Statistical inference (MVE155/MSG200)

Hypothesis testing

Scientific method

- Observation/question, for example "Are children in Sweden taller today than 50 years ago?"
- Collect information: What is known about the topic? Has somebody else posed the same question?
- Construct a hypothesis: I believe that children in Sweden today are taller than children in Sweden 50 years ago.

  $\rightarrow$ Researcher's hypothesis (alternative hypothesis)

  The null hypothesis is that there is no difference between the heights today and 50 years ago.

Scientific method (continues)

- Make experiments: Collect data from, say, 5 and 55 year olds on their height at the age of 5.
- Analyze the data: How? Which method? Here, it could be a two sample t-test.
- Report the results: Did you reject the null hypothesis? What are the concequences if you did/did not?
- New observations/questions.

Set-up: Two hypotheses

H_0: The effect of interest is zero
(e.g. there is no difference between the average heights)

H_1: The effect of interest is not zero
(e.g. the average height of 5 year olds today is larger than the average height of 5 year olds 50 years ago)

To decide whether we can reject the null hypothesis (and believe that our own is true), we collect data, $x_1, ..., x_n$, and from that, compute a test statistic $t = t(x_1, ..., x_n)$, which is chosen so that the distribution of the corresponding random variable T (stochastic version of t) is known.

Rejection region

Find a rejection region $\mathcal{R}$ so that the null hypothesis H_0 is rejected if and only if the test statistic $t \in \mathcal{R}$. Then,

	Do not reject H_0	Reject H_0 in favour to H_1
H_0 true	True negative	False positive (type I error)
H_1 true	False negative (type II error)	True positive

Some important conditional probabilities

- Significance level α

$$\alpha = \mathrm{P}(T \in \mathcal{R} | H_0)$$

which is also the probability of the type I error.

- Specificity of the test

$$1 - \alpha = 1 - \mathrm{P}(T \in \mathcal{R} | H_0) = \mathrm{P}(T \notin \mathcal{R} | H_0)$$

- Probability of the type II error

$$\beta = \mathrm{P}(T \notin \mathcal{R} | H_1)$$

- Power or sensitivity of the test

$$1 - \beta = 1 - \mathrm{P}(T \notin \mathcal{R} | H_1) = P(T \in \mathcal{R} | H_1)$$

How to choose α and β?

We would like both α and β be small and specificity and sensitivity (power) large. However, when we decrease α, β increases.
$\rightarrow$ Have a large sample size.

We fix α, typically 5%, and choose $\mathcal{R}$ based on

$$\alpha = \mathrm{P}(T \in \mathcal{R} | H_0).$$

Then, compute

$$\beta = \mathrm{P}(T \notin \mathcal{R} | H_1)$$

and the power of the test (given n).

p-value

Let us have a test statistic T and its observed value (computed from the data) t_{obs} and we test the null hypothesis H_0. Then, the p-value is the probability that the test statistic takes the observed value t_{obs} or even a more extreme value when H_0 is true.

If the p-value is small (say, less than 0.05), the null hypothesis can be rejected.

Example

Your friend claims that they have exceptional skills and can guess the suit of a card very well. You show 100 cards to your friend and they guess correctly the suit of 30 of the cards. Does this indicate that they have exceptional skills?

We can use

- binomial test with the number of right guesses $C \sim Bin(100, 0.25)$ under H_0 of pure guessing
- large sample test based on normal approximation

Large sample test for proportion

We would like to test $H_0 : p = p_0$, where p is the proportion of successes (1's) in n independent Bernoulli trials.

We have a random sample of size n from the Bernoulli distribution $\text{Bin}(1, p)$ with unknown p. The number of successes C has the distribution $\text{Bin}(n, p)$ and p can be estimated by the sample proportion $\hat{p} = c/n$, which is the ML estimate for p.

Given $H_0 : p = p_0$, if n is large enough, we can use the test statistic

$$Z = \frac{C - np_0}{\sqrt{np_0(1 - p_0)}} = \frac{n\hat{p} - np_0}{n\sqrt{p_0(1 - p_0)/n}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

which is approximately $N(0, 1)$-distributed.

Large sample test for proportion: rejection region

The rejection region depends on H_1 and α:

H_1	Rejection region
$p > p_0$	$\mathcal{R} = \{Z \geq z(\alpha)\}$
$p < p_0$	$\mathcal{R} = \{Z \leq -z(\alpha)\}$
$p \neq p_0$	$\mathcal{R} = \{Z \leq -z(\alpha/2)\} \cup \{Z \geq z(\alpha/2)\}$

where $z(\alpha)$ can be found using $\Phi(z(\alpha)) = 1 - \alpha$.

To test $H_0 : p = p_0$ against $H_1 : p \neq p_0$, the same conclusion can be drawn based on a $100(1-\alpha)\%$ confidence interval for p: If the confidence interval does not cover p_0, H_0 can be rejected.

For large samples, we can test $H_0 : \mu = \mu_0$ by using the test statistic

$$T_0 = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$$

which is approximately $N(0, 1)$-distributed ($s_{\bar{x}}$ is the sample standard deviation of $\bar{x}$) since

- $\bar{X}$ is approximately normally distributed
- sample sample deviation S is a consistent estimator for the population standard deviation σ, σ can be replaced by $s_{\bar{x}}$.

For small samples, we have to assume that $X_1, ..., X_n \sim N(\mu, \sigma)$ and use the test statistic (one sample t-test)

$$T_0 \sim t_{n-1}.$$

To test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, the same conclusion can be drawn based on a $100(1 - \alpha)\%$ confidence interval for μ: If the confidence interval does not cover μ_0, H_0 can be rejected.

Likelihood ratio test: simple hypotheses

We can test simple hypotheses

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1$$

by using the ratio of the two likelihood functions

$$\frac{L(\theta_0)}{L(\theta_1)}$$

as the test statistic.

If the ratio is large, θ_0 is more likely and if it is small, θ_1 is more likely given the observed data.

Likelihood ratio test: nested hypotheses

The hypotheses do not have to be simple: We can also test

$$H_0 : \theta \in \Omega_0 \text{ against } H_1 : \theta \notin \Omega_0 \ (\theta \in \Omega \setminus \Omega_0),$$

where Ω is a parameter set and $\Omega_0 \subset \Omega$. Often, we use nested hypotheses

$$H_0 : \theta \in \Omega_0 \text{ against } H_1 : \theta \in \Omega.$$

Likelihood ratio test: nested hypotheses

We find two ML estimates

- $\hat{\theta}_0$ which maximizes $L(\theta)$ when $\theta \in \Omega_0$
- $\hat{\theta}$ which maximizes $L(\theta)$ when $\theta \in \Omega$

and compute the likelihood ratio

$$w = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}.$$

The null hypothesis is rejected if w is small or

$$-ln(w) = ln(L(\hat{\theta})) - ln(L(\hat{\theta}_0)) = l(\hat{\theta}) - l(\hat{\theta}_0)$$

is large.

Note that $0 < w \leq 1$ and therefore, $-ln(w) \geq 0$.

Likelihood ratio test: nested hypotheses

It can be shown that under H_0,

$$-2ln(W) \approx \chi^2_{df},$$

where $df = \dim(\Omega) - \dim(\Omega_0)$ and dim refers to the number of free parameters.

Example: Let us have a random sample from a $N(\mu, \sigma)$, where σ known. We test

$$H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0.$$

Then, under H_0, there are no free parameters and under H_1, there is one. Therefore, $df = \dim(\Omega) - \dim(\Omega_0) = 1 - 0 = 1$.

χ^2 goodness-of-fit test

We have a random sample of size n, where each observation belongs to one of the J classes with probabilities $(p_1, ..., p_J)$. The joint distribution of the corresponding counts is multinomial, $(C_1, ..., C_J) \sim Mn(n; p_1, ..., p_J)$, with

$$P(C_1 = c_1, ..., C_J = c_J) = \frac{n!}{c_1! \cdots c_J!} p_1^{c_1} \cdots p_J^{c_J}.$$

The general parameter space is

$$\Omega = \{(p_1, .., p_J) : \ p_1 + ... + p_J = 1, \ p_i \geq 0 \text{ for each } i\}$$

which has the dimension $\dim(\Omega) = J - 1$.

χ^2 goodness-of-fit test

Test

$$H_0 : (p_1, ..., p_J) \in \Omega_0 \text{ against } H_1 : (p_1, ..., p_J) \in \Omega,$$

where

$$\Omega_0 = \{(p_1, .., p_J) \in \Omega : \ (p_1, ..., p_J) = (p_1(\theta), ..., p_J(\theta))\},$$

which has the dimension $\dim(\Omega_1) = 1$ if the parameter θ needs to be estimated and $\dim(\Omega_0) = 0$ if θ is fixed, i.e (p_i's are given constants.