AN INTRODUCTION TO DETECTION THEORY

Olivier Michel

Univ. Grenoble-Alpes, GIPSA-Lab, F-38000 Grenoble

Outline

A- The detection problem
1 : formulation
2 : Detection Errors
3 : Example
4 : The min error receiver
5 : Multiple hypothesis testing
6 : Neyman Pearson

B- Characterization, performances (simple hypothesis)
1- ROC Curves
2- Evaluating Signal detectability
3- Link to information theory, large deviations

C- Composite hypothesis testing
1-UMP test
2- Alternate Strategies

A1. The detection problem : formulation

Observe $x \in \Omega$ where $x \overset{d}{\simeq} Q$, $Q \in \{P_1, P_0\}$
or
Observe : $X = \{x_1, \dots, x_n\} \in \mathcal{X}$, n i..i.d. realizations,
thus $\mathcal{X} = \Omega^n$ and $X \overset{d}{\simeq} Q^n \in \{P_1^n, P_0^n\}$

Problem statement : Design a "receiver" that makes as few errors as possible, in deciding
for the SIMPLE hypothesis framework (no unknown parameters) :

$$\begin{cases} H_0 : & X \overset{d}{\simeq} P_0^n \\ H_1 : & X \overset{d}{\simeq} P_1^n \end{cases}$$

or for the COMPOSITE hypothesis framework : (unknown parameters)

$$\begin{cases} H_0 : & X \overset{d}{\simeq} P_0^n(X, \theta_0), \quad \theta_0 \in \Theta_0 \\ H_1 : & X \overset{d}{\simeq} P_1^n(X, \theta_1), \quad \theta_1 \in \Theta_1 \end{cases}$$

- The choices H_0, H_1 are mutually exclusive
- The receiver ALWAYS makes a choice

$\rightarrow$ Design a decision function ϕ, which expresses a partition of $\mathcal{X}$

$$\mathcal{X}_0 = \{x : \phi(x) = 0 : \quad \text{decide } H_0\} \quad \text{Rejection region}$$
$$\mathcal{X}_1 = \{x : \phi(x) = 1 : \quad \text{decide } H_1\} \quad \text{Acceptance region}$$

where

$$\mathcal{X}_1 = \mathcal{X}_0^c \quad \text{and} \quad \mathcal{X}_1 \cup \mathcal{X}_0 = \mathcal{X}$$

A.2. Detection errors

For the binary hypothesis testing problem, 2 kinds of errors :

False Alarm (FA) and Miss (M)

$$P_{FA}(\theta_0) = \int_{\mathcal{X}_1} P_0^n(X, \theta_0) dX = \mathsf{E}_{P_0}[\phi] \ , \theta_0 \in \Theta_0$$

$$P_M(\theta_1) = \int_{\mathcal{X}_0} P_1^n(X, \theta_1) dX = 1 - \int_{\mathcal{X}_1} P_1^n(X, \theta_1) dX = \mathsf{E}_{P_1}[1 - \phi] \ , \theta_1 \in \Theta_1$$

The correct detection probability is expressed by $P_D = 1 - P_M, \theta \in \Theta_1$.

A.3. Example : test on the mean of a gaussian observation

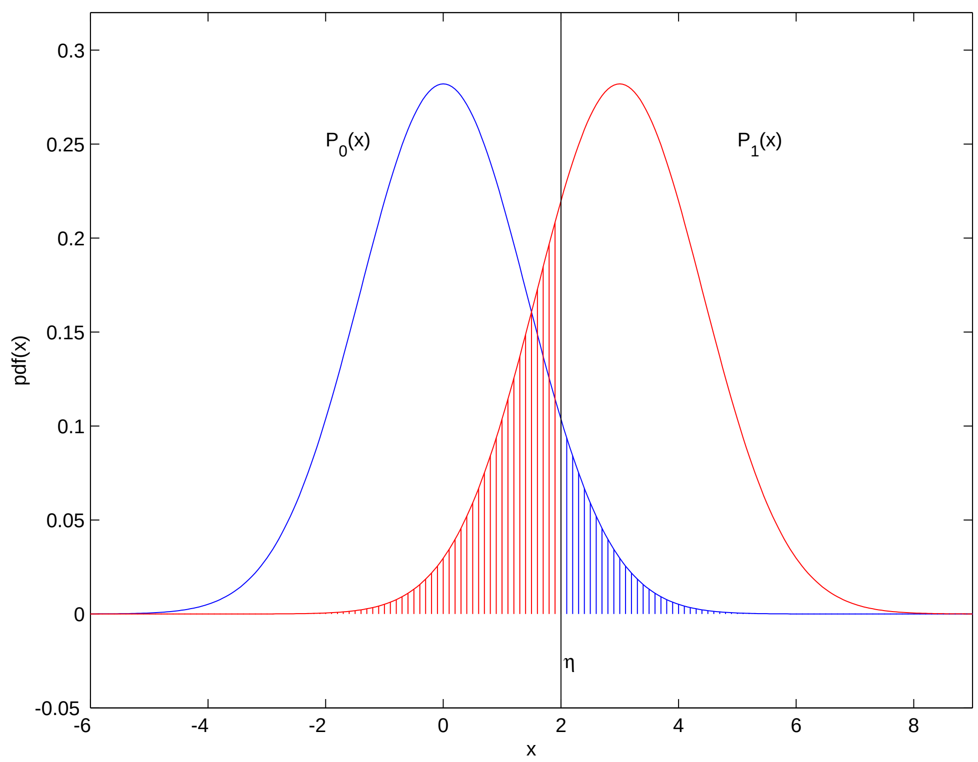
Let x be a normal R.V. wich has pdf $p_{\theta_i}(x)$ under H_i, $i \in \{0, 1\}$

$$\begin{cases} P_0(x|\theta_0) = p_{\theta_0}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2}) & \text{under} \quad H_0 \\ P_1(x|\theta_1) = p_{\theta_1}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2}) & \text{under} \quad H_1 \end{cases}$$

where σ^2 and m are known.

