Null Hypothesis Significance Testing

Shubham Gupta

May 22, 2019

1 Introduction

- Goal of NHST: Determine if a particular value for a parameter can be rejected.
- Probability of getting an outcome from the null hypothesis that is as extreme as the observed outcome is called the 'p' value.
- If p value is very small, we reject the outcome.
- Single outcome can have multiple p values depending on the sampling intention.
- Stopping at a fixed number of flips or after a fixed duration **does not** bias the data. Stopping after getting a fixed number of flips **does** bias the data.

2 p value

- We will compute p value with the bayes formula:

$$\text{pvalue} = likelihood * prior.$$

- **Likelihood function**: Probability for a single measurement AND the intended sampling process that defines space of all possible outcomes.
- **Null hypothesis**: Likelihood function with a specific value of θ.
- All possible values are defined by I.
- Each sample from the null hypothesis is given by $D_{\theta,I}$.
- For coin null hypothesis, $D_{\theta,I}$ will be $\frac{z}{N}$.
- **Sampling distribution**: Probability distribution over all possibilities i.e $p(D_{\theta,I}|\theta, I)$.
- **Expected value**: Typical value of $p(D_{\theta,I}|\theta, I)$ i.e $E[D_{theta,I}]$

$$pvalue = p(D_{\theta,I} \ fancysymbol \ D_{actual}|\theta, I).$$

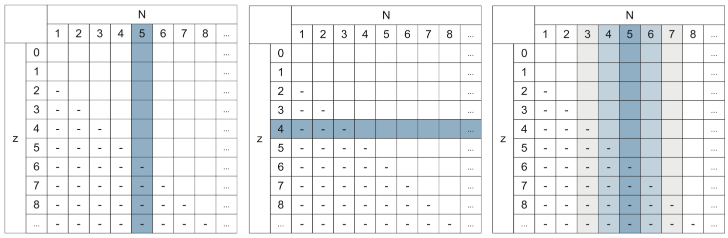

Figure 1: Space sample for coin flips. Left: Fixed N. Middle: Fixed z Right: Fixed time duration

2.1 Intention to fix N

- When N is fixed, it will represent the left table.
- **Aim**: What is the probability of z when N is fixed?
- **Answer**: Binomial distribution

$$p(z|N,\theta) = \binom{N}{z}\theta^z(1-\theta)^{N-z}.$$

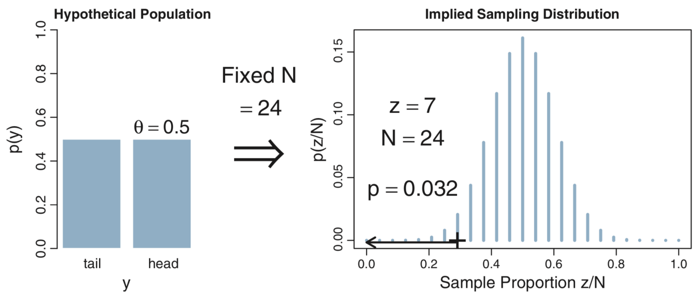

Figure 2: Binomial distribution i.e Sampling distribution

- We will get the above sampling distributions where there are infinitiely many samples performed.
- Sampling distribution is probability distribution over samples of data, **not over values of theta.**
- p-value is conventially set to 5%.
- We will compute p-value using the following:

$$p(\text{right tail}) = p((z/N)_{\theta,I} \geq (z/N)_{actual}|\theta, I).$$

$$p(\text{left tail}) = p((z/N)_{\theta,I} \leq (z/N)_{actual}|\theta, I).$$

- For the given example, we get the one-tailed p-value as 3.2%. Since it is larger than 2.5%(p-value for one tailed distribution. For two tailed it will be the conventional 5%), we **do not** reject the null hypothesis.

2.2 Intention to fix z

- Calculate probability of the process taking N flips to get z heads.
- We know Nth flip got the z head.
- $N-1$ flips had $z-1$ heads.
- Using binomial distribution, probability of $z-1$ heads in $N-1$ flips is:

$$\binom{N-1}{z-1}\theta^{z-1}(1-\theta)^{N-z}.$$

- Probability of getting heads in last flip is θ.
- Therefore, we get:

$$\binom{N-1}{z-1}\theta^{z-1}(1-\theta)^{N-z} * \theta.$$

$$= \frac{z}{N}\binom{N}{z}\theta^{z}(1-\theta)^{N-z}.$$

- This is called the **negative binomial distribution**

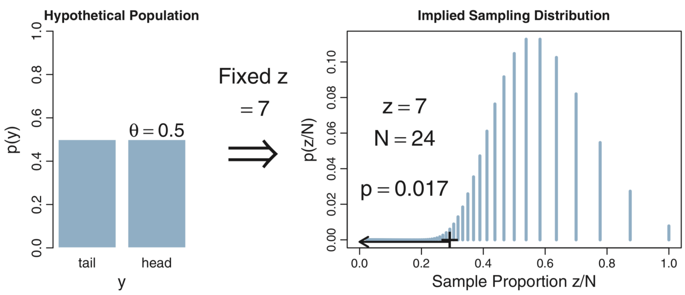

Figure 3: Sampling distribution for fixed z

- When $N = 7$ and $z = 7$, we get the spike at 1.0
- When $N = 8$ and $z = 7$, we get the spike at 0.875
- Spikes at left tail become dense and short as the value of N increases.
- Since $pvalue = 0.017$ is less than threshold of 2.5%, we reject the null hypothesis.
- **pvalue varies for different sampling scenarios.**

2.3 Intention to fix duration

- Neither z or N will be fixed.
- We need to specify how various combos of N and z.
- N can be small or large depending on the speed of sampling. Hence, we will use a **Poisson distribution**.
- Poisson distribution has a parameter λ, which is the mean value(and also the variance).
- For every value of N, the z values are binomial distributions. Hence, a Poisson distribution is a **mixture of binomial distributions**

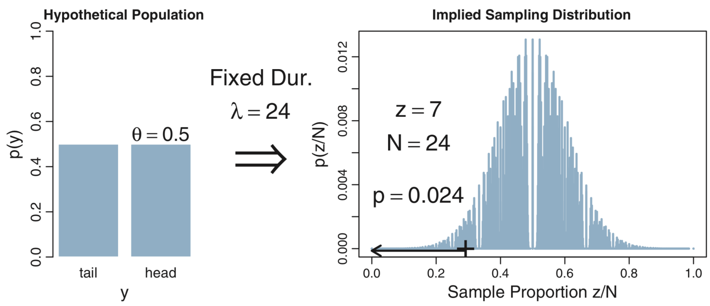

Figure 4: Sampling distribution for fixed duration

- We can see that the pvalue is about 0.024, which is slightly less than our one-tail pvalue of 0.025. Hence, we can call this result **marginally significant**.

2.4 Intention to run multiple tests

-