1 Hypothesis testing

1.1 Motivating example with definitions

Example 1 Suppose that a friend says that she can predict the outcome of a flip of a coin. How can you test if this is true?

Toss a coin n times and count the number of times t that your friend has predicted the right outcome.

The **Test statistic** t is an outcome of $T \in Bin(n, p)$. If your friend can NOT predict the outcome of a coin toss then $p = 1/2$.

Null hypothesis (this is a statement about a parameter)

$$H_0 : \quad p = \frac{1}{2}$$

Alternative hypothesis

$$H_1 : \quad p > \frac{1}{2}$$

It would seem reasonable to reject H_0 if t is "big" (your friend has many correct answers), say $t \geq c$.

Test: Reject H_0 if

$$t \in C = \{c, c+1, \ldots, n\} = \textbf{critical region}$$

How should you choose c? You want the **significance level** or the error level α to be small, i.e.

$$P(\text{reject } H_0 \text{ if } H_0 \text{ true}) \leq \alpha = \text{significance level}$$

You should therefore choose c such that

$$P(T \geq c \text{ if } H_0 \text{ true}) \leq \alpha.$$

If n=15 then (use Table 6 in the compiled formulae or a calculator)

$$\begin{aligned}
P(T \geq 10) &= 1 - P(T < 10) = 1 - P(T \leq 9) = 1 - 0.84912 = 0.15088 \\
P(T \geq 11) &= 1 - P(T \leq 10) = 1 - 0.94077 = 0.05923 \\
P(T \geq 12) &= 1 - P(T \leq 11) = 1 - 0.98242 = 0.01758 \\
P(T \geq 13) &= 1 - P(T \leq 12) = 1 - 0.99631 = 0.00369 \\
P(T \geq 14) &= 1 - P(T \leq 13) = 1 - 0.99951 = 0.05923 \\
P(T \geq 15) &= 1 - P(T \leq 14) = 1 - 0.99997 = 0.00003
\end{aligned}$$

In order for the significance level to be 5% you should choose $c = 12$ (which results in $C = \{12, 13, 14, 15\}$).

If $t = 13$ then $P(T \geq 13) = 0.00369$ is called the **P-value** or the **observed significance level** (the probability to have the observation you have actually gotten or an even more extreme observation when H_0 is true). If $t = 13$ one would reject H_0 at the significance level 5% (but also at the significance level 1%). To compute the level of significance you can compute

$$P(T \geq t \text{ if } H_0 \text{ true}) = \sum_{i=t}^{15} \binom{15}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{15-i} = \sum_{i=t}^{15} \binom{15}{i} \left(\frac{1}{2}\right)^{15}$$

Then you compare this to the chosen significance level and if the P-value $\leq \alpha$ then H_0 is rejected, otherwise not. This is called the **P-value method**. □

1.2 General case

State a **null hypothesis**

$$H_0 : \quad \theta \in \Theta_0$$

and an **alternative hypothesis**

$$H_1 : \quad \theta \in \Theta_1$$

Using your observed sample $x_1, \ldots, x_n$ you form a test statistic

$$t = t(x_1, \ldots, x_n)$$

which is an outcome of a random variable $T = t(X_1, \ldots, X_n)$. The distribution of T should be known if H_0 is assumed to be true. You should now choose a critical (or rejection) region C such that

$$P(T \in C) = \alpha \quad \text{if } H_0 \text{ true}$$

(if the distribution of T is discrete then you will have to settle for $\lesssim \alpha$)
Decision rule: Reject H_0 if an outcome $t \in C$ is observed. Thus:

		H_0 true	H_0 false
$t \notin C$	Do not reject H_0	OK	Type 2 error
$t \in C$	Reject H_0	Type 1 error	OK

We have that

$$P(\text{Type 1 error}) = P(T \in C) \leq \alpha \quad \text{if } H_0 \text{ true}$$

So far we have not looked at type 2 errors.

