ECS647U / ECS773P

Bayesian Decision and Risk Analysis (BDRA)

Semester B, 2021

Coursework 2 – “Too Good to be True”

Mariya Pavlova

Queen Mary University of London

MSc Artificial Intelligence

Student ID 170703132

Question 1

a) **Show a screen shot of your BN identifying the relevant variables to model this problem, where the Sanhedrin can be independent or dependent (biased). [15marks]**

In this assignment, we shall explore the problem of unanimous support being too good to be true due to a systemic failure. We shall prove it via developing a Bayesian Network (BN thereof) with different scenarios.

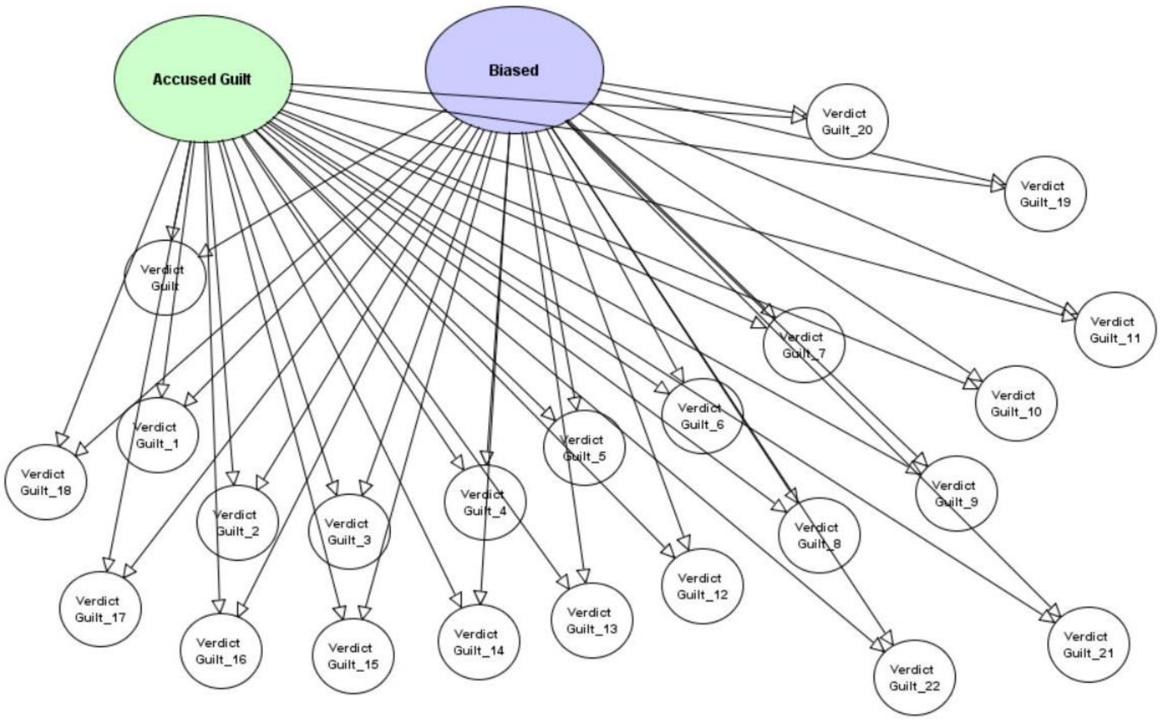


Fig.1 Bayesian network model representing the Sanhedrin situation

My model is a discrete BN with separate Bernoulli trial nodes for each judge. As we saw in Assignment 1, the Bernoulli distribution is a special case of the binomial distribution where n=1. Thus, Bernoulli nodes can be used to build our model as this approach will generate similar results.

As illustrated in Fig.1, the BN consists of the following standard nodes (not simulation ones):

- Accused Guilt node – summarises the Sanhedrin's judgement. This is a Boolean node.
- Biased node – the probability of the Sanhedrin being biased. This is also a Boolean node.
- Verdict Guilt nodes – there is a total of 23 nodes. They represent the verdict of each individual judge.

Fig.2 below shows the risk graphs after adding the prior probabilities provided in the assignment pdf. No observations were added.