Here, $n = 1$, $\Theta_0 = \{0\}$ and $\Theta_1 = \{m\}$
$p(\theta_0) = \delta(\theta_0)$ and $p(\theta_1) = \delta(\theta_1 - m)$

$\rightarrow$ The test resumes to compare x to a threshold η :
$\mathcal{X}_0 =]-\infty, \eta]$ and $\mathcal{X}_1 =]\eta, \infty[$;



P0(x)
P1(x)
pdf(x)
x
η
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-6
-4
-2
0
2
4
6
8

Another example

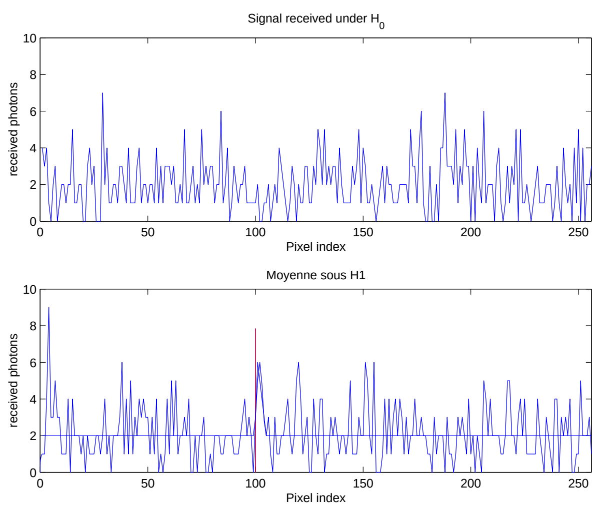
Detecting a change in the rate of a random Poisson process

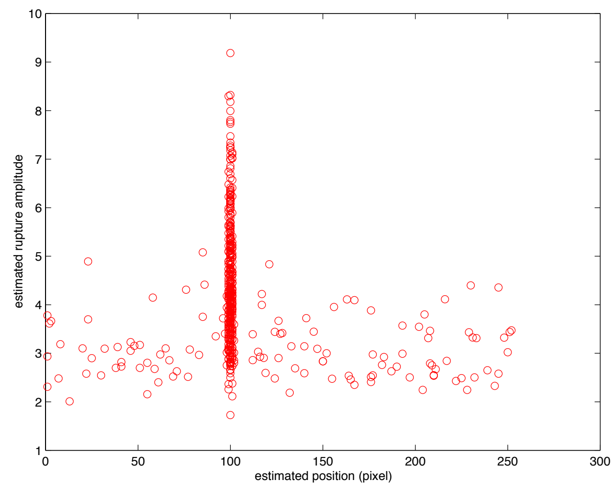
— CCD , 256 pixels on a unique line

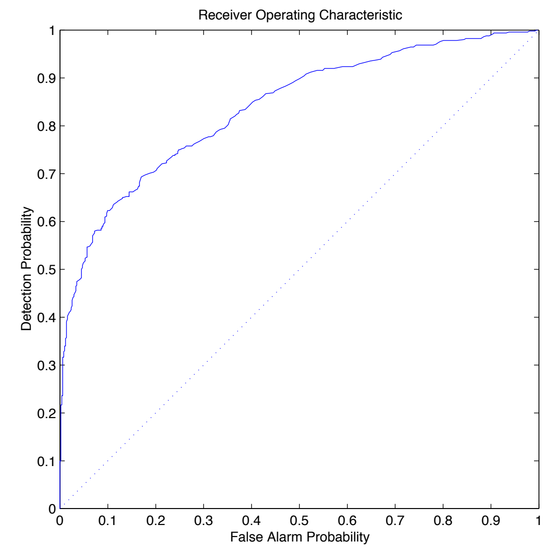
— Background noise : Poisson process of rate $f = 2$

— *Random* Signal : Poisson process of rate $\lambda = h \times$"Airy PSF" , $r = 5$, $\frac{h}{f} = 2$;

Illustration *Generalized Max likelihood*







Signal model (H0, H1) Perf. (position,amplitude) ROC

H0 : λ = cste = f = 2 ; H1 : r=5,n1=100,amp=4

A. 4. Binary hypothesis testing strategy : Bayes approach for designing of the minimum error receiver

Maximize the probability of correct classification, P_C

!! PRIORS !! : ($\{P(H_0), P(H_1)\}$), or prior density $p(\theta_i)$ and $p_{\theta_i}(x)$

$$P_C = \sum_{i=0,1} P(H_i) \int_{\mathcal{X}_i} \int_{\Theta_i} p_{\theta_i}(x) p(\theta_i) d\theta_i dx$$

Rk : Maximizing P_C amounts to define Bayes cost functions
$C_{11}(\theta_1) = C_{00}(\theta_0) = 0$ and $C_{10}(\theta_0) = C_{01}(\theta_1) = 1, \forall\theta_1, \forall\theta_0$
where C_{ij} is the cost of deciding H_i when H_j is true.

This leads to select the hypothesis with the largest posterior probability
$\rightarrow$ evaluate

$$p(H_i|x) = \frac{P(H_i).p(x|H_i)}{p_{\mathcal{X}}(x)}$$

hence, the MAP test is expressed by a LR Test

$$L(x) = \frac{p(x|H_1)}{p(x|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)}$$

A.5. Case of Multiple hypothesis testing-1- (skip next 3 slides ?)

$$H_0: \quad \theta_0 \in \Theta_0 \quad [x \simeq p_{\theta_0}(x), \quad \theta \in \Theta_0]$$
$$\vdots$$
$$H_M: \quad \theta_M \in \Theta_M \quad [x \simeq p_{\theta_M}(x), \quad \theta \in \Theta_M]$$

The decision function $\phi(x) = [\phi_1(x), \ldots, \phi_M(x)]^T$ verifies

$$\phi(x) \in \{0, 1\} \forall x \in \mathcal{X}$$
$$\sum_{i=1}^{M} \phi_i(x) = 1 \forall x \in \mathcal{X}$$

Let C_{ij} = cost to decide H_i when H_j is true, and $p(\widehat{H}_i|H_j)$ the proba of such decision

$$\boxed{\overline{C} = \sum_{i,j=1}^{M} C_{ij} p(\widehat{H}_i|H_j) p(H_j)}$$

Consider the case

$$\begin{array}{ll} C_{ii} = 0 & i \in \{1, \ldots, M\} \\ C_{ij} = 1,\ i \neq j, & i, j \in \{1, \ldots, M\} \end{array}$$

The P_{err} equals Bayes's risk :

$$\begin{aligned} \overline{C} &= \sum_{i \neq j = 1}^{M} C_{ij} p(\widehat{H}_i | H_j) p(H_j) \\ &= 1 - \sum_{i=1}^{M} C_{ii} p(\widehat{H}_i | H_i) p(H_i) \\ &= 1 - \sum_{i=1}^{M} C_{ii} p(H_i) \int_{\mathcal{X}_i} p(x|H_i) dx \end{aligned}$$

where $p(x|H_i) = \frac{\int_{\Theta_i} p(x,\theta) d\theta}{p(H_i)}$

$\overline{C}$ is minimal if $\forall i \in \{1, \ldots, M\}$, $\mathcal{X}_i$ are such that

$$x \in \mathcal{X}_i \Leftrightarrow p(H_i) p(x|H_i) \geq p(H_j) p(x|H_j)\ j \neq i$$

