Fairness of Coin

1 Problem

You toss a coin 100 times. 60 heads turned up. What can you say about the fairness of the coin?

2 Frequentist

The number of heads follows the binomial distribution. This is a bell shaped distribution with mode and mean at np and standard deviation of $\sqrt{np(1-p)}$ (n is the number of tosses and p is the probability of getting head).

Suppose our hypothesis is $p = 0.5$ (fair). Then the mean and standard deviation for the binomial distribution (of the number of heads) are 50 and 5 respectively. The outcome of 60 heads sits 2 standard deviations away from the mean. Calculating the numbers, this is within the 2,5% upper tail of the bell shape. So the $p = 0.5$ hypothesis is **rejected** under this approach.

3 Bayesian

Suppose the prior belief on the head probability p is $f(p)$, the probability density function of p. Then after seeing the outcome, the posterior distribution of p is

$$f(p|60H) = \frac{f(p) \cdot P(60H|p)}{\int_{-\infty}^{\infty} f(p) \cdot P(60H|p)dp} \tag{1}$$

If our prior is 'a fair coin', i.e.

$$f(p) = \begin{cases} 1 & p = 0.5 \\ 0 & otherwise \end{cases} \tag{2}$$

Then the Bayesian equation gives the posterior

$$f(p|60H) = \frac{f(p) \cdot P(60H|p)}{f(p=0.5) \cdot P(60H|p=0.5)} = \begin{cases} 1 & p = 0.5 \\ 0 & otherwise \end{cases} \tag{3}$$

Note that the posterior is exactly the same as prior (i.e the coin is fair). So the outcome of 60 heads **does not change our belief** on the fairness of the coin. Also note that this derivation is independent of what outcome we see (equation holds valid for any outcome from any experiment, rather than just '60H').

Slightly relax the strict prior density to, say, a density that peaks at 0.5 and diminishes at 0.45 and 0.55. Putting it through the machinary of Bayesian, we will get a *slightly* lower probability at $p = 0.5$ compare to the prior, but we are nowhere near rejecting the coin is fair.

4 Which Way?

The two approaches contradict. Apparently this is a case of the general Jeffrey-Lindley paradox.

But which way is more 'correct' in this particular case?

First have another look at the Bayesian method. It was shown that if the prior is zero probability at $p \neq 0.5$, then the posterior is guaranteed to be the same. Basically it is saying, if something is absolutely impossible, no amount of evidence will change that. When the prior is certainty then Bayesian method is powerless. Therefore it is only logical to assume there're some other possibilities, i.e. there's some distribution around $p = 0.5$. Looking in another way, it is contradictory to ask 'how the 60H result would change our belief' and to assume 'our belief is absolutely no other possibilities' at the same time.

Therefore it is only reasonable to assume a distribution $f(p)$ that is not certainty. But still, as it was pointed out, by assuming a reasonable distribution around 0.5, Bayesian method would change the probability density at $p = 0.5$ only slightly. The conclusion is still opposite to that of frequentiest approach (which strongly rejects the fairness hypothesis).

Next have another look at the frequentist method. The reasoning is based on hypothesis testing using the test statistics of the number of heads. For this reasoning to work, there has to be a concept of 'extremeness' of the outcome, or, more specifically, the observed outcome sits at a 'extreme' position in the distribution of the outcomes. In this case this is conveniently the *tail* of the *bell shaped* binomial distribution. However, this is a specious argument. In fact it is difficult to define the 'extremeness' in the distribution of outcomes. The issue is that, the 'outcome' of our experiment is not just '60 heads', but is a full sequence of heads and tails. Moreover, any such sequence has *equal* probability to happen (under the hypothesis of fair coin), i.e. any sequence has a probability 0.5^{100}. It's a *flat* distribution for the outcome, and there's no extreme position defined.

The frequentiest approach is actually discarding some information of the outcome of the experiment (only using the 60 heads, discarding the full sequence information). However the information is *inseparable*. Even if we are only told of 60 heads but not told of the actual sequence, we still should have known, by the nature of the experiment, that the outcome is a particular realisation of heads-tails sequence. The distribution is still flat and we cannot define extremeness as requested by the hypothesis reasoning. What happened happened. The change of belief shall base on what happened, not just partial summary of information.

To view it another way, suppose instead of the original '60H' question, we are asked 'if 100 heads turned up by tossing 100 times, what can you say about the fairness of the coin?'. Many would find easier to understand the answer 'no change to the fairness hypothesis', because 100 heads is just a realisation of a sequence, equally likely as any other realisation of a sequence. The clearer reasoning in the '100H' question is because in this case saying '100 heads turned up' implicitly contains the information that the actual sequence is head, head, head, ... head (100 times). The 60 heads case is no difference.

5 Conclusion

To conclude, for this particular problem, I am more convinced by the Bayesian approach with the caveat that the prior has to be a distribution that is not certainty. With a reasonable prior distribution, the fairness of the coin will not be rejected.