CentraleSupélec

Statistics and Learning

Lecturers (alphabetic order):

Julien Bect, Gilles Faÿ, Ziad Kobeissi, Laurent Le Brusquet,
Vincent Lescarret, Arshak Minasyan, Arthur Tenenhaus† & Xujia Zhu

† Course coordinator

Lecture 4/9

Hypothesis testing

Course objectives

- make (binary) decisions through hypothesis testing,
- choose and construct a test,
- define and compute risks of error of the first and second kind.

Lecture outline

Lecture outline

Lecture outline

Example: component reliability

Reminder: $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, $\theta > 0$.

Problem

The manufacturer considers offering a one-year warranty...

➠ is it a good idea ?

Formalization

The manufacturer considers that it is a "good idea" if:

$$\text{the return rate is lower than } 10\%$$
$$\Updownarrow$$
$$\mathbb{P}_\theta(X_1 \le 1) = 1 - \exp(-\theta) < 0.1$$
$$\Updownarrow$$
$$\theta < \theta_0 = -\ln(0.9)$$

Example: component reliability

Therefore, the manufacturer wants to know if $\theta < \theta_0$ or $\theta \ge \theta_0$.

➠ hypothesis to be tested: $H_0 : \theta \ge \theta_0$
(component quality is not sufficient)

Making (binary) decisions from data

We want to evaluate the "compatibility" between H_0 and $\underline{x}$:

- if a strong incompatiblity is detected,
  ➠ H_0 is rejected (and the warranty proposed);
- otherwise, H_0 is accepted.

Note the asymmetry between the two scenarios
(H_0 = is retained by default)

Hypothesis tests make it possible to formalize this decision making.

Another example / construction of a first test

Goal: test the mean parameter of a Gaussian distribution.

- $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \sigma_0^2)$ (σ_0 known; $n = 10$, $\sigma_0 = 2.5$)
- hypothesis to be tested $\rightarrow H_0 : \theta = \theta_0$ (fixed),
- alternative hypothesis $\rightarrow H_1 : \theta = \theta_1$ (fixed, and $\theta_0 < \theta_1$).

Approach. Making a decision about H_0 means estimating if it is

- either true ➠ $\delta = 0$,
- or false ➠ $\delta = 1$.

Constraint. We want δ to be such that, if $\theta = \theta_0$ (H_0 true),

$$\mathbb{P}_{\theta_0}(\delta = 1) = 5\% \; (= \alpha).$$

Intuitive construction of a test: $\delta = \mathbb{1}_{\bar{X} > t}$

- where t is chosen such that $\mathbb{P}_{\theta_0}(\delta = 1) = \mathbb{P}_{\theta_0}(\bar{X} > t) = 5\%$.

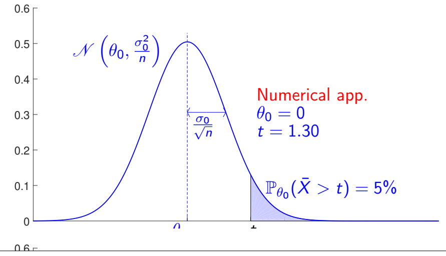
If H_0 is true ($\theta = \theta_0$)H_1 is true ($\theta = \theta_1$): $\bar{X} \sim \mathcal{N}\left(\theta_0\theta_1, \frac{\sigma_0^2}{n}\right)$,
therefore

$$t \;=\; \theta_0 + q_{0.95}\,\frac{\sigma_0}{\sqrt{n}}$$

where q_r is the $\mathcal{N}(0,1)$ quantile of order r.

$$\mathbb{P}_{\theta_1}(\delta = 0) = \mathbb{P}_{\theta_1}(\bar{X} \leq t) = \Phi\left(\frac{t-\theta_1}{\sigma_0/\sqrt{n}}\right)$$

where Φ is the cdf of the $\mathcal{N}(0,1)$ distribution.



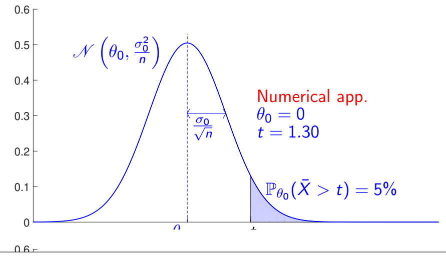
If H_0 is true ($\theta = \theta_0$)H_1 is true ($\theta = \theta_1$): $\bar{X} \sim \mathcal{N}\left(\theta_0\theta_1, \frac{\sigma_0^2}{n}\right)$,
therefore

$$t \;=\; \theta_0 + q_{0.95}\,\frac{\sigma_0}{\sqrt{n}}$$

where q_r is the $\mathcal{N}(0,1)$ quantile of order r.

$$\mathbb{P}_{\theta_1}(\delta = 0) = \mathbb{P}_{\theta_1}(\bar{X} \leq t) = \Phi\left(\frac{t-\theta_1}{\sigma_0/\sqrt{n}}\right)$$

where Φ is the cdf of the $\mathcal{N}(0,1)$ distribution.