Thus the decision rule

$$\widehat{H}_i = ArgMax_{H_j} \left[p(H_j) p(x|H_j) \right] = ArgMax_{H_j} \left[p(H_j|x) \right]$$

Decision rules are non transitive in general

For example

$$\begin{aligned} H_1 \text{vs} H_2 &\rightarrow \widehat{H}_2 \\ H_2 \text{vs} H_3 &\rightarrow \widehat{H}_3 \\ H_1 \text{vs} H_3 &\rightarrow \widehat{H}_1 \end{aligned}$$

But, as the LR for the thest H_i vs H_j is $L_{ij} = \frac{f(x|H_i)}{f(x|H_j)}$,

$$L_{ik}(x) = L_{ij}(x) L_{jk}(x)$$

The decision rule $L_{ij}(x) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ij}$, shows that transitivity is obtained if

$$\eta_{ik} = \eta_{ij} \eta jk \Leftrightarrow \frac{(C_{ji} - C_{ii})}{(C_{ij} - C_{jj})} \frac{(C_{kj} - C_{jj})}{(C_{jk} - C_{kk})} = \frac{(C_{ki} - C_{ii})}{(C_{ik} - C_{kk})}$$

BUT .. if no available prior ($\{P(H_0), P(H_1)\}$), and densities p_{θ_i})

A.6. The neymann pearson framework

Constrain $\max|_{\theta\in\Theta_0} P_{FA}(\theta) \leq \alpha$ and maximize $P_D(\theta)|_{\theta\in\Theta_1}$ maximal.

neymann Pearson Lemma *The most powerfull test (MP) of level $\alpha \in [0, 1]$ is a randomized LRT of the form*

$$\phi(x) = \begin{cases} 1, & p_{\theta_1}(x) > \eta p_{\theta_0}(x) \\ q, & p_{\theta_1}(x) = \eta p_{\theta_0}(x) \\ 0, & p_{\theta_1}(x) < \eta p_{\theta_0}(x) \end{cases}$$

where the parameters η and q are choosen to satisfy the constraint $\mathsf{E}_{\theta_0}[\phi] = \alpha$

$\Rightarrow$ Likelihood ratio tests

Summarizing :

If known priors $p(H_i)$ and C_{ij} $\rightarrow$ Bayes approach $\rightarrow$ LRT with "well defined threshold" η : *Min-error or min-cost test*

If unknown priors $p(H_i)$ but known C_{ij} $\rightarrow$ Bayes approach $\rightarrow$ LRT with $\eta(p_0)$ where p_0 *minimizes the maximal (worst case) bayes cost*.

Otherwise $\rightarrow$ NP approach $\rightarrow$ LRT with η satisfying constraints on P_{FA} :
Most powerfull test of level $\alpha = P_{FA}$

Evaluation of the performances ?

B. Evaluating signal detectability, or the performances of a test

Assume that n indep. observations x_i are available ;
Take for the test statistics e.g ;

$$G = \sum_{i=1}^{n} \log L(x_i)$$

ROC computation requires the evaluation of

$$P_{FA} = \int_{\eta}^{\infty} p_{G,\theta_0}(g)dg \quad \text{and} \quad P_D = \int_{\eta}^{\infty} p_{G,\theta_1}(g)dg$$

where $p_{G,\theta_i}(g)dg$ = pdf of the test statistics G under hypothesis H_i.

In most cases : no closed form expressions for these integrals...

Note : For next section : Both hypothesis are simple

Signal detectability ?

a- ROC, either by direct calculation or by numerical methods

(not in this course) b- Methods relying upon limited expensions of the stats. around the normal distribution : $\sim$ *more accurate for large* n *;*
rely upon CLT $\rightarrow$ *accuracy obtained only if* $\eta \simeq \overline{G}$

c- Minimal performances Chernoff Bounds :
Give lower bound for P_D *and upper bound for* P_{FA}
Interesting relations with the frame of information theory

d- Getting some hint on the error decay rate when n increases ?
Stein's Lemma and Chernoff information

B.1. : Receiver operating characteristics : ROC curves

A good receiver : HIGH P_D, LOW P_{FA}

P_D, P_{FA} depends upon the threshold η only.

Definition : the ROC curve is the parametric curve
$P_D(\eta, q)$ plotted versus $P_{FA}(\eta, q)$.

Example : test on the mean of a normal RV (known identical variances under H_i)

The log-LRT is simply expressed by $x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$

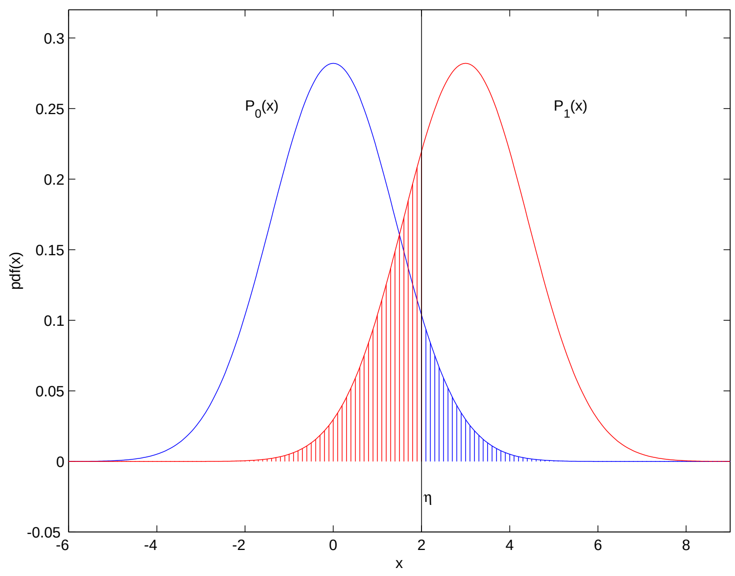
Hence,

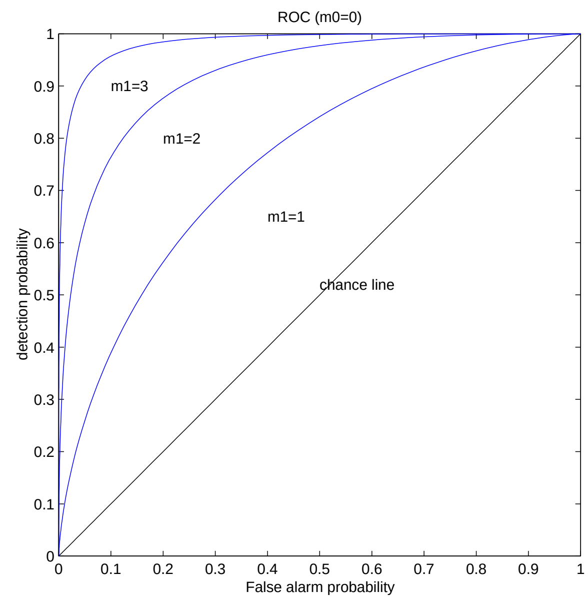
$$\begin{cases} P_{FA}(\eta) = \frac{1}{2}\left[1 - \text{erf}(\frac{\eta}{\sigma\sqrt{2}})\right] \\ P_D(\eta) = \frac{1}{2}\left[1 - \text{erf}(\frac{\eta - m}{\sigma\sqrt{2}})\right] \end{cases}$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$

