CS 112: Modeling Uncertainty in Information Systems
Midterm Solutions 2012

Problem 1 (25 points)

Every Wednesday before CS112, you meet your friends for lunch. On any given week, independent of your lunch choices on other weeks, you go to the food court at Ackerman with probability 0.6. Otherwise, you go to the Court of Sciences. When you go to Ackerman, there are lots of available options, but the tacos are pretty good, so you order tacos with probability 0.2. There are fewer options at the Court of Sciences, so you end up ordering tacos with probability 0.3 when you go there.

(a) What is the probability that you order tacos on any particular Wednesday?

(b) Given that you did not order tacos on a particular Wednesday, what is the probability that you went to Ackerman?

(c) If you continue this routine for all ten weeks of the quarter, what is the expected number of times you will go to Ackerman?

(d) If you continue this routine for all ten weeks of the quarter, what is the expected number of times that you will order tacos at the Court of Sciences?

(e) If you continue this routine for all ten weeks of the quarter, what is the probability that you will order tacos at the Court of Sciences exactly twice?

Solution

Let A_i be the event that I go to Ackerman on week i and T_i be the event that I order tacos on week i. We are given that:

$$\begin{aligned} P(A_i) &= .6 \\ P(T_i|A_i) &= .2 \\ P(T_i|A_i^c) &= .3 \end{aligned}$$

(a)

$$\begin{aligned} P(T_i) &= P(T_i|A_i)\,P(A_i) + P(T_i|A_i^c)\,P(A_i^c) && \text{By law of total probability} \\ &= .2 \times .6 + .3 \times .4 \\ &= .24 \end{aligned}$$

(b)

$$\begin{aligned}
P(A_i|T_i^c) &= \frac{P(T_i^c|A_i)\,P(A_i)}{P(T_i^c)} && \text{By Bayes rule} \\
&= \frac{P(T_i^c|A_i)\,P(A_i)}{1 - P(T_i)} \\
&= \frac{.8 \times .6}{1 - .24} && \text{From part (a)} \\
&= \frac{48}{76} \\
&= \frac{12}{19}
\end{aligned}$$

(c) Let X be a random variable that denotes the number of times I will go to Ackerman in the 10 week period. It is clear that X is a binomial random variable with parameters $n = 10$ and $p = .6$. The expected value of a binomial random variable with parameters n and p is simply np, so we have:

$$\begin{aligned}
E(X) &= np \\
&= 10 \times .6 \\
&= 6
\end{aligned}$$

(d) On a given week the probability that I will go to to the Court of Sciences and have tacos is:

$$\begin{aligned}
P(A_i^c \cap T_i) &= P(T_i|A_i^c)\,P(A_i^c) && \text{By the multiplication rule} \\
&= .3 \times .4 \\
&= .12
\end{aligned}$$

Let Y be a random variable the denotes the number of times I will go to the Court of Sciences and have tacos in the 10 week period. This is also a binomial random variable but with parameters $n = 10$ and $p = .12$ as calculated above. So:

$$\begin{aligned}
E(Y) &= np \\
&= 10 \times .12 \\
&= 1.2
\end{aligned}$$

(e) In part (d) I showed that Y is a binomial random variable with parameters $n = 10$ and $p = .12$. The PMF of Y is therefore given by:

$$P(Y = y) = \binom{n}{y} (1-p)^{n-y}\, p^y$$

So we have:

$$\begin{aligned}
P(Y = 2) &= \binom{10}{2} (1 - .12)^8\, .12^2 \\
&= \frac{10!}{8!2!} .88^8 .12^2 \\
&= .233
\end{aligned}$$

Problem 2 (20 points)

A device used by the Lottery Commission can generate any number between 2 and 30, inclusive. The numbers are not necessarily equiprobable and the probability of any individual number is a closely-guarded secret. The following probabilities are known, however:

- A is the event that the number is prime; $P(A) = 0.4$
- B is the event that the number is less than 15; $P(B) = 0.5$
- C is the event that the number is a prime less than 15; $P(C) = 0.3$

You might find it useful to draw Venn diagrams for intuition.

(a) Are events A and B independent? Give a convincing justication (e.g., proof, counterexample, theorem, etc.) for your answer.

(b) Event D is the event that the number generated is a prime greater than or equal to 15. Express event D in terms of the other events that have been defined and calculate $P(D)$.

(c) Event E is the event that the number generated is either a prime or less than 15 or both. Express E in terms of the other events that have been defined and calculate $P(E)$.