Lecture outline

How to formulate an hypothesis testing problem

Recall that we have a statistical model parameterized by θ :

$$\mathscr{P}^{\underline{X}} \;=\; \left\{\mathbb{P}_\theta^{\underline{X}},\ \theta \in \Theta\right\}.$$

Statistical hypothesis

A statistical hypothesis is represented by a subset of $\mathscr{P}^{\underline{X}}$, and thus by a subset of Θ.

Notation. Let $\Theta_j \subset \Theta$ denote the subset representing H_j

➠ $H_j : \theta \in \Theta_j$

Parametric / non-parametric test
A testing problem is called parametric if Θ is finite-dimensional.

How to formulate an hypothesis testing problem (cont'd)

Null hypothesis

We call the null hypothesis the hypothesis $H_0 : \theta \in \Theta_0$

- that we "want to test", and
- that will be retained "by default" unless it is clearly at odds with the data.

Legal analogy: presumption of innoncence

Alternative hypothesis

We call alternative hypothesis the hypothesis $H_1 : \theta \in \Theta_1$

- that will be chosen if H_0 is rejected.
- We assume that $\Theta_1 \cap \Theta_0 = \varnothing$.

Remark : we can assume wlog that $\Theta_0 \cup \Theta_1 = \Theta$.

Examples of parametric tests

Example 1. Exercice

- $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$, with $\theta \in \Theta = [0, +\infty[$,
- $\Theta_0 = \{\theta \geq \theta_0\}$; $\Theta_1 = \{\theta < \theta_0\}$ with $\theta_0 > 0$ a given threshold.

Example 2. Same example, with :

- $\Theta_0 = \{\theta_0\}$ (singleton) ; $\Theta_1 = \{\theta \neq \theta_0\}$,
- or $\Theta_0 = \{\theta_0\}$; $\Theta_1 = \{\theta < \theta_0\}$.

Definitions: simple / composite hypotheses

An hypothesis H_j is called simple if Θ_j is a singleton.
Otherwise, it is called composite.

Other examples of (non-parametric) tests

Goodness-of-fit tests for a distribution or family of distributions

- voir section 3

Other types of tests

- testing the independence of two variables
- testing the symmetry of a distribution
- ...

Test procedures

Definition: test (procedure)

A test is a statistic $\delta = \delta(\underline{X})$ with values in $\{0, 1\}$:

$$\begin{aligned} \delta : \quad \underline{\mathcal{X}} \quad &\mapsto \quad \{0,1\}, \\ \underline{x} \quad &\to \quad \begin{cases} 0 & \text{if } H_0 \text{ is accepted,} \\ 1 & \text{if it is rejected (in favour of } H_1). \end{cases} \end{aligned}$$

Definition: critical region of a test

The critical region $\mathscr{R}_\delta$ of a test δ is the region of rejection

$$\mathscr{R}_\delta \;=\; \{\, \underline{x} \in \underline{\mathcal{X}} \;\text{ such that }\; \delta(\underline{x}) = 1 \,\}\,.$$

Quantifying the risks of error

Definition: risk (of error) of the first kind

We call the risk of the first kind, or risk of type I error, the probability of rejecting H_0 when it is true :

$$\mathbb{P}_\theta(\delta = 1) \;=\; \mathbb{E}_\theta(\delta), \qquad \theta \in \Theta_0.$$

(⚠ This risk depends on the value of θ, for $\theta \in \Theta_0$.)

Definition: risk (of error) of the second kind

We call the risk of the second kind, or risk of type II error, the probability of accepting H_0 when it is false :

$$\mathbb{P}_\theta(\delta = 0) = 1 - \mathbb{E}_\theta(\delta), \qquad \theta \in \Theta_1.$$

(Note the asymmetry of terminology
$\to$ more emphasis is put on H_0.)

Definition: power of a test

We define power as the probability to reject H_0 when it is wrong:

$$\mathbb{P}_\theta(\delta = 1) = \mathbb{E}_\theta(\delta), \qquad \theta \in \Theta_1.$$

Remark: equal to "1 - risk of type II error".

Usual approach† for the construction of tests.
Let $0 < \alpha < 1$ be a level of risk. We will look for tests s.t.

