PHSX815_Project2
Stress, Sleep, and Sparsity: Simulating Hair Loss to Determine its Cause

Berg Dodson

May 2023

1 Introduction

Project 1 focused on determining whether balding was the result of natural factors or gnomes stealing your hair in the middle of the night. The results were inconclusive based on simply using a single rate for each both the null and our hypothesis. This was mostly due to the rate assumptions for both theories being too close to each other to be distinguishable, resulting in a power of the test that was less than 5%. Further reading and study has lead to an update of the methods and the assumptions used to simulate these two hypotheses.

The consensus in the literature (hypothesis H_0 is that increased stress levels and decreased sleep contribute to premature hair loss [1]. However, the simulations we are creating for this project test whether these factors are the cause or merely a cover for the real cause of hair loss: gnomes. Based on observations [2], gnomes come out while you sleep to steal your hair for their nests they build in the walls. The rate at which they are able to collect your hairs depends on how relaxed they are and how well rested they are. These two quantities are inversely proportional to how much sleep and stress the person they are stealing the hair from has had. If this H_1 theory is correct it would look like people who stress all night about the work they need to do would lose hair much faster than people who are able to sleep through the night with an unburdened conscious. A similar result would be observed if the null hypothesis were the correct hypothesis. The only difference between the two theories will be the distributions they produce based on the amount of sleep the participants get.

This paper seeks to determine if the cause of hair loss can be considered natural or supernatural. This paper is organized as follows: Sec. 2 explains the finer details of the two hypotheses. It also explains the statistical analyses that will be applied to make our conclusion. A description of the computer simulation developed to simulate these possibilities is provided in Sec. 3. This will also include any necessary acknowledgements for inspiration for the code. Analysis of the outputs of the code are included in Sec. 4. Finally, conclusions are presented in Sec. 5 and we should be able to answer confidently what causes hair loss.

2 Hypotheses to Explain Hair Loss

One explanation for going bald is that stress and lack of sleep cause hair loss. We will refer to this as our null hypothesis (H_0). The other theory is that hair loss is a result of gnomes stealing our hair. This is our new hypothesis (H_1). H_1 assumes that gnomes sleep while a person is awake - it's difficult to steal from someone while they're awake and alert. The more that sleep the gnomes get the more hairs they can steal since they are more rested and energetic. However, since gnomes don't like to be seen and especially don't like to be watched while they sleep, Our model makes a distribution of how much sleep the gnomes get each night, i.e. how many hairs per day the gnomes are able to take on average, and then uses these simulated rate parameters to inform the distribution of hairs taken each day in the actual experiment.

Similar to the distribution of project 1, the actual histogram of the hair lost each day should be a Poisson distribution and is modeled as such here. This means that for each simulated measurement a random number and a rate parameter (λ) is needed.

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{1}$$

Where this project distinguishes itself from the previous project is that instead of this rate parameter being fixed, this λ is pulled from a different distribution. This distribution is a gamma distribution which is the conjugate prior to the Poisson distribution. There are two variables that affect the gamma distribution, k and θ, the shape and scale respectively. These two parameters shape the gamma distribution by making the mean of the distribution $k\theta$ and the variance $k\theta^2$ [3]. Both k and θ and greater than zero while the gamma distribution itself takes the form

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \tag{2}$$

For both H_0 and H_1 a k and θ was chosen as described below:

1. Null hypothesis H_0: $k = 2$, $\theta = 1.5$
2. Alternative hypothesis H_1: $k = 6$, $\theta = 1.75$

The way this was done is explain the following section.

3 Code Explanation

The code for this project is broken into two parts. The first part creates and records the gamma and Poisson distributions and is done in `Proj2_Gen+Write.py`. `Proj2_Gen+Write.py` creates these distributions by making an instance of the random class in `MyRandom.py` which is capable of sampling from these distributions. `Proj2_Gen+Write.py` then records the sampled gamma and Poisson distributions into `.txt` with a filename specified by the user. Specifically, the user can set the following flags for this portion of the project.

- -seed: the seed used to initialize the Random class.
- -rateFile: The filename where the rates sampled from the gamma distribution are recorded.
- -Nmeas: the number of measurements in the experiment.
- -Nexp: the number of experiments to be performed.
- -alpha: setting the k parameter for the gamma distribution. The mismatch between flag and function was a result of early code misconception and was not identified before it had become an integral part of the code. It has been left uncorrected to avoid potentially breaking the code.

- -beta: setting the theta parameter for the gamma distribution. The mismatch here is the same as with alpha above.
- -output: the name of the file where the data sampled from the Poisson distribution is recorded.

The second part imports the simulated data for both the gamma and Poisson distributions and plots them. It then calculates the log-likelihood ratios and other relevant statistics such as the mean, median, and confidence intervals. This is done in `Proj2_Read+Plot_v2.py`. `Proj2_Read+Plot_v2.py` creates an instance of the sorting class found in `MySort.py` and then imports the data files and rate files specified by the user flags. The script then sorts the imported data and calculates the values mentioned above. Once these are calculated, the script plots the sorted data and saves the resulting plots. The script finally calculates the log-likelihood ratios and power of the test. The user can control this entire process with the following flags

- -inputH0: The name of the text file that contains the data sampled from the Poisson distribution for the H_0 theory.
- -inputH1: The name of the text file that contains the data sampled from the Poisson distribution for the H_1 theory.
- -ratesH0: The name of the text file that contains the data sampled from the gamma distribution for the H_0 theory.
- -ratesH1: The name of the text file that contains the data sampled from the gamma distribution for the H_1 theory.

