Chi-squared test

April 25, 2019

Applied STAtistics group at AAU

Department of Mathematical Sciences

Aalborg University

AALBORG UNIVERSITY

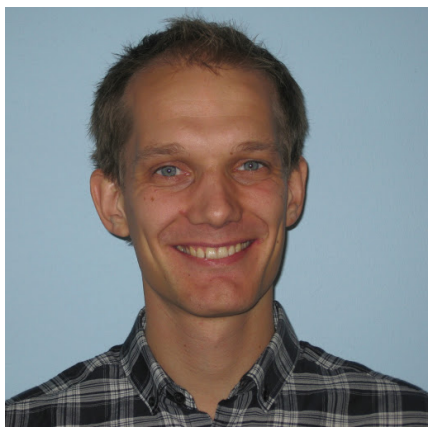
DENMARK

Introduction

Outline of session:

- Calculation of test statistic
- χ^2-distribution

Lecturer for this session is Ege Rubak, Dept. of Math. Sciences, AAU



Chi-squared (χ^2) test statistic

- We have an **observed table** and an **expected table** if H_0 is true:

	Grades	Popular	Sports
Rural	57	50	42
Suburban	87	42	22
Urban	103	49	26

	Grades	Popular	Sports
Rural	77.0	44.0	28.1
Suburban	78.0	44.5	28.4
Urban	92.0	52.5	33.5

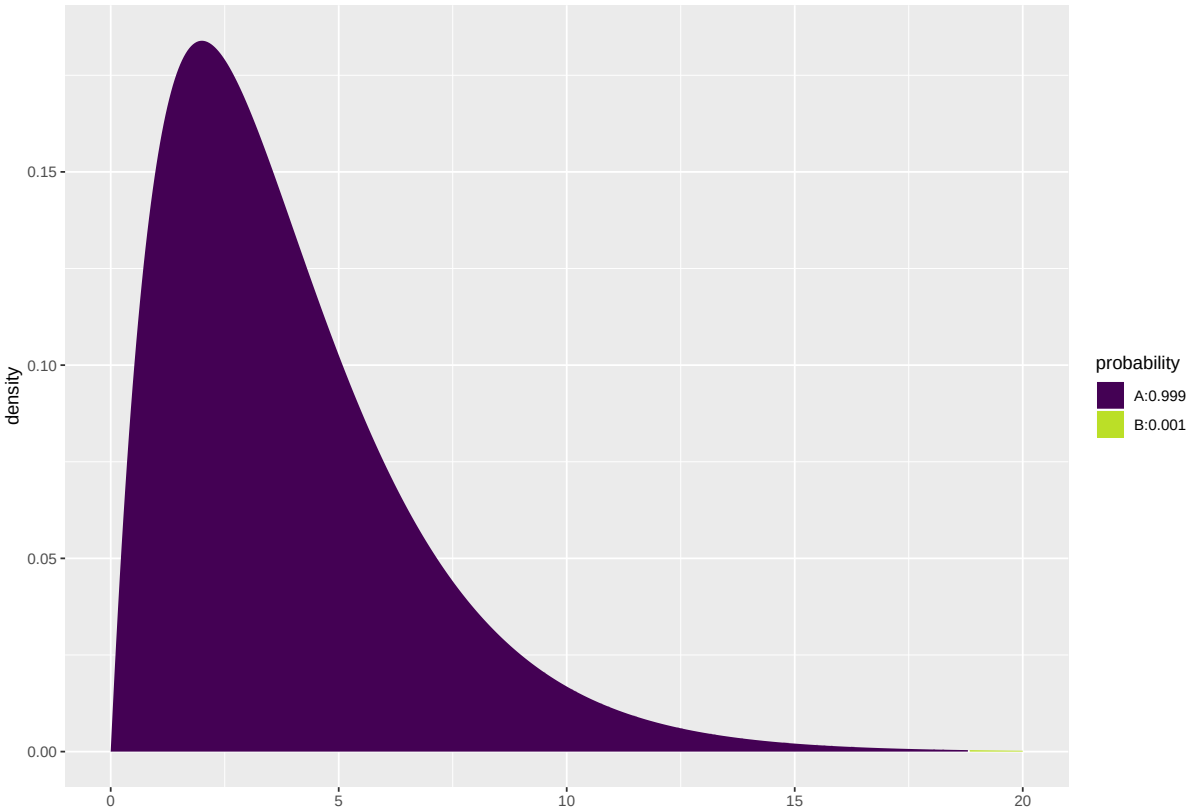
- If these tables are "far from each other", then we reject H_0. We want to measure the distance via the Chi-squared test statistic:
  - $X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$: Sum over all cells in the table
  - f_o is the frequency in a cell in the observed table
  - f_e is the corresponding frequency in the expected table.
- We have:

$$X^2_{obs} = \frac{(57-77)^2}{77} + \ldots + \frac{(26-33.5)^2}{33.5} = 18.8$$

- Is this a large distance??

χ^2-test template.

- To test the H_0 of independence in a table with r rows and c columns:
  - For a sample we calculate the observed value of the test statistic: X^2_{obs}.
  - p-value: What is the prob. of getting a larger X^2 than X^2_{obs}, if H_0 is true?
- This can be approximated by the χ^2-**distribution** with $df = (r-1)(c-1)$.
- For `Goals` and `Urban.Rural` we have $r = c = 3$, i.e. $df = 4$ and $X^2_{obs} = 18.8$, so the p-value is 0.001:



The χ^2-distribution

- The χ^2-distribution with df degrees of freedom:
  - Is never negative. And $X^2 = 0$ only happens if $f_e = f_o$.
  - Has mean $\mu = df$
  - Has standard deviation $\sigma = \sqrt{2df}$
  - Is skewed to the right, but approaches a normal distribution when df grows.