4 Hypothesis testing

4.1 Simple hypotheses

A computer tries to distinguish between two sources of signals. Both sources emit independent signals with normally distributed intensity, the signals of the first source are distributed as $N(0,1)$, the second source has the same variance, but a higher mean - $N(2,1)$. The computer has to decide after 10 signals.

- The computer is deciding between to hypotheses
  H_1*: Signal comes from source 1* in H_2*: Signal comes from source 2.*

  Write the test statistic that shall be used by the computer (Try being a bit more general - denote the variance of both sources by σ^2, let the mean intensity of the second source be denoted by a, $a > 0$, and use n for the sample size)
  Hint: Use densities to decide what is more likely to happen

- Say we wish the computer reacts only if it is very confident that the signal does not come from source 1. We proclaim the hypothesis H_1 as the null hypothesis and see the hypothesis H_2 as the alternative. We set the decision rule so that the probability of a wrong decision if the null hypothesis holds is at most $\alpha = 0.05$.

  - The test statistic is a random variable (denote it by Y). What can we say about its distribution under the null hypothesis?
    Hint: Use the logarithm to make things simpler
  - Set the critical value at which the computer should react.

- What is the probability that the computer reacts if the signal indeed comes from source 2? (This probability is referred to as the power of the test)
  Hint: What is the distribution of the test statistic under the alternative hypothesis?

- Transform the test statistic Y so that it shall be a standard normal variable under the null hypothesis.

- Summarize: does the critical value c depend on a? Try explaining intuitively. Is the value of a at all relevant?

Understanding the ideas with R:

- Generate data in two steps. First choose between the values of a with equal probability (a equals 0 or 2), then generate 10 values from the distribution $N(a, 1)$. Calculate the value of the test statistic from the first item and choose a hypothesis depending on whether the value is above or below 1. Repeat the procedure many times and calculate the proportion of cases when you decide the for each of the hypotheses and the proportion of cases in which this decision is correct.

- Replace the probability in the first step (let $a = 2$ be the more probable value). How do the proportion from the previous item change?

- Generate data, so that a constantly equals 0 and check that the value of α at the calculated value of c indeed equals 0.05.

- Generate the data so that $a = 2$ and check the power of the test.

4.2 Simple hypotheses, a generalization

We repeat the previous exercise in a more general form (this is an example of exam assignment of prof. Perman).
Assume that the observed values are i.i.d. random variables $X_1, X_2, ..., X_n$. Assume we only have two options: the density equals $f(x)$ or $g(x)$, where $f(x)$ and $g(x)$ are known positive densities. We formally set:

$$H_0 : \textit{density equals} f(x) \text{ against } H_1 : \textit{density equals } g(x).$$

- Propose a test statistic to test the null hypothesis if your given values equal $x_1, ..., x_n$.

- When will you reject at a given level of significance α? Express the approximate critical value with quantities

$$a = \int \log\left(\frac{g(x)}{f(x)}\right) f(x)dx \text{ and } b = \int \log\left(\frac{g(x)}{f(x)}\right)^2 f(x)dx$$

4.3 The Neyman-Pearson paradigm

We wish to check whether a coin is fair. We've performed an experiment by tossing the coin 10 times. We get 7 heads.

- Write the null hypothesis for your example. Is the null hypothesis simple of composite? Write the test statistic, denote it by X - what is its distribution under the null hypothesis?
- Say your alternative hypothesis is $H_A : p > 0.5$. Is this a simple or composite hypothesis? What kind of values speak in favour of the alternative hypothesis? Is the alternative hypothesis one-sided or two-sided?
- In our case, $X = 7$. What is the probability of seing this event on our sample under the null hypothesis?
- Say that the rejection region consists of the value $\{10\}$. What is the level of significance α in this case? What is the level of significance if the rejection region equals $\{6, 7, 8, 9, 10\}$?
- Determine the rejection region so that $\alpha = 0.05$. Can you reject the null hypothesis at this level?
- What is the power of the test for this α, if we assume that the true value of parameter equals $p = 0.6$ or $p = 0.7$? What is the value of the type II error?
- Assume that your alternative hypothesis equals $H_A : p \neq 0.5$. Is this a simple or a composite hypothesis? Is it one or two-sided?
- What is the rejection region if $\alpha \leq 0.05$? What is the exact significance level for this rejection region?
- Calculate the power of the test in this example.

Understanding the ideas with R:

- Repeat an experiment with 10 coins 1000 times. Check the probabilities of chosen rejection regions.
- Change the probability of heads and check the power of the test.

p\k	0	1	2	3	4	5	6	7	8	9	10
0.5	0.001	0.011	0.055	0.172	0.377	0.623	0.828	0.945	0.989	0.999	1
0.6	0.000	0.002	0.012	0.055	0.166	0.367	0.618	0.833	0.954	0.994	1
0.7	0.000	0.000	0.002	0.011	0.047	0.150	0.350	0.617	0.851	0.972	1

Table 1: The cumulative probabilities for the binomial distributions with $n = 10$ ($P(X \leq k|p)$)

- Increase the sample size (for example take 20 coins) and check how the power changes.

4.4 Test power

The available literature claims that the mean haemoglobin value of an athlete that stays at least 14 days at the altitude above 1500m increases for 2 g/l, but no change in variance is expected. At normal altitudes, the values are approximately normally distributed $X \sim N(\mu_1, 5^2)$, where μ_1 is the athlete's individual mean.
An athlete often trains on altitude, but in shorter time intervals. He wishes to know whether his mean haemoglobin values nevertheless increases. He made 12 (independent) measurements during the season, 8 of these were during the altitude training and 4 otherwise. The goal of this exercise is to determine the power of his test for significance level $\alpha = 0.05$.

- What is the null and what the alternative hypothesis?
- Propose a test statistic. Calculate its distribution under the null hypothesis.
- Calculate the power of the test, i.e. the probability that he can reject the null hypothesis if his mean indeed increases for 2 g/l?
- How would the power change if he had an equal number of measurements in each of the phases?
- How does the power of the test depend on the variance of individual's measurements? How does it depend on the true differences in the population?

4.5 Generalized likelihood ratio test

An assumption that is used in the Athlete Biological Passport is that haemoglobin varies equally in all athletes. We wish to test this assumption on a sample of k athletes. Let the values of i-th athlete be normally distributed ($i = 1, \dots, k$), i.e. $X_{ij} \sim N(\mu_i, \sigma_i^2)$, where $j = 1, \dots, n_i$ denote the individual's measurements. Assume that all measurements are independent.

- Write the null and alternative hypothesis
- Consider first the case of only one athlete with n measurements. How would we estimate his parameters μ and σ^2 with the method of maximum likelihood?
- Show that under the alternative hypothesis (for k athletes), the parameter estimates equal

$$\begin{aligned} \hat{\mu}_i &= \frac{1}{n_i} \sum x_{ij} \\ \hat{\sigma}_i^2 &= \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2 \end{aligned}$$

- How do we estimate means under the null hypothesis?
- What is the variance estimate under the null hypothesis?
- How would you test the null hypothesis with the generalized likelihood ratio test?