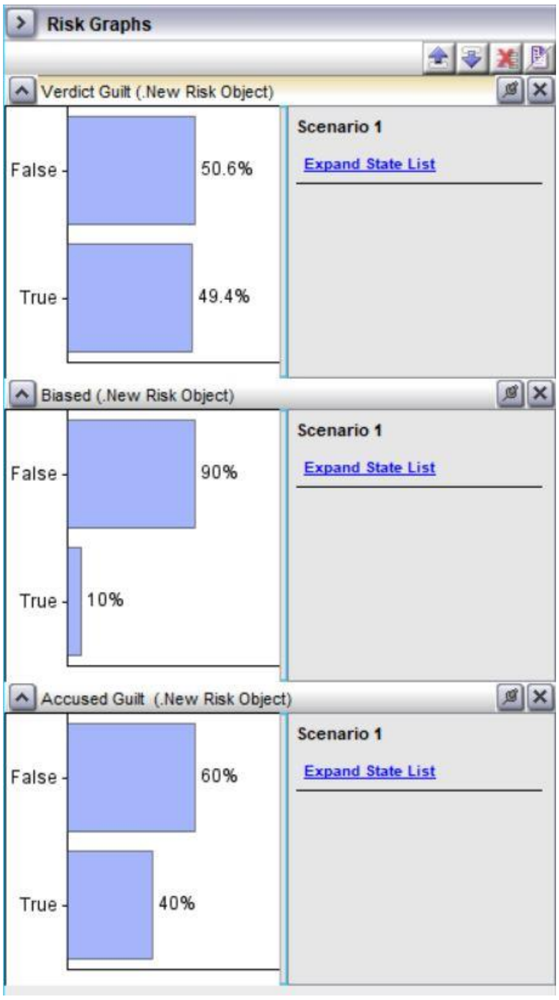
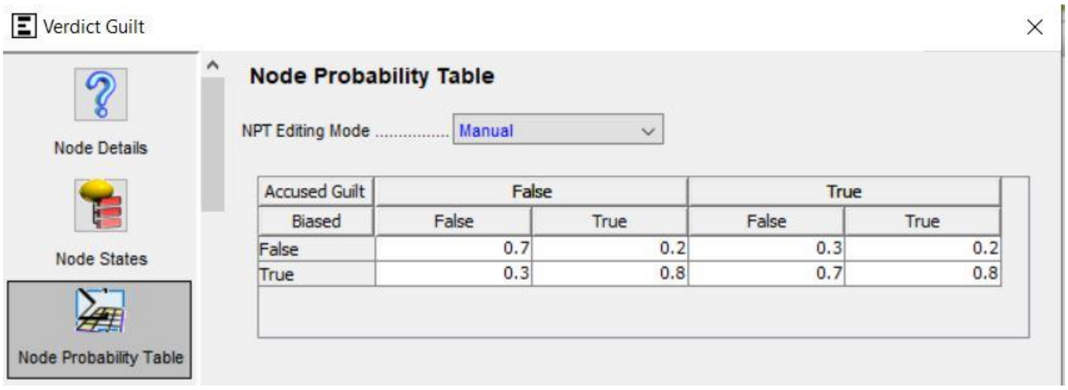


Fig.2 Risk graphs of each node after adding the prior probabilities

b) Show the conditional probability tables for all relevant variables. [15 marks]

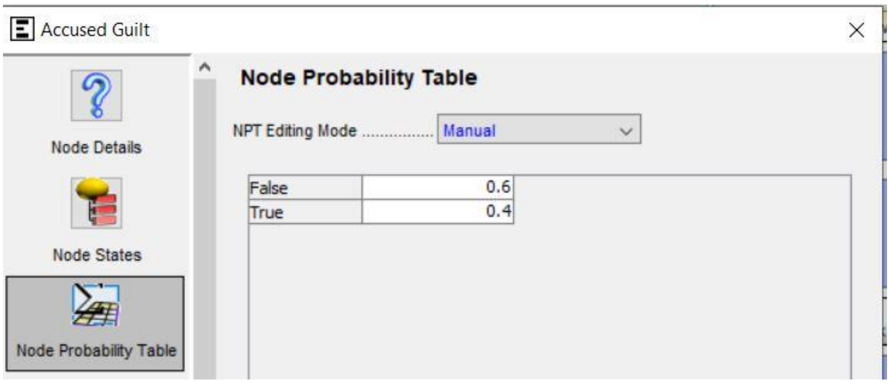
First, I shall display the conditional probability table of the Verdict Guilt node. Please note that the rest of Verdict Guilt nodes are identical to the first one because they were duplicated. Hence, they have the same probability table and parameters. Please see an illustration in Fig.2 below.



