K300 (4392) Statistical Techniques (Fall 2007)

Assignment 3: Probability (Due October 1)

Instructor: Hun Myoung Park
kucc625@indiana.edu, (317) 274-0573

Please read the following instructions carefully to do this homework (120 points) successfully.

- You MAY NOT use a wordprocessor (e.g., Microsoft Word and WordPerfect); **Use separate sheets and write down answers by hand.**
- Answer all questions. Do not skip any one.
- Show how you obtain the answers. In most questions, a single number may not be accepted as the answer.
- Once you finish, print out your assignment; put it into an envelope; and then hand in on Monday, October 1.
- **Deadline and late penalty are strictly applied** as described in the syllabus. Do not wait until the last minute. Hurry up to meet the deadline.
- You MAY NOT discuss with other classmates when answering questions.

If you have any problem with any of the questions, please email the instructor or come and see the instructor during office hour MW 2:00-3:00 P.M. Or you may make an appointment with the instructor.

Read questions carefully. Do not omit any question. Do not solve wring questions. This is a Type III error! For a question "George, can you hear me now?", do not answer as "No, I cannot see you well." #%$@*&^%#@$!!!

1. (2 points) Identify the sample space for tossing a coin four times. The coin is fair so that we have 50-50 percent chance of getting head or tail. You may use (but do not report) a tree diagram described in Example 4-4 on page 181. *See page 179-180*.

HHHH; HHHT; HHTH; HHTT; HTHH; HTHT; HTTH; HTTT;
THHH; THHT; THTH; THTT; TTHH; TTHT; TTTH; TTTT

Hun: sample space include all possible outcomes. Since you toss a coin four times, you will have 16=2 ×2 ×2 ×2 outcomes. Alternatively, you select one out of two choices (head and tail) in each trial. So there are ${}_2C_1$=2 outcome. Each trial is independent and thus multiplied to get the total number of outcomes. After listing the outcomes, you'd better check if you have all 16 outcomes. Unfortunately, some of you failed to list all of them.

2. (5 points) Using the sample space you identified in question 1 above, find the following probabilities. These are classical probabilities. *See page 182-185*.

1) Probability of getting all head. 1/16, HHHH
2) Probability of getting all tail. 1/16, TTTT

3) Probability of getting two heads and two tails. 6/16, HHTT, HTHT, HTTH, THHT, THTH, TTHH
4) Probability of getting three tails and one head. 4/16, HTTT, THTT, TTHT, TTTH
5) Probability of landing on its edge (neither head nor tail). 0/16, 0 percent probability of getting outcome other than head and tail.

3. (5 points) Solve question 15 on page 192. You MUST show how you get the answers. Do not simply copy the answers in the textbook. In addition, find the probability that the customer get neither money nor a coupon (2 points). *See page 185-187*.

a. There are four 1. 4/24=1/6, there are four 1.
b. There are four 1 and eight 2. 12/24=1/2, there are four 1 and eight 2.
c. There are eight 3 and four 4. 12/24=1/2, there are four 4 and eight 3.
d. No number other than 1 through 4. 0/24=0, none.

4. (10 points) Solve question 40 on page 194. You MUST show how you get the answers. Do not forget to find the sample space such as (0, 0), (0, 1)… Odd numbers are 1 and 3, while even numbers are 0, 2, and 4. Identify all relevant outcomes first.

Sample space (25): (0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (1, 0), (1, 1), (1, 2), (1, 3), (1, 4), (2, 0), (2, 1), (2, 2), (2, 3), (2, 4), (3, 0), (3, 1), (3, 2), (3, 3), (3, 4), (4, 0), (4, 1), (4, 2), (4, 3), (4, 4)

a. (1, 0), (1, 2), (1, 4), (3, 0), (3, 2), (3, 4). 6/25
b. (1, 4), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4). 10/25=2/5
c. (0, 0), (0, 2), (0, 4), (2, 0), (2, 2), (2, 4), (4, 0), (4, 2), (4, 4). 9/25
d. (0, 1), (0, 3), (1, 0), (1, 2), (1, 4), (2, 1), (2, 3), (3, 0), (3, 2), (3, 4), (4, 1), (4, 3). 12/25
e. (0, 0), (1, 1), (2, 2), (3, 3), (4, 4). 5/25=1/5

Hun: You have five outcomes per trial. Since each trial is independent of other trials, the total number of possible outcomes is $25=5\times5=5^2$. I would like to list the proper outcomes first in order to avoid counting errors. Some of you committed such a silly mistake.