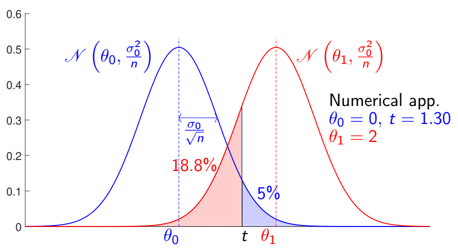
- $\forall \theta \in \Theta_0$, $\mathbb{P}_\theta(\delta = 1) \leq \alpha$;
  ➠ control of the risk of type I errors.
  The test δ is said to have level (at most) α.
- $\forall \theta \in \Theta_1$, $\mathbb{P}_\theta(\delta = 1)$ "as large as possible";
  ➠ capacity to reject H_0 when it is false.

Typical values: $\alpha = 5\%,\ 1\%,\ 1‰\ldots$

† a.k.a. Neyman's

Back to the introductory example

- type I error: blue area
- type II error: red area



Probability density function of $\bar{X}$ under H_0 and H_1

Definition: size of a test

We say that δ has a level *exactly* α, or size α, if

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta = 1) = \alpha.$$

Definition: comparing two tests

Let δ and δ' be two tests with a level (at most) α. We say that δ' is uniformly more powerful than δ if

$$\forall \theta \in \Theta_1, \quad \mathbb{P}_\theta\left(\delta' = 1\right) \;\geq\; \mathbb{P}_\theta\left(\delta = 1\right).$$

(Some authors require a strict inequality at one or all $\theta \in \Theta_1$.)

Remarks :

- this is a partial order on power functions,
- whenever possible, we will look for the uniformly most powerful (UMP) test at level α (i.e., a test with α, that is uniformly more powerful than all other tests with level α).

Lecture outline

Lecture outline

1 – Examples and first definitions

2 – Parametric tests
- 2.1 – Simple null vs simple alternative
- 2.2 – Composite hypotheses
- 2.3 – Asymptotic tests

3 – Goodness-of-fit testing: Pearson's χ^2 test

4 – Standard exercises (with solutions)

5 – Annexes

Likelihood ratio test

Assume two simple hypotheses : $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$.

Denote by $\mathcal{L} : (\theta, \underline{x}) \mapsto \mathcal{L}(\theta, \underline{x})$ the likelihood function†.

Definition: likelihood ratio test

We call the likelihood ratio (LR) test the test

$$\delta^{\mathrm{LR}} = \begin{cases} 1 & \text{if } T^{\mathrm{LR}} > c, \\ 0 & \text{otherwise,} \end{cases}$$

built using the likelihood ratio statistic:

$$T^{\mathrm{LR}} = \frac{\mathcal{L}(\theta_1, \underline{X})}{\mathcal{L}(\theta_0, \underline{X})}.$$

† It can be proved that the family $\left\{\mathbb{P}^{\underline{X}}_{\theta_0}, \mathbb{P}^{\underline{X}}_{\theta_1}\right\}$ is always dominated (Radon-Nikodym).

Fundamental result

Let $\alpha \in (0, 1)$.

Theorem: Neyman-Pearson "lemma"

Assume that there exists⊛ a threshold $c = c_\alpha$ such that
- the associated LR test δ^{LR} has a level exactly α (i.e., has size α).

Then δ^{LR} is most powerful† at the level α:
- for any test $\tilde{\delta}$ with a level (at most) α, δ^{LR} is more powerful than $\tilde{\delta}$.

➠ The LR test is optimal in this setting.

⊛ Always true if the cdf of T^{LR} is continuous.

† No need to specify "uniformly" since H_1 is simple.

Back to the Gaussian example

Likelihood ratio :

$$T^{\mathrm{LR}} = \frac{\frac{1}{(\sqrt{2\pi}\sigma_0)^n}\exp\left(-\frac{\sum_{i=1}^n (X_i-\theta_1)^2}{2\sigma_0^2}\right)}{\frac{1}{(\sqrt{2\pi}\sigma_0)^n}\exp\left(-\frac{\sum_{i=1}^n (X_i-\theta_0)^2}{2\sigma_0^2}\right)}$$

$$= \exp\left(-\frac{n(\theta_1^2-\theta_0^2)}{2\sigma_0^2}\right)\exp\left(\frac{(\theta_1-\theta_0)}{\sigma_0^2}\sum_{i=1}^n X_i\right).$$

LR test at level α: since $\theta_1 > \theta_0$, we have

$$\delta^{\mathrm{LR}} = 1 \iff T^{\mathrm{LR}} > c_\alpha \iff T = \bar{X} > t_\alpha$$

➠ the test that was constructed in introduction is optimal.

Test statistic and p-value

The result of a test can be expressed using the concept of p-value.

Definition: p-value

Let T be the test statistic of a test of the form $\delta = \mathbb{1}_{T>t_\alpha}$.

Definition. We call p-value the statistic

$$\mathrm{pval}\,(\underline{x}) = \mathbb{P}_{\theta_0}\left(T(\underline{X}) > T(\underline{x})\right)$$

taking values in $(0,1)$. ⚠ Function of the data!

Let F_0 denote the cdf of T under H_0. Then:

$$\mathrm{pval}(\underline{x}) = 1 - F_0(T(\underline{x})).$$

Interpretation of the p-value

Assume that F_0 is continuous and strictly increasing:

$\forall\alpha \in (0,1), \quad \exists! t_\alpha \in \mathbb{R}, \quad \delta = \mathbb{1}_{T>t_\alpha}$ has level exactly α

Proposition ➠ Proof

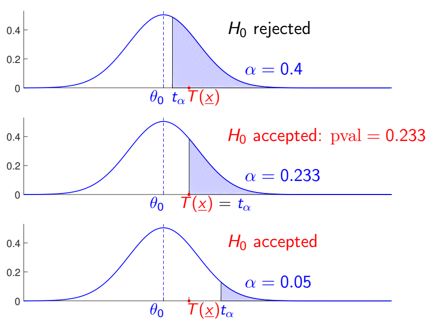
H_0 is rejected at the level α $\Leftrightarrow$ $T > t_\alpha$ $\Leftrightarrow$ $\mathrm{pval} < \alpha$.

t_α is called the critical value for the test statistic T.

Interpretation: p-value = mesure of evidence against H_0

p-value	evidence against
$\mathrm{pval} < 0.01$	very strong evidence
$0.01 \leq \mathrm{pval} < 0.05$	strong evidence
$0.05 \leq \mathrm{pval} < 0.10$	weak evidence
$0.1 < \mathrm{pval}$	no evidence

Back to the Gaussian example, where $T(\underline{X}) = \bar{X}$



(pval is the maximal level α at which H_0 is accepted.)

Lecture outline

Examples of problems with composite hypotheses

Simple null / composite alternative
- $\Theta_0 = \{\theta_0\}$ / $\Theta_1 = \{\theta > \theta_0\}$ (one-sided test),
- $\Theta_0 = \{\theta_0\}$ / $\Theta_1 = \{\theta \neq \theta_0\}$ (two-sided test),
- ...

Composite null / composite alternative
- $\Theta_0 = \{\theta \leq \theta_0\}$ / $\Theta_1 = \{\theta > \theta_0\}$ (one-sided test),
- $\Theta_0 = \{\mu = \mu_0\}$ / $\Theta_1 = \{\mu = \mu_1\}$,
  where $\theta = (\mu, \sigma^2)$ with unknown σ^2 (nuisance parameter),
- $\Theta_0 = \{\theta^{(1)} = \theta^{(2)}\}$ / $\Theta_1 = \{\theta^{(1)} \neq \theta^{(2)}\}$,
  where $\theta \in \Theta = \mathbb{R}^2$ (egality of two parameters),
- ...

Differences with the case of simple hypotheses

- Test with a level (at most) α, when Θ_0 is composite :

$$\forall \theta \in \Theta_0,\ \mathbb{P}_\theta(\delta = 1) \leq \alpha \quad \Leftrightarrow \quad \underbrace{\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\delta = 1)}_{\text{size of the test}} \leq \alpha.$$

- If Θ_1 is composite, the power is a function of $\theta \in \Theta_1$:

$$\begin{array}{rcl} \Theta_1 & \to & [0,1] \\ \theta & \mapsto & \mathbb{P}_\theta(\delta = 1). \end{array}$$

- p-value for a test of the form $\delta = \mathbb{1}_{T > t_\alpha}$:

$$\mathrm{pval} = \sup_{\theta \in \Theta_0} (1 - F_\theta(T)).$$

  where F_θ is the cdf of T under $\mathbb{P}_\theta$.

Back to the Gaussian example / testing the mean

- **Simple hypothesis testing**

$$H_0 : \theta = \theta_0 / H_1 : \theta = \theta_1, \quad \text{with } \theta_0 < \theta_1$$

- **Reminder of the optimal test.**

$$\delta(\underline{X}) = 1 \iff \bar{X} > t_\alpha, \quad \text{with } t_\alpha = \theta_0 + q_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}$$

  Following the Neyman-Pearson lemma, δ is UMP among tests of level α.