With that explained, the data from the simulations is presented in the following section.

4 Analysis

To rehash briefly, both the H_0 and H_1 theories, the natural and gnome assisted hair loss, were simulated by sampling from a gamma distribution and using the results of that distribution to shape the simulation of the Poisson distributions.

$$f(x) = \frac{\lambda^x}{x!}e^{-\lambda} \tag{3}$$

From this information we expect that both distributions should have peaks at λ = $k\theta$.For the actual simulated rates we see in fig. 1

The sampled data is presented in the following figure.

From 2 we can see that these distributions both have a mean of 2.9996 and 10.5077 for H_0 and H_1 and a standard deviation of 2.1213 and 4.2866, respectively. In this case both theories much more distinct than they were in project 1.

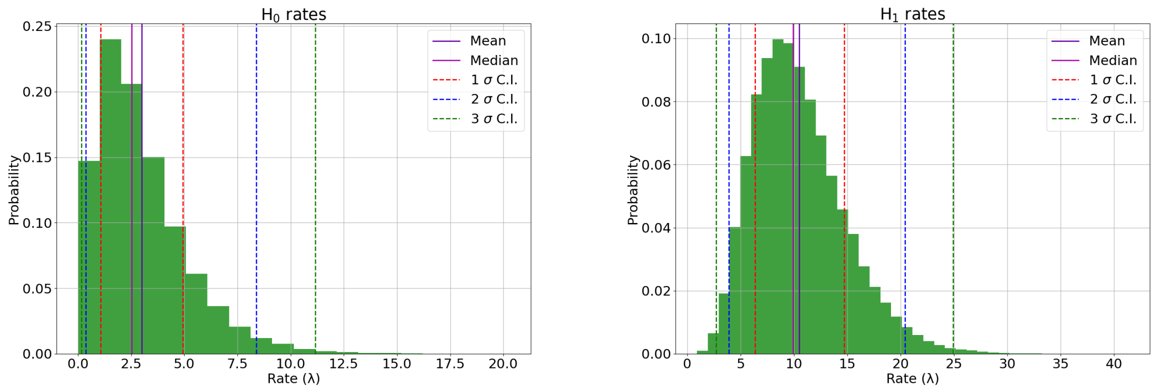

Figure 1: Simulations of the rate parameter for H_0, left, and H_1,right.

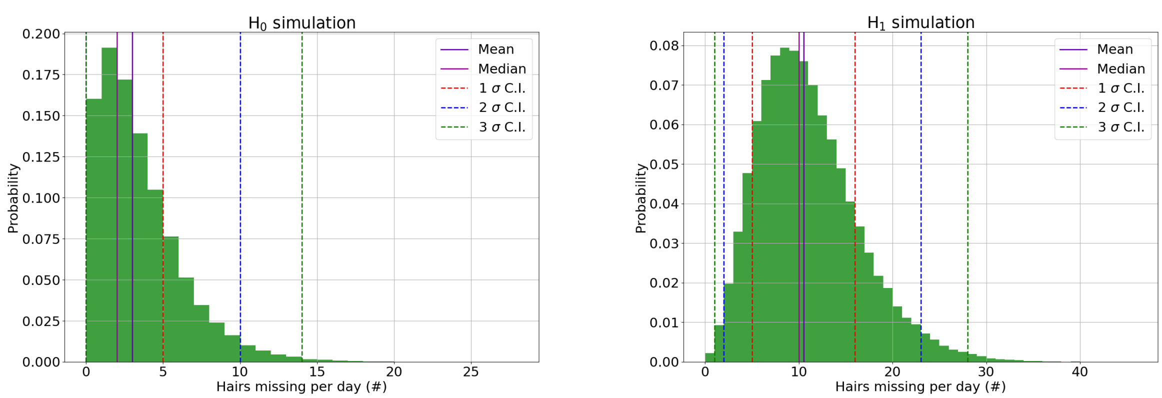

Figure 2: Simulations of the outcomes for H_0, left, and H_1,right.

5 Conclusion

According to our results, it is fairly easy to distinguish between our H_0 and H_1 in this project. Thus by making a few measurements it should be possible to distinguish between whether hair loss is due to gnomes or natural causes.

Here are the references

[1] D. Williamson, M. Gonzalez, and A. Finlay, *The effect of hair loss on quality of life*, *Journal of the European Academy of Dermatology and Venereology* **15** no. 2, (2001) 137–139, `https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1468-3083.2001.00229.x`.

[2] B. Dodson, *Gnomes Observations* I hear the pitter-patter of their feet as i sleep, March, 2023.

[3] *Gamma Distribution*.