11.1 NHST for the Bias of a Coin

Zongyi Liu

2023-06-24

11.1 NHST for the Bias of a Coin

In the previous chapters we have seen a thorough introduction to Bayesian inference involving a Bernoulli likelihood function.

Here we will compare Bayesian inference with 20-th century Null Hypothesis Significance Testing (NHST) of binomial data, which can be helpful to decide which parameter to be rejected.

Example: suppose we have a coin that we want to test for fairness; we decide that we will conduct an experiment wherein we flip the coin $N = 26$ times, and we observe how many times it comes up heads. If it is fair, it should have 13/26 heads. Then we conduct our experiment: we flip the coin $N = 26$ times and happen to see $z = 8$ heads. All we need to do is figure out the probability of getting that few heads if the coin were truly fair. If the probability of getting so few heads is sufficiently tiny, then we doubt that the coin is truly fair.

The problem with NHST is that the interpretation of the observed outcome depends on the space of possible outcomes when the experiment is repeated. N might be 26, 23, or 32, or whatever. On the other hand, the experimenter might have intended to flip the coin until observing 8 heads, and it just happened to take 26 flips to get there. In this case, the space of possibilities is all samples that have the $8th$ head as the last flip.

This chapter mainly talks about details of NHST.

11.1 NHST for the Bias of a Coin

11.1.1 When the Experimenter Intends to Fix N

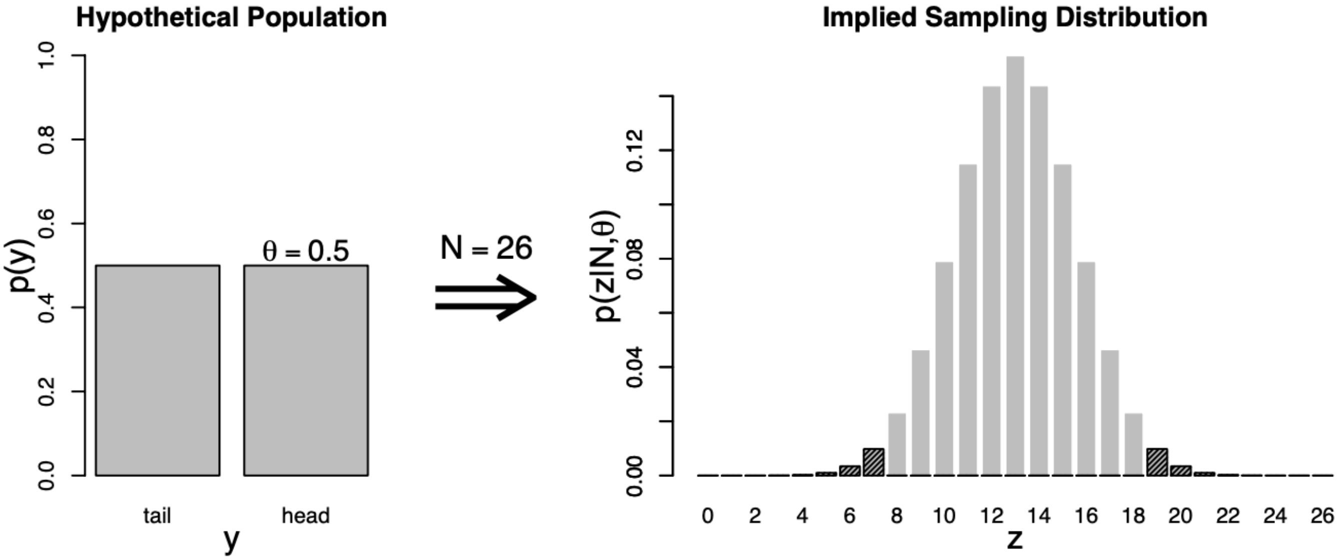
Here we still use the old example, which has $N = 26$ and we happen to observe $z = 8$ heads, which is not true if the coin is fair.

Thus we need to test if the coin is fair.

The probability of getting a particular number of heads when N is fixed can be given by the binomial probability distribution, which states that the probability of getting z heads out of N flips is

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

The plot below shows the binomial sampling distribution; the left graph shows the hypothesis whereas the right graph shows the real distribution.



We can come up the conclusion for our particular case. The actual observation had $z = 8$, and so we would not reject the null hypothesis that $\theta = 0.5$. In NHST parlance, we would say that the result "has failed to reach significance". This does not mean we accept the null hypothesis; we merely suspend judgment regarding rejection of this particular hypothesis. Notice that we have not determined any degree of belief in the hypothesis that $\theta = 0.5$. The hypothesis might be true or might be false; we suspend judgment.