Remark 1 The most powerful thing you can do with a hypothesis test is to reject H_0 (the data does not support the null hypothesis). In general you will not have proved H_1 to be true when you reject H_0. What you have best control over are type 1 errors, i.e. to reject H_0 although H_0 is true. If you can it is therefore usually good to choose your hypotheses so that you have control over the most "severe" type of error. If you choose

$$H_0 : \quad \text{innocent} \qquad \text{and} \qquad H_1 : \quad \text{guilty}$$

then the type 1 error becomes to sentence an innocent", which is not desirable, whereas if you choose

$$H_0 : \quad \text{guilty} \qquad \text{and} \qquad H_1 : \quad \text{innocent}$$

then the type 1 error becomes to äcquit a guilty", which is often considered to be less severe. Note that "innocentänd guiltyåould have to be related to a parameter, like an expected value or so.

The function

$$h(\theta) = P(\text{reject } H_0) = P(T \in C) \quad \text{if } \theta \text{ is the correct parameter value}$$

is called the **power (function)** of the test. Assume that

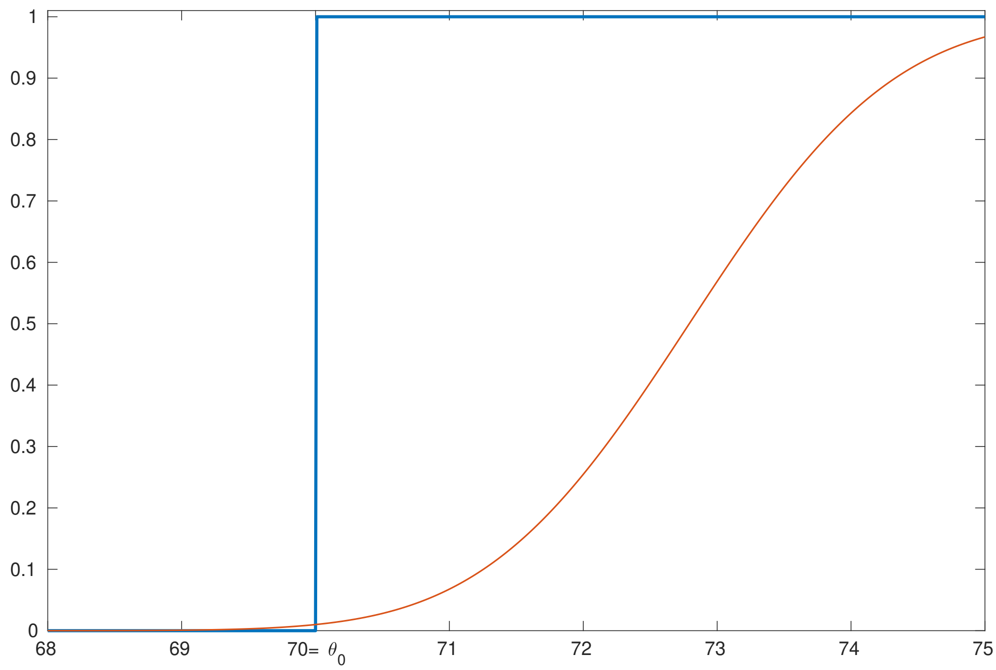
$$H_0: \quad \theta \leq \theta_0$$

$$H_1: \quad \theta > \theta_0$$

then ideally the power function would be the following indicator function

$$I\{\theta > \theta_0\} = \begin{cases} 1 & \text{if } \theta > \theta_0 \\ 0 & \text{if } \theta \leq \theta_0 \end{cases}$$

In general this will not be the case, we already know for instance that the level of significance is an upper bound of the probability to reject H_0 although H_0 is true.



Figur 1: The ideal power function of a test and the actual power function of the test in Example 2.

