

### Solutions: Problem Set #1

(1) The following table gives the joint probability distribution  $p(X, Y)$  of random variables  $X$  and  $Y$ .

	X		
Y	1	2	3
1	.02	.04	.12
2	.03	.18	.04
3	.00	.02	.10
4	.09	.18	.18

Determine the following:

(a) Do the entries of the table satisfy the conditions for a bivariate density function?

ANSWER: Yes, since all the entries are non-negative, and they sum to unity (one).

(b) The marginal (or unconditional) probability distributions of  $X$  and  $Y$ . [Note: These will be a collection of probabilities: the probabilities associated with the 3 values of  $X$  and the probabilities associated with the 4 values of  $Y$ ]. ANSWER

$$\Pr(X = 1) = .14, \Pr(X = 2) = .42, \Pr(X = 3) = .44$$

and

$$\Pr(Y = 1) = .18, \Pr(Y = 2) = .25, \Pr(Y = 3) = .12, \Pr(Y = 4) = .45.$$

(c) The conditional probability distributions  $p(X|Y = 3)$  and  $p(Y|X = 1)$ . (Note: The first conditional probability distribution is the collection of three numbers,  $\Pr(X = 1|Y = 3), \Pr(X = 2|Y = 3), \Pr(X = 3|Y = 3)$ .)

ANSWER:Applying our formulas for calculating a conditional from a joint:

$$\Pr(X = 1|Y = 3) = 0, \Pr(X = 2|y = 3) = .02/.12 = 1/6, \Pr(X = 3|Y = 3) = .10/.12 = 5/6.$$

as for the remaining conditional

$$\Pr(Y = 1|X = 1) = .02/.14, \Pr(Y = 2|X = 1) = .03/.14$$

and

$$\Pr(Y = 3|X = 1) = 0, \Pr(Y = 4|X = 1) = .09/.14.$$

(2) (Based on a newspaper article I recently read). Suppose you want to study the impact of a certain *trait*  $T$  on the probability of having a heart attack. Let  $D = 1$  denote the event of a heart attack and  $D = 0$  denote the event of no heart attack. The variable  $T$  is also binary (i.e., it takes on only one of two possible values), and  $T = 1$  indicates that an individual has the trait, and  $T = 0$  indicates that he or she does not have the trait. (You might think of  $T$  as representing if you have a particular gene, if you have high cholesterol, etc).

The article claimed that the trait  $T$  had no effect on the prevalence of heart attacks ( $D = 1$ ) because “among those individuals who had heart attacks, 50 percent of them had the trait ( $T = 1$ ) while 50 percent of them did not ( $T = 0$ ).”

Critically evaluate this statement. Does this result necessarily imply that the probability of having a heart attack is unrelated to having the trait  $T$ ? (Hint: Think about what the statement means formally in terms of conditional probability. You might try and construct some  $2 \times 2$  probability tables to investigate this issue).

ANSWER: What does the statement: “among those individuals who had heart attacks, 50 percent of them have the trait while 50 percent do not” actually mean?

Formally speaking, it means

$$\Pr(T = 1|D = 1) = \Pr(T = 0|D = 1) = .5$$

Since

$$\Pr(T = 1|D = 1) = \Pr(T = 1, D = 1)/\Pr(D = 1)$$

and

$$\Pr(T = 0|D = 1) = \Pr(T = 0, D = 1)/\Pr(D = 1),$$

and both of these conditional probabilities are equal, it follows that

$$\Pr(T = 0, D = 1) = \Pr(T = 1, D = 1).$$

This is the only restriction imposed on the joint distribution of  $D$  and  $T$  that arises from the statement in the article.

So, consider the following  $2 \times 2$  probability table: (this is just my example - there are lots of tables that illustrate the same point!)

	D	
T	0	1
0	.79	.1
1	.01	.1

This is a valid joint distribution that is consistent with the evidence provided in the article.

