1.

Intro

In this question, we will derive the maximum likelihood estimator for the Bernoulli / binomial model. Let's get a quick reminder of those models. The Bernoulli model says that when we have a dataset $\mathcal{D}$ formed by the result of N coin tosses, with N_1 heads and N_0 tails, the likelihood of that event is $p_{Ber}(D|\theta) = \theta^{N_1}(1-\theta)^{N_0}$. On the other hand, the binomial model says that the probability of gettin N_1 heads in N trials is given by $p_{bin}(D|\theta) = \binom{N}{N_1}\theta^{N_1}(1-\theta)^{N-N_1}$. The binomial coefficient $\binom{N}{N_1}$ was introduced to take into account the lack of order in the result. The Bernoulli model gives the probability of a specific sequence of events, while the binomial model gives the probability of all sequence of events that satisfy a given property (in this case, the number of heads). Since, their expression are proportional, maximizing one is the same as maximizing the other. That being said, we will work with the Bernoulli model in the next section.

Solution

As said in the intro, the likelihood of the Bernoulli model is given by:

$$p(D|\theta) = \theta^{N_1}(1-\theta)^{N_0} \tag{1}$$

Since we want to find the argument that maximizes this function, we can also find the argument that maximizes any increasing monotonic function $f(p)$. As the author suggest, we will use the log function.

$$l(p) = log(p) = N_1 log(\theta) + N_0 log(1-\theta) \tag{2}$$

Now, we just have to differentiate this function and equal it to zero:

$$\begin{aligned} \frac{dl(p)}{d\theta} &= 0 \\ N_1\frac{1}{\theta} - N_0\frac{1}{1-\theta} &= 0 \end{aligned} \tag{3}$$

Putting θ in evidence in (2), we arrive at:

$$\theta_{MLE} = \frac{N_1}{N_0 + N_1} = \frac{N_1}{N} \tag{4}$$

Conclusion

As we can see, the maximum likelihoond estimator is just the empirical count of heads $\frac{N_1}{N}$, which makes sense. Nevertheless, be aware of the dangers of MLE estimators. The MLE estimator can suffer from the **zero count problem**. For example, if we toss a coin 20 times and all results are tails, then $\theta_{MLE} = 0$. This result says that there is no chance of getting heads on this coin, which is absurd. We solve this by beign a good Bayesian and using priors . This will be discussed in more depth in other exercises on this chapter.