Example 2 This example is about speeding. Let $x = 72.7$ km/h be an observation of X where

$$X = \text{speed} + \text{measurement error} = v + Z$$

Here $Z \in N(0, \sigma)$ denotes the measurement error with a known standard deviation of $\sigma = 1.2$ km/h. Therefore $X \in N(v, \sigma)$. We would now like to test the null hypothesis

$$H_0 : v \leq 70 = v_0 \qquad \text{innocent of speeding}$$

(from a statistical point of view you can think of $v = 70 = v_0$ just as well) against the alternative

$$H_1 : v > v_0 \qquad \text{guilty of speeding}$$

at the significance level $\alpha = 0.01$. We will reject H_0 for large values of

$$t = \frac{x - v_0}{\sigma}$$

If H_0 is true, i.e. $X \in N(v_0, \sigma)$ then t is an observation of $T \in N(0,1)$ and in order for the significance level to become 1% we should choose to reject H_0 if we observe a value of t greater than $\lambda_{0.01} = 2.3263$. Here we have that

$$t = \frac{72.7 - 70}{1.2} = 2.25 < 2.3263$$

so in this case we can NOT reject H_0, i.e. the data does not contradict that the driver kept the speed limit (at the chosen level of significance). $\square$

Remark 2 The reason that you can just as well think of $H_0 : v = v_0$ is that for composite hypotheses which include several parameter values (like $H_0 : v \leq v_0$) we have that

$$\alpha = \max_{\theta \in H_0} h(\theta)$$

and in the example above the maximum is obtained for v_0.

1.3 Relation to confidence intervals

The decision rule in Example 2 can be written as: do not reject H_0 if

$$v_0 \geq x - \lambda_\alpha \cdot \sigma$$

or

$$v_0 \in I_v = (x - \lambda_\alpha \cdot \sigma, \infty)$$

where the interval on the right hand side is a one-sided confidence interval for v with confidence level $1 - \alpha$. Alternatively the decision rule can be written as: reject H_0 if

$$v_0 \notin I_v = (x - \lambda_\alpha \cdot \sigma, \infty)$$

In general, if you assume that x_i is an observation of $X_i \in N(\mu, \sigma)$, $i = 1, \ldots, n$, where σ is known then you have the following table over the null hypothesis, the alternative hypothesis, and the decision rule.

Hypotheses H_0	Hypotheses H_1	Decision rule: Reject H_0 if
$\mu = \mu_0$	$\mu \neq \mu_0$	$\mu_0 \notin I_\mu = \left(\bar{x} - \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\mu_0 \notin I_\mu = \left(\infty, \bar{x} + \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\mu_0 \notin I_\mu = \left(\bar{x} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty\right)$

The above will result in a test at the significance level α (that is with an error level of α).

Remark 3 In the table above it is stated what you should do if you have observations of $X_i \in N(\mu, \sigma)$ with a known σ, but should you find yourself in a position where σ is unknown you should construct a confidence interval using the t-method and the same decision rules will apply. Should your observations come from an approximately normal distribution then you construct a confidence interval using the approximate method and then your test will have an approximate significance level of α.

Example 3 Do problem 1412 from [2]. $\square$

Example 4 The power function of the test in Example 2 can be computed in the following way

$$\begin{aligned} h(v) &= P((H_0 \text{ is rejected when } v \text{ is the correct parameter value}) \\ &= P(v_0 \notin I_v(X)) \\ &= P(v_0 < X - \lambda_\alpha \cdot \sigma) = \{X \in N(v, \sigma)\} \\ &= P\left(\frac{X - v}{\sigma} > \frac{v_0 - v}{\sigma} + \lambda_\alpha\right) = 1 - P\left(\frac{X - v}{\sigma} \leq \frac{v_0 - v}{\sigma} + \lambda_\alpha\right) \\ &= 1 - \phi\left(\frac{v_0 - v}{\sigma} + \lambda_\alpha\right) \end{aligned}$$

This function has been plotted in Figure 1. $\square$

Referenser

[1] Blom, G., Enger, J., Englund, G., Grandell, J., och Holst, L., (2005). Sannolikhetsteori och statistikteori med tillämpningar.

[2] Blom, Gunnar, (1989). Probability and Statistics. Theory and Applications.