5. (5 points) Determine whether two events are mutually exclusive. *See pages 195-197 and question 2 on page 199*.

1) Rolling a die: get a prime number (2, 3, 5) and get an even number. No
2) A student is a junior and a student is a girl student. No
3) A student registers K300 and a student takes a public management course. No
4) Indianapolis Colts beat Jacksonville Jaguars and Jaguars beat Colts on October 22 Yes
5) You watch a movie and drink a coke. No

6. (5 points) See the car information in question 13 on page 200. Find 1) probability that you select a domestic or compact car; 2) probability that you select a foreign SUV; 3) probability that you select a foreign or SUV; 4) probability that you select a compact or mid-size car; 5) probability that you select a foreign or domestic car. Show how you get the answers. *See pages 197-198.*

	SUV	Compact	Mid-sized	Total
Foreign	20	50	20	90
Domestic	65	100	45	210
Total	85	150	65	300

1) (65+100+45)/300+(100+50)/300-100/300=210/300+150/300-100/300=260/300
2) 20/300
3) (20+50+20)/300+(20+65)/300-20/100=90/300+85/300–20/300 =155/300
4) (50+100)/300+(20+45)/300=150/300+65/300=215/300
5) (20+50+20)/300+(65+100+45)/300=90/300+210/300=300/300=1

Hun: I counted the total number of cars. It is 300. And then I count cars in proper cells. In order to get the number of domestics or compact cars, for example, I got the sum of all domestic car (210) and the sum of all compact cars (150) and then subtract domestic compact cars (100, see the second row and second column). Keep in mind event of foreign/domestic and SUV/compact/mid-sized are not mutually exclusive.

7. (3 points) See the example 4-25 on pages 206-207. Note that a ball selected is replaced. Find 1) probability that you select 2 red balls; 2) probability that you select 1 white ball and then 1 red ball; 3) probability that you select 1 blue and then 1 red ball. *See pages 64-65.*

1) 3/10 X 3/10 = 9/100
2) 5/10 X 3/10 = 15/100
3) 2/10 X 3/10 = 6/100

Hun: the probabilities of selecting red balls, blue balls, and white balls are 3/10, 2/10, and 5/10, respectively. For example, 3/10 means there are a total of 10 balls in an urn and three of them are red balls. Since a ball selected is replaced, each trial is independent, these probabilities do not change at all. The events in questions involve "AND" (e.g., "select 1 blue **and** then 1 red ball") operation; you have to multiply the two probabilities.

8. (2 points) Solve question 6 on page 215 by changing prison inmates selected from 2 to 3. Show how you get the answer. *See 205-208.*

.25 X .25 X .25 = .0156

Hun: the probability of selecting non-U.S. citizens is 25%, P(x=non-U.S. citizen)=.25. Do not write as "P(.25)" because this means the probability of event .25. Random selection of each trial is independent. In order word, the probability remains unchanged. Why? This sampling involves large N. Think about the total number prison inmates. Taking one of two does not affect the overall probability significantly. You'd better interpret as taking 3 random samples at a time (as opposed to serially: take one and check if he/she is U.S. citizen; and then pick another; and select the third). This event involves AND operation: select non-U.S. citizen, and select non-U.S. citizen again, and again select non-U.S. citizen. The probability .25 do not necessarily mean that there are 25 successes among 100 trials (N=100).

9. (2 points) Solve question 13 on page 215 by selecting 3 children instead of 4. Show how you get the answer. *See 205-208.*

The probability that a child is not covered is .12 = 1-.88.
.12 X .12 .12 = .0017
Hun: The probability that U.S. children are not covered by some type of health insurance is .12=1-.88. Note that .88 is the probability that U.S. children are not covered. This is a complementary probability. The each trial is independent; the probability remains unchanged. Again this event involves AND operation.

10. (5 points) Solve question 35 on page 217. Show how you get the answer. You should use the formula on page 210.