note that for n multiple indep. observations, the test reads $\frac{1}{n}\sum_{i=1}^{n} x_i \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$ where $\frac{1}{n}\sum_{i=1}^{n} x_i$ is a normal R.V. with same mean as x and variance $\frac{\sigma^2}{n}$.

Example : ROC, test on mean of a normal R.V.





($\mu = m.\delta_{1i}$ under H_i).

ROC properties

• If $\eta \to \infty$, $P_D = 0$ and $P_{FA} = 0$; If $\eta \to -\infty$, $P_D = 1$ and $P_{FA} = 1$.

• ROC curve for a coin flip detector : diagonal line with unit slope ($\phi(x) = q$ indep. of the data) .

• ROC curve always lies above the chance line ; otherwise, flipping a coin has better performances.

• ROC curve of any LRT is concave (if the ROC was convex a randomized test would perform better.)

• For differentiable ROC $P_D(P_{FA})$, the threshold for the MP-LRT which lead to the perf. $P_{FA}(\eta), P_D(\eta)$ is given by (letting $g(x)$ stand for the LR in x)

$$g(\eta) = \frac{d}{dP_{FA}} P_D(P_{FA})$$

Testing the increase in rate λ of a Poisson R.V.

"H_1 : the rate is λ_1" vs "H_0 : the rate is λ_0" ($\lambda_1 > \lambda_0$)

$$L(x) = \left(\frac{\lambda_1}{\lambda_0}\right)^x \exp(\lambda_0 - \lambda_1)$$

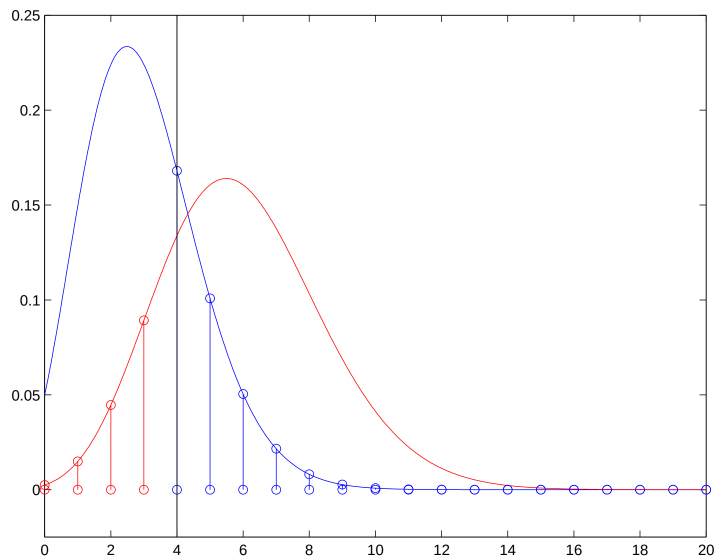
Taking the log-likelihood functin for the test statistics, one gets $x \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$

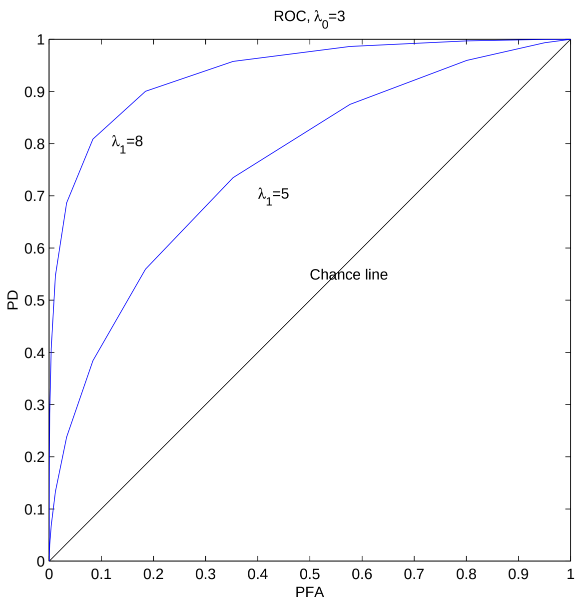
and

$$\begin{cases} P_{FA}(\eta) = 1 - \lambda_0 \sum_{x=0}^{\eta-1} \frac{\lambda_0^x}{x!} \\ P_D(\eta) = 1 - \lambda_1 \sum_{x=0}^{\eta-1} \frac{\lambda_1^x}{x!} \end{cases}$$

(notice that here both λ_0 and λ_1 are assumed to be known...)

test on the rate of Poisson R.V. : ROC





$\lambda_0 = 3$ (H_0) and $\lambda_1 = 5$ (H_1) ; left plot : ROC for two values of λ_1

B.2. Detectatibility and min performances, Chernoff Bounds

Let

$$P_{FA}(\eta) = \int_{\eta}^{\infty} p_{G,\theta_0}(g)dg = \int_{-\infty}^{\infty} U(g-\eta)p_{G,\theta_0}(g)dg$$

then, for $s \geq 0$

$$P_{FA} \leq \int_{-\infty}^{\infty} \exp((g-\eta)s)p_{G,\theta_0}(g)dg = \exp(\eta s)h(s)$$

where $h(s) = \mathrm{E}_G[\exp(-gs)]$: moment generating function of the pdf of the test statistics G.

Minimizing the rhs wrt s gives after some algebra the Chernoff Bound

$$\boxed{P_{FA} \leq \exp(\eta s_0)h(s_0) \quad \text{where} \quad \eta = \frac{h'(s_0)}{h(s_0)}}$$

Chernoff Bounds interpretation

Let $\mu(s) = \log(h(s))$, then

$$\eta = \frac{h'(s_0)}{h(s_0)} = \frac{d\mu(s)}{ds} = \mu'(s)$$

hence, taking for the test statistics $G(x) = \log(\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)})$, we get after some calculation

$$\begin{cases} P_{FA} \leq & \exp(\mu(s) - s\mu'(s)), \quad s \geq 0 \\ \mu(s) = & \log \int_{\mathcal{X}} \left(\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}\right)^s p_{\theta_0}^s(x)dx = (1-s)\mathsf{D}_s(p_{\theta_1}||p_{\theta_0}) \end{cases}$$

where D_s is the Rényi info. divergence of order s.