Accused Guilt	False		True	
Biased	False	True	False	True
False	0.7	0.2	0.3	0.2
True	0.3	0.8	0.7	0.8

Fig.2 Conditional probability table of the Verdict node (prior probability)

Fig.3 shows the conditional probability table of the Accused Guilt node.



False	0.6
True	0.4

Fig.3 Conditional probability table of the Accused Guilt node (prior probability)

Fig.4 displays the conditional probability table of the Biased node.

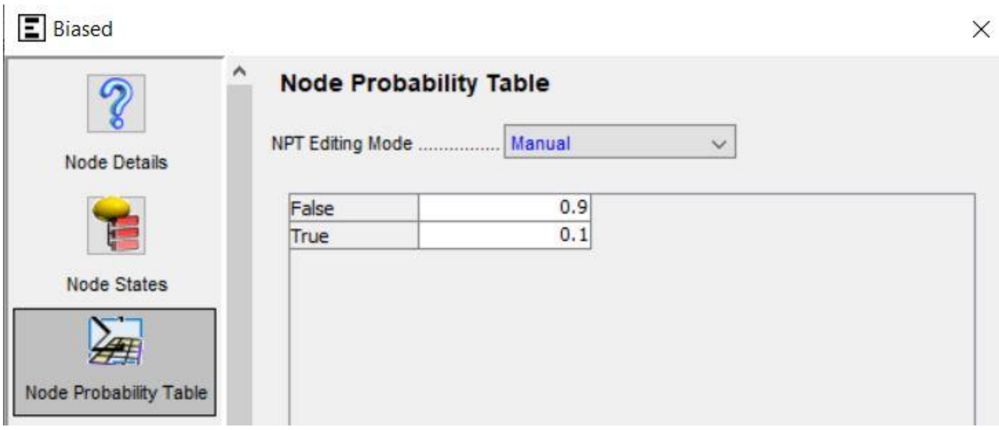


Fig.4 Conditional probability table of the Biased node (prior probability)

The probability table information was obtained in the following way.

- Distribution of guilt and bias when a judge verdict is true

AG

Biased

	No	Yes
Yes	0.8	0.8
No	0.3	0.7

Distributions

P(VG=True|AG, B)

We can derive that when the Sanhedrin is not biased, there is a division of opinions (0.3 vs 0.7) on whether the defendant is guilty or not. However, the probability of a guilty verdict increases when the judges are biased. This is illustrated by the fact that both distributions increase to 0.8.

- Accused Guilt and Biased: joint prior beliefs

The following equation describes this situation:

$P(AG|VG = True,B) = P(VG|AG = True, B)\ P(AG = True,B)\ /P(VG) = \sum_b P(VG|AG = True,B = b)P(AG = True, B = b)\ /\sum_{b,AG} P(VG|AG,B = b)P(AG,B = b)$

$P(VG|AG,B)$ follows from a binomial distribution where n=23 and x=number of judges who give a positive verdict (VG = True).

Question 2

<u>Show a screenshot of the 'Risk Table' in AgenaRisk showing the observations entered in all scenarios (Look in the AgenaRisk manual for information on the Risk Table). [10 marks]</u>

Seven scenarios were created as per the assignment requirements. Observations were added to each judge (Verdict Guilt nodes) for each scenario accordingly. Fig.5 below illustrates the updated BN, followed by Fig.6 which shows the risk table.