	Corporation	Government	Individual	Total
United States	70,894	921	6,129	77,944
Foreign	63,182	104	6,267	69,553
Total	134,076	1,025	12,396	147,497

a. (63182/147497)/(134076/147497) = .4712
b. (6129/147497)/(77944/147497) = .0786

Hun: First, I counted all patents to get 147,497. P(foreign|corporation)=P(foreign AND corportation)/P(corporation). I found 63182 patents granted to foreign corportation (see second row and first column). The total number of patents given to corporation is 134076. So, P(foreign|corporation)= (63182/147497)/(134076/147497). You may simply get 63182/134076 in this case. But computation of a conditional probability is not always simple and intuitive.

11. (2 points) Solve question 27 on page 216. Show how you get the answer. *See 210-212.*

P(stolen and found)=.0009
P(stolen)=.0015
P(found|stolen)=.0009/.0015=.6

Hun: the question is the conditional probability that a car will be found within one week given the car is stolen). P(found|stolen)=P(stolen AND found)/P(stolen)=.0009/.0015.

12. (5 points) See example 4-34 on pages 211-212. Find 1) the probability that the respondent answered no, given that the respondent was a male; 2) the probability that the respondent was a male, given that the respondent answered no. 3) Do you have the same probability in 1) and 2)? *See pages 210-211.*

1) P(No|Male)=(18/100)/(50/100)=18/50
2) P(Male|No)=18/100/(60/100)=18/60
3) No

Hun: in order to compute P(no|male), you need to get P(no AND male) and P(male). There are 50 male people out of 100. P(male)=50/100. There are 18 male who answered as Yes. P(no AND male)=18/100. Thus P(no|male)=(18/100)/(50/100). There are 60 people who answered as No. P(no)=60/100.

13. (5 points) Solve question 32 on page 216. Show how you get the answer. Suppose that 70 percent of the customers order salad. Do you think ordering pizza and salad are statistically independent events? *See 208, 210-212*.

P(pizza)=.95
P(pizza and salad)=.65
P(salad|pizza)=.65/.95=.6842
P(salad)=.70
Since P(salad) is not equal to P(salad|pizza), ordering pizza and ordering salad are not statistically independent. P(pizza) and P(salad) are not independent since p(salad|pizza)=.6872≠.70=P(salad).

Hun: question is the conditional probability of ordering salad, given the he orders pizza. P(salad|pizza)=P(salad and pizza)/p(pizza). P(salad and pizza) is the probability of ordering both pizza and salad, not of ordering pizza OR salad. P(salad AND pizza)≠P(salad OR pizza). Alternatively, P(pizza|salad)=P(pizza and salad)/p(salad)=.65/.70=.93≠.95=P(pizza). Therefore, P(salad) and P(pizza) are not independent.

14. (10 points) According to the Statistical Abstract of the United States, 70 percent of females age 20 to 24 have never been married. Let us call this event S. Event M is that females have been married at least once. 1) Compute P(M) + P(S). Are events S and M mutually exclusive? Why? 2) Compute P(M|S), the conditional probability of M given S. 3) Show whether events S and M are statistically independent. Are they statistically independent? Do not simply answer as yes or no. *See question 37 on page 217 and question 53 on page 218*.

1) P(M) + P(S) = .3 + .7 = 1. Yes. A female cannot be S and M at the same time.
2) P(M|S)=0/.7=0
3) Since P(M) is not P(M|S), events M and S are not statistically independent.

Hun: P(x=S)=.3, the probability that females age 20 to 24 have never been married. P(x=M)=.7=1-.3 since event S and M are mutually exclusive. P(single AND married) is the probability of being single and married at the same time. No one can be a single and get married at a particular point. How about your sister and mom? Two events are mutually exclusive. Therefore, P(M and S)=0. P(M|S)=P(M and S)/P(S)=0/.7=0. P(M|S)=0≠.3=P(M), so M and S are not independent.

15 (2 points) Solve question 34 on page 236. In addition, find the number of options if an employee selects two health care plans and one retirement plans without expense accounts. Show your work using formula. *See 229-231*.