- **Analysis of the test.** δ is the same for any $\theta_1 > \theta_0$ (it only depends on α and θ_0); therefore δ is also UMP for a test of the form:

$$H_0 : \theta = \theta_0 / H_1 : \theta > \theta_0.$$

It can be proved that δ **is also UMP for a test of the form:**
$H_0 : \theta \leq \theta_0$ / $H_1 : \theta > \theta_0$

exercise

Lecture outline

1 – Examples and first definitions

2 – Parametric tests
- 2.1 – Simple null vs simple alternative
- 2.2 – Composite hypotheses
- 2.3 – Asymptotic tests

3 – Goodness-of-fit testing: Pearson's χ^2 test

4 – Standard exercises (with solutions)

5 – Annexes

Context : $X_1, X_2, \dots \overset{\text{iid}}{\sim} \mathrm{P}_\theta$

When the distribution of $T_n(\underline{X}_n)$ is hard to determine

➠ use of the limit distribution for $n \to \infty$.

Example: component reliability

$$\mathcal{R}_{\alpha,n} = \{\underline{x}_n \text{ such that } T_n(\underline{x}_n) = \bar{x}_n \; > \; \tilde{t}_{\alpha,n}\}.$$

with $\tilde{t}_{\alpha,n}$ chosen in such a way that :

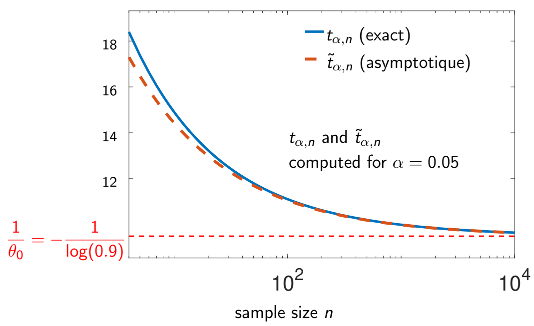
$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}\left(T_n(\underline{X}_n) > \tilde{t}_{\alpha,n}\right) = \alpha.$$

By the CLT under H_0 : $\sqrt{n}\left(\bar{X}_n - \frac{1}{\theta_0}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \frac{1}{\theta_0^2}\right)$, therefore

$$\tilde{t}_{\alpha,n} \;=\; \frac{1}{\theta_0} + \frac{1}{\theta_0\sqrt{n}}\, q_{1-\alpha}$$

where q_r is the $\mathcal{N}(0,1)$ quantile of order r.

Example: component reliability (cont'd)



Lecture outline

1 – Examples and first definitions

2 – Parametric tests

3 – Goodness-of-fit testing: Pearson's χ^2 test

4 – Standard exercises (with solutions)

5 – Annexes

Goodness-of-fit test for a single distribution

Context: $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathrm{P}$ with unknown P (can be anything)
➠ $\theta = \mathrm{P}$, $\Theta = \{$ probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $\}$.

Statistical hypotheses to be tested

For a given probability P_0, we consider the hypotheses:

$H_0 : \mathrm{P} = \mathrm{P}_0$
$H_1 : \mathrm{P} \neq \mathrm{P}_0$

Component reliability example:

- The component manufacturer knows, from past analyses, that the component lifetimes should follow a $\mathcal{E}(\theta_0)$ distribution.
- In order to check that the production line is still properly working, he wants to test if $H_0 : \mathrm{P} = \mathcal{E}(\theta_0)$ is still true.

➠ back to slide 12

Pearson's χ^2 test statistic

Let $(A_1, \ldots, A_K)$ be a partition of P_0's support, and

- $N = (N_1, \ldots, N_K)$ with
  $N_k = \sum_{i=1}^n \mathbb{1}_{A_k}(X_i)$ $\rightarrow$ observed frequencies (counts),
- $p = (p_1, \ldots, p_K)$ with
  $p_k = \mathrm{P}_0(X_1 \in A_k)$ $\rightarrow$ np_k = expected frequ. under H_0.

Proposition

Under hypothesis H_0, N follows a multinomial $\mathrm{Multi}(n, p)$ distribution, and

$$T_n = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} \xrightarrow[n\to\infty]{\text{d}} \chi^2(K-1).$$

(χ^2 distribution with $K-1$ degrees of freedom)

➠ Complement: the family of multinomial distributions
➠ Complement: the family of χ^2 distributions

Pearson's chi-squared test (χ^2)

Recall that we want to test $H_0 : \mathrm{P} = \mathrm{P}_0$ against $H_1 : \mathrm{P} \neq \mathrm{P}_0$.