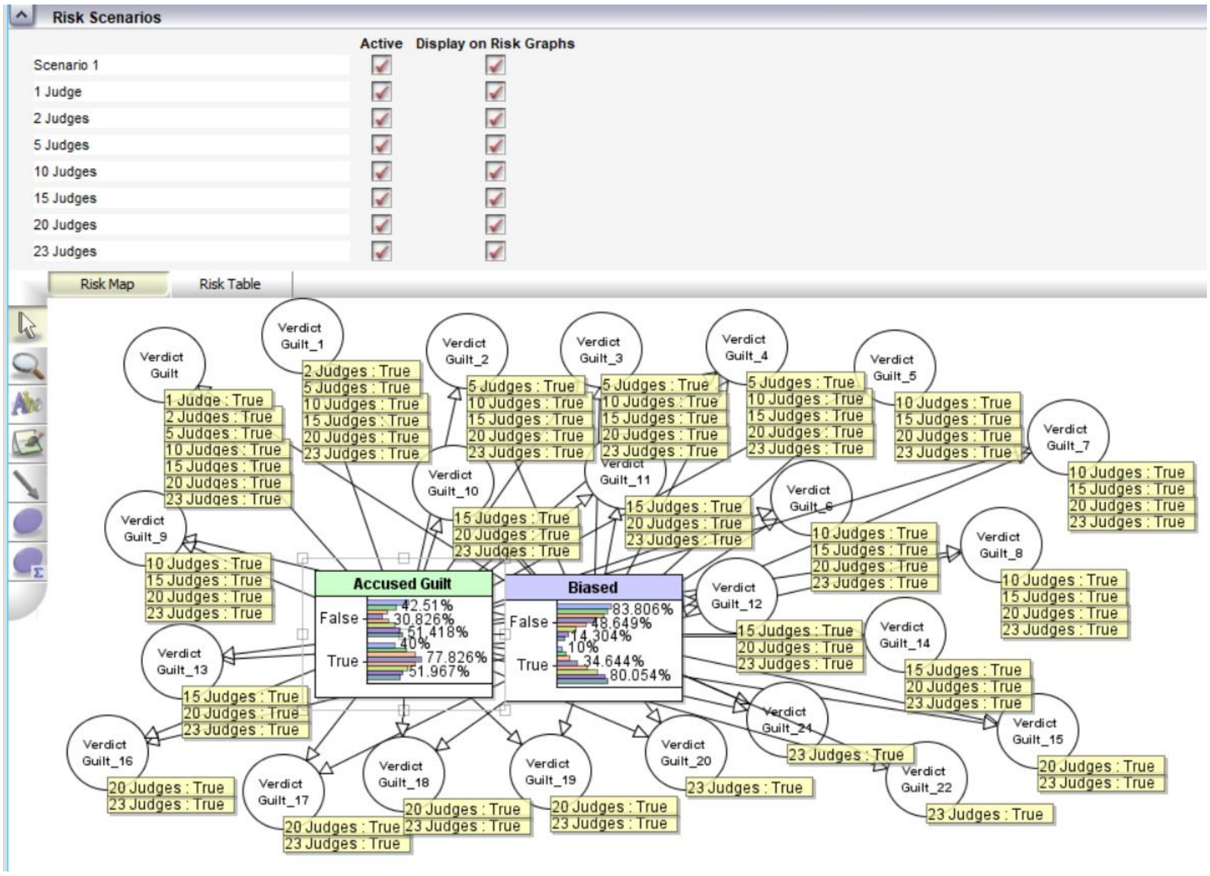


Fig.5 Updated BN model after adding scenario observations

Risk Scenarios

Risk Map | Risk Table

New Risk Object	Scenario 1	1 Judge	2 Judges	5 Judges	10 Judges	15 Judges	20 Judges	23 Judges
Accused Guilt	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer
Biased	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer
Verdict Guilt	No Answer	True	True	True	True	True	True	True
Verdict Guilt_1	No Answer	No Answer	True	True	True	True	True	True
Verdict Guilt_2	No Answer	No Answer	No Answer	True	True	True	True	True
Verdict Guilt_3	No Answer	No Answer	No Answer	True	True	True	True	True
Verdict Guilt_4	No Answer	No Answer	No Answer	True	True	True	True	True
Verdict Guilt_5	No Answer	No Answer	No Answer	No Answer	True	True	True	True
Verdict Guilt_6	No Answer	No Answer	No Answer	No Answer	True	True	True	True
Verdict Guilt_7	No Answer	No Answer	No Answer	No Answer	True	True	True	True
Verdict Guilt_8	No Answer	No Answer	No Answer	No Answer	True	True	True	True
Verdict Guilt_9	No Answer	No Answer	No Answer	No Answer	True	True	True	True
Verdict Guilt_10	No Answer	No Answer	No Answer	No Answer	No Answer	True	True	True
Verdict Guilt_11	No Answer	No Answer	No Answer	No Answer	No Answer	True	True	True
Verdict Guilt_12	No Answer	No Answer	No Answer	No Answer	No Answer	True	True	True
Verdict Guilt_13	No Answer	No Answer	No Answer	No Answer	No Answer	True	True	True
Verdict Guilt_14	No Answer	No Answer	No Answer	No Answer	No Answer	True	True	True
Verdict Guilt_15	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True	True
Verdict Guilt_16	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True	True
Verdict Guilt_17	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True	True
Verdict Guilt_18	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True	True
Verdict Guilt_19	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True	True
Verdict Guilt_20	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True
Verdict Guilt_21	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True
Verdict Guilt_22	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	No Answer	True

