Mini-Project 1 – Part 2

ECE 471

Fall 2024

Zach Larson (zlarson2), Matthew Paul (mjpaul3), Maximo Rojas Pelliccia (mgr9)

Note: small decimal numbers were rounded to 3 significant figures or the best representation.

For complete decimal numbers, refer to the Jupyter notebook.

Task 3

Q1: Based on the above assumptions, answer the following questions on basic probability.

a) The assumption of at most one disengagement per mile allows us to treat the occurrence of a disengagement in a mile as a random variable with a Bernoulli distribution

b) Based on the above assumptions, calculate the probability of disengagement per mile on a cloudy day.

Probability of Disengagement Per Mile: **0.00203**

Probability of Disengagement Per Mile on a Cloudy Day: **0.00590**

c) Based on the above assumptions, calculate the probability of disengagement per mile on a clear day.

Probability of Disengagement Per Mile on a Clear Day: **0.000520**

Task 3

Q1: Based on the above assumptions, answer the following questions on basic probability

d) Similarly, calculate the probability of an automatic disengagement per mile on a cloudy day, and the probability of an automatic disengagement per mile on a clear day.

Probability of Automatic Disengagement Per Mile: **0.000976**

Probability of Automatic Disengagement Per Mile on Cloudy Day: **0.00281**

Probability of Automatic Disengagement Per Mile on Cloudy Day: **0.000264**

e) How likely is it that there are 100 or more disengagements in 10,000 miles under cloudy conditions? (**hint**: use Central Limit Theorem)

Probability of 100 or more disengagements in 10000 miles = **8.85e-08**

Task 3

Q2: Assuming that the disengagement per mile is a random variable with the distribution you answered in Task 3.1.a, and the weather condition is *cloudy* .

a) What is the distribution of “the number of miles until the next disengagement”? Explain your reasoning. Calculate and state the values of the parameters of the distribution

This would follow a geometric distribution, as this distribution captures the waiting time until the first occurrence of an event -- in this case, a disengagement. Our 'waiting time' in this scenario is miles driven. After each occurrence, we 'reset' with a new random variable, so that our random variable always represents the time before the first disengagement in its lifetime.

$X \sim Geom\ (p = 0.0059)$

μ = $\frac{1}{p}$ = 169.4915

σ = $\sqrt{\frac{1-p}{p^2}}$= 168.9908

Task 3

b) What is the distribution of “the number of disengagements in 10,000 miles”? (hint: this is equivalent to drawing n=10,000 independent trials from the distribution of disengagement per mile you calculated from Task 3.1.a). Calculate and state the values of the parameters of the distribution.

This would follow a binomial distribution, as it is simply a yes/no decision, but with $n = 10000$ trials.

$X \sim Binom\ (n = 10000, p = 0.0059)$

c) Notice that the number of disengagements “$\boldsymbol{n}$" in Task 3.2.b is large while the probability of disengagement per mile "$\boldsymbol{p}$" is very small, what distribution does your answer in Task 3.2.b approximate? Calculate and state the values of the parameters of the distribution.

