PHSX815_Project2:
How does complexity affect inaccuracy?

Vinayak Jha

March 2023

1 Introduction

In statistics, we model an experiment (or a set of experiments) through a hypothesis model. Before the experiment, we have two different hypotheses models - the test hypothesis (referred to as H_2 in this paper) and the null hypothesis (referred to as H_1 in this paper). The experiment's goal is to determine if the test hypothesis can be accepted or not, within some given certainty. In this paper, I am using the log-likelihood ratio test to determine if the test hypothesis can be accepted.

The hypotheses models can be simple or complicated.

Simple: A simple model is a model where the data can be generated in one single step directly from the hypothesis model given the model's parameters.

Complex: A complex model is a model where there are multiple intermediary hierarchical steps to generating data. At each level of the intermediary draw, the random values drawn are used to control the parameters of the next model, thus controlling the distribution of the subsequent draw.

In the class, it was observed that if either of test hypothesis or the null hypothesis is complex, then it becomes harder to distinguish between these hypotheses. Subsequently, more measurements are needed to resolve between the null and test hypothesis. A metric that can be used to quantify the 'resolution' between the hypotheses is the false negative rate. The false negative rate refers to the probability of failing to reject the null hypothesis in favor of the test hypothesis. If the false negative rate is high, it means that the test is not powerful and more measurements per experiment will be needed to resolve (or separate the hypotheses).

In this project, I want to further investigate the relationship between false negative rate and the complexity of the hypotheses. To do this, I model my null and test hypothesis as a complex hypothesis with the complexity parameterized, i.e. I can change the complexity with a parameter. To study the relationship, I increase the complexity and measure the false negative rate. Additionally, I also study how the number of measurements per experiment also affects the false negative rate, given that hypotheses are complex.

I expect that as the complexity of hypotheses increases, the false negative rate will also increase. Further, as the number of measurements per experiment increases, the false negative rate decreases.

2 Experiment and Algorithm Design

To model complex hypotheses with the "complexity" parameterized, I will use nested Gaussian distributions as my hypotheses with a depth d. Thus, my hypothesis takes the form

$$
\begin{gathered}
X_1 \sim \mathcal{N}(\mu_0,\, {\sigma_0}^2) \\
X_2 \sim \mathcal{N}(X_1, {X_1}^2) \\
X_3 \sim \mathcal{N}(X_2, {X_2}^2) \\
\vdots \\
X_d \sim \mathcal{N}(X_{d-1}),\; {X_{d-1}}^2)
\end{gathered}
$$

Here, if the depth is 1, then the algorithm is stopped after the first iteration, for simplicity. In the initial setup, we have two hypotheses H_1 and H_2, and H_1 corresponds to null hypothesis with $\{\mu_0 = \gamma_0, \sigma_0 = 0.5\}$ and H_2 corresponds to test hypothesis $\{\mu_0 = \gamma_1, \sigma_0 = 0.5\}$, and the complexity (or depth) is d. It is also assumed that the complexity of both the test and the null hypothesis is the same, and thus the only difference between them is in the initial μ_0, which is γ_0 for the null hypothesis and γ_1 for the test hypothesis.

Additionally, $N_{measurements}$ refers to the number of measurements per experiment and $N_{experiments}$ refers to the total number of experiments.

The goal is to calculate the false negative rate of hypothesis 1 and hypothesis 2. The first step is to calculate the probability of obtaining a sample under hypothesis 1 - $P(\langle X \rangle | H_1)$ and the probability of obtaining a sample under hypothesis 2 - $P(\langle X \rangle | H_2)$.

2.1 Calculating $P(\langle X \rangle | H_1)$ and $P(\langle X \rangle | H_2)$

1. Generate $N_{measurements}$ samples from the initial distribution $\mathcal{N}(\mu_0,\, {\sigma_0}^2)$. Let the sample generated be represented by the set X_1, where $X_1 \in R^{N_{measurements}}$.
2. Consider a sample $X_{1,i}$ in X_1. For each i, generate another sample from $\mathcal{N}(X_{1,i},\, {X_{1,i}}^2)$. Let the final samples generated from all i be the set X_2.
3. Perform step 2, d times such that the depth will be d. The final set will be X_d and the dimension will still be $R^{N_{measurements}}$. Now, perform an average over the set X_d, and the result will be the outcome of this experiment.
4. Perform the above steps $N_{experiments}$ times and $N_{experiments}$ samples will be obtained.

