PHSX815_Project1:
Experiments with Gaussian Hypotheses

Vinayak Jha

February 2023

1 Introduction

The goal of performing experiments is to quantify how likely (or unlikely) one hypothesis is over another. Generally, the initial hypothesis is called the null hypothesis (referred to as H_1 in this paper), and the second hypothesis is called the test hypothesis (referred to as H_2 in this paper).

During an experiment, the false negative rate refers to the probability of failing to reject the null hypothesis in favor of the test hypothesis. If the test statistic is the log-likelihood ratio (LLR), the false negative rate is the probability that the observed LLR lies below the λ_α of the LLR distribution of the null hypothesis but still lies inside the LLR distribution of the test hypothesis, where α is a measure of the significance of the test.

In class, it was noticed that if the initial hypotheses take the form of exponential distributions, then the number of measurements needs to be increased to decrease the false negative rate.

The aim of this project is to perform a similar experiment using Gaussian distributions as initial hypotheses. Since a Gaussian distribution is parameterized by two parameters, the mean and the standard deviation, I am making a simplifying assumption that the hypotheses differ only in the mean and they have the same standard deviation. Given the two means and the standard deviation, I calculate the number of measurements needed to get the false negative rate of the experiment below a certain threshold. Further, I vary the standard deviation to evaluate how the number of measurements needed changes. In other words, I am evaluating if I need to make more measurements if the standard deviation of the initial hypothesis increases.

2 Algorithm Design

The first step is to calculate the false negative rate given a set of $(\mu_1, \mu_2, \sigma, N_{meas})$. Here, it is assumed that μ_1 refers to the mean of the null hypothesis (H_1) and μ_2 refers to the mean of the test hypothesis (H_2). σ is the standard deviation (which is assumed to be the same for the two hypotheses). N_{meas} refers to the number of measurements per experiment. $N_{experiment}$ refers to the number of experiments conducted. The following algorithm is used to calculate the false negative rate.

- Draw $N_{experiment} \times N_{meas}$ samples from $\mathcal{N}(\mu_1,\, \sigma^2)$ (acts as null hypothesis). Thus, I have N_{meas} samples in $N_{experiments}$ experiments. I am using the standard NumPy library to draw random variables out of the Gaussian distribution.

- For each experiment, average the N_{meas} samples to calculate the average value in that experiment $\langle X \rangle$.
- For each average value from the previous step $\langle X \rangle$, calculate the log-likelihood ratio (LLR) assuming it was drawn from H_2 or H_1. Thus, the LLR equals $log(\frac{P(\langle X \rangle | H_2)}{P(\langle X \rangle | H_1)})$. Since the average of Gaussian distributions also follows a Gaussian distribution, $P(\langle X \rangle | H_1)$ and $P(\langle X \rangle | H_2)$ equal to $\mathcal{N}(\mu_1, \frac{\sigma^2}{N_{meas}})$ and $\mathcal{N}(\mu_2, \frac{\sigma^2}{N_{meas}})$ respectively.
- From the previous step, after averaging, there will be $N_{experiment}$ samples of log-likelihood ratios. Sort these samples and find the $(1-\alpha)\%$ percentile to calculate the λ_α. I am using the in-built sort library and NumPy's percentile function to get the percentile.
- The false negative rate is given by
$$\int_{-\infty}^{\lambda_\alpha} P(\lambda | H_2) d\lambda$$
- To approximate the above integral, follow the step 1 through 3 again but this time draw $N_{experiment} \times N_{meas}$ samples from $\mathcal{N}(\mu_2, \sigma^2)$. Then, sort the LLR samples and count the samples where the LLR was less than λ_α. Then, false negative rate equals $\frac{count}{N_{experiments}}$.

Note: To optimize the calculating the LLR, I directly drew the $N_{experiment}$ samples from $\mathcal{N}(\mu_1, \frac{\sigma^2}{N_{meas}})$ and $\mathcal{N}(\mu_2, \frac{\sigma^2}{N_{meas}})$ instead of drawing $N_{experiment} \times N_{meas}$ samples from $\mathcal{N}(\mu_1, \sigma^2)$ or $\mathcal{N}(\mu_2, \sigma^2)$ and then performing an average.

Following these steps, I get the false negative rate for a set of $(\mu_1, \mu_2, \sigma, N_{meas})$. This concludes the first step of the algorithm (calculating the false negative rate)

To calculate the number of measurements needed to get the false negative rate below a threshold β, I repeat the process starting from $N_{meas} = 1$, and keep increasing N_{meas} till the false negative rate becomes less than β. Once I calculated N_{meas} for a particular σ, I vary the σ and calculate the N_{meas} for the new σ again, following these steps from the start.

After following the steps, I have a set of σ_i and $N_{meas,i}$ such that the false negative rate for each pair of σ_i and $N_{meas,i}$ is less than β.