(d) If the number generated is greater than or equal to 15, what is the probability that it is not prime?

Solution

We observe that $C = A \cap B$.

(a)

$$\begin{aligned} P(A \cap B) &= P(C) && \text{Since } C = A \cap B \\ &= .3 \\ P(A)\,P(B) &= .4 \times .5 \\ &= .2 \end{aligned}$$

So $P(A \cap B) \neq P(A)\,P(B)$ which means that A and B are **not** independent.

(b) $D = A \cap B^c$, therefore

$$\begin{aligned} P(D) &= P(A \cap B^c) \\ &= P(A - B \cap A) \\ &= P(A - C) && \text{Since } C = A \cap B \\ &= P(A) - P(C) && \text{Since } P(A) = P(A - C) + P(C) \\ &= .4 - .3 \\ &= .1 \end{aligned}$$

(c) $E = A \cup B$. So we have

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) && \text{Inclusion exclusion rule} \\ &= P(A) + P(B) - P(C) && \text{Since } C = A \cap B \\ &= .4 + .5 - .3 \\ &= .6 \end{aligned}$$

(d)

$$
\begin{aligned}
P\left(A^c|B^c\right) &= \frac{P\left(A^c\cap B^c\right)}{P\left(B^c\right)} \\
&= \frac{P\left(A^c\cap B^c\right)}{1-P\left(B\right)} \\
&= \frac{P\left(\left(A\cup B\right)^c\right)}{1-P\left(B\right)} \\
&= \frac{1-P\left(A\cup B\right)}{1-P\left(B\right)} \\
&= \frac{1-P\left(E\right)}{1-P\left(B\right)} && \text{From part (c)} \\
&= \frac{1-.6}{1-.5} \\
&= .8
\end{aligned}
$$

Problem 3 (15 points)

Suppose that a system consists of 20 disks organized into 4 groups of 5 disks each. Data is replicated within each group such that data becomes inaccessible if 2 or more disks fail within the same group. (Another way of saying the same thing is that all data remains accessible if no group has more than one failed disk.) All disks are equally likely to fail.

Find the probability that some data is inaccessible given that 3 disks have failed.

Solution

We will compute this as 1 minus the probability that all the data is accessible. Let A be the event that all the data is accessible. Since all the outcomes are equally likely, we can use a counting argument to calculate $P\left(A\right)$.

The number of outcomes in A can be calculated as follows. If all the data is accessible, the three failed disks must be in different groups. There are $\binom{4}{3}$ different ways to pick the 3 groups that have the failed disk. For each of those groups there are 5 ways to pick a which disk failed. Since there are 3 groups with failed disks, this means that there are 5^3 ways to pick which disks have failed. So:

$$|A| = \binom{4}{3}5^3$$

Since there are 20 disks, there are $\binom{20}{3}$ ways for 3 disks to fail so,

$$|\Omega| = \binom{20}{3}$$

Therefore:

$$\begin{aligned} P(A) &= \frac{|A|}{|\Omega|} \\ &= \frac{\binom{4}{3}5^3}{\binom{20}{3}} \\ &= \frac{\frac{4!}{3!}5^3}{\frac{20!}{3!17!}} \\ &= \frac{4!17!}{20!}5^3 \end{aligned}$$

Probability that some data is inaccessible $= P(A^c) = 1 - P(A) = 1 - \frac{4!17!}{20!}5^3$.

Problem 4 (10 points)

30 of the 90 students in an intro programming class are sloppy programmers. On expectation, a program written by one of these students has 10 bugs. The rest of the class are more diligent, with a mean of only 3 bugs per program. An instructor chooses a random student in the class and evaluates a program she has written. Let X be a random variable denoting the number of bugs in this program. What is $E[X]$?

Solution

Let A be the event that the randomly chosen student is a sloppy programmer. We are given that $E[X|A] = 10$, $E[X|A^c] = 3$, $P(A) = \frac{30}{90}$, and $P(A^c) = \frac{60}{90}$.

$$\begin{aligned} E[X] &= E[X|A]\,P(A) + E[X|A^c]\,P(A^c) && \text{By total expectation} \\ &= 10 \times \frac{30}{90} + 3 \times \frac{60}{90} \\ &= \frac{16}{3} \\ &= 5.33 \end{aligned}$$

Problem 5 (20 points)

There are five identical computers in a lab. Two of them have old software, and both of these have a probability p_{old} of crashing each time a student uses them. The rest have new software, and have a smaller probability p_{new} of crashing when they are used.

