{-x-y} dy dx$$

(e) For convenience, let $Z = \frac{X}{Y}$. The cdf of Z can be found as follows:

$$\begin{aligned}
F(z) &\equiv P(Z \leq z) = P\left(\frac{X}{Y} \leq z\right) = P(X \leq zY) \\
&= \int_0^\infty \int_{x/z}^\infty e^{-x-y} dy dx = \int_0^\infty e^{-x} \int_{x/z}^\infty e^{-y} dy dx \\
&= \int_0^\infty e^{-x} e^{-x/z} dx = \int_0^\infty e^{-x(1+1/z)} dx \\
&= -\left.\frac{e^{-x(1+1/z)}}{1+\frac{1}{z}}\right|_0^\infty \\
&= -\left(0 - \frac{1}{1+\frac{1}{z}}\right) \\
&= \frac{z}{z+1}
\end{aligned}$$

5(a)