Statistical Inference 2 Summer Term 2025

Lecture: Prof. Dr. Göran Kauermann
Exercises: Sergio Buttazzo, Jan Anders
Tutorial: Eugen Gorich

Repetition: Statistical Inference 1

Exercise 0.1 - MLE

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a sample of n independent identically distributed random variables, following a negative binomial distribution $X_i \sim NB(r, p)$. The Negative Binomial distribution is a discrete probability distribution that counts the number of failures observed in a sequence of independent Bernoulli trials, until r successes are observed.
The probability function for a negative binomial variable is:

$$f(x \mid r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x \ , \qquad x \in \mathbb{N}, r \in \mathbb{N}, 0 < p < 1.$$

(a) Write down the log-likelihood of the sample.

(b) Assume that the r parameter is known and show that the maximum likelihood estimator (MLE) for p is $\hat{p} = \frac{r}{r+\bar{x}}$, where $\bar{x}$ is the sample mean. Show that it is indeed a maximum.

(c) Show that $\frac{1}{\hat{p}}$ is an unbiased estimator for $\frac{1}{p}$.

(Hint: the expected value for $X \sim NB(r,p)$ is $\mathbb{E}[X] = \frac{r(1-p)}{p}$.)

(e) Write down the asymptotic distribution of the MLE, and construct an asymptotic 95% confidence interval for it.

(f) Define the concept of a sufficient statistic. Still assuming that r is known, show that the sample sum $\sum_{i=1}^{n} X_i$ is a sufficient statistic for the parameter p.

Exercise 0.2 - Testing

(a) Consider the type I error and type II error in the context of a defendant in court. How should the hypotheses “guilty“ and “innocent“ be assigned to H_0 and H_1? What do P(Type I error), P(Type II error) and the power mean in this context?

(b) We will now assume a test setting for μ at $\alpha = 0.1$, where $H_0 : \mu \leq -0.5$ vs. $H_1 : \mu > -0.5$, with the collected data with $n = 25$ following this distribution: $X \sim N(0.5, 3^2)$.

Replicate this setting in the following web app and explain it’s function:
https://istats.shinyapps.io/power/

How do P(Type I error), P(type II error) and the power develop when you change the sample size, the true mean and the significance level? Explain the behaviour.

(c) Assume the distribution of each individual X_i in a sample X is $X_i \sim N(0, 0.8^2)$ under H_1 and $X_i \sim N(-0.5, 0.8^2)$ under H_0. H_1 holds with a probability of $P(H_1) = 0.2$. Given the sample size, hypotheses and α-level from b), we want to conduct a Gauss-test on the sample. What is the sensitivity, specificity and accuracy of this test?

Exercise 0.3 - True or False

Are the following statements true or false? Motivate and discuss your answers in your own words. If necessary, add some context to your thoughts.

(a) A sufficient statistic contains all the information in the sample about the parameter being estimated, and no other data/statistic can provide additional information.

(b) The Fisher information can be seen as a measure of the amount of information carried by the MLE about the real parameter value, since it is positively correlated with the asymptotic variance of the MLE.

(c) If the 95% confidence interval for a population parameter includes zero, we can conclude that the parameter is not significantly different from zero.

(d) In a statistical test, the p-value is the probability that the null hypothesis is true.

(e) A model with a lower AIC is usually simpler than a model with a higher AIC.

(f) If an interaction effect is included in a linear model, the model is not linear anymore.