(a) A bored lab supervisor chooses a computer at random and observes n students using this computer during his shift. The computer crashes for k of these users. Given these observations, what is the probability that the randomly chosen computer has the old software? (You can leave your answer in terms of p_{old}, p_{new}, k, and n.)

(b) One day the lab supervisor is extra bored and decides to conduct the following experiment. He first chooses one of the five computers at random. Then he watches students using that computer until it fails. If he watches at least ten students without seeing it fail, he'll give up. Let X be a random variable denoting the number of students he observes during this experiment. What is the PMF of X?

Solution

(a) Let A be the event that the randomly chosen computer has the old software.

Let K be the random variable specifying the number students for whom the computer crashes.

We are given that $P(A) = \frac{2}{5}$. We observe that $P(K|A)$ has a binomial PMF with parameters n and p_{old} and $P(K|A^c)$ has a binomial PMF with parameters n and p_{new}.

$$
\begin{aligned}
P(A|K=k) &= \frac{P(K=k|A)\,P(A)}{P(K=k)} && \text{By Bayes rule} \\
&= \frac{P(K=k|A)\,P(A)}{P(K=k|A)\,P(A) + P(K=k|A^c)\,P(A^c)} && \text{By law of total probability} \\
&= \frac{\binom{n}{k}(1-p_{old})^{n-k}\,p_{old}^k \times \frac{2}{5}}{\binom{n}{k}(1-p_{old})^{n-k}\,p_{old}^k \times \frac{2}{5} + \binom{n}{k}(1-p_{new})^{n-k}\,p_{new}^k \times \frac{3}{5}} \\
&= \frac{\frac{2}{5}(1-p_{old})^{n-k}\,p_{old}^k}{\frac{2}{5}(1-p_{old})^{n-k}\,p_{old}^k + \frac{3}{5}(1-p_{new})^{n-k}\,p_{new}^k}
\end{aligned}
$$

(b) Let A be the event that the randomly chosen computer has the old software.

We consider two cases separately.

Case 1: $1 \le x \le 9$

When $1 \le x \le 9$, $X = x$ is the event that the computer did not crash on the first $x-1$ students and then crashed on the x^{th} student. It follows that

$$P(X=x|A) = (1-p_{old})^{x-1}\,p_{old}$$

and

$$P(X=x|A^c) = (1-p_{new})^{x-1}\,p_{new}$$

So we have:

$$
\begin{aligned}
P(X=x) &= P(X=x|A)\,P(A) + P(X=x|A^c)\,P(A^c) && \text{By law of total probability} \\
&= (1-p_{old})^{x-1}\,p_{old} \times \frac{2}{5} + (1-p_{new})^{x-1}\,p_{new} \times \frac{3}{5}
\end{aligned}
$$

Case 2: $x = 10$

$X = 10$ is the event that the computer did not crash on the first 9 students, so

$$P(X=10|A) = (1-p_{old})^9$$

and

$$P(X=10|A) = (1-p_{new})^9$$

So we have:

$$
\begin{aligned}
P(X=10) &= P(X=10|A)\,P(A) + P(X=10|A^c)\,P(A^c) && \text{By law of total probability} \\
&= (1-p_{old})^9 \times \frac{2}{5} + (1-p_{new})^9 \times \frac{3}{5}
\end{aligned}
$$

Problem 6 (10 points)

Suppose that X and Y are random variables with the same variance, i.e., $Var(X) = Var(Y)$. Show that $X+Y$ and $X-Y$ are uncorrelated, i.e., that $Cov(X+Y, X-Y) = 0$.

Solution

$$\begin{aligned}
Cov\,(X+Y, X-Y) &= E\left[(X+Y)(X-Y)\right] - E\left[X+Y\right]E\left[X-Y\right] \\
&= E\left[(X+Y)(X-Y)\right] - \left(E\left[X\right] + E\left[Y\right]\right)\left(E\left[X\right] - E\left[Y\right]\right) \\
&= E\left[\left(X^2 - Y^2\right)\right] - \left(E\left[X\right]^2 - E\left[Y\right]^2\right) \\
&= \left(E\left[X^2\right] - E\left[X\right]^2\right) - \left(E\left[Y^2\right] - E\left[Y\right]^2\right) \\
&= Var\,(X) - Var\,(Y) \\
&= 0 \qquad \text{since } Var\,(X) = Var\,(Y)
\end{aligned}$$