Master of Management Analytics

MMA 863

Intro to Analytic Modelling

Dr. Keith Rogers

MMA 863 Assignment 1

May 14th @ 11:59PM

Team Stirling

Student Name	**Student Number**
Janice Tian	
Mike Hetherington	
Nazia Khan	
Paul Dong	
Richard Dam	
Ricky Ye	
Sally Cheung	
Shashir Kumar	

Assignment One

Question 1.

1.Kingston is becoming increasingly known for the potholes in its streets. It seems to me that driving along you encounter on average about one every 50 m of driving, though they are randomly distributed. Most of the potholes are small, but about 10% of them are large enough to be a bit dangerous. Those dangerous ones will flatten a tire about 0.1% of the time (i.e. if hit 'just right', which itself is, of course, random.)

a. I live 1 km from the Kingston campus. If I drive to school and back home, what is the probability that I will encounter more than 3 dangerously large potholes? Be sure to document your relevant assumptions.

Assumptions:

- The number of events that occur in any interval is independent of the number of events that occur in any other interval.
- The probability of an event in an interval is the same for all equal-sized intervals.
- The probability of an event is proportional to the size of the interval.
- The probability of more than one event in an interval approaches 0 as the interval becomes smaller.

Using R:

```
driving_distance <- 1e3*2
rate_large_pothole <- driving_distance / 50 * 0.1
```

the random variable can be described by:

$$X \sim Poisson(4)$$

the probability to encounter more than 3 dangerously large potholes is $P(X > 3)$

```
1-sum(dpois(0:3,lambda = rate_large_pothole))

## [1] 0.5665299
```

Alternatively, using the Binomial Distribution Solution:

Total distance covered = 2km = 2000m
Average no. of potholes = 2000/50 = 40
Large potholes (dangerous ones) = 10% of total potholes = 4

Here, probability of large pothole = 4/40 = 0.1
Probability to encounter more than 3 potholes = P(X>3) = 1-P(X<=3)

Probability to encounter more than 3 potholes
= 1-BINOM.DIST(3,40,0.1,TRUE)
= **0.576869347**

b. Suppose I drive to school and back home 200 times in a year. Develop a model to estimate the probability that I will flatten a tire due to a pothole. Be sure to list any necessary and reasonable assumptions, if any.

Using Poisson Solution:

Assumptions:

- The number of events that occur in any interval is independent of the number of events that occur in any other interval.
- The probability of an event in an interval is the same for all equal-sized intervals.
- The probability of an event is proportional to the size of the interval.
- The probability of more than one event in an interval approaches 0 as the interval becomes smaller.

```
annual_drive <- driving_distance*200
flat_tire_rate <- annual_drive/50*0.1*0.001
```

the random variable can be described by:

$$X \sim Poisson(0.8)$$

The probability to encounter a flat tire in any year is $P(X \geq 1)$

```
1-sum(dpois(0,lambda = flat_tire_rate))
```

[1] 0.550671

Using Binomial Solution:

Total distance covered = 2km * 200 = 2000m * 200 = 400000m
Average no. of potholes encountered= 400000/50 = 8000
Large potholes (dangerous ones) = 10% of total potholes = 800
Probability of tire getting flattened = 0.1% of the time = Tire flattening potholes = 0.1% of 800
= 0.8
Probability of flattening a tire = P(X>=1) = 1- P(X=0)
Probability of flattening a tire = 1-BINOM.DIST(0,800,0.001,TRUE) **= 0.550851**

Question 2.