Fig.6 Risk table for all seven scenarios

a) **<u>Show screenshots of AgenaRisk showing the posterior marginal distributions for $P(Accused\ Guilt = True \mid scenario\ data)$ and $P(All\ Judges\ Baised = True \mid scenario\ data)$ for all scenarios. [20 marks]</u>**

$P(Accused\ Guilt = True \mid scenario\ data)$ is shown in Fig.7, and $P(All\ Judges\ Baised = True \mid scenario\ data)$ is shown in Fig.8. Data is for all seven scenarios.

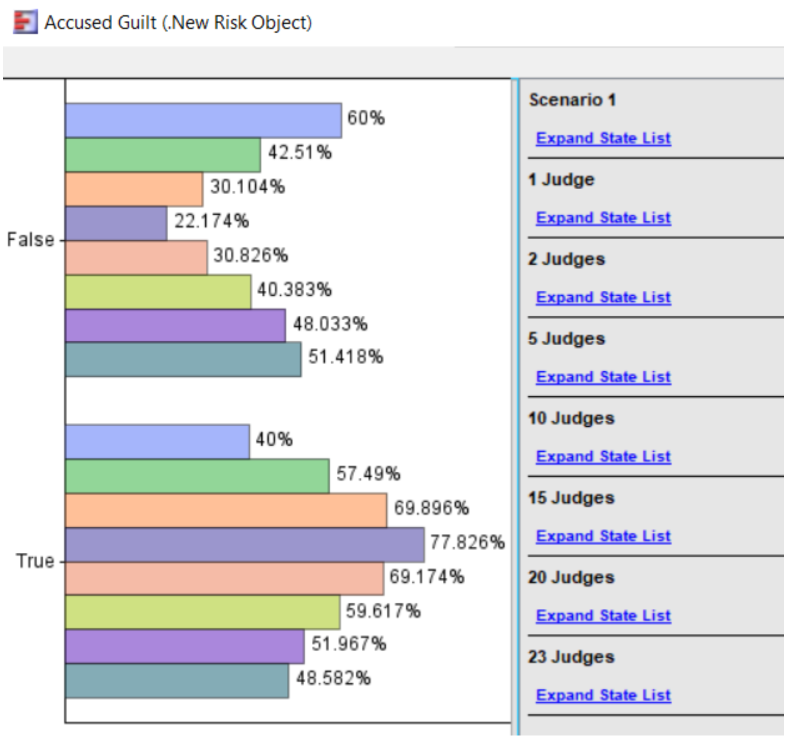


Fig.7 Posterior Marginal Distribution for Accused Guilt, all scenarios

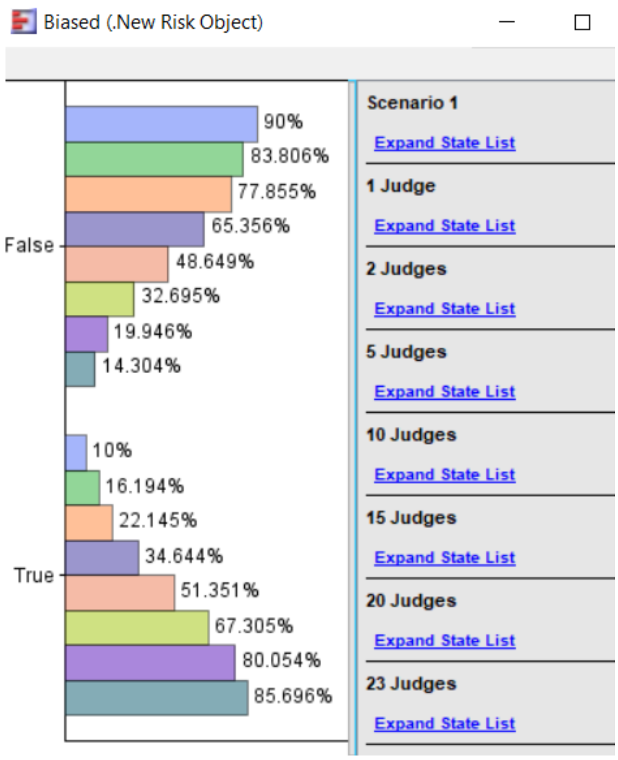


Fig.8 Posterior Marginal Distribution for Biased, all scenarios

b) **<u>Take the posterior marginals for $P(Accused\ Guilt = True \mid scenario\ data)$ and $P(All\ Judges\ Baised = True \mid scenario\ data)$ and plot, using Excel or similar, the 3 response curve with x-axis equal to the number of judges in each scenario and y-axis equal to these posterior marginals. [20 marks]</u>**

