Data Valorization: Hypothesis Testing

Lionel Fillatre

fillatre@unice.fr

Topics

- Hypothesis Testing
- Neyman-Pearson Test between simple hypotheses
- Maximum Likelihood Ratio Test
- Conclusion

1 Hypothesis Testing

Hypothesis

- **Hypothesis:** A statement that something is true

- **Statistical Hypothesis Test:** A process by which a decision is made between two opposing hypotheses. The two opposing hypotheses are formulated so that each hypothesis is the negation of the other.

  - **Null Hypothesis, H_0:**
    - The hypothesis to be tested.
    - Assumed to be true.
    - Usually a statement that a population parameter has a specific value.
    - The “starting point” for the investigation.

  - **Alternative Hypothesis, H_1:**
    - A statement about the same population parameter that is used in the null hypothesis.
    - Generally this is a statement that specifies the population parameter has a value different, in some way, from the value given in the null hypothesis.

Example

- ✓ **Example:** Suppose you are investigating the effects of a new pain reliever. You hope the new drug relieves minor muscle aches and pains longer than the leading pain reliever.

  State the null and alternative hypotheses.

Solution:

- H_0: The new pain reliever is no better than the leading pain reliever
- H_1: The new pain reliever lasts longer than the leading pain reliever

- Note: the solution is not unique! It depends on the interpretation of the problem.

Example

- **Example:** You are investigating the presence of radon in homes being built in a new development. If the mean level of radon is greater than 4 then send a warning to all home owners in the development.

State the null and alternative hypotheses.

Solution:

- H_0: The mean level of radon for homes in the development is 4 (or less)
- H_1: The mean level of radon for homes in the development is greater than 4

Example

- **Example:** A calculator company has just received a large shipment of parts used to make the screens on graphing calculators. They consider the shipment acceptable if the proportion of defective parts is 0.01 (or less). If the proportion of defective parts is greater than 0.01 the shipment is unacceptable and returned to the manufacturer.

State the null and alternative hypotheses.

Solution:

- H_0: The proportion of defective parts is 0.01 (or less)
- H_1: The proportion of defective parts is greater than 0.01

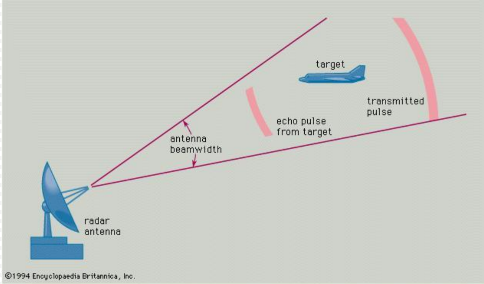
Example: Radar Detection

- You are observing the sky with a radar
- You are counting each result:

Decision	H_0: no target (no plane) True	False
Accept H_0	90	5
Reject H_0	10	15

- In terms of probability

Decision	H_0: no target (no plane) True	False
Accept H_0	$0{,}9 = 1 - \alpha$	$0{,}25 = \beta$
Reject H_0	$0{,}1 = \alpha$	$0{,}75 = 1 - \beta$



- Conditional probabilities:

$\alpha = \Pr(\text{reject } H_0 \mid H_0 \text{ is true})$

$\beta = \Pr(\text{accept } H_0 \mid H_0 \text{ is false})$

$1 - \alpha = \Pr(\text{accept } H_0 \mid H_0 \text{ is true})$

$1 - \beta = \Pr(\text{reject } H_0 \mid H_0 \text{ is false})$

Errors in a nutshell

- Since we make a decision based on a **random sample**, there is always the chance of making an error (**the decision is a random variable**):
  - Probability of a type I error = α
  - Probability of a type II error = β

- Decisions:

Error in Decision	**Type**	**Probability**	
Rejection of a true H_0	I	α	→ False alarm
Failure to reject a false H_0	II	β	→ Miss detection

Correct Decision	**Type**	**Probability**	
Failure to reject a true H_0	A	$1-\alpha$	
Rejection of a false H_0	B	$1-\beta$	→ Power

