# 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 a continuous distribution, 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

## 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?

Beispiel: Würfe auf Basketball

Person x wirft N1 = 5 Bälle auf Baskeballkorb. Wir wissen nicht ob es sich bei dieser Person um einen Amateur oder Profi handelt, denken jedoch, dass sich dies darin widerspiegelt, wie viele Bälle diese Person x „reinmacht“. Der Fragestellung ob Amateur oder Profi wird sich in dem Ergebnis des Versuchs widerspiegeln. Wenn wir nun anstatt der fünf Bälle 50 Bälle werfen lassen würde sich die Wahrscheinlichkeit für eine richtige Einschätzung zu Amateur oder Profi deutlich steigen. Die Likelihood Funktion wird dabei schmaler und höher.

Herleitung:

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

In the next step we need to plug this into the score function. What is the score function by the way and why do we do so?

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 drivate 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 do the math on those four individual terms. This results to:

So we already got the result of the score function here. We now can calculate our peaks position. 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 saw on a plot that a expressiveness tranaslate in a more curved likelihood. 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: