

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 or composite? Write the test statistic, denote it by X - what is its distribution under the null hypothesis?

$H_0 : p = 0.5$. This determines the distribution of the random variable under the null hypothesis and the hypothesis is thus simple. The test statistic X is simply the number of heads: the distribution of the null test statistic under the null hypothesis equals $B(0.5, 10)$.

- 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?

The hypothesis is composite since it encompasses different values of the parameters of the distribution. We shall reject for high values of X . The alternative hypothesis is one-sided, since we are interested only in the right tail of the distribution.

- In our case, $X = 7$. What is the probability of seeing this event on our sample under the null hypothesis?

If the null hypothesis holds ($p = 0.5$), $P(X = 7) = 0.117$.

- 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 $\{6, 7, 8, 9, 10\}$?

If the null hypothesis holds ($p = 0.5$), $P(X = 10) = 0.001$.

If the null hypothesis holds ($p = 0.5$), $P(X \geq 6) = 0.377$.

- Determine the rejection region so that $\alpha = 0.05$. Can you reject the null hypothesis at this level?

The smallest value of k , for which $P(X \geq k) \leq 0.05$, equals 9. The significance level in this case equals 0.01. The null hypothesis cannot be rejected in our example, since $X < 9$.

- 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?

The power of the test is low - 0.046 or 0.149. In such a small sample and with such a small significance level, we shall only rarely be able to reject the null hypothesis. The type II error equals 1-power.

- 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?
The alternative hypothesis is still composite, it is now also two-sided.
- What is the rejection region if $\alpha \leq 0.05$? What is the exact significance level for this rejection region?
The rejection region shall consist of values $\{0, 1, 9, 10\}$. The level of significance for this region equals $\alpha = 0.02$.
- Calculate the power of the test in this example.
When calculating the power, we have to take into account that the null hypothesis shall be rejected also if X equals 0 or 1. Since the probability of the two values for $p = 0.6$ and $p = 0.7$ is low, the power changes only slightly.

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 $n = 10$ ($P(X \leq k|p)$)

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.
- 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?

Denote the distribution of the haemoglobin values in the time of altitude training as $N(\mu_2, 5^2)$. The null hypothesis is:

H_0 : The mean haemoglobin value is equal in both phases, $\mu_1 = \mu_2$.

The alternative hypothesis is $\mu_2 > \mu_1$, so the one-sided test is of interest.

- Propose a test statistic. Calculate its distribution under the null hypothesis.

The athlete shall calculate the mean difference between the samples, i.e.

$$R = \bar{X}_2 - \bar{X}_1 \sim N\left(\mu_2 - \mu_1, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

The test statistic

$$Z = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

under the null is thus distributed as $N(0, 1)$. Since only the one-sided alternative hypothesis is of interest, he shall reject the null hypothesis when $Z > z_\alpha$, i.e. $Z > 1.64$.

- 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?

We are interested in $P(Z > 1.64)$, i.e. $P(R > 1.64 \cdot \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$, in our case $P(R > 5.02)$. Under the alternative hypothesis $R \sim N\left(2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$ and therefore

$$P\left(R > 1.64 \cdot \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = P\left(\frac{\bar{X}_2 - \bar{X}_1 - 2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} > 1.64 - \frac{2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}\right)$$

$$P\left(U > 1.64 - \frac{2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right),$$

where $U \sim N(0, 1)$. In our case:

$$P\left(U > 1.64 - \frac{2}{5\sqrt{\frac{1}{8} + \frac{1}{4}}}\right) = P(U > 0.99) = 0.16 \quad (1)$$

The power of the test is very low - with such a small number of measurements the probability of rejecting the null hypothesis is very low even if his average indeed increases for 2 g/l.

- How would the power change if he had an equal number of measurements in each of the phases?

If he had 6 measurements in each phase, the power would equal

$$P\left(U > 1.64 - \frac{2}{5\sqrt{\frac{1}{6} + \frac{1}{6}}}\right) = P(U > 0.95) = 0.17$$

- 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?

As we can see from (1), a larger difference implies higher power - if the actual difference between the phases is larger, it is easier to see it on the data despite variability.

If the variance of individual measurements was smaller, the standard error would be smaller as well and the power thus higher.

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

Null hypothesis:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

Alternative hypothesis:

$$H_1 : \sigma_i^2 \text{ are not all equal}$$

- 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?