Chi-square (χ^2) goodness-of-fit test

Let $0 < \alpha < 1$ and let T denote Pearson's statistic:

$$T = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k}.$$

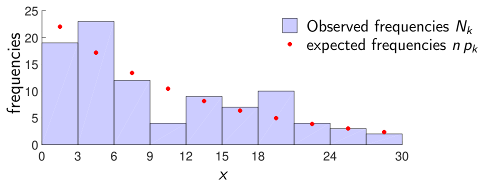
The chi-squared (χ^2) test is

$$\delta = \mathbb{1}_{T > t_\alpha},$$

where t_α is the $\chi^2(K-1)$ quantile of order $1-\alpha$.

⚠ In practice: choose $A_1, \ldots, A_K$ such that $np_k \geq 5$, $\forall k$.

The χ^2 test for goodness-of-fit: "component reliability"

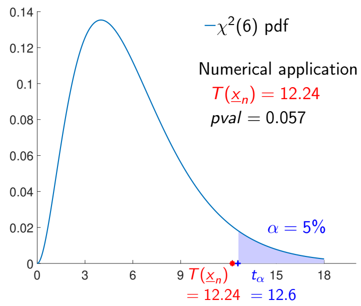


class	$[0,3[$	$[3,6[$	$[6,9[$	$[9,12[$	$[12,15[$	$[15,18[$	$[18,\infty[$
N_k	19	23	12	4	9	7	19
$n\,p_k$	25.90	19.2	14.2	10.5	7.8	5.8	11.6

$$T(\underline{X}_n) = \sum_{k=1}^7 \frac{(N_k - n\,p_k)^2}{n\,p_k} \xrightarrow[n\to\infty]{\text{d}} \chi^2(7-1)$$

The χ^2 test for goodness-of-fit: "component reliability"

Numerical application. $n = 100$, $T(\underline{x}_n) = 12.24$



➠ at the 5% level, H_0 is accepted

More on goodness-of-fit testing...

- Pearson's χ^2 test for a family of distributions
  - extension of the test just presented to the case where some parameters must be estimated under H_0

complement

- Kolmogorov-Smirnov test
  - another test, based on the cumulative distribution function,
  - without requiring the choice of a partition

complement

Lecture outline

Lecture outline

Exercise 1 (Testing a proportion)

solution

In the context of a coin toss game, we want to test if the coin is balanced.

Questions

1. Propose a statistical experiment to test this hypothesis. Specify the underlying statistical model, and define the null and alternative hypotheses.
2. Propose a test at the asymptotic level α.

Exercise 2 (Component reliability testing)

solution

A manufacturer wishes to offer its customers a guarantee on light bulbs. It is assumed that the lifetime of a bulb follows an exponential distribution with parameter $\theta > 0$.

Questions

Propose a UMP test for the following test:

$H_0 : \Theta_0 = \{\theta \geq \theta_0\}$ (bulb insufficiently reliable)

$H_1 : \Theta_1 = \{\theta < \theta_0\}$ (bulb sufficiently reliable)

with a given threshold $\theta_0 > 0$.

back to slide 11
back to slide 26

Lecture outline

Solution of exercise 1

back to questions

i) n "coin toss" experiments are carried out, and the outcomes are modeled as n independent, identically distributed random variables $X_1, \ldots, X_n$ according to a $Ber(\theta)$ distribution.
We want to test if

$$H_0 : \theta = \frac{1}{2}, \quad \text{i.e.,} \quad \Theta_0 = \left\{\frac{1}{2}\right\} \quad \text{(simple hypothesis)},$$

vs.

$$H_1 : \theta \neq \frac{1}{2} \quad \text{therefore} \quad \Theta_1 = \left]0, \frac{1}{2}\right[\cup \left]\frac{1}{2}, 1\right[\quad \text{(two-sided hypothesis)}.$$

Solution of exercise 1

back to questions

ii) Let $\hat{\theta}_n = \bar{X}_n$ be the empirical mean of the sample. By direct application of CLT, it follows that:

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\theta(1-\theta)/n}} \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$$

To construct a two-sided asymptotic test of level α, we place ourselves under H_0. We obtain the following convergence in distribution:

$$2\sqrt{n}\left(\hat{\theta}_n - \frac{1}{2}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1).$$

We consider a critical region of the form: $2\sqrt{n}|\hat{\theta}_n - \frac{1}{2}| > c_\alpha$. where c_α is chosen so that the Type I error rate is equal to α.

Solution of exercise 1

back to questions

ii) Let

$$\lim_{n\to\infty} \mathbb{P}\left(2\sqrt{n}\left|\hat{\theta}_n - \frac{1}{2}\right| > c_\alpha\right) = \alpha.$$

We deduce that $c_\alpha = q_{1-\frac{\alpha}{2}}$, the $(1-\frac{\alpha}{2})$-th quantile of a standard normal distribution $\mathcal{N}(0,1)$.