After obtaining $N_{experiments}$ samples from hypothesis 1, I generate the histogram for the samples. The normalized height of that histogram acts as the probability density of obtaining a particular sample under hypothesis 1. The probability density obtained is referred to as $P(\langle X \rangle | H_1)$. Now, I perform all the steps again with hypothesis 2 and obtain the probability density for obtaining a sample under hypothesis 2, and this probability density is referred to as $P(\langle X \rangle | H_2)$. Note that $P(\langle X \rangle | H_1)$, $P(\langle X \rangle | H_2)$ are numerical approximations and are not absolute results.

The method described previously will act as a numerical approximation and might not give complete coverage over the domain of values possible. Thus, if $\langle X \rangle$ lies out of the observed domain, I assign it a very low probability 10^{-50}. This is done to have a positive value for the probability when calculating the log-likelihood ratio.

2.2 Calculating False Negative Rate

The following algorithm is used to calculate the false negative rate.

- From calculating the probabilities conditioned on hypothesis 1 and hypothesis 2, we also obtained $N_{experiments}$ samples for each hypothesis (step 4). For each sample (out of $N_{experiments}$ samples) from the null hypothesis (hypothesis 1), calculate the log-likelihood ratio which is given by $log(\frac{P(\langle X\rangle|H_2)}{P(\langle X\rangle|H_1)})$ where $P(\langle X \rangle | H_1)$, $P(\langle X \rangle | H_2)$ are numerical approximations from the previous subsection (2.1). The result is denoted as the set λ_1, and $\lambda_1 \in R^{N_{experiments}}$
- Sort the set λ_1 and calculate the $(1-\alpha)\%$ percentile, which is denoted by $\lambda_{1,\alpha}$.
- To approximate false negative ratio, repeat step one but now take samples from hypothesis 2 and calculate the set λ_2.
- The false negative rate is given by

  $$\int_{-\infty}^{\lambda_{1,\alpha}} P(\lambda_2|H_2) d\lambda_2$$

  .
- To approximate false negative ratio, I count the number of times the value in λ_2 was less than $\lambda_{1,\alpha}$. Divide this count with $N_{experiments}$ and the result will be the false negative rate.

Following the steps, I get the false negative rate for the set of $(\gamma_0, \gamma_1, N_{measurements}, d, N_{experiments})$.

3 Analysis

Figure 1 (top left) below shows the log-likelihood ratio's probability distribution where the null hypothesis (hypothesis 1) corresponds to $\gamma_0 = 2$ and the test hypothesis (hypothesis 2) corresponds to $\gamma_1 = 3$. The depth is 1 and the number of measurements is 7. Visually, it can be observed that the test hypothesis can be clearly separated from the null hypothesis. This statement is also supported statistically as a false negative rate was calculated to be 2%.

Now, I changed the depth of both of the initial hypotheses from 1 to 4 and left the γ_0 and γ_1 unchanged for both of the hypotheses. The resulting log-likelihood ratio distribution is shown in Figure 1 (top right). It can be seen that it is harder to separate the test hypothesis from the null hypothesis which is also supported numerically since the false negative rate was estimated to be 85%.

Figure 1 (bottom left) demonstrates a similar analysis but now both of the hypotheses have depth 7. The false negative is close to 99% which means the hypothesis cannot be separated from one another at all, and the log-likelihood ratio distribution of the test hypothesis is very close to the distribution of the null hypothesis.