This task was completed using Excel. Fig.9 shows the input data, and Fig.10 illustrates the output. The x-axis represents each guilty verdict for all scenarios. The y-axis shows the posterior probability.

Risk Scenarios	P(AG=True|Risk Scenarios)	P(All Judges Biased=True|Risk Scenarios)
1 Judge Guilty Verdict	57.490	16.194
2 Judges Guilty Verdict	69.896	22.145
5 Judges Guilty Verdict	77.826	34.644
10 Judges Guilty Verdict	69.174	51.351
15 Judges Guilty Verdict	59.617	67.305
20 Judges Guilty Verdict	53.298	77.836
23 Judges Guilty Verdict	48.582	85.696

Fig.9 Input data for Posterior Marginals for (Accused Guilt = True) and (All Judges Biased=True)

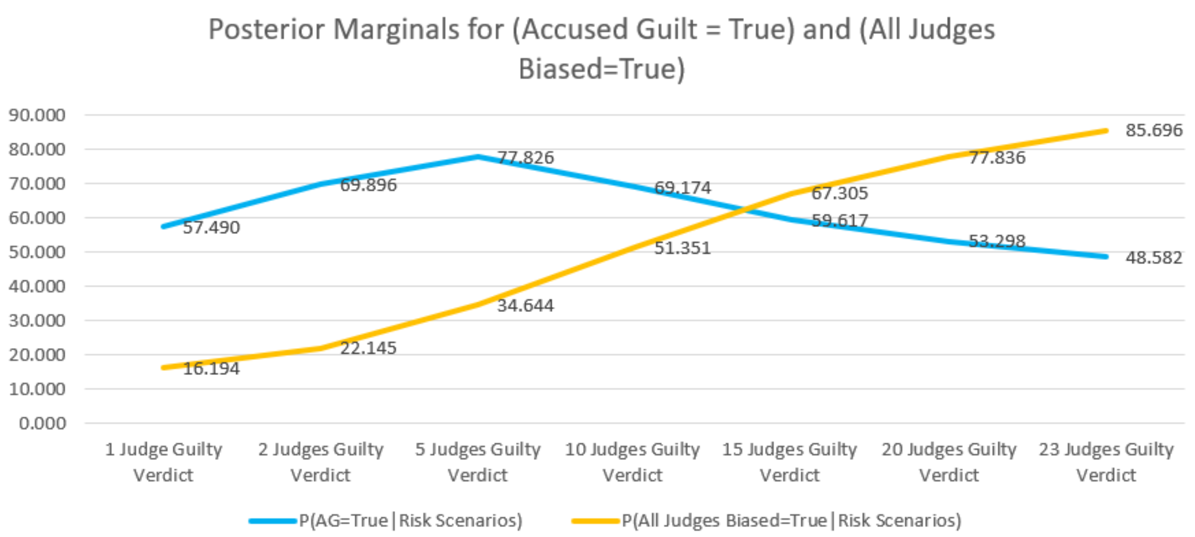


Fig.10 Posterior marginals plot

c) **<u>From your answer to Question c):</u>**
   i. **<u>When is the posterior belief in Guilt at its maximum? [5 marks]</u>**
   ii. **<u>When is the posterior belief in Bias greater than the posterior belief in Guilt? [5 marks]</u>**

We can derive the answers to these questions from Fig. 10. The posterior belief is at its maximum in scenario 3. That is, when 5 out of 23 judges announce a guilty verdict. It has a probability of 77.826%.