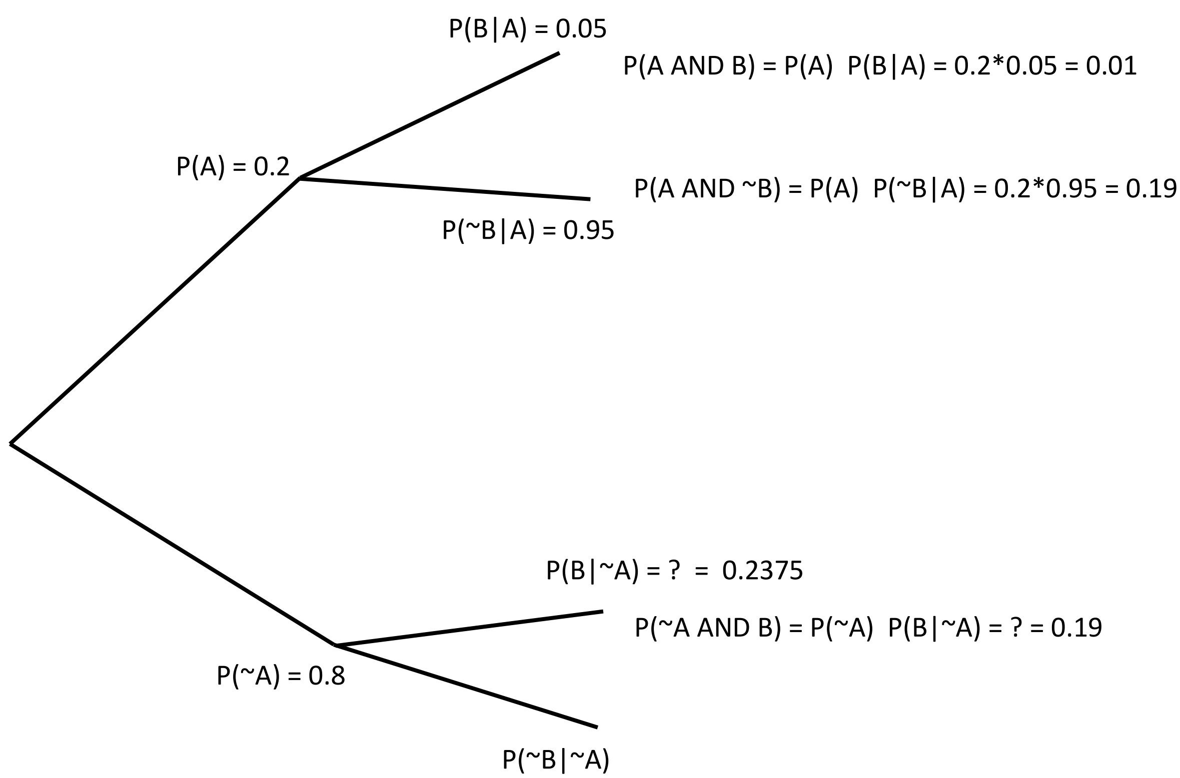
Where I live it rains about 20% of days, but it seldom rains two days in a row – if it rains on one day, there is only a 5% chance it will rain on the next. It is Friday morning and it looks like it will rain today, in fact, I'd guess there is an 80% chance that it will rain. The rain should not really bother me as I am going to fly to Zurich for two days, so I will not actually see it rain.

a. What is the probability it will rain tomorrow?
b. What is the probability it will rain both today and tomorrow?
c. Suppose I come back and my neighbor tells me it rained on Saturday, what is the probability it rained on Friday?

On an average day scenario (Where I live it rains about 20% of days, but it seldom rains two days in a row – if it rains on one day, there is only a 5% chance it will rain on the next):

Event A = Rain on any given day
Event B = Rain on the 2nd dayWith the Probability Tree:

Note relevant variables:
P(A) = P(B) = 0.2 P(~A) = P(~B) = 0.8
P(B|A) = 0.05 P(~B|A) = 0.95