2 Neyman-Pearson Test between simple hypotheses

Hypothesis Testing: common case

- We have a sample $\boldsymbol{x} = (x_1, \dots, x_n)$ belonging to a set $\mathcal{X}$ (typically $\mathbb{R}^n$)

- The probability density (or mass) function of $\boldsymbol{x}$ is denoted $f(\boldsymbol{x}; \theta)$ or $f_\theta(\boldsymbol{x})$ where θ belongs to the parameter space Ω

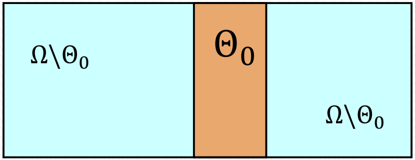
- The statement is about the unknown parameter θ

  $H_0: \theta \in \Theta_0$

  $H_1: \theta \in \Theta_1 = \Omega - \Theta_0 = \Omega \backslash \Theta_0$

  where Θ_0 is a well-defined subset of the parameter space Ω

Example:



Example with simple hypotheses

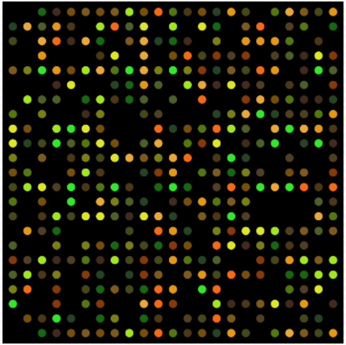
- Let X be the temperature measured with a thermometer

- Two hypotheses:
  - H_0 : X is a normal random variable with mean $\theta = 37$°C and standard deviation 0,5°C
  - H_1 : X is a normal random variable with mean $\theta \neq 37$°C and standard deviation 0,5°C

- Mathematical statement about the unknown parameter θ:
  - $H_0 : \theta \in \Theta_0 = \{37\}$
  - $H_1 : \theta \in \Theta_1 = \mathbb{R} \backslash \{37\}$

Example with composite hypotheses

- Let X be the expression level of a gene in gene microarray experiments.



- Two hypotheses:
  - H_0 : X is a normal random variable with mean $\theta = 0$ and standard deviation 1
  - H_1 : X is a normal random variable with mean $\theta > 0$ and standard deviation 1

- Mathematical statement about the unknown parameter θ:
  - $H_0: \theta \in \Theta_0 = \{0\}$
  - $H_1: \theta \in \Theta_1 = (0, +\infty)$

Terminology

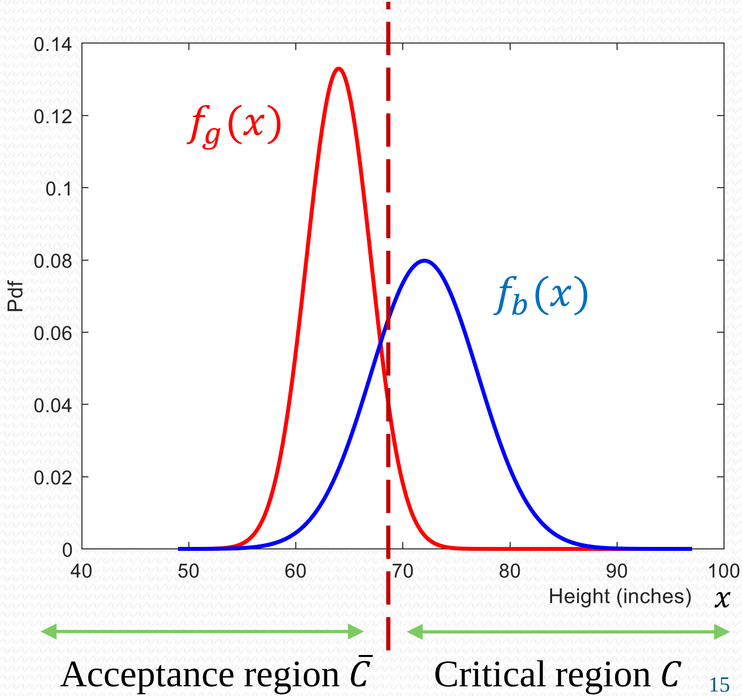
- Hypotheses:
  - Simple hypothesis: Θ_0 (or Θ_1) contains only one point (one single parameter)
  - Composite hypothesis: the opposite of simple hypothesis (many parameters)

- Acceptance:
  - Critical region (or rejection region) : a subset C of the sample space for the random sample $\boldsymbol{x} = (x_1, \dots, x_n)$ such that we reject H_0 if $\boldsymbol{x} \in C$
  - The complement of C, i.e. $\bar{C}$ will be referred to as the acceptance region

Example: Girl or Boy?

