Hey guys,

welcome to our explain video. Last week already saw a really nice video about the fishers.

We want to derive the observed fisher information for the binomial distribution with you.

But what was it all about last week, and wasn’t it something different? Yes, there is a difference, but still there is a quite important connection between them!

Lets have a look at that.

Our friend has a coin, and when she tossed it, there is a probability of μ that it lands head, and a probability of 1- μ that it lands tail. There are no other possibilities of outcome, so you call this a Bernoulli random variable.

Your friend now wants to play a game with you. She says: I will toss the coin 10 times! Each time it lands head I will give you 1€, each time it lands tails, you give me 1€. You are a smart guy and go into business. After all, you have the chance to win up to 10€.

You know, each toss is independent and identically of the others. Tossing 10 times, mean you can win up to 10€ or loose up to 10€, and you also know the probability to win 1€ is μ, so if you conclude the probability to win 10€ is μ^10.

But what is the probability that you and your friend win nothing? Means tossing 10 times and getting 5 heads and 5 tails. This can be achieved by the probability of m under the condition N and μ. This results in the vector 10, 5 times (5 heads to the power of μ) times (10 tosses minus 5 heads) to the power of (1 minus μ) in brackets).

If you write it down with variables, it will result in the probability for m, given N and μ equals the vector N, m times (m to the power of μ) times (N-m to the power of (1 minus μ) in brackets)

You call this the binomial distribution. It’s the sum of independent and identically distributed Bernoulli random variables.

Okay now we know what is the binomial distribution. To derive the observed fisher information for that, we need a short recap back to your favourite subject at school – Maths!

A small recap about derivatives and there meanings.

If we have a function like, x squared. We can easily build the first derivative equals 2 times x. If we have a look at the graph you see, the first derivative describes the gradient of the tangent at the point x.

We can also build the second derivative, which is the first derivative of 2 times x. So it will be 2. What does this two means for us? Lets have a look at the graph. The second derivative describes the curviness of the graph.

Ok. Lets keep that interpretation in our minds and start with the derive of the observed fisher information.

As we already know from last week the fisher information is just the second derivative of the distribution we are looking at. Okay, lets go back to the binomial distribution and build the derivatives.

Hmm? Does it really have to be that complicated? Nobody of us is motivated to use the chain rule twice and going on with that.

But, we know that applying a logarithm does not change a function's information. Lets try this.

Hmm, okay, knowing a product inside the logarithm can be changed to the sum of the logarithms of each factor. Oh wow, now it’s really much more easier to derive it!

Lets call our first derivative S of μ.

**… Description of step by step derive of S(μ) + What is the meaning of the Score function?**

Okay, but as we know we need the second derivative of the binomial distribution. So lets derive the Score Function.

**… Description of step by step derive of I(μ) + What is the meaning of the fisher information?**