Similar derivation for P_D leads to

$$(1 - P_D) = P_M \leq \exp(\mu(s) + (1-s)\mu'(s)), \quad s \leq 1$$

Chernoff bounds are loose, alternative : study error decay rate

B.3. Hypothesis testing in an Information theoretic framework

Reformulate NP lemma (Simple hypothesis framework) :

Let $x \overset{d}{\simeq} Q$ and $X = \{x_1, \ldots, x_n\}$ a set of n i.i.d. samples, test $Q = P_0$ vs $Q = P_1$, and for $\eta \geq 0$

$$\mathcal{X}_1 = \left\{ \frac{P_1(X)}{P_0(X)} > \eta \right\}$$

Let $\alpha = P_{FA} = P_0(\mathcal{X}_1) = \int_{\mathcal{X}_1} P_0(X)dX$
and $\beta = P_M = P_1(\mathcal{X}_0) = P_1(\mathcal{X}_1^c) = \int_{\mathcal{X}_0} P_1(X)dX$

Define another acceptance region

$$\mathcal{X}_1^\star = \left\{ \frac{P_1(X)}{P_0(X)} > \eta^\star \right\}$$

with corresponding $(\alpha^\star, \beta^\star)$, then

$$\boxed{\alpha^\star \leq \alpha \Rightarrow \beta \leq \beta^\star \quad \text{equiv. to} \quad (\pi^\star = 1 - \beta^\star) \leq (\pi = 1 - \beta)}$$

where $\pi \overset{\text{def}}{=}$ power of the test

Alternate interpretation for LR test : IT Approach

Let $X \in \mathcal{X} = \Omega^n$ be a set of n i.i.d. samples ($x_i \in \Omega$), with probability distribution $P(x, \theta_j) = P_j(x)$ under H_j.

$$\log \frac{P_1(X)}{P_0(X)} = \log \prod_{i=1}^{n} \frac{P_1(x_i)}{P_0(x_i)} = \sum_{i=1}^{n} \log \frac{P_1(x_i)}{P_0(x_i)}$$

but for any function h of the r.v.

$$\sum_{i=1}^{n} h(x_i) = n \sum_{x \in \Omega} P_{x_i}(x) h(x_i)$$

where $P_{x_i}(x)$ is defined by

$$P_{x_i}(x) = \sum_{i=1}^{n} \frac{\delta(x - x_i)}{n}$$

- $P_{x_i}(x)$ is the empirical histogram

Alternate interpretation for LR test -continued-

From the previous expression of the average of any function of the r.v,

$$\begin{aligned}
\log \frac{P_1(X)}{P_0(X)} &= n \sum_{x\in\Omega} P_{x_i}(x) \log \frac{P_1(x)}{P_0(x)} \\
&= n \sum_{x\in\Omega} P_{x_i}(x) \log \frac{P_1(x) P_{x_i}(x)}{P_0(x) P_{x_i}(x)} \\
&= n \sum_{x\in\Omega} P_{x_i}(x) \log \frac{P_1(x)}{P_{x_i}(x)} - n \sum_{x\in\Omega} P_{x_i}(x) \log \frac{P_0(x)}{P_{x_i}(x)} \\
&= n\mathsf{D}(P_{x_i}(x) \parallel P_0(x)) - n\mathsf{D}(P_{x_i}(x) \parallel P_1(x))
\end{aligned}$$

and finally

$$\boxed{\mathsf{D}(P_{x_i}(x) \parallel P_0(x) - \mathsf{D}(P_{x_i}(x) \parallel P_1(x)) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{n} \log \eta}$$

• The accepted hypothesis is chosen according to the minimal divergence between the empirical law and the law under each hypothesis.

Insights from IT

Reminder : Sanov and conditional limit theorems

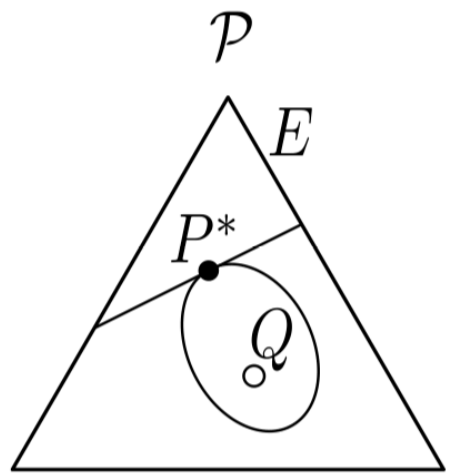
Let $x \overset{d}{\simeq} Q$ and $E \subseteq \mathcal{P}$ and

$P^\star = \operatorname*{argmin}_{P \in E} \mathsf{D}(P \parallel Q)$

Sanov : $Q^n(E) \doteq \mathrm{e}^{-n\mathsf{D}(P^\star \| Q)}$

Cond Limit : $\Pr(x = a | P_{x_i} \in E) = P^\star(a)$

in propability for large n

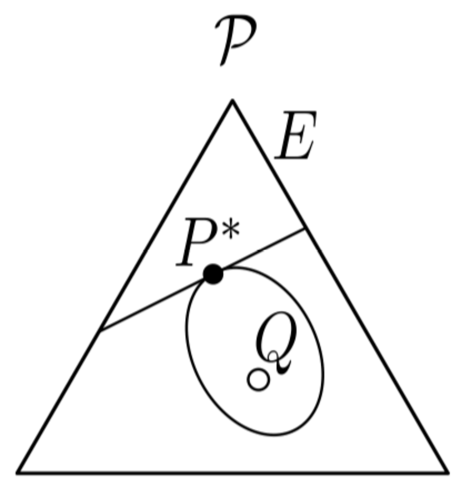


This means

- Probability of E is close to proba. of $P^\star$ (Sanov)
- Total probability of $P_{x_i} \in E$ far from $P^\star$ is negligible (Cond.Lim)
- With high probability, P_{x_i} is close to $P^\star$ (Cond. Lim.)

