

3 Parameter estimation - maximum likelihood method

3.1 Estimation of the proportion

Let x_1, \dots, x_n be independent realizations of a Bernoulli distributed random variable X . We wish to estimate the parameter p .

- Say that $n = 5$ and that we got the following 5 values: 1, 0, 1, 1, 1. What would be the probability of this event if $p = 0.2$? What if $p = 0.75$? Plot the curve of these probabilities for various values of p . How would you calculate its peak?
- The data obtained on a sample are denoted as x_1, \dots, x_n (in the above case, $n = 5$, $x_1 = 1$ in $x_2 = 0$). Write the value of $P(X_i = x_i|p)$, i.e. the probability that the event, we have seen, has happened. Write the likelihood function.
- Find the estimator of p using the maximum likelihood method
- Is this estimate unbiased?
- How can you estimate the standard error?
- We wish to estimate the proportion of voters for a certain candidate. In a sample of $n = 500$, he gets 29 % of the votes. Give the 95 % confidence interval for this estimate.

Understanding the ideas in R:

- Use R to plot the Figure :

```
> p <- seq(0,1,length=100)      #for 100 values p between 0 in 1
> y <- p^4*(1-p)                #calculate the probability for each value
> plot(p,y,type="l")            #plot them as a curve
```

- Generate a sample of size 500, in which every individual has the probability 0.3 to vote for a certain candidate. Estimate the probability using the sample proportion. Repeat this procedure 1000x and look at the distribution of sample estimates.

- Add the estimated 95% confidence interval for each sample. What is the proportion of the samples, on which the interval encompasses the true value (0.3)?

3.2 The association of two random variables

We wish to know, how the revenue of a company in a certain branch depends on the number of employees. Assume that the income is randomly distributed with the average $\beta_0 + \beta_1 X$, where X is the logarithm of the number of employees. Say that we have data on a sample of companies and would like to estimate the parameters β_0 and β_1 .

- What is the density of the company revenues if we know the variance equals σ^2 ?
- Write the likelihood function. What is the function to maximize?
- Estimate β_0 and β_1 using the maximum likelihood method
- Calculate the standard error of both estimates.