11.1.2 When the Experiem
enter Intends to Fix z

Suppose the experimenter did not intend to stop flipping when N flips were reached; instead, the intention was to stop when z heads were reached. This scenario can happen in many real life; for example, widgets on an assembly line can be checked for defects until z defective widgets are identified.

The probability that it takes N flips to get z heads is

$$
\begin{aligned}
p(N|z,\theta) &= \binom{N-1}{z-1}\theta^{z-1}(1-\theta)^{N-z} \times \theta \\
&= \binom{N-1}{z-1}\theta^{z}(1-\theta)^{N-z}
\end{aligned}
$$

And we can see this in the plot below.

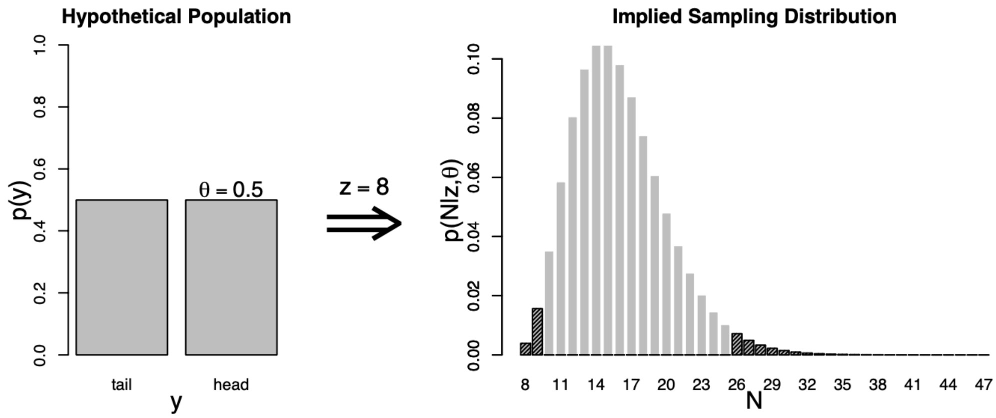


Figure 11.2: Sampling distribution of N when $p(y = \text{head}) = \theta = .5$ and z is fixed (Eqn. 11.2). The total probability of the dark bars does not exceed 5%.

This distribution is sometimes called the "negative binomial". Notice that values of N start at z and rise to infinity, because it takes at least z flips to get z heads, and it might take a huge number of flips to finally get the z-th flip.

If the coin is biased to come up heads rarely, then it will take a large number of flips until we get z heads. If the coin is biased to come up heads frequently, then it will take a small number of flips until we get z heads.

11.1.3 Soul Searching

Here is the situation. We see the same results as the experimenter, and we observe $z = 8$ heads out of $N = 26$ flips. According to NHST, if the intention of the experimenter was to stop when $N = 26$, then we do not reject the null hypothesis. If the intention of the experimenter was to stop when $z = 8$, then we do reject the null hypothesis.

Actual observed events are the same regardless of how the experimenter decided to stop flipping coins; In either case we observe z heads in N flips. An outside observer of the flipping experiment, who is not privy to the covert intentions of the flipper, simply sees N flips, of which z were heads. It could be that the flipper intended to flip N times and then stop. Or it could be that the flipper intended to keep flipping until getting z heads. Or it could be that the flipper intend to flip for one minute.

In all of these scenarios, the coin itself has no idea what the flipper's intention is, and the propensity of the coin to come up heads does not depend on the intentions of the flipper. Indeed, we carefully design experiments to insulate the coins from the intentions of the experimenter. Therefore our inference about the coin should not depend on the intentions of the experimenter.

11.1.4 Bayesian Analysis

The Bayesian interpretation of data doesn't depend on the covert intentions of the data collector. For data that are independent across trials, the probability of the conjoint set of data is simply the product of the probabilities of the individual outcomes.

Thus for $z = \sum_{i=1}^{N} y_i$ heads in N flips, the likelihood is $\Pi_{i=1}^{N} \theta^{y_i} (1-\theta)^{1-y_i}$, regardless of the experimenter's private reasons for collecting those data.

The likelihood function captures everything we assume to influence the data. In the case of the coin, we assume that the bias of the coin is the only influence on its outcome, and that the flips are independent.