Sanov and conditional limit theorems : some consequences

$\log \mathrm{LRT} = \log \frac{P_1(X)}{P_0(X)}$
$= \sum_{x\in\Omega} P_{x_i}(x) g(x)$
where $g(x) = n \log \frac{P_1(x)}{P_0(x)}$



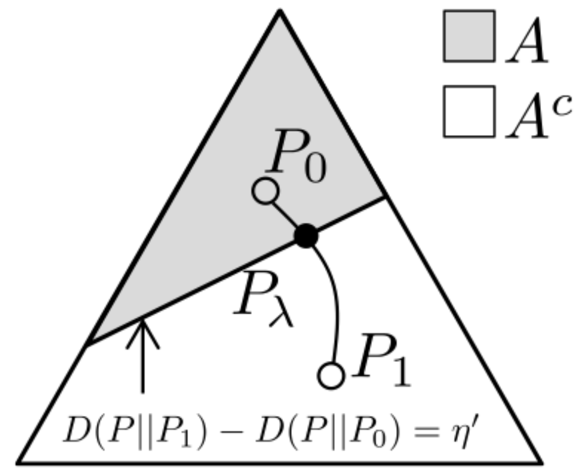
Let $E = \left\{P : P^T g = \sum_{x\in\Omega} P(x) g(x) \geq \eta\right\}$, then, using Lagrange multipliers

$$P^\star(x) = \arg \min_{\substack{P^T g \geq \eta \\ \sum_a P(a)=1}} \mathrm{D}(P \parallel Q) = c Q(x) \mathrm{e}^{\lambda g(x)}$$

where λ satisfies $P^T g = \eta$ and c is a normalizing constant.

$$\mathsf{LRT}(X) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \Leftrightarrow P_{x_i}^T g \underset{H_0}{\overset{H_1}{\gtrless}} \eta'$$
$$A = \left\{P : P_{x_i}^T g < \eta'\right\}$$

From Sanov th. (note A is convex) :



$$\alpha = P_0(P_{x_i} \in A^c) \doteq \mathrm{e}^{-n\mathsf{D}(P_0^\star||P_0)}$$
$$\beta = P_1(P_{x_i} \in A) \doteq \mathrm{e}^{-n\mathsf{D}(P_1^\star||P_1)}$$

and from result of the previous slide, we get

$$P_1^\star = cP_1\mathrm{e}^{\lambda' g} = cP_1\mathrm{e}^{\lambda' \log \frac{P_1}{P_0}} = cP_1^{1+\lambda'} P_0{}^{\lambda'} = cP_1^{\lambda} P_0^{1-\lambda} = P_0^\star = P_\lambda$$

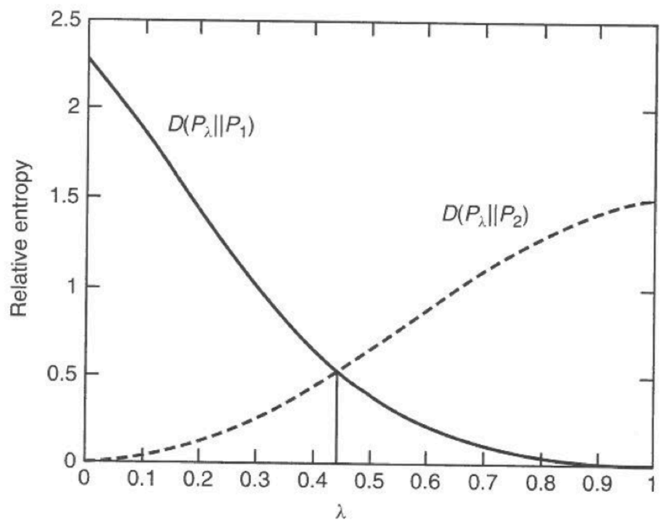
Chernoff information

Question : what is the best error exponent for the Bayes criterion ? (assume that $p(H_0)$ is known)

$$\begin{aligned} P_E &= p(H_0)P_{FA} + p(H_1)P_M \\ &= p(H_0)\alpha + (1 - p(H_0)\beta \end{aligned}$$

We have seen that LRT is optimal in this case,
so previous results apply :

$$\begin{cases} P_E &\doteq p(H_0)\mathrm{e}^{-n\mathrm{D}(P_\lambda \| P_0)} + (1 - p(H_0))\mathrm{e}^{-n\mathrm{D}(P_\lambda \| P_1)} \\ &\doteq \mathrm{e}^{-n\min\{\mathrm{D}(P_\lambda \| P_0), \mathrm{D}(P_\lambda \| P_1)\}} \\ P_\lambda &= cP_1^\lambda P_0^{1-\lambda} \end{cases}$$



Min in the decay rate exponent occurs when

$$\mathrm{D}(P_\lambda \parallel P_1) = \mathrm{D}(P_\lambda \parallel P_0) \stackrel{\text{def}}{=} C(P_1, P_0)$$

Equivalently

$$C(P_1, P_0) = -\min_{0 \le \lambda \le 1} \log \left(\sum_{a \in \Omega} P_1^\lambda(a) P_0^{1-\lambda}(a) \right)$$

$P(H_0)$ unknown : non Bayesian case, the Stein Lemma

Reminder

$X = \{x_1, \dots, x_n\}$ where $x_i \overset{d}{\simeq} Q$, i.i.d and $Q \in \{P_1, P_0\}$

$\alpha = P_{FA} = P_0(\mathcal{X}_1)$, $\beta = P_M = P_1(\mathcal{X}_0)$

<u>Stein Lemma</u> (csq of Sanov th.) :

$\forall 0 < \varepsilon < \frac{1}{2}$,

define $\beta_\varepsilon = \min\limits_{\substack{\alpha<\varepsilon \\ \mathcal{X}_1}} \beta$, then

$$\lim_{n\to\infty} \log \beta_\varepsilon = -\mathrm{D}(P_0 \parallel P_1)$$

1. $\mathrm{D}(P_0 \parallel P_1)$ is the best β exponent
2. if instead $\beta < \varepsilon$ then
$$\lim_{n\to\infty} \log \alpha_\varepsilon = -\mathrm{D}(P_1 \parallel P_0)$$
3. the most different P_1 is from P_0, the easier is the pb
4. !! $\quad \mathrm{D}(P_0 \parallel P_1) \neq \mathrm{D}(P_1 \parallel P_0)$

Summary bounds on exponential decay rate

Distribution	KL (Stein) $\mathsf{D}(P_0 \parallel P_1)$ $\beta \doteq \mathrm{e}^{-n\mathsf{D}(P_0\|P_1)}$	Chernoff $C(P_1, P_0)$ $P_E \doteq \mathrm{e}^{-nC(P_1,P_0)}$
Gaussian (m_1, m_0, σ^2)	$\frac{(m_1-m_0)^2}{2\sigma^2}$	$\frac{(m_1-m_0)^2}{8\sigma^2}$
Poisson $r = \frac{\lambda_0}{\lambda_1}$	$\lambda_1(1 - r + r\log r)$	$\lambda_0 \frac{(r-1)\left\{\log\left(\frac{r-1}{\log r}\right)-1\right\}+\log r}{\log r}$

C - Composite hypothesis testing

Up to now : θ_0 or θ_1 could assume only one value : conditional pdfs were precisely known under each H_i.

this is however not the case in many applications

Example : Test presence vs absence of a sinusoïdal signal with unknown phase embedded in noise.