P(B) = P(A AND B) + P(~A AND B)
0.2 = 0.01 + P(~A AND B)
P(B AND ~A) = 0.2 – 0.01 = 0.19

P(~A AND B) = P(~A) P(B|~A)
0.19 = 0.8 P(B|~A)
P(B|~A) = 0.2375

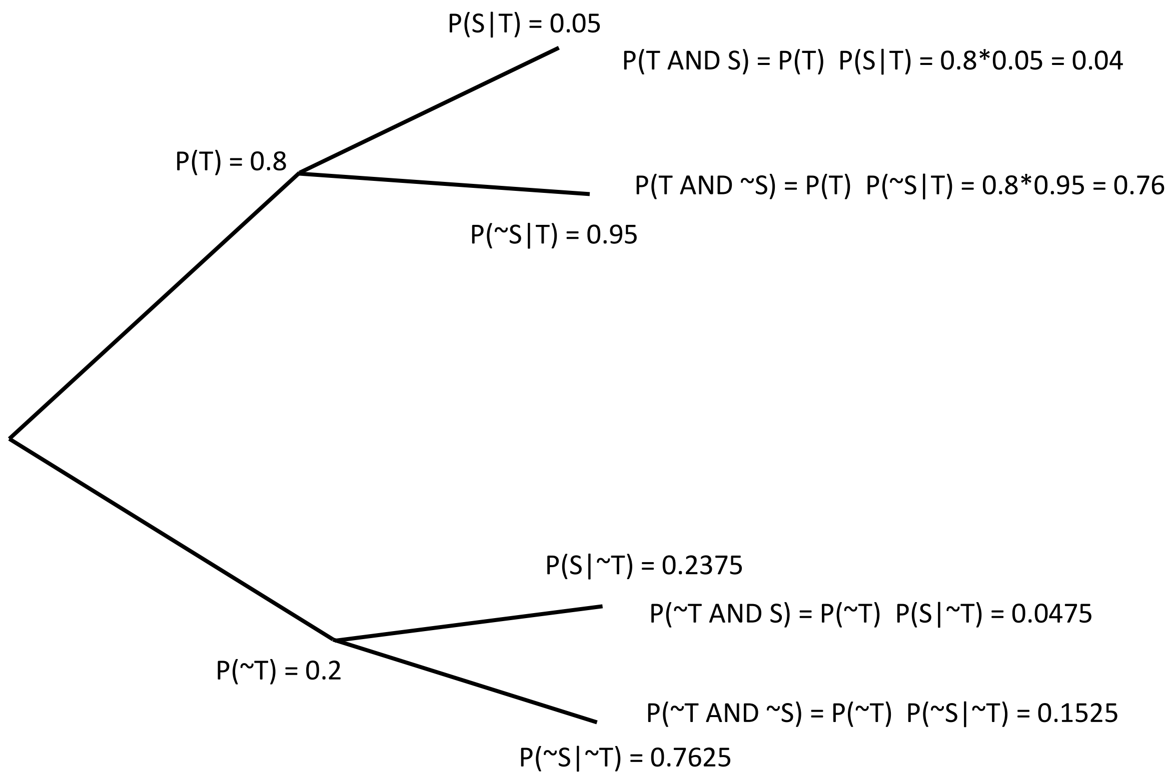
The probability of raining on the 2nd day given the first day did not rain is 0.2375, then apply it to the Friday (80%) scenario.

Friday morning scenario (with new information):

Event T = Rain on Friday (today)
Event S = Rain on Saturday (tomorrow)

Note relevant variables:

P(T) = 0.8   P(~T) = 0.2
P(B|~A) = P(S|~T) = 0.2375
Probability of raining on Sat is conditioned on the **outcome** of the previous day. The probability of raining on Saturday given Friday has rained P(S|~T) remains the same as an average date given the previous day has rained P(B|~A).



Using the above Probability Tree:

a. The probability it is raining tomorrow is P(S) = P(T AND S) + P(~T AND S) = 0.04 + 0.0475 = **0.0875**

b. The probability it will rain both today and tomorrow is P(T AND S) = **0.04**

c. The probability it rained on Friday given Saturday has rained is
P(T|S) **=** P(T AND S) / P(S) = 0.04 / 0.0875 = **0.4571**

Question 3

Winter Wonderland: The driveway at my cottage is literally 1 km long. It is quite a peaceful drive in the summer but in the winter it can be a treacherous frozen nightmare. Last week, after the big ice storm, I drove up to the lake and realized I couldn't drive down the driveway because it was too icy for the car. It was even dangerous for walking and I anticipated a 5% chance I'd slip and fall during the first 100 m alone. Nevertheless, it was important to go in because I needed to know how much propane we had left for heating the place. (You may be interested in knowing that the daily consumption of propane is distributed with a mean of 6.5 L with a standard deviation of 2 L.) Clearly I was going to have to walk all the way in and then all the way back out later that afternoon – I was not happy!

d. SOLVE THIS TWO DISTINCT WAYS AND COMPARE ANSWERS: So I started walking in, only to realize, half way to the cottage, that I'd forgotten my keys and had to go back to the car to get them. Given that I 'got lucky' and had not fallen by the time I realized I had forgotten my keys, what is the probability I would finish the day without slipping?

   Binomial Solution:

   1st Method: BINOM.DIST(0,25,0.05,TRUE) = 0.277389573 or
   2nd Method: BINOM.DIST(25,25,0.95,FALSE) = 0.27739 = **27.7%**

   Using the binomial distribution, the probability of finishing the day without slipping is approximately **27.7%**

   Logic:
   One trial = 100m of walking
   25 trials to complete the days walking. (500 back to car, 1 km to cottage, 1 km back out)
   5% chance of falling = 95% chance of not falling
   Number of successes = 25
   Number of trials = 25
   Probability of success = .95
   Probability mass function (FALSE) used to solve for exactly 25 successes.