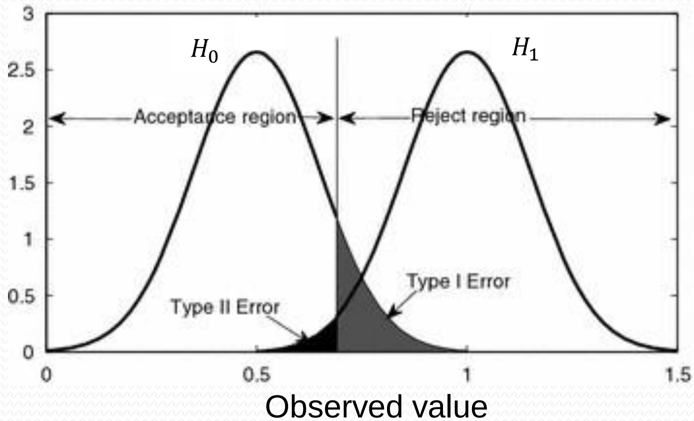
- X: height of an adult
- Hypotheses:
  - $H_0: \{X \sim f_g\}$
  - $H_1: \{X \sim f_b\}$

- Decision (thresholding):
  - Girl if $X \leq 69$
  - Boy otherwise



Balance the Errors

- We would like α and β to be as small as possible but α and β are inversely related
- We usually set α and minimize β
- Regardless of the outcome of a hypothesis test, we never really know for sure if we have made the correct decision



Most Powerful Test

- Probability of errors:
  - $\alpha = \Pr(X \in C \mid H_0)$
  - $\beta = \Pr(X \notin C \mid H_1) = \Pr(X \in \bar{C} \mid H_1)$

- Neyman-Pearson approach:
  - Fix α and then find a test that makes β desirably small

- Most powerful test of size α :
  - A test which minimizes β for a fixed value of α is called a **most powerful** test (or best test) of size α

Likelihood function (recall)

- For a sample $\boldsymbol{x} = (x_1, \dots, x_n)$
  - The likelihood function for a parameter θ is defined as

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i\,; \theta)$$

  - The log-likelihood function is

$$l(\theta; \boldsymbol{x}) = \ln\big(L(\theta; \boldsymbol{x})\big) = \sum_{i=1}^{n} \ln f\,(x_i\,; \theta)$$

  - It measures how likely (or expected) the sample is

Neyman-Pearson Lemma

- $\boldsymbol{x} = (x_1, \dots, x_n)$ a random sample from a distribution with pdf $f(x;\theta)$

- We wish to test:

$H_0 : \theta = \theta_0$ (simple hypothesis)

versus $H_1 : \theta = \theta_1$ (simple hypothesis)

- Then, the most powerful test of size α has a critical region of the form

$$\frac{L(\theta_1;\boldsymbol{x})}{L(\theta_0;\boldsymbol{x})} \geq h$$

where h is some non-negative constant called the threshold.

Example

- Samples :
  - $\boldsymbol{x} = (x_1, \dots, x_n)$ a random sample from $Exp(\theta)$ with pdf $f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0$
- Hypotheses:
  - Test H_0: $\theta = \theta_0$ vs. H_1: $\theta = \theta_1$ where $\theta_1 > \theta_0$ with a test of size α
- Most Powerful Test:
  - Recall that the likelihood is

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}$$

  - The most powerful test is

$$\frac{L(\theta_1; \boldsymbol{x})}{L(\theta_0; \boldsymbol{x})} = \frac{\theta_1^{-n} e^{-\frac{1}{\theta_1}\sum_{i=1}^{n} x_i}}{\theta_0^{-n} e^{-\frac{1}{\theta_0}\sum_{i=1}^{n} x_i}} \geq h \quad \Rightarrow \quad e^{-\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum_{i=1}^{n} x_i} \geq h\left(\frac{\theta_1}{\theta_0}\right)^n$$

Example (cont.)

