COMP 680

Holly Do

Problem Set 03

In the problems that follow, you do not need to write down a very large (or very small) number. It is acceptable to have your answer incorporate factorial and/or binomial coefficients as the answer.

1. [12 pts] Among employees of a certain firm, 70% know Python, 60% know Java, and 50% know both languages. What portion of programmers

1.1. [2pts] does not know Python?

Since 70% among the employees know Python, 30% does not know Python.

1.2. [2pts] does not know Python and does not know Java?

The portion of programmers that knows Python or Java is

$P(P \cup J) = P(P) + P(J) - P(P \cap J) = 70\% + 60\% - 50\% = 80\%$

So the portion that does not know Python and does not know Java is

$100\% - 80\% = 20\%$

1.3. [2pts] knows Java but not Python?

Portion that knows Java but not Python is the portion that knows Java, 60%, minus the portion that knows both so 60% - 50% = 10%

1.4. [2pts] knows Python but not Java?

Portion that knows Python but not Java is the portion that knows Python, 70%, minus the portion that knows both so 70% - 50% = 20%

1.5. [2pts] If someone knows Python, what is the probability that he/she knows Java too?

$P(Java \mid Python) = P(Java \cap Python) / P(Python) = 50\% / 70\% = 5/7$

1.6. [2pts] If someone knows Java, what is the probability that he/she knows Python too?

$P(Python \mid Java) = P(Java \cap Python) / P(Java) = 50\% / 60\% = 5/6$

2. [12pts] Suppose there is a recently discovered new (hopefully rare) disease called LiveSession Pneumonia (LSP). A careful study shows that 5% of the student population actually has LSP. Recently, a new test has been developed for LSP. Since it is a new test, there is a 2% false positive rate. The false negative rate, however, is 0.

2.1. [4pts] Suppose a student tests positive for LSP. What is the probability that the student actually has LSP?

$P(lsp+ \mid t+) = (P(t+ \mid lsp+) * P(lsp+)) / P(t+) = (1 * 5\%) / ((1*5\%)+(2\%*95\%)) = 0.05 / (0.05+0.019) = 0.7246$

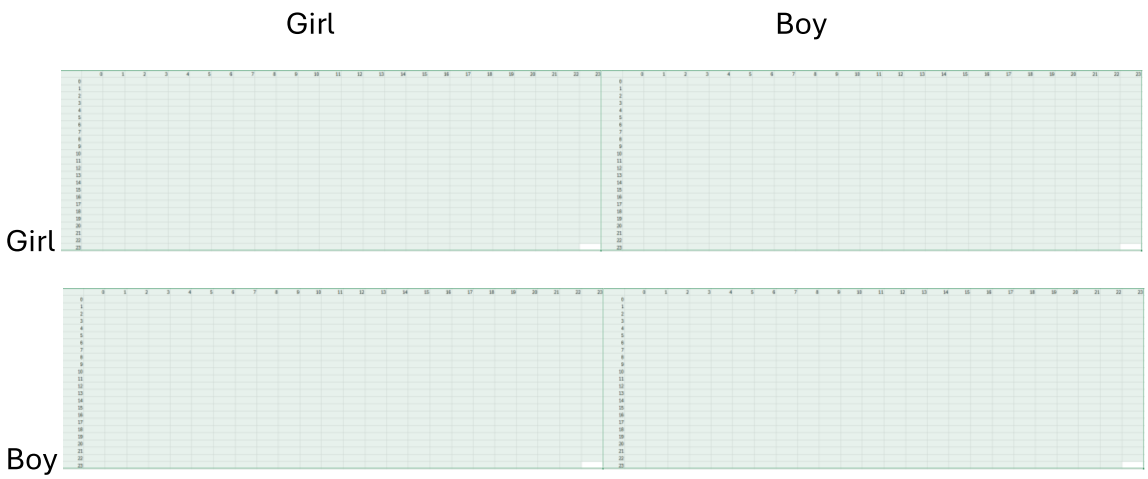
2.2. [4pts] Suppose a new test is developed. The new test is found to have an improved 1% false positive rate. The new test, however, also has a false negative rate of 1%. When using the new test, if a student tests positive for LSP, what is the probability that the student actually has LSP?

