# 1: Repeat task:

Derive the observed Fisher information for the Binomial distribution. Discuss the general steps to derive the Fisher Information.

# 2: Definition of important task aspects:

## Binomial Disribution:

<https://www.investopedia.com/terms/b/binomialdistribution.asp>

The binomial distribution is a probability distribution that summarizes the likelihood that a value will take one of two independent values under a given set of parameters or assumptions.

The underlying assumptions of the binomial distribution are that there is only one outcome for each trial, that each trial has the same probability of success, and that each trial is mutually exclusive or independent of each other.

The binomial distribution is a common discrete distribution used in statistics, as opposed to continuous distributions, such as the normal distribution.

m… number of trials, that are rated as success  
µ… probability of m being the outcome of a trial  
N… number of total trials

This definition gives us the probability p of m successive trials given the number total trials and the probability of m.

## Observed fisher Information:

The observed fisher information is a way of measuring the amount of information that an observable random variable m carries about an unknown parameter µ upon which the probability of m depends.

So what does this mean for binominal distributions?

Example: Coin toss

Some famous gambler gave us a coin, but we don´t know if this coin is a fair one, so we can´t expect µ to be 0.5. Anyway, we want to know this, because we would like to gamble a bit. In our case head is a positive outcome. So, we ask ourselves how long we need to flip the coin and gather data until we have a solid prediction.

* 1. Verteilungen aufzeichnen
* 2. Zeigen, dass die Glockenkurve schmaler wird desto mehr Versuche gemacht wird, und dass die Schätzung daher zuverlässiger ist desto steiler die Kurve ist
* <https://matheguru.com/stochastik/binomialverteilung.html>
* Vielleicht sowas oder ähnliches
* Score function erklären (eigentlich ist das einfach nur die erste Ableitung -> fancy Name)
* Observed fisher function erklären:

*“The observed Fisher information is a way of measuring the amount of information*

*that an observable random variable X carries about an unknown parameter”*

Am konkreten Beispiel coin toss nur sagen, dass die fisher information bei steigender Anzahl von Versuchen (engere Glocke und stärkere Rechtskrümmung der log likelyhood) zunimmt -> Begründung dieses Verhaltens am Ende anhand der hergeleiteten Formel (hier nur die logische Begründung: höhere Aussagekraft bei mehr Wiederholungen!)

Überleitung zur mathematischen Herleitung in der Art: Lässt sich dieses Verhalten/ diese Vermutung mathematisch bestätigen? Und dann Abfahrt!

# 3: Derivation:

We use the binomial disribution as starting point for the derivation:

In the next step we need plug this into the score function.

First of all, we later want to know where the maxima of our function lays and thus need to set the first derivative to zero. So this makes clear why we derivate the function. But why do we need the logarithm? This seem to be of some cosmetic reason. Through this ln-funtion we don´t shift the maxima in that we are interested, but it becomes easier to derivate the function. We will see this in a second:

So instead of using product rule and chain rule to solve this algebraic problem we can simply derive those three individual terms. This makes things easier here. This results to:

So, we already got the result of the score function here. We now can calculate the maxima. For this we set s(µ)=0

This shows now, that the probability of a success in a trial equals the number of successive trials divided by the total trials -> Makes sense.

So now we got our peak-point. What we next are interested in is the negative derivative of the score function.

Why? Because we explained beforehand that expressiveness translates in a more curved score function. The magnitude of curvature is equivalent to the expressiveness of the estimate. The curvature now equals the second derivative of the likelihood and hence the first derivative of the score function. Because both functions are always “right-curved” we look at the negative derivatives. To our score function this results to:

After this we can plug in the position of the peak in this formula:

What does the observed fisher information for the binominal distribution give us?

With rising numbers of N the Fisher information also becomes greater in a linear manner.

For the values of µ the following plot shows the results from µ=0.01 to 1

The fisher information has its minimum if m/N equals 0.5 and rises if the probability of m increase or decrease.