However, note that

$$\Pr(D = 1|T = 1) = \Pr(T = 1, D = 1)/\Pr(T = 1) = .1/.11 = .91$$

and

$$\Pr(D = 0|T = 1) = \Pr(T = 1, D = 0)/\Pr(T = 1) = .01/.11 = .09.$$

So, for this particular joint distribution, having the trait  $T$  clearly makes it far more likely that the individual will have the heart attack  $D$ !

(3) Extra Credit: (Worth 2 percentage points on the first exam).

*(Based on The Monte Hall Problem)* In a popular game show, a contestant is asked to pick one of three boxes. One of the three boxes contains a prize, while the other two are empty. After the contestant makes her initial selection, the game show host opens one of the two boxes which was not chosen, and shows her that this box is empty. (The host knew that this box was empty all along). The contestant then has the option to either stay with her

original choice, or switch to the other unopened box. If the contestant wishes to maximize her probability of getting the prize, what should she do? (Stay with her original choice, or switch?)

(Hint: This is all conditional probability! To get credit for your answer, you MUST support your argument using laws of conditional probability. Guesses or intuition will not be sufficient to earn the extra credit).

ANSWER: Let  $\hat{C}$  denote the box that is chosen by the contestant and let  $C$  denote the actual location of the prize. Let  $M$  denote the box that Monte actually reveals as being empty. Without loss of generality, let us suppose that  $\hat{C} = 1$  and  $M = 2$ . We also assume

$$\Pr(\hat{C} = j) = 1/3, \quad j = 1, 2, 3.$$

and

$C$  and  $\hat{C}$  are independent.

These assumptions are reasonable - the contestant randomly chooses among the three doors, and the choice made is independent of the actual location of the prize. The second of these assumptions implies

$$\Pr(\hat{C} = j|C) = \Pr(\hat{C} = j) = 1/3, \quad j = 1, 2, 3.$$

To evaluate if the contestant should switch her choice, we need to consider and quantify the following two probabilities:

$$\Pr(C = 3|M = 2, \hat{C} = 1) \quad \text{and} \quad \Pr(C = 1|M = 2, \hat{C} = 1).$$

Since

$$\Pr(C = 3|M = 2, \hat{C} = 1) = [\Pr(M = 2, \hat{C} = 1|C = 3)\Pr(C = 3)]/\Pr(M = 2, \hat{C} = 1).$$

and

$$\Pr(C = 1|M = 2, \hat{C} = 1) = [\Pr(M = 2, \hat{C} = 1|C = 1)\Pr(C = 1)]/\Pr(M = 2, \hat{C} = 1).$$

it follows that

$$\frac{\Pr(C = 3|M = 2, \hat{C} = 1)}{\Pr(C = 1|M = 2, \hat{C} = 1)} = \frac{\Pr(M = 2, \hat{C} = 1|C = 3)}{\Pr(M = 2, \hat{C} = 1|C = 1)}$$

$$\begin{aligned}
&= \frac{\Pr(M = 2|\hat{C} = 1, C = 3)\Pr(\hat{C} = 1|C = 3)}{\Pr(M = 2|\hat{C} = 1, C = 1)\Pr(\hat{C} = 1|C = 1)} \\
&= \frac{\Pr(M = 2|\hat{C} = 1, C = 3)}{\Pr(M = 2|\hat{C} = 1, C = 1)}
\end{aligned}$$

The numerator of this last expression must be 1, since Monty has no choice but to reveal box 2 if the contestant actually selected box 1 and the location of the prize is in box 2. We are not certain of the value of the denominator, but we know it must be less than or equal to one. It is reasonable to assume that Monty randomly selects between the two empty boxes if the contestant actually correctly selects the location of the prize. If so, the probability in the denominator of the above is  $(1/2)$  and thus the entire ratio reduces to 2!

This means that the contestant is twice as likely to win the prize by switching her choice when given the chance!