   Poisson Solution:

1st Method: POISSON.DIST.DIST(0,1.25,TRUE) = 0.286504797 or
2nd Method: POISSON.DIST(0,1.25,FALSE) = 0.286505 = **28.7%**

Using the Poisson distribution, the probability of finishing the day without slipping is approximately **28.7%.**

Logic:
(.05) (falls) per (100m)
=(1.25)(falls) per (2.5km)
x = 0
mean = 1.25
Probability mass function (FALSE) used to solve for exactly 0 occurrences.

e. What assumptions do I need to make in the questions above? How reasonable are they / which ones do you suspect are violated?

Binomial assumptions:

- Specific number of trials. This condition is met.

- Identical trials. This assumption is violated. It is very likely that some portions of the driveway are relatively easy to walk (increasing the chance of success), while other portions are more challenging (decreasing the chance of success).

- Independent trials. This assumption is violated. Dr Rogers will develop skill at not falling on the treacherous driveway with experience, making the chance of success increase on later trials. On the other hand, Dr Rogers may become overconfident and/or fatigued as the journey progresses, decreasing the chance of success on later trials.

- Two outcomes. This is a reasonable assumption, as only one of the two outcomes will occur. However, the assumption does not account for the chance of multiple falls in one trial. As a result, this model would be less accurate in predicting the probability of more than one fall during the day.

- Maximum number of occurrences is known. This assumption is violated, as Dr Rogers could fall an unknown amount of times in each trial.

Poisson assumptions:

- The Poisson distribution deals with situations where things happen randomly at a given rate over time or space. The rules for the Poisson are that events occur independently over time at a constant random rate and only one event can occur at a time
- Independent events. Violated as per binomial assumptions above.
- Constant random rate of time. Violated, as per the identical trial reasoning above.
- Only one event can occur at a time.

Question 4:

As the owner of a growing online company, I have considerable intellectual property tied up in various bits of data. My technology team has proposed a system involving 3 identical hard drives. If all three drives fail at the same time, my company goes offline. That would be bad.This may sound like a lot of technology, but we feel we need this kind of redundancy to ensure that our system is available 24 hrs. per day for 365 days a year.

Drive failures are rare with new equipment, but we bought some refurbished equipment so each drive has a 5% chance of failing per day.

If it fails, it can be repaired when the system goes into maintenance mode just before midnight and be ready for the next day.

a.Using language your manager is likely to understand, explain why this could be modelled as a binomial, develop and use that model to determine the probability of having more than 50 drive failures in a year?

b.What is the probability that I will go offline at least once in a year?

c.Would it be better to have a system with 2 brand new identical drives, each of which only has a 2% chance of failure?"

A).

Binomial Solution:

Using binomial model best fits in the cases, where we have only 2 possible outcome, “Success” or “Failure”. In this experiment, the hard drive can either fail or not fail on any given day. Also, each occurrence of “Success” and “Failure” are identical and independent from each other, and each trial has the same probability of “Success” and “Failure” as any other trials.

As we have a limited number of trials(days), a binomial distribution method will give us a better approximation.

N = 365*3 (as we have 3 hard drives, the total number of trials will be total days in year * 3)

X > 50

P = 0.05

P(X>50) = 1 – P(X<=50)

P(X<=50) = (BINOM.DIST(50,365*3,0.05,TRUE))
P(X>50) = 1 – P(X<=50) = **0.71755**

The probability of more than 50 drive failures in the year is **71.76%**

B).
Probability that all the 3 hard drives will go down on a single day = (0.05)^3

N = 365
X>= 1
P = 0.000125

P(X>=1) = 1 – P(X=0)

P(X>=1) = 1 - (BINOM.DIST(0,365,0.0000125,TRUE)) = **0.04460**

The probability of going offline once a year **is 4.46%**

C).

Probability that both the new drives will go down on a single day is = (0.02)^2

N = 365
X>= 1
P = 0.0004

P(X>=1) = 1 – P(X=0)

P(X>=1) = 1 - (BINOM.DIST(0,365,0.0004,TRUE)) = **0.135867537**

The probability of going offline once a year is with having a new system with 2 brand new hard drives is **13.58%**

Having a system with two brand new hard drives will not be a good idea as our probability of system being down for at least a day increases **to 13.58% from 4.46%**

<u>Assumption:</u>

For questions asked in b) and c), we are making an assumption that the only reason that could bring the system down is the hard drives failure and no other event can contribute to system outage.