This would approximate a Poisson distribution where $\lambda = np = 59$

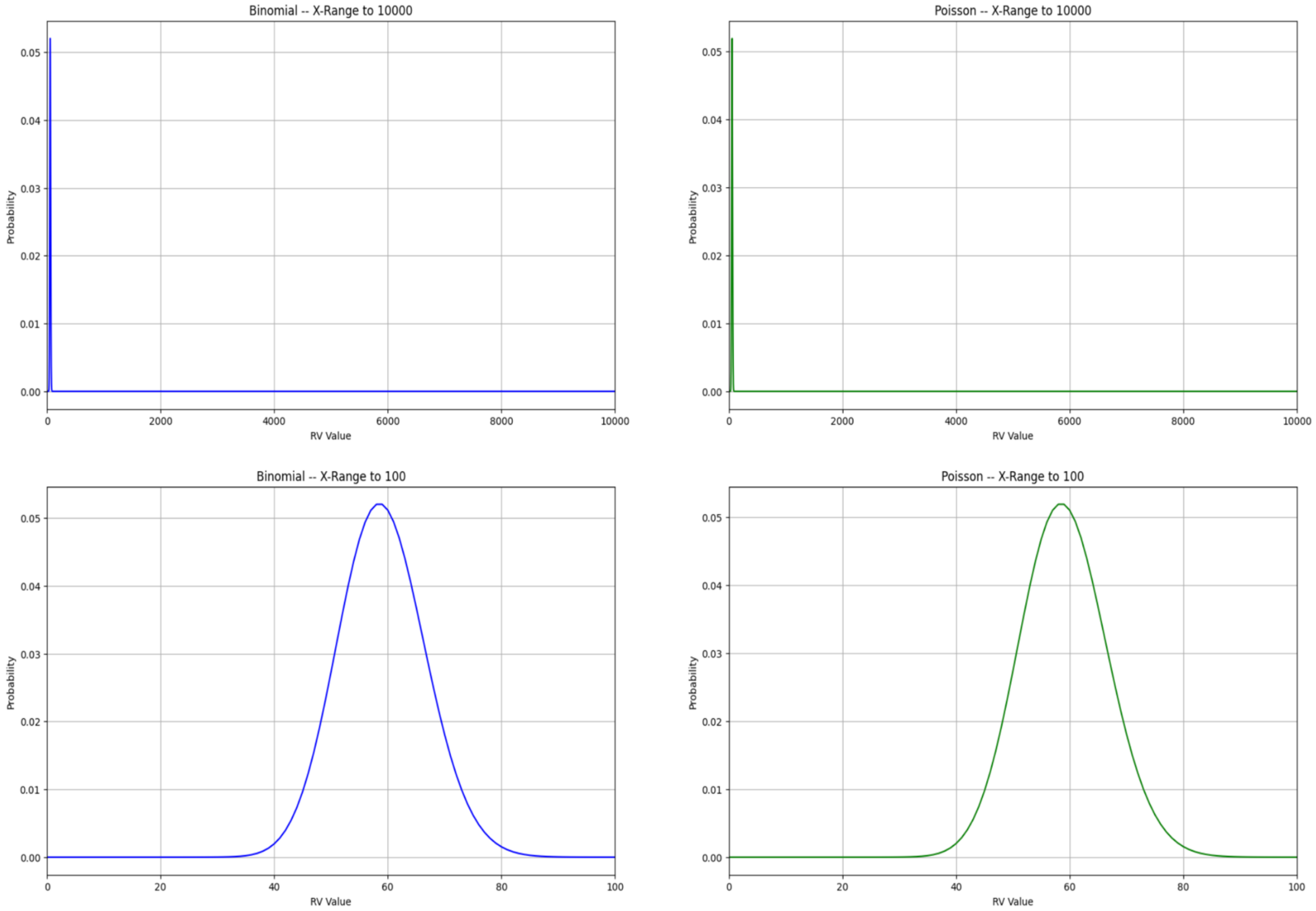
Task 3

Q2: Assuming that the disengagement per mile is a random variable with the distribution you answered in Task 3.1.a, and the weather condition is *cloudy* .

d) Plot the probability mass function (PMF) of the distribution in Task 3.2.b and Task 3.2.c for:

1. x-axis ranging between 0 and 10000.

2. The x-axis ranging between 0 and 100



Task 3

Describe the plots and discuss your findings

The CDF plots for the binomial distribution, and the approximated Poisson distribution, are essentially identical. The mean of both is 59, and the spread, i.e. variance/bell curvature are similar as well. This goes to show that when modelling for a large number of miles, the binomial and Poisson distributions both adequately represent the distribution for the number of disengagements.

e) Solve Task 3.1.e by using the cumulative distribution function (CDF) of the distribution you computed in Task 3.2.c and compare the results. Discuss your findings

Self-Calculated Value: **8.85e-08**

Distribution-Calculated Value: **3.103e-07**

These findings make sense, as the expectation for disengagements in 10000 miles driven, considering both the Poisson and binomial distributions, was 59. One hundred disengagements is well outside of the range of reasonable values.

Task 3

Q3: What's the conditional probability that the reaction time is:

a) Greater than 0.4s given that the weather was cloudy? Reaction time is measured only in cases where there was an **automatic disengagement**.

Probability of reaction time greater than 0.4 seconds, given cloudy weather: **0.612**

b) Greater than 0.7s given that the weather was clear? Reaction time is measured only in cases where there was an **automatic disengagement**.

Probability of reaction time greater than 0.7 seconds, given clear weather: **0.385**

Task 3

Q4: A study found that an **automatic AV disengagement** will result in an accident if the human driver is *slow* in reacting. Following reactions are considered slow: (i) a reaction time greater than 0.4s under cloudy conditions and, (ii) a reaction time greater than 0.7s under clear conditions. Find the probability of an accident per mile due to automatic AV disengagement and slow reaction.

P(accident) = P(slow reaction) * P(autodisengagement)
ASSUMING: disengagement + slow reaction ==> accident (always)

P(slow reaction) = P(slow reaction|cloudy)*p(cloudy) + P(slow react|clear)*P(clear)

Probability of an accident: **0.00335**

Task 3

Q5: You will investigate how to diagnose the cause of an AV disengagement based on new observations:

a) An AV had a disengagement with a reaction time greater than 0.4s on a cloudy day. What is the posterior probability that the root cause of the disengagement was “Software Froze”?

