# Advanced statistical methods

Frédéric Pascal

CentraleSupélec, Laboratory of Signals and Systems (L2S), France
frederic.pascal@centralesupelec.fr
http://fredericpascal.blogspot.fr

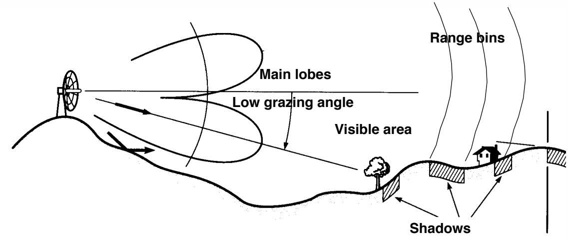
**MSc in Data Sciences & Business Analytics**

CentraleSupélec / ESSEC

Oct. $2^{\text{nd}}$ - Dec. $20^{\text{th}}$, 2017

# Contents

# Toy Example




Does a target is present if some received data?

Similar problem for testing if a person carries a disease or not...
After a blood test, an X-ray, a sonogram or whatever the procedure...

# Toy Example

1. Observations $\rightsquigarrow x_1, \ldots, x_n$
2. Statistical model $\rightsquigarrow$ Gaussian model, $\mathcal{N}(\mu, \sigma^2)$
3. Unknown parameters $\rightsquigarrow \sigma^2$ (could be extremely more complex...)
4. Data to Decision $\rightsquigarrow$ Binary hypothesis test

$$\begin{cases} \text{Hypothesis } H_0\text{:} & \mu = 0, \text{ i.e., no target} \\ \text{Hypothesis } H_1\text{:} & \mu > 0, \text{ i.e. a target is present} \end{cases}$$

5. How to exploit the data $\rightsquigarrow$ Parameter estimation

$$\hat{\mu}_n = \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_n)^2 \text{ or } \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_n)^2$$

Discussion on parameter estimation, choice, properties, ...

# Toy Example

Discussion on parameter estimation, choice, properties, ...

- convergence: much more data implies better accuracy?
- biais: in expectation, can we find the true value?
- error, variance: what is the best we can do?
- estimate characterisation: do we know some properties? e.g., the estimators distribution...
- identifiability: are we sure to find the correct value of the parameter? (likelihood approach)
- Confidence Interval

# Toy Example

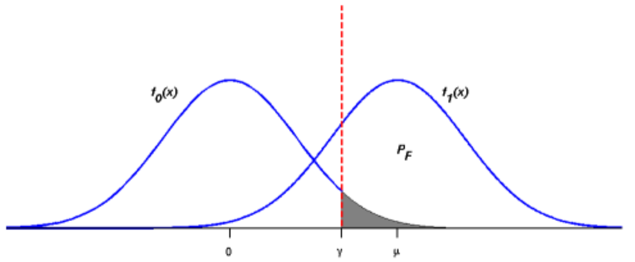



Figure: PDF of $\hat{\mu}_n$ under $H_0$ and $H_1$

Intuition:

- if $\hat{\mu}_n < \gamma$, one decides $H_0$,
- if $\hat{\mu}_n \geq \gamma$, one decides $H_1$.

Some natural questions:

- How to choose $\gamma$?... under which criterion...
- What are the errors? Are they equivalent? if yes, $\gamma = \mu/2$ but if not...

# Toy Example

Beginning of understanding ans answers on the problem

- Two types of errors:

| Truth \ Decision | $H_0$ | $H_1$ |
|---|---|---|
| $H_0$ | OK | Type-I error=PFA |
| $H_1$ | Type-II error=PND | power = PD |

- Type-II error extremely serious!
  Missing a target (e.g., a missile!) is more dangerous than detecting a false alarm... Claiming that a person is contaminated is more serious than the opposite: no treatment ...

Problem! impossible to minimize both errors at the same time!!!!
Explanations and details
Solution: Fixe the less serious error and minimize the other one
($\Leftrightarrow$ maximize the power of the test)
In practice, e.g., in radar (depending on the applications), PFA $= 10^{-2}$ to $10^{-5}$, resulting in PND of $10^{-7}$ or more...