$\rightarrow$ The pdfs under H_i depends on the values taken by θ_i :
These random parameters are included in the hypotheses to be tested.

Let us consider the composite hypothesis testing problem

$$H_0 : f(x|H_0) = f_0(x|\theta) = f_{\theta_0}(x), \theta \in \Theta_0$$
$$H_1 : f(x|H_1) = f_1(x|\theta) = f_{\theta_1}(x), \theta \in \Theta_1$$

where θ_0 and θ_1 are unknown parameters vectors (possibly with comon comp.)

$\Rightarrow$ HOW DO BAYES, nP, LRT approaches generalize ?

Bayes cost for cht

$$\begin{aligned}\overline{C} = \ & p(H_0)\int_{\mathcal{X}_0}\int_{\Theta_0} f_{\theta_0}(x)p_0(\theta_1)C_{00}(\theta)d\theta dx \\ & + p(H_0)\int_{\mathcal{X}_1}\int_{\Theta_0} f_{\theta_0}(x)p_0(\theta)C_{10}(\theta)d\theta dx \\ & + p(H_1)\int_{\mathcal{X}_0}\int_{\Theta_1} f_{\theta_1}(x)p_1(\theta)C_{01}(\theta)d\theta dx \\ & + p(H_1)\int_{\mathcal{X}_1}\int_{\Theta_1} f_{\theta_1}(x)p_1(\theta)C_{11}(\theta)d\theta dx\end{aligned}$$

but ϕ verifies

$$\begin{aligned}\int_{\mathcal{X}_1} f_{\theta_0}(x)dx = \ & 1 - \int_{\mathcal{X}_0} f_{\theta_0}(x)dx \\ \int_{\mathcal{X}_1} f_{\theta_1}(x)dx = \ & 1 - \int_{\mathcal{X}_0} f_{\theta_1}(x)dx\end{aligned}$$

thus

$$\begin{aligned}\overline{C} = \ & p(H_0)\int_{\Theta_0} p_0(\theta)C_{10}(\theta)d\theta + p(H_1)\int_{\Theta_1} p_1(\theta)C_{11}(\theta)d\theta \\ & + \int_{\mathcal{X}_0}\Big[p(H_1)\int_{\Theta_1} f_{\theta_1}(x)p_1(\theta)\left[C_{01}(\theta) - C_{11}(\theta)\right]d\theta \\ & - p(H_0)\int_{\Theta_0} f_{\theta_0}(x)p_0(\theta)\left[C_{10}(\theta) - C_{00}(\theta)\right]d\theta\Big]dx\end{aligned}$$

and the Bayes optimal test is

$$\boxed{\frac{\int_{\Theta_1} f_{\theta_1}(x)p_1(\theta)[C_{01}(\theta)-C_{11}(\theta)]d\theta}{\int_{\Theta_0} f_{\theta_0}(x)p_0(\theta)[C_{10}(\theta)-C_{00}(\theta)]d\theta} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)}}$$

notice : if C_ij do not depend upon θ , as

$$f_{\theta_1}(x)p_1(\theta) = f_1(x|\theta)p_1(\theta) = f_1(x,\theta)$$

and

$$f_{\theta_0}(x)p_0(\theta) = f_0(x|\theta)p_0(\theta) = f_0(x,\theta)$$

the test takes the bayes sht expression, after the density marginals over θ are computed.

Example Let $x \in \mathcal{X}$ be a set of n i.i.d. samples from a normal r.v. with mean θ_1 and variance σ^2 :

Test the simple null hypothesis H_0 : $\theta_1 = 0$:

$$H_0 : f(x|H_0) = f_0(x) = \frac{1}{(\sigma\sqrt{2\pi})^n} \mathrm{e}^{-\frac{\sum_{i=1}^{n}(x_i)^2}{2\sigma^2}}$$

against

H_1 : $\theta_1 \neq 0$:

$$H_1 : f(x|H_1) = f_{\theta_1}(x) = \frac{1}{(\sigma\sqrt{2\pi})^n} \mathrm{e}^{-\frac{\sum_{i=1}^{n}(x_i-\theta_1)^2}{2\sigma^2}}$$

Bayes decision costs are $C_{00} = C_{11} = 0, , C_{01}(\theta_1) = 1, C_{10}(\theta_1) = k, \forall \theta_1$

A priori knowledge about θ_1 is given by

$$p_1(\theta_1) = \frac{1}{m}, -m \leq \theta_1 \leq m$$

and we know $p(H_0) = p(H_1) = 1/2$

Solution

The sufficient statistics $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is used. $\overline{x}$ is a normal r.v. with variance $\frac{\sigma^2}{n}$ and mean θ_1 under H_1, zero under H_0.

Applying the preceding general results :

$$\frac{\int_{\Theta_1} \frac{\sqrt{n}}{(\sigma\sqrt{2\pi})} \mathrm{e}^{-\frac{n(x_i-\theta_1)^2}{2\sigma^2}} \frac{1}{m} d\theta}{k \frac{\sqrt{n}}{(\sigma\sqrt{2\pi})} \mathrm{e}^{-\frac{n x_i^2}{2\sigma^2}}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