Posterior probability (Slow, Software Froze, Cloudy): **0.0535**

b) What is the posterior probability that the root cause of the disengagement was “Software Froze if the disengagement happened on a clear day with reaction time greater than 0.7s. Based on the probabilities calculated in Tasks 3.5.a and 3.5.b, discuss your findings.

Posterior probability (Slow, Software Froze, Clear): **0.243**

The posterior probability of a disengagement being caused by frozen software is over 4 times higher when the weather is clear and reaction time greater than 0.7 seconds, than when the weather is cloudy and reaction time only greater than 0.4 seconds.

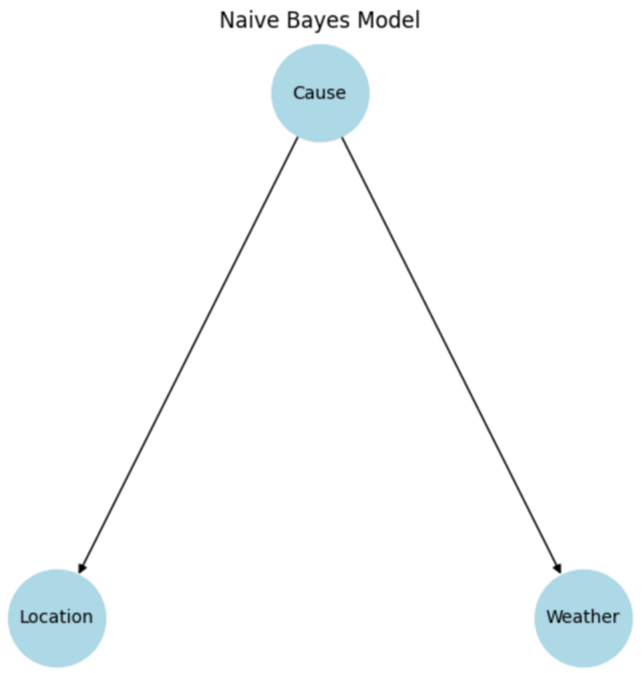
$0.2432 \,/\, 0.0535 \approx \mathbf{4.55}$ (multiplicative factor)

This makes sense intuitively, as a longer reaction time can indicate software issues, especially on a clear day where a longer lag time would not be expected, whereas in other conditions a longer time may be understandable.

Task 3

Q6: In this question, you will construct a Naive Bayes model to infer the root cause of disengagement scenarios of AVs. Naive Bayes assumes that the factors are class conditionally independent. We assume that both Location (urban-street or highway) and Weather (cloudy or clear) are factors related to the Cause (consider the Cause has 3 different values, “Software Froze”, “Hardware Fault” or “Other”), and Location and Weather are independent given the Cause. Answer the following questions

a) Draw a graph for the Naive Bayes model described in the question.



Task 3

b) Count the number of parameters needed to define the Naive Bayes model (including the prior and the conditional probability distributions

Arbitrary Joint Distribution:

$(3 * 2 * 2) - 1 = 11$ parameters

Naive Bayes Model:

P(C,L,W) = P(L|O)P(W|O)P(O)

P(L = urban-street | O = Software Froze), P(L = urban-street | O = Hardware Fault), P(L = urban-street | O = Other),

P(W = cloudy | O = Software Froze), P(W = cloudy | O = Hardware Fault), P(W = cloudy | O = Other),

P(O = Software Froze), P(O = Hardware Fault)

8 Parameters

Task 3

Q6 **(continued):**

c) Based on the number of parameters needed, derive, and show the conditional probability tables and prior probability from the given dataset to infer the Cause .

Cause (O)	P(O)
Software Froze	0.099391
Hardware Fault	0.101420
Other	0.799189

Location (L)	Cause (O)	P(L|O)
urban-street	Software Froze	1.00
urban-street	Hardware Fault	0.98
urban-street	Other	1.00
highway	Software Froze	0.00
highway	Hardware Fault	0.02
highway	Other	0.00

Weather (W)	Cause (O)	P(L|O)
cloudy	Software Froze	0.387755
cloudy	Hardware Fault	0.400000
cloudy	Other	0.908629
clear	Software Froze	0.612245
clear	Hardware Fault	0.600000
clear	Other	0.091371

d) According to the conditional probability tables you derived, what is the most probable root cause of disengagement given the Weather was **cloudy** and the Location was **urban-street**

Software Freeze: 0.0385
Hardware Fault: 0.0398
Other: **0.726**

Therefore **“Other”** root cause is most probable given Weather was cloudy and the Location was urban-street