- We get

$$-\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)\sum_{i=1}^{n} x_i \geq \ln\left(h\left(\frac{\theta_1}{\theta_0}\right)^n\right) = \ln h + n(\ln\theta_1 - \ln\theta_0)$$

- Note that $\theta_1 > \theta_0$ involves that $\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right) < 0$. It follows that

$$\sum_{i=1}^{n} x_i \geq -\frac{\ln h + n(\ln\theta_1 - \ln\theta_0)}{\frac{1}{\theta_1}-\frac{1}{\theta_0}} = h'$$

- If $\theta_1 > \theta_0$, the test has the form:

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \geq h'$$

- **Note:** if $\theta_1 < \theta_0$, the test has the form:

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i \leq h'$$

- This structure is consistent since $\theta = E(X)$.

Example (cont.)

- How to choose the threshold h'?

- If $\theta_1 > \theta_0$, then the threshold h' must satisfy

$$\Pr\left(\sum_{i=1}^{n} x_i \geq h' \;\middle|\; \theta = \theta_0\right) = \alpha$$

- Recall that a sum of $Exp(\theta)$ distributed variables is Gamma distributed $Gamma(n, \theta)$

- Let $F_{T,\theta_0}(t)$ be the cdf of the $Gamma(n, \theta_0)$ variable $T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$, then

$$\alpha = \Pr\left(\sum_{i=1}^{n} x_i \geq h' \;\middle|\; \theta = \theta_0\right) = 1 - \Pr\left(\sum_{i=1}^{n} x_i < h' \;\middle|\; \theta = \theta_0\right) = 1 - F_{T,\theta_0}(h')$$

- It follows that

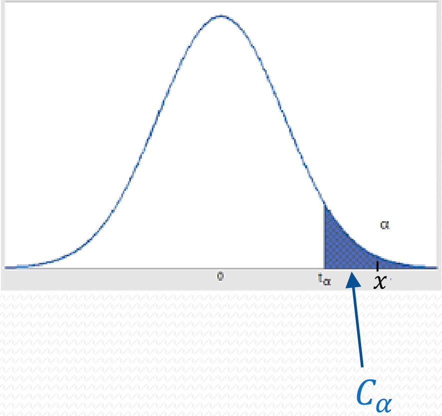
$$h' = F_{T,\theta_0}^{-1}(1 - \alpha)$$

Significance test (p-value)

- For varying α, the most powerful test provides an example of the typical situation in which the rejection regions C_α are nested in the sense that

$$C_\alpha \subset C_{\alpha'}, \text{ if } \alpha < \alpha'$$

(more severe on α, smaller rejection region)



- It is good practice to determine not only whether the hypothesis is accepted or rejected at the given significance level, but also to determine the smallest significance level for a given sample $\boldsymbol{x}$

$$\hat{p} = \hat{p}(\boldsymbol{x}) = \inf\{\alpha : \boldsymbol{x} \in C_\alpha\}$$

- This number, the so-called **p-value**, gives an idea of how strongly the data $\boldsymbol{x}$ contradicts the **null hypothesis**

3 Maximum Likelihood Ratio Test

Maximum Likelihood Ratio Test

- Consider again the composite hypotheses:

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

$$H_1 : \theta \in \Theta_1 = \Omega - \Theta_0 = \Omega \backslash \Theta_0$$

- The Maximum Likelihood Ratio Test (MLRT) is defined as rejecting H_0 if

$$\Lambda(\boldsymbol{x}) = \frac{\max_{\theta \in \Theta_1} L(\theta; \boldsymbol{x})}{\max_{\theta \in \Theta_0} L(\theta; \boldsymbol{x})} \geq h$$

Example

- Samples : $\boldsymbol{x} = (x_1, \dots, x_n)$ a random sample from $Exp(\theta)$ with pdf $f(x,\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x > 0$