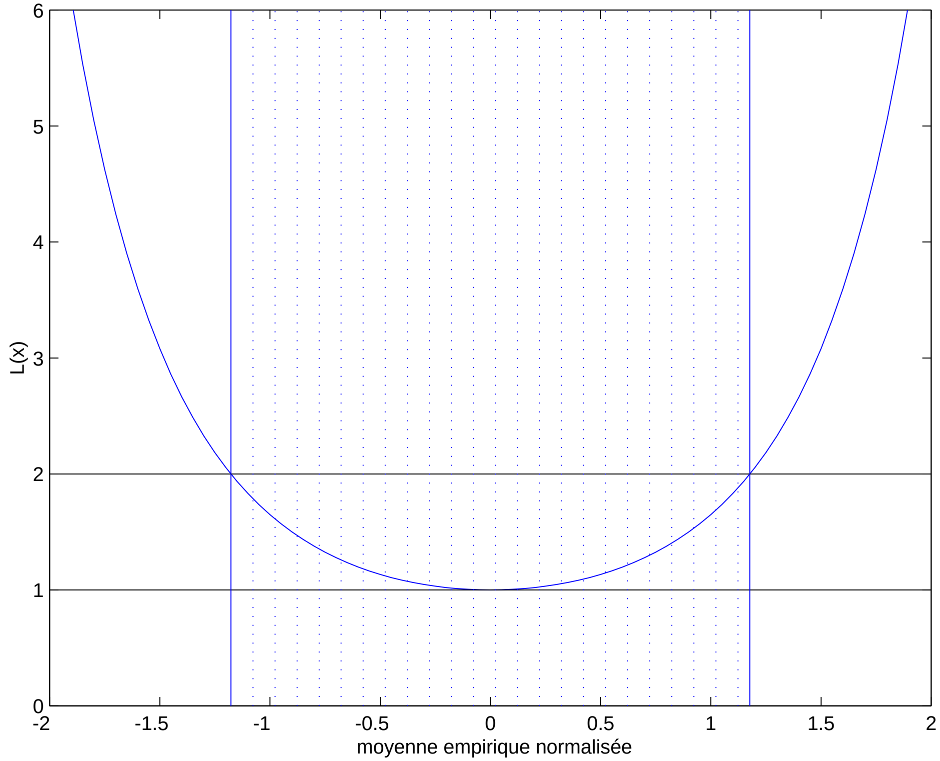
and after some algebra

$$L(\overline{x}) = \mathrm{e}^{\frac{x^2}{2}} \left(F(\frac{\sqrt{n}(m-\overline{x})}{\sigma}) - F(\frac{\sqrt{n}(m+\overline{x})}{\sigma}) \right) \underset{H_0}{\overset{H_1}{\gtrless}} k$$

where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} dt$$

is the Laplace Gauss pdf.



Rapport de vraisemblance pour le test optimal de Bayes trait dans l'ex. précédent. ($\sigma^2 = 2; n = 8$,$k = C_{10} = 2$). La zone grisée correspond à la région de décision $\theta_1 = 0$. Pour ce problème, le test de minimum de probabilité d'erreur de décision ($k = 1$) conduit à ne retenir H_0 que si $\overline{x} = 0$.

C.1. UMP Test : definitions, existence

In general, no basis for picking θ_1 : UMP is a procedure to test composite hypothesis

Def : ϕ is the *Most Powerful Test of size* $\alpha(\phi)$ if $\forall \phi'$ of size α, then $\beta(\phi) \leq \beta(\phi')$
Def : ϕ is the *Most Powerful Test of level* α, i.e. such that $\alpha(\phi) \leq \alpha$, if $\forall \phi'$ of level α (i.e. $\alpha(\phi) \leq \alpha$), then $\beta(\phi) \leq \beta(\phi')$

def : $\phi^\star$ is UMP of level α if for any level α test or decision function ϕ, the power $(P(\widehat{H}_1|H_1) = 1 - \beta)$ of the test verifies for all $\theta \in \Theta_1$,

$$\mathcal{P}^\star_\alpha = 1 - \beta^\star(\theta_1) = \mathsf{E}_{(\theta_1)}[\phi^\star] \geq \mathsf{E}_{(\theta_1)}[\phi] = 1 - \beta(\theta_1)$$

$\rightarrow$ the test maximizes the detection probability (or power) independant of the value θ_1.

An UMP test does not always exist (ex. testing non-zero mean of a R.V. :the sign cannot be integrated in any decision function).

• Necessary condition for existence of UMP : the likelihood must be a monotone increasing function of a sufficient test statistics.

def : Monotone Likelihood Ratio (MLR)
A model $P(X|\theta)$ (with θ is real valued), has Monotone Likelihood Ratio if there exists a real valued function $u(X)$ such that $\forall \theta_1 > \theta_0$,
$LR = \frac{P(X|\theta_1)}{P(X|\theta_0)}$ is monotone increasing fonction of $u(X)$ on
$\{X : L(X|\theta_1) > 0 \text{ and } L(X|\theta_0) > 0\}$

Then the higher the observed $u(X)$, the more likely X was drawn from $P(X|\theta_1)$ rather than from $P(X|\theta_0)$

Example (MLR), the exponential family

$$P(X, \theta) = \exp\{\sum_i C(x_i) - A(\theta) \sum_i T(x_i) + nB(\theta)$$

$$\text{then } \log LR = \sum_i T(x_i)[A(\theta_1) - A(\theta_0)] + n[B(\theta_1) - B(\theta_0)]$$

Let $u(X) = \sum_i T(x_i)$, then

$$\frac{\partial \log LR}{\partial u} = [A((\theta_1) - A(\theta_0)] > 0$$

if $A(.)$ is monotonic in θ. Note that $u(X)$ is a sufficient statistic.

• Normal (known σ) , Poisson, binomial, exponential are MLRP with $u(X) = \sum_i x_i$.

Example Consider the preceding problem studied as an example, but now, no a priori knowledge about θ_1 is available.

The log-LRT for this problem is

$$\sigma^2 n^{-1} \log L(x) = \frac{\theta_1}{n} \sum_{i=1}^{n} x_i - \frac{\theta_1^2}{2} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where the threshold η depends upon the adopted strategy for the test. Equivalently, the test is summarized by

$$\theta_1 \overline{x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta'$$

with $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, gaussian r.v. $\mathcal{N}(\theta, \frac{\sigma^2}{n})$ ($\theta_1 \neq 0$ under H_1).

Single sided test

The test is derived with the assumption that $\theta_1 > 0$

From the preceding, and by normalizing the statistics $\overline{x}$, one gets

$$T = \frac{\sqrt{n}\overline{x}}{\sigma} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

The most powerful test of level at most α allows to determine γ by

$$P_{FA} = \alpha = P_0(\frac{\sqrt{n}\overline{x}}{\sigma} > \gamma)$$

and therefore

$$\alpha = 1 - F(\gamma), \quad \gamma = F^{-1}(1-\alpha),$$

Finaly

$$\frac{\sqrt{n}\overline{x}}{\sigma} \underset{H_0}{\overset{H_1}{\gtrless}} F^{-1}(1-\alpha)$$

- The decision function obtained does not depend upon θ_1, and the power is maximal (among all test of level at most α) : this test is UMP under assumption $\theta_1 > 0$.

- $d = \frac{\theta_1 \sqrt{n}}{\sigma}$ appears as a correction to the threshold for the test applied to normalized statistics. d is sometimes referred to as *the detectability index*