The posterior belief in Bias is greater than the posterior belief in Guilt when 15 out of 23 judges make a guilty verdict (67.305% vs 59.617% accordingly).

d) **<u>Explain, in your own words, the results from Q2 c) and explain why the verdicts of the judges may not be independent. [5 marks]</u>**

The results from question 2 c) are an example of the phenomenon that anonymous support can be "too good to be true". Normally, a prevailing number of guilty verdicts would be considered as undeniable truth. However, we assume independence of opinion and rarely consider whether a systemic failure has occurred.

The ancient Jewish law practice accounted for this phenomenon. Our Sanhedrin experiment proves that large sequence of observations which support a hypothesis can counterintuitively decrease our confidence. It also demonstrated that with relatively low error rates, large sample sizes are not required for unanimous results to start becoming indicative of systematic failure. It also indicates that the requirement of a dissenting opinion would have provided a substantial increase in the probability of guilt required to secure a conviction (2). We saw that after a certain number of guilty verdicts the posterior probability of guilt diminishes. The peak posterior probability was reached with only five guilty verdicts, completely counter to intuition. It starts to decrease afterwards, as illustrated by the blue line in Fig.10.

We can also see in Fig.10 the tendency of the judges to be biased. This is illustrated by the yellow line. The more judges give a guilty verdict, the bigger the bias. We should also note how the probability of guilt decreases. As mentioned, the peak posterior guilt probability was reached when 5 judges gave guilty verdicts. At that point, the probability of the defendant being guilty was 77.8 %, and the probability of judges being biased was 34.6%. In contrast, when 10 judges gave guilty verdicts the guilt probability started to decrease for the first time. It reached 69.2%. At the same time, the probability of the Sanhedrin being biased increased to 51.4%. This was described mathematically in the first part of the assignment.

As seen in (1) and (2), the proper use of Bayesian reasoning can improve the efficiency, transparency, and fairness of the criminal justice system and the accuracy of its verdicts by enabling the relevance of evidence to be meaningfully evaluated and communicated.

e) **Identify an area of scientific, social or political discourse, where there is near unanimity of opinion but where a suspicion of systematic bias might be reasonably entertained. [5 marks]**

I believe stock market and financial analysts are subject to such a bias. This has been documented by numerous studies (3) and is easy to prove given that there always is ex-post data to compare the actual stock performance to analyst predictions. Analyst bias was also evident during the credit crunch in 2008. Indeed, this is what caused the catastrophe. All analysts and even central bank governors and committees deemed the US housing market too good to fail. Almost nobody doubted it, and everyone relied on the consensus forecast. It was the same “too good to be true” situation we explored in this assignment. Speaking of central bank decision making, it is another example of a political discourse where a systematic bias might be reasonably entertained.

Analysts cover multiple firms and often need to produce many forecasts in a single day. Forecast accuracy is said to decline over the course of a day as the number of forecasts the analyst has already issued increases. Furthermore, the more forecasts an analyst issues, the higher the probability that they rely on the consensus forecast, or reissuing their own previous forecast), or forecasting a round number (3).

Given the above, I believe the “too good to be true” bias should be paid more attention to by investment professionals. After all, the financial industry is prone to huge failure modes where the consequences of an incorrect decision are large.

References

1. Fenton. N, and M. Neil. 2019. Risk Assessment and Decision Analysis with Bayesian Networks: Vol. Second edition. Chapman and Hall/CRC.
2. J. Gunn, L., Chapeau-Blondeau, F., D. McDonnell, M., R. Davis, B., Allison, A. and Abbott, D., 2021. Too good to be true: when overwhelming evidence fails to convince | Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. [online] Doi.org. Available at: <https://doi.org/10.1098/rspa.2015.0748 > [Accessed 22 April 2021].
3. Hirshleifer, D., 2020. Behavioral Biases of Analysts and Investors. [online] The Harvard Law School Forum on Corporate Governance. Available at: <https://corpgov.law.harvard.edu/2020/07/22/behavioral-biases-of-analysts-and-investors/> [Accessed 22 April 2021].