3 Analysis

Figure 1 (right) shows the log-likelihood ratio probability distribution with $\mu_1 = 2.0$ (null hypothesis: hypothesis 1), $\mu_2 = 3.0$ (test hypothesis: hypothesis 2) and $\sigma = 1.0$. The number of measurements (N_{meas}) is 20, $\alpha = 0.95$. The false negative rate was calculated to be 0.002387 (0.23%). The X-axis has the log-likelihood ratio (LLR) calculated after each set of N_{meas} samples, and on the Y-axis we have the absolute probability (not log probability). Figure 1 (left) shows the result of the same setup with 10 measurements instead of 20. The false negative rate, in this case, was calculated to be 0.064 (6.4%) - this can also be visually explained since the overlap between the two log-likelihood ratio densities is greater in Fig 1 (left) than in Fig 1 (right) thus it is harder to resolve. In Fig 1 (right), the overlap is less thus the probability that the test hypothesis is rejected in favor of the null hypothesis is less (hence the false negative rate is low).

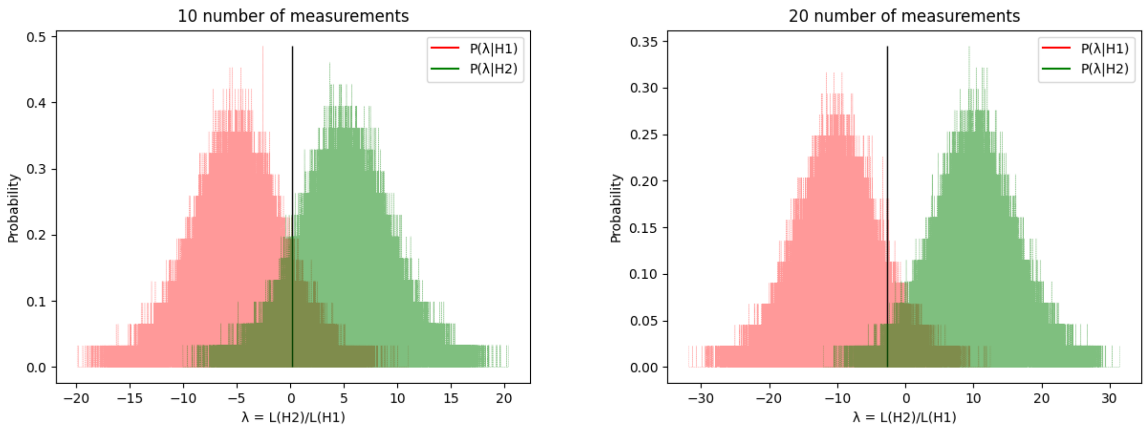

Figure 1: Histogram for log-likelihood density ratio for two Gaussian hypotheses under a different number of measurements per experiment. The left experiment was conducted with 10 measurements per experiment and the right experiment was conducted with 20 measurements per experiment. The black vertical line represents the λ_α at $(1-\alpha)\times 100\%$ percentile. The portion of the test hypotheses lying before the λ_α is greater in the first figure thus we are more likely to reject the test hypothesis (hypothesis 2) in favor of the null hypothesis (hypothesis 1), which makes the false negative rate higher in the first figure (6.4%)

Now, I calculate the number of measurements needed to get the false negative rate below a threshold given that $\mu_1 = 2.0$ (hypothesis 1), $\mu_2 = 3.0$ (hypothesis 2), and $\sigma \in \{0.5, 0.75, 1.0, 1.25, 1.50, 1.75, 2\}$. The threshold for the false negative rate is set as $\beta = 10^{-3}$.

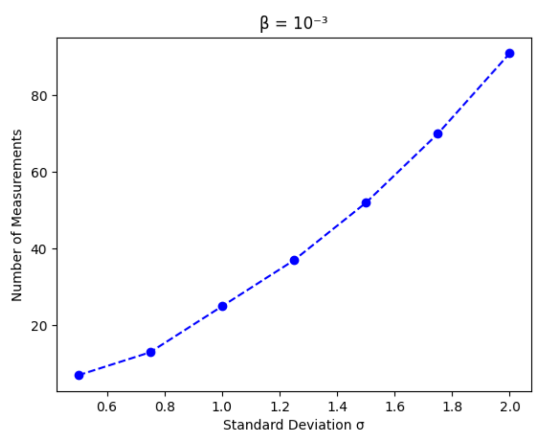

Figure 2: Simulation of the number of measurements needed to decrease false negative rate below a certain threshold ($\beta = 10^{-3}$) as standard deviation (σ) of hypotheses H_1 and H_2 keeping means same. When $\sigma = 0.5$, the number of measurements needed equals 7, but as σ is increased, the number of measurements needed also increases. For example, for $\sigma = 2$, the number of measurements equals 91

It can be seen in Figure 2 that the number of measurements needs to be increased as the standard deviation increases. This can be explained by the fact that as the standard deviation increases, the hypotheses overlap more (since they are Gaussian). Hence, more measurements need to be made to resolve the hypotheses.

4 Conclusion

According to numerical solutions, I conclude that the false negative rate of the experiment decreases as the number of measurements per experiment increases (as seen in Figure 1).
Further, if the means of the initial hypotheses are kept constant and only the standard deviation is varied, then the number of measurements needed to decrease the false negative rate below a fixed threshold also increases (as seen in Figure 2) as the standard deviation increases.