$P(lsp+ \mid t+) = (P(t+ \mid lsp+) * P(lsp+)) / P(t+) = (99\% * 5\%) / ((99\%*5\%)+(1\%*95\%)) = 0.0495 / (0.0495+0.0095) = 0.0495/0.059 = 0.838983$

2.3. [4pts] Using the new test of problem 2.2, if a student tests negative for LSP, what is the probability that the student does not have the disease?

$P(lsp- \mid t-) = (P(t- \mid lsp-) * P(lsp-)) / P(t-) = (99\% * 95\%) / ((1\%*5\%)+(99\%*95\%)) = 0.9405 / (0.0005+0.9405) = 0.9405/0.941 = 0.99946865$

3. [14 pts] Jack and Jill Fertile have 2 children of different ages. Assume the probability of a boy (or girl) is 0.5. We are interested in the birth hour, numbered from 0 to 23. The birth hour is computed by noting the (military) time hour digits (there is no rounding). Further suppose that a child is equally likely to be born at a given birth hour. Now suppose you are told that the Fertiles have a girl with birth hour 11.



From counting the blocks on this graph:

G11 = Girl born at 11

C = Any child

[O,Y] = older and younger order

The Fertiles have a girl with birth hour 11 so:

P(1 G11) = P(G11-C U C-G11) = P(G11-C) + P(C-G11) - P(G11-G11) = (24+24+24+24-1)/(48x48) = 95/2304

The probability that the oldest child is a girl given the Fertiles have a girl with birth hour 11

P(G-C | 1 G11) = P (G-C $\cap$ 1 G11) / P(1 G11) = (24+24+24-1) / 95 = 71/95

4. [14 pts] An emission test is being performed on n individual automobiles.

Each car can be tested separately, but this is expensive. Pooling (grouping) can decrease the cost: The emission samples of k cars can be pooled and analyzed together. If the test on the pooled sample is negative, this 1 test suffices for the whole group of k cars and no more tests are needed for this group. If the test on the pooled sample is positive, then each of the k automobiles in this group must be tested separately. This strategy is referred to as a (n,k)-pooling strategy.

Suppose that we create n/k disjoint groups of k automobiles (assume n is divisible by k) and use the pooling method. Assume the probability that a car tests positive is p, and that each of the n individuals' autos are "independent," i.e., their tests are independent of one another.

Finally suppose that the cost for testing an emission sample is C, no matter how many individual elements are pooled in the sample.

4.1 [6 pts] Given a pooled sample of k autos, what is the expected cost to test the sample so that results are known for each individual auto?

p = probability of positive

q = probability of negative = 1 - p

n = total number of items

k = pool size

c = cost

For each pool, if it is tested negative, then every item is negative. If it is tested positive, then at least one item is tested positive.

The probability that all n items are negative is: q^k

The probability of at least one item tested positive is: 1 - q^k

The expected number of test is:

E(T) = 1 + (1-q^k)*k

The expected cost is: E(T)*C

The expected cost per car is: (E(T)*C) / k

4.2 [4 pts] Compute the testing cost per car for n = 1000, p = 0.02, k = 10, C = $100.00

p = probability of positive = 0.02

q = probability of negative = 1 - 0.02 = 0.98

n = total number of items = 1000

k = pool size = 10

c = cost = $100

The expected number of tests is:

E(T) = 1 + (1-q^k)*k = 1 + (1-(0.98^10))*10 = 2.83

The expected cost is:

E(T)*c = 2.83 *100 = 283

The expected cost per car is:

E(T)C/k = 283/10 = 28.3

4.3 [4 pts] Compute the testing cost per car for n = 1000, p = 0.02, k = 5, C = $100.00

p = probability of positive = 0.02

q = probability of negative = 1 - 0.02 = 0.98

n = total number of items = 1000

k=pool size = 5

c = cost = $100

The expected number of tests is:

E(T) = 1 + (1-q^k)*k = 1 + (1-(0.98^5))*5 = 1.4804

The expected cost is:

E(T)*c = 1.4804*100 = 148.04

The expected cost per car is:

(E(T)*c)/k = 148.04/ 5 = 29.608

5. [12 pts] 2 fair 4-sided “die” are tossed. Let S be the smaller number of points shown on a die. Let L be the larger number of points shown on a die. If both dice show the same number, say, n points, then S = L = n.