We reject the null hypothesis H_0 in favor of H_1 at the level α when:

$$\left|\hat{\theta}_n - \frac{1}{2}\right| > q_{1-\frac{\alpha}{2}}\frac{1}{2\sqrt{n}}.$$

Thus, the difference between $\hat{\theta}_n$ and $1/2$ is considered significant at the level α if it exceeds $q_{1-\frac{\alpha}{2}}\frac{1}{2\sqrt{n}}$.

Solution of exercise 2

back to questions

Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$

$H_0 : \Theta_0 = \{\theta \geq \theta_0\}$ (component is not reliable enough)
$H_1 : \Theta_1 = \{\theta < \theta_0\}$ (component is reliable enough)

By the Neyman-Pearson lemma, the LRT is UMP for

$$H_0 : \Theta_0 = \{\theta_0\} \quad / \quad H_1 : \Theta_1 = \{\theta_1\}, \quad \text{with} \quad \theta_1 < \theta_0$$

$$\begin{aligned} T^{\mathrm{LR}}(\underline{X}) &= \frac{\theta_1^n \exp\left(-\theta_1 \sum_{i=1}^n X_i\right)}{\theta_0^n \exp\left(-\theta_0 \sum_{i=1}^n X_i\right)} \\ &= \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left((\theta_0 - \theta_1)\sum_{i=1}^n X_i\right) \end{aligned}$$

Solution of exercise 2

back to questions

We then define the **critical region** of this test at level α:

$$\mathcal{R}_\alpha = \left\{\underline{x} \mid T^{\mathrm{LR}}(\underline{x}) > t_\alpha^{\mathrm{LR}}\right\} = \{\underline{x} \mid T(\underline{x}) = \bar{x} > t_\alpha\}.$$

Reminder : if $\theta = \theta_0$, then $\theta_0 \bar{X} \sim \Gamma(p = n, \lambda = n)$.

➠ $t_{\alpha,n} = \frac{1}{\theta_0} q_{1-\alpha}$
where q_r is the $\Gamma(p = n, \lambda = n)$ quantile of order r.

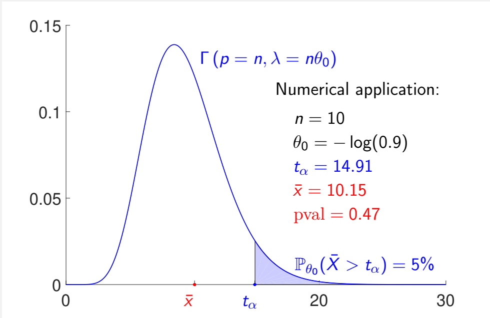
This test is also UMP for its composite version, indeed :

- the likelihood ratio test is the same for any $\theta_1 < \theta_0$,
- the function $\theta \mapsto \mathbb{P}_\theta(\delta = 1)$ is strictly ↘.

Summary. The test that we have built is UMP at the level α.

Solution of exercise 2

back to questions



- at the 5% level, H_0 is not rejected
- out of precaution, the manufacturer will not propose a warranty

Lecture outline

Lecture outline

1 – Examples and first definitions

2 – Parametric tests

3 – Goodness-of-fit testing: Pearson's χ^2 test

4 – Standard exercises (with solutions)

5 – Annexes
5.1 – Proof & complements

Proof

Note that t_α is, by construction, such that

$$F_0(t_\alpha) = 1 - \alpha.$$

Thus we have

$$\begin{aligned} \delta = 1 \quad &\Leftrightarrow \quad T > t_\alpha \\ &\Leftrightarrow \quad F_0(T) > F_0(t_\alpha) = 1 - \alpha \\ &\Leftrightarrow \quad \text{pval} < \alpha \end{aligned}$$

$\square$

back to slide ??

Generalized likelihood ratio test

It enables the construciton of a test when Θ_0 and/or Θ_1 are/is composites.

- Test statistic :

$$T(\underline{X}) = \frac{\sup_{\theta\in\Theta_1} \mathcal{L}(\theta; \underline{X})}{\sup_{\theta\in\Theta_0} \mathcal{L}(\theta; \underline{X})}.$$

- The test is not, in general, uniformly most powerful (UMP) at level α.

The multinomial family of distributions

Parameters

- n integer, ≥ 1,
- K integer, ≥ 2 and $p \in (\mathbb{R}_\star^+)^K$ such that $\sum_{k=1}^K p_k = 1$.