${}_5C_1$X ${}_3C_1$ X ${}_2C_1$ = 30
${}_5C_2$X ${}_3C_1$ = 30

Hun: Combination itself is not a probability. Combination tells you the number of ways that select *r* out of *N*. Imagine there are three bags; the first bag has 5 balls (five different health care plans), the second bag has three balls (three different retirement plans), and the last bag has two balls (two different expense accounts). How many ways do you have if you select 2 balls from the five in the first bag? Combination gives you the answer. It is ${}_5C_2$=5!/(2!3!)=10. Do not write as 5C2; you have to use subscripts. Taking balls from each bag is independent and involves AND operation. The first question is to take a health care plan out of five from the first bag, and take a retirement plan out of three from the second bag, and then take a expense account out of two from the last bag. That is, ${}_5C_1\times{}_3C_1\times{}_2C_1$. Second question can be answered as ${}_5C_2\times{}_3C_1\times{}_2C_0={}_5C_2\times{}_3C_1$. ${}_2C_0$ is to select none out of 2; there is only one way of not selecting at all.

16 (10 points) Solve question 10 on page 233. Show your work using formula. Fractions such as 1/91 and 5/9 will do for your answers. *See 229-230*.

a. ${}_8C_3/{}_{15}C_3$ =8/65
b. [${}_8C_1$ + ${}_5C_1$]/[${}_{15}C_3$]=1/35
c. ${}_5C_3/{}_{15}C_3$ =2/91
d. [${}_8C_1$ X ${}_5C_1$ X ${}_2C_1$]/[${}_{15}C_3$]=16/91
d. [${}_8C_2$ X ${}_5C_1$]/[${}_{15}C_3$]=4/13

Hun: There are 15 (=8+5+2) choices. The total number of ways of selecting 3 out of 15 is ${}_{15}C_3$. Why? The combination dictates exactly what you want to get (selecting 3 out of 15). We have three bags again. This time, the first bag has 8 balls with different numbers. The second and third respectively have 5 and 2 balls. The first question is the probability of selecting all 3 from the life policies. The number of ways of selecting 3 out of 8 is ${}_8C_3$. Therefore, the probability is ${}_8C_3/{}_{15}C_3$. The second question is to select two homeowner policies and one policy from either life policy or automobile policy. This process involves both AND and OR operations. In fact, AND operation can be skipped since there is only one way of selecting two homeowners out of two. (${}_8C_1$ + ${}_5C_1$) ${}_2C_2$=${}_8C_1$+${}_5C_1$. Therefore, the probability is (${}_8C_1$+${}_5C_1$)/${}_{15}C_3$. For the last question, you may ignore homeowner policies because is ${}_2C_0$=2!/(0!2!)1. Note that 0!=1≠0.

17 (5 points) Solve question 35 on page 239. Show your work. You may take advantage of complementary probability, which asks the probability that a person does not suffer from a heart attack. *See 229-231*.

1-.45^6=.9917

Hun: You have two approaches. First, consider the all possible outcomes and sum their probabilities up. “at least 1 of 6” means 1, 2, 3, 4, 5, 6. In other word, you have to compute P(x=1)+P(x=2)+P(x=3)+P(x=4)+P(x=5)+P(x=6), where x is the number of peoples who have suffered a heart attack. For example, P(x=1) is ${}_6C_1(.55)^1(.45)^5$. Use your calculator to compute the probability. I know this is not fun at all. For approximate probably, look at page 626 for the binomial distribution table. It is about .094 when using p=.5. Similarly, you may get the approximate P(x=2) of .234. And so on. Thus, the sum of P(x=1) through P(x=6) is 1-P(x=0)=.984. This is not correct answer but rough estimation because p is not .5 but .55. This approach is not recommended. Why? It is time consuming to push buttons of your calculator. This is the reason why I recommend you use the complementary probability. (Of course, it is up to you to take approaches. If you are quite comfortable with the first approach, feel free to go there.) Let us compute the complementary probability first. The probability of a person has NOT suffered a heart attack is 1-.55=.44. The question can be rewritten as “1-the probability that all 6 people have not suffered from a heart attack.” Why? Compare this statement to “at least 1 of 6 will have had a heart attack.” Two statements are equivalent. The probability that all 6 people have not suffered is .45^6 since this event is statistically independent. Select one not-suffered person, and select another not-suffered person, and select another not-suffered again, and then select the last not-suffered. .45^6=.45×.45×.45×.45. Therefore, the answer is 1-.45^6.