The likelihood function equals

$$L(x, \mu, \sigma) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x_j - \mu)^2}{2\sigma^2},$$

the part of its logarithm that contains the parameters to be estimated equals

$$\log L(x, \mu, \sigma) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2.$$

We find the maximum with respect to μ :

$$\begin{aligned}\frac{\partial \log L(x, \mu, \sigma)}{\partial \mu} &= 0 \\ -\frac{1}{2\hat{\sigma}^2} \sum_{j=1}^n (x_j - \hat{\mu})(-2) &= 0 \\ \sum_{j=1}^n (x_j - \hat{\mu}) &= 0 \\ \hat{\mu} &= \frac{1}{n} \sum_{j=1}^n x_j\end{aligned}$$

The variance:

$$\begin{aligned}\frac{\partial \log L(x, \mu, \sigma)}{\partial \sigma} &= 0 \\ -\frac{n}{\hat{\sigma}} - \frac{1}{2} \sum_{j=1}^n (x_j - \hat{\mu})^2 \frac{-2}{\hat{\sigma}^3} &= 0 \\ -\hat{\sigma}^2 n + \sum_{j=1}^n (x_j - \hat{\mu})^2 &= 0 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{j=1}^n (x_j - \hat{\mu})^2\end{aligned}$$

- Show that under the alternative hypothesis (for k athletes), the parameter estimates equal

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

The likelihood function under the alternative hypothesis equals

$$L(x, \mu, \sigma) = \prod_{i=1}^k \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma_i} \exp -\frac{(x_{ij} - \mu_i)^2}{2\sigma_i^2}$$

After taking the logarithm of the function, we are left with a sum of terms, each of them containing parameters of one individual only. If interested in individual i , we thus do not need any information about the other individuals and the result is equal to considering each individual separately.

- How do we estimate means under the null hypothesis?
Under the null hypothesis, σ_i equals for all i , and thus does not affect our estimate of the individual averages. The estimates of the averages are thus equal to those under the alternative hypothesis.
- What is the variance estimate under the null hypothesis?
The part of the logarithm of the likelihood function that we are interested in equals

$$\log L(x, \mu, \sigma) = - \sum_{i=1}^k n_i \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2.$$

After equalling the derivative with respect to σ to 0, we get

$$\hat{\sigma}_0^2 = \frac{1}{\sum_{i=1}^k n_i} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2$$

- How would you test the null hypothesis with the generalized likelihood ratio test?

Wilks' Λ equals (the maximum of the likelihood function under the alternative hypothesis in the numerator, the maximum under the null

in the denominator):

$$\begin{aligned}
\Lambda &= \frac{\prod_{i=1}^k \prod_{j=1}^{n_i} \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_i}} \exp\left\{-\frac{(x_{ij}-\hat{\mu}_i)^2}{2\hat{\sigma}_i^2}\right\}\right)}{\prod_{i=1}^k \prod_{j=1}^{n_i} \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_0}} \exp\left\{-\frac{(x_{ij}-\hat{\mu}_{0i})^2}{2\hat{\sigma}_0^2}\right\}\right)} \\
&= \frac{\left(\prod_{i=1}^k \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\hat{\sigma}_i}} \right) \prod_{i=1}^k \exp\left\{-\frac{\sum_{j=1}^{n_i} (x_{ij}-\hat{\mu}_i)^2}{2\hat{\sigma}_i^2}\right\}}{\left(\prod_{i=1}^k \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\hat{\sigma}_0}} \right) \exp\left\{-\sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(x_{ij}-\hat{\mu}_{0i})^2}{2\hat{\sigma}_0^2}\right\}}
\end{aligned}$$

We insert the estimates of the variance in the exponent and get $\exp\{-\frac{1}{2} \sum_{i=1}^k n_i\}$ in the numerator as well as the denominator. The two terms cancel out, leaving the logarithm of Λ equal to

$$\begin{aligned}
\log \Lambda &= - \left(\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\hat{\sigma}_i) \right) + \left(\sum_{i=1}^k \sum_{j=1}^{n_i} \log(\hat{\sigma}_0) \right) \\
&= \left(\sum_{i=1}^k n_i \log(\hat{\sigma}_0) \right) - \left(\sum_{i=1}^k n_i \log(\hat{\sigma}_i) \right) \\
&= \sum_{i=1}^k n_i [\log(\hat{\sigma}_0) - \log(\hat{\sigma}_i)]
\end{aligned}$$

The $2 \log \Lambda$ is distributed as χ_{k-1}^2 , since we estimate $k-1$ more parameters under the alternative hypothesis.