Let $n_1, \ldots, n_K$ entiers ≥ 0 such that $\sum_{k=1}^K n_k = n$:

If $N \sim Multi(n, p)$, $\mathbb{P}(N_1 = n_1, \ldots, N_K = n_k) = \dfrac{n!}{n_1! \ldots n_K!} p_1^{n_1} \ldots p_K^{n_K}$

Moments

- expectation : $\mathbb{E}_p(N) = np$
- covariance matrix : $\mathrm{cov}_p(N_i, N_j) = n(p_i\delta_{ij} - p_ip_j)$

Marginal distributions

- Marginal distributions are binomial : $N_j \sim \mathrm{Bin}(n, p_j)$.

back to slide ??

The χ^2 family of distributions

Parameters

- q integer, ≥ 1 : number of "degrees of freedom".

Definition. If $Y_1, \ldots, Y_q \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ then

$$T = \sum_{k=1}^q Y_k^2 \sim \chi^2(q)$$

The χ^2 distribution is a special case of the Γ distribution :

$$\chi^2(q) = \Gamma\left(p = \frac{q}{2}, \lambda = \frac{1}{2}\right)$$

➠ The properties of the χ^2 follow from those of the Γ distribution.

Expectation

- $\mathbb{E}_q(T) = q$

Variance

- $\mathrm{var}_q(T) = 2q$

back to slide ??

The χ^2 test with parameter estimation

Does the lifetime of a component follow an exponential distribution?

➠ **Null hypothesis** H_0**:** $\exists\theta > 0,\ \mathrm{P} = \mathrm{P}_\theta = \mathcal{E}(\theta)$.

Two-step approach

1. Construction of a consistent estimator of $\theta \to \hat{\theta}$.
2. Test the goodness of fit to $\mathrm{P}_{\hat{\theta}}$.

Details

$\hat{p}_k = \mathrm{P}_{\hat{\theta}}(X_1 \in A_k)$

$$T(\underline{X}_n) = \sum_{k=1}^K \frac{(N_k - n\hat{p}_k)^2}{n\hat{p}_k} \xrightarrow[n\to\infty]{\text{d}} \chi^2(K-1-q) \text{ with } q = \mathrm{card}(\theta)$$

Rejection of H_0 if $T(\underline{x}_n) > t_{1-\alpha}$ ($t_{1-\alpha}$ being the quantile of order $1-\alpha$ of a $\chi^2(K-1-q)$ distribution).

back to slide 34

The Kolmogorov-Smirnov test

Goodness-of-fit test for a single distribution : $H_0 : \mathrm{P} = \mathrm{P}_0$.

Kolmogorov-Smirnov distance

The Kolmogorov-Smirnov distance is defined as

$$D_n \; = \; \sup_x \left| \hat{F}_n(x) - F_0(x) \right| ,$$

with F_0 the cdf of P_0 and $\hat{F}_n$ the empirical cdf: $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$.

Kolmogorov-Smirnov test, with asymptotic level α

Under the null hypothesis H_0, if F_0 is continuous:

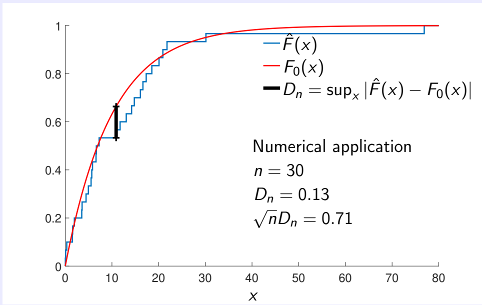
$$T(\underset{\sim}{X}_n) = \sqrt{n} D_n \xrightarrow[n\to\infty]{d} \mathcal{K},$$

where $\mathcal{K}$ is the Kolmogorov-Smirnov distribution.

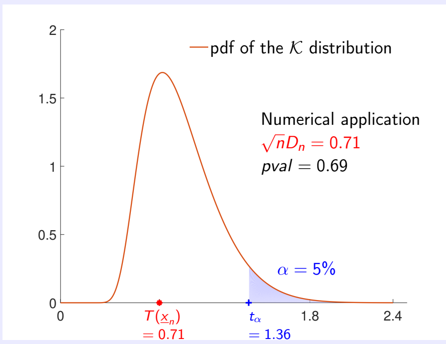
➠ H_0 is rejected if $T_n > t_\alpha$, with t_α the $(1-\alpha)$-quantile of $\mathcal{K}$.

The Kolmogorov-Smirnov test

"Component reliability example": $H_0 : \mathrm{P} = \mathcal{E}(\theta_0)$ with $\theta_0 = 0.1$



The Kolmogorov-Smirnov test



➠ at the 5% level, H_0 is accepted

back to slide 34