- Hypotheses: Test H_0: $\theta = \theta_0$ vs. H_1: $\theta \neq \theta_0$

- Maximum Likelihood Estimate

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \Rightarrow \hat{\theta} = \overline{\boldsymbol{x}}$$

- Maximum Likelihood Ratio Test:

$$\Lambda(\boldsymbol{x}) = \frac{(\overline{\boldsymbol{x}})^{-n} e^{-(\overline{\boldsymbol{x}})^{-1}\sum_{i=1}^{n} x_i}}{\theta_0^{-n} e^{-\theta_0^{-1}\sum_{i=1}^{n} x_i}} = \left(\frac{\overline{\boldsymbol{x}}}{\theta_0}\right)^{-n} e^{\theta_0^{-1}\sum_{i=1}^{n} x_i - n} = \left(\frac{\overline{\boldsymbol{x}}}{\theta_0}\right)^{-n} e^{n(\theta_0^{-1}\overline{\boldsymbol{x}} - 1)}$$

$$\Lambda(\boldsymbol{x}) \geq h' \Leftrightarrow \ln \Lambda(\boldsymbol{x}) \geq \ln h' \quad \Rightarrow \quad n\left(\frac{\overline{\boldsymbol{x}}}{\theta_0} - 1\right) - n \ln\left(\frac{\overline{\boldsymbol{x}}}{\theta_0}\right) \geq \ln h' = h''$$

Distribution of $\Lambda(\boldsymbol{x})$

- Sometimes $\Lambda(\boldsymbol{x})$ has a well-defined sampling distribution but, often, this is not the case.

- Asymptotic result as the number of samples $n \to \infty$ (under suitable assumptions):

  - If H_0 is simple, it can be shown that $2\ln\big(\Lambda(\boldsymbol{x})\big)$ is asymptotically χ^2-distributed with d degrees of freedom under H_0, where d is the dimension of Ω (under suitable assumptions on the probability distribution of $\boldsymbol{x}$)

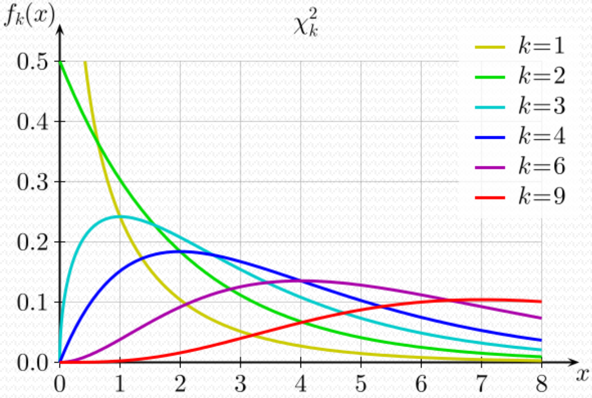
  - This result could be extended to composite H_0 (Wilks' theorem)

  - This convergence is useful to determine the threshold $h = F^{-1}_{\chi^2,d}(1-\alpha)$ where $F_{\chi^2,d}(.)$ is the χ^2 cdf with d degrees of freedom

Example (cont.)

- $\ln \Lambda(\boldsymbol{x}) = n\left(\frac{\bar{\boldsymbol{x}}}{\theta_0} - 1\right) - n \ln\left(\frac{\bar{\boldsymbol{x}}}{\theta_0}\right)$
- $d = 1$ as we have a simple null hypothesis and 1 unknown parameter in the numerator
- $2 \ln \Lambda(\boldsymbol{x})$ is asymptotically χ_1^2-distributed when $\theta = \theta_0$ (H_0 is true)

Pdf of chi-squared distribution



4 Conclusion

Conclusion

- Hypothesis testing is a famous tool to decide between two classes
- It is also very useful for model reliability (linear regression, etc.)
- Many applications can be adressed: biology, medicine, computer engineering, artificial intelligence, etc.
- Classification in Machine Learning is largely based on hypothesis testing (SVM, logistic regression, etc.)
- Tests between simple hypotheses are well known but it is more difficult to design tests for composite hypotheses
- Asymptotic optimality comes to our rescue!