18 (5 points) Let *x* the number of heads you get when tossing a fair coin four times. 1) Construct a table for this probability distribution. The first row is labeled as “x” and the second row as “P(x)”. Note you have five outcomes, 0 through 4. And then 2) compute the expected value and variance of *x*. *See 251-254, 266.*

X	0	1	2	3	4
P(x)	1/16	4/16	6/16	4/16	1/16

N=4, p=1/2
Expected value is 2=4 X ½ or 0 X 1/16 + 1 X 4/16 + 2 X 6/16 + 3 X 4/16 + 4 X 1/16
Variance is 1=4 X ½ X ½

Hun: First, find the all possible outcomes (sample space). Since you toss a coin four times, there are five outcomes. You may get no head out of four trials (x=0); get one head out of four trials (x=1); and so on. Keep in mind that this probability distribution is different from the example in question 1. Of course, you may get probabilities from question 1. For example, P(x=0)=1/16. There is only one case of TTTT. Alternatively, you may get the probability using the binomial distribution:$1/16={}_4C_0(.5)^0(.5)^4$, $P(x=1)=4/16={}_4C_1(.5)^1(.5)^3$, $P(x=2)=6/16={}_4C_2(.5)^2(.5)^2$, … See Example 264 as well. Once you know the N=4 and p=1/2, it is a piece of cake to get the expected value (mean) and variance! Why? As shown in page 266, mean=np and variance=np(1-p). Of course, you may use the formula in pages 252, 254. But it is time consuming and painful sometimes.

19 (2 points) Solve question 9 on page 259. Show your work. You should report at least two digits below the decimal point. *See 251-254.*

Expected value is 13.86=12 * .15 +13 * .20 + 14 * .38 + 15 * .18 + 16 * .09
Variance is 1.3204 = (12-13.86)^2*.15+(13-13.86)^2*.20+(14-13.86)^2*.38+(15-13.86)^2*.18+(16-13.86)^2*.09
Standard deviation is 1.1491=sqrt(1.3204)

20 (5 points) See example 5-16 on pages 264-265. 1) Find the probability that exactly 2 people visited a doctor last month. You may take advantage of the binomial distribution table on pages 626-627 to get the probability. 2) What is the probability that less than 4 people (none, 1, 2, and 3) have visited a doctor. 3) compute the expected value and variance of this probability distribution. *See 263-264 and 266-267.*

1) P(2) is ${}_{10}C_2\,(1/5)^2(4/5)^8$=.3020
2) .8791= P(0)+P(1)+P(2)+P(3)=.107+.268+.302+.201
3) N=10, p=1/5, q=4/5. Expected value is 10*1/5=2. Variance is 10*1/5*4/5=8/5

Hun: First, I would get information of N and p. N=10, p=1/5=.2, q=1-p=4/5=.8. P(2)=${}_{10}C_2\,(1/5)^2(4/5)^8$ I would look up the binomial distribution table to get the probability. I am a kind of "lazy bird." Try to find the probability for N=10, x=2, and p=.2. It is .201. "Less than 4" means 0 through 3. Thus, you have to get P(0), P(1), P(2), and P(3), and then sum them up. Using the table is much easier to compute the binomial distribution function.

21 (5 points) Solve question 10 on page 269. In addition, find the probability when the fraud rate is changed from .6 to .7. Is this event likely? *See 263-264.*

1) .0425= ${}_{10}C_3\,(.6)^3(.4)^7$
2) .0090= ${}_{10}C_3\,(.7)^3(.3)^7$ This event is not likely.

Hun: Let us check the information first. p=.60 and N=10. q=1-p=.40. What you need to do is plug this information in the formula. P(x=3)= ${}_{10}C_3\,(.6)^3(.4)^7$.

22 (5 points) See question 5 on page 269. Let us assume that the probability that a student know a correct answer is .70 instead of .50, and that this probability does not change across questions. Yes, this is a strong assumption. A student will pass the exam when he misses fewer than 2 questions (none or only one) out of the total 10. What is the probability that a student passes this exam? Is it likely? You may take advantage of the complementary probability, which is the probability that a student misses none or one question. *See 263-264.*

The probability that a student misses none of the 10 questions is .0282= ${}_{10}C_0\,(.3)^0(.7)^{10}$
The probability that a student misses only one of the 10 questions is .1211= ${}_{10}C_1\,(.3)^1(.7)^9$

Thus, the probability of passing the exam is .1493=.0282+.1211. Yes, it is not a tough exam. You have about 15 percent chance to pass the exam.