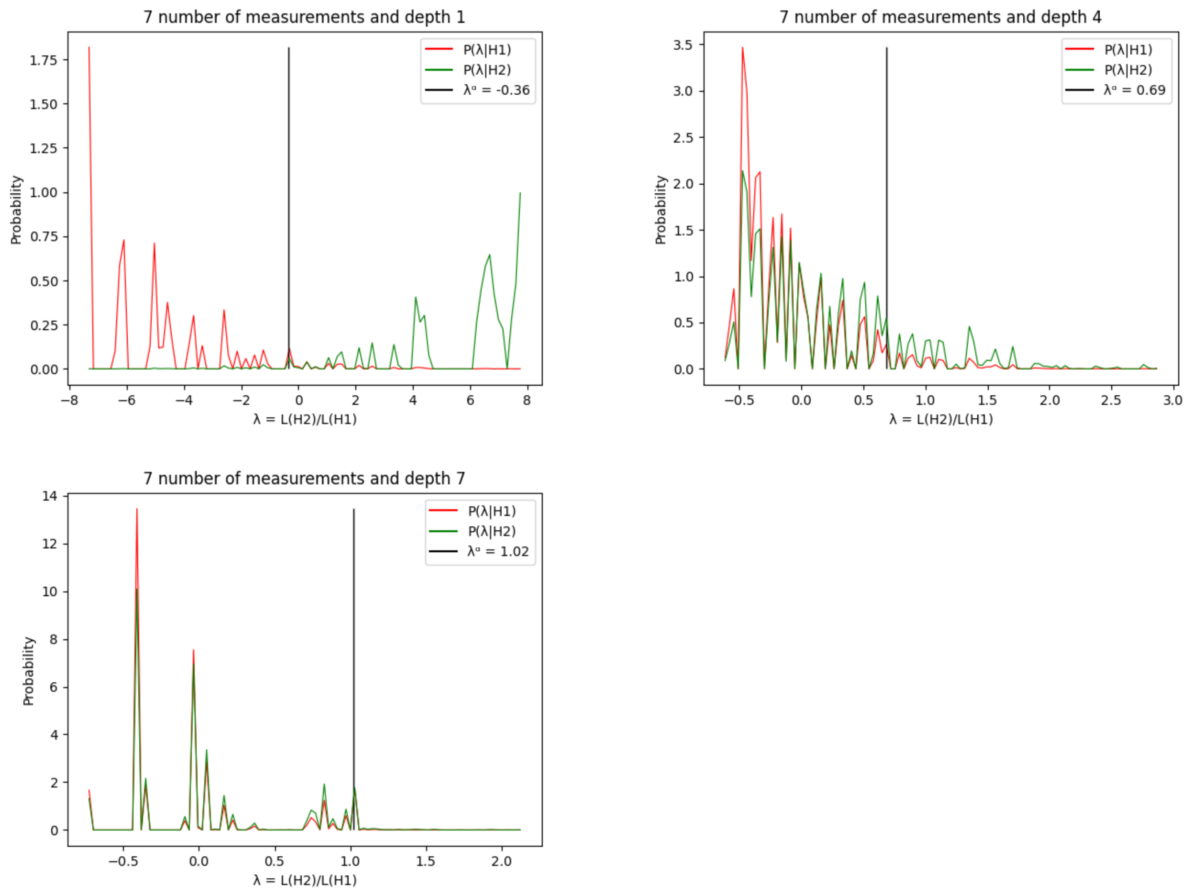

Figure 1: Demonstrates the log-likelihood ratio's probability distribution for varying complexity of the test and null hypothesis. The black vertical line represents the λ_α at $(1-\alpha)\times 100\%$ percentile. In each of the histograms, we reject the null hypothesis if the log-likelihood ratio of the observed sample (in an experiment) is more than λ_α. In the first figure, the test and null hypothesis are clearly separated and thus we are less likely to fail to reject the null hypothesis over the test hypothesis - thus the false negative rate is low (2%). However, as the complexity (depth) increases, the test and null hypothesis become less separated and thus we will fail to reject the null hypothesis - thus the false negative rate is very high $(85\%$ and 99%)

Figure 2 provides a grid of false negative rates for a set of a number of measurements and depth of both of the models. The false negative rate is written inside the grids as fractions. It can be seen that as the depth increases, the false negative rate increases very rapidly. Additionally, the false negative rate decreases very slowly as the number of measurements increases.

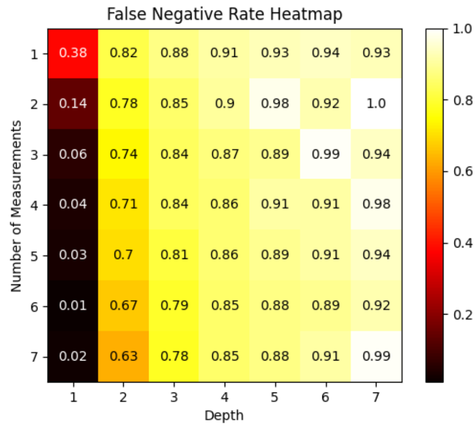

Figure 2: The heat map demonstrates the sensitivity of false negative rate on the complexity (or depth). Even if the depth increases from just 1 to 2, the false negative increases on average 22.8 times. The most affected models are the ones that have the most complexity. When the depth is high, the false negative rate tends to be high and doesn't decrease much even if the number of measurements is increased

The increase in complexity (or depth) of the models increases the false negative rate significantly. As the number of measurements increases, going down the map, the false negative rate decreases. However, as the depth increases (going right of the map), the increase in the number of measurements doesn't decrease the false negative rate by a significant amount.

4 Conclusion

My project provides evidence for the conclusion that as the complexity of the model (the intermediate distribution) increases, the resolution between the test and the null hypothesis decreases, and the false negative rate increases very rapidly. Additionally, the rate at which false negative decreases for a given complexity (depth in my project) is low, thus the number of measurements needed to get a good resolution (decrease false negative rate below a small threshold) increases. Circling back to the topic, as the complexity increases, the inaccuracy in our analysis also increases since we are more likely to reject the test hypothesis over the null hypothesis.