5.1. [3 pts] Find the joint probability mass function of (S, L).

5.2. [3 pts] Are S and L independent? Explain.

5.3. [3 pts] Find the probability mass function of S.

5.4. [3 pts] If S = 2, what is the probability that L = 4?

5.1.

Joint probability mass function of (S, L) has distributions:

	1	2	3	4
1				
2				
3				
4				

P(S=1,L=1) = 1/16

P(S=1,L=2) = 2/16

P(S=1,L=3) = 2/16

P(S=1,L=4) = 2/16

P(S=2,L=2) = 1/16

P(S=2,L=3) = 2/16

P(S=2,L=4) = 2/16

P(S=3,L=3) = 1/16

P(S=3,L=4) = 2/16

P(S=4,L=4) = 1/16

Marginal Probability of S

P(S=1) = 7/16

P(S=2) = 5/16

P(S=3) = 3/16

P(S=4) = 1/16

Marginal Probability of L

P(L=1) = 1/16

P(L=2) = 3/16

P(L=3) = 5/16

P(L=4) = 7/16

5.2.

S and L are not independent because if we take a sample in P(S,L) with S=1, L=1

P(S=1, L=1) is not equal to P(S=1) * P(L=1)

1/16 is not equal to 7/16 * 1/16

5.3.

The probability mass function of S

$P(S=1) = 7/16$

$P(S=2) = 5/16$

$P(S=3) = 3/16$

$P(S=4) = 1/16$

5.4.

If S = 2, what is the probability that L = 4?

$P(L=4 \mid S=2) = P(S=2 \cap L=4) / P(S=2) = (2/16) / (5/16) = 2/5$

6. [12 pts] Tossing a fair 6-sided die is an experiment that can result in any

integer number from 1 to 6 with equal probabilities. Let X be the number of

dots on the top face of a die. Compute E(X) [6 pts] and Var(X) [6 pts].

Because this is a probability distribution with equal probability of 1/6 for each value of X, the expected value is:

$E(X) = 1/6 * (1+2+3+4+5+6) = 21/6 = 3.5$

$E(X^2) = 1/6 * (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = 1/6 * (1+4+9+16+25+36) = 91/6$

$E(X)^2 = (3.5)^2 = 12.25$

$Var(X) = E(X^2) - ((E(X)^2) = 91/6 - 12.25 = 91/6 - 147/12 = 35/12$

7. [12 pts] Let E(A) be the expected value of any rv A.

Assume the following properties of E

1. $E(aX) = aE(X), \forall a$

2. E(X + Y) = E(X) + E(Y)

Show that E(aX + bY) = aE(X) + b E(Y)

Hint You: can prove the result just using properties 1 and 2 above. You need not resort to the definitions.

E(aX + bY) = E(aX) + E(bY) because property#2

E(aX) + E(bY) = aE(X) + bE(Y) because property #1

So E(aX + bY) = aE(X) + bE(Y)

8. [12 pts] It takes an average of 40 seconds to download a certain file, with a standard deviation of 5 seconds. The actual distribution of the download time is unknown. What can be said about the probability of spending more than 1 minute for this download?

u = 40 sec sd = 5 sec

P(X>60) = P((X-40) >20))

P(|X-40|) >20) = P((X-40) >20) + P((X-40) < - 20)

So: P((X-40) >20) <= P(|X-40|) >20)

According to Chebyshev inequality

P(|X-u| > k*sd) < 1/(k^2) with sd = 5 & k*sd=20 so k=4

P(|X-u| > k*sd) < 1/(4^2)

P(|X-40|) >20) < 0.0625

So:

P((X-40) >20) < 0.0625

Or the probability of spending more than 1 minute (or 60 seconds) for this download is less than 0.0625 or 6.25%.