Hun: p=.70, q=.3, and N=10. The "fewer than 2 questions missed" means that he/she hits nine or ten questions. So you need to compute P(x=9) and P(x=10). $P(x=9)={}_{10}C_9 (.7)^9(.3)^1={}_{10}C_1 (.3)^1(.7)^9$ Again, I would look up the binomial probability table on page 627. P(x=9)=.121 and P(x=10)=.028. Depending on subjective criterion, the sum of two probabilities may or may not likely. When conventional significance level of .05 is applied, the probability is likely.

23 (5 points) Solve question 20 on page 270 by changing the seating capacity from 80 to 100. In addition, find the probability that at least 2 out of 10 randomly select patrons will smoke. You may take advantage of the complementary probability. *See 263-264.*

N=100, p=.42, q=.58
Expected value is 42=100*.42. Variance is 24.36=100*.42*.58. Standard deviation is 4.9356. The restaurant may have at least 42 seats for smoking customers.

$p(1)=.0312={}_{10}C_1 (.42)^1(.58)^9$
$p(0)=.0043={}_{10}C_0 (.42)^0(.58)^{10}$
p(2)+p(3)…p(8)=1-.0312-.0043= .9645

Hun: Identify the necessary information of N, p, and q, and then plug them into the formula. "At least 2 out of 10" means 2, 3, 4, 5, 6, 7, 8, 9, or10 out of 10. Therefore, you have to compute P(x=2)+P(x=3)…P(x=10). We know that P(0)+P(1)+…P(10)=1 because the sum of all probability is 1. Therefore, we can compute 1-P(0)-P(1) instead of P(2)+P(3)…P(10). What is P(1)? $p(x=1)={}_{10}C_1 (.42)^1(.58)^9$

24 (10 points) Indiana Pacers won 35 games out of 82 games last season. The winning rate is about .4. One of your friends insists that Pacers become much powerful this season and will win at least 6 games (6 and 7) among the first 7 games. You are stroked by the binomial distribution we discussed and decide to show how likely his forecast is. 1) Provide N (the number of trials), r (the number of wins), p(probability of wining a game), and q(probability of losing a game); 2) compute probability that Pacers win at least 6 out of the first 7 games; 3) Is his forecast likely? Based on this result, what would you tell your friend? 4) (5 points) Now, you get to remember the key features of a binomial experiment that the Instructor emphasized. For example, the outcomes of each trial must be independent of each others and the probability of success must remain the same for each trial. Is there any problem in your reasoning about forecasting the likelihood? Is the winning rate of .4 still valid for this season? If yes, explain why and how your reasoning is problematic. Otherwise, defend your reasoning and interpretation.

1) N=7, r=6, 7, p=.4, q=.6
2) $P(6)+P(7)={}_7C_6 (.4)^6(.6)^1+{}_7C_7 (.4)^7(.6)^0=.0172+.0016=.0188$
3) Not likely. I do not agree with you. Show me solid evidence that Pacers changed dramatically.

4) The rate of .4 is not valid. If Pacers become powerful and other teams remain unchanged, the rate should be higher than that. But it is not possible to measures exactly relative strengths of all NBA teams and compute new winning rates. The most important thing you have to remember is that the rate may change depending on the outcomes of previous games. Suppose Pacers win the first two games. Probably players may become more excited about the next game and they are more likely to win the next game assuming other things remain unchanged. If Pacers is defeated consecutively, probably the rate may decrease. What if some players got injured, what if the weather is too bad, what if… There are so many factors that may influence the winning rate, but that we cannot control. Therefore, it is not reasonable to forecast the first 7 games using the binomial distribution even if the correct new rate is available. The data generation process is not the Bernoulli process! Consequently, we do not know whether your friend's assertion is right or wrong since your analysis is neither valid nor reliable. We do not have any reliable information. This is the fundamental nature of sports games.

Hun: "At least 6 out of 7" means 6 and 7 wins in 7 games. So you have to get both $P(x=6)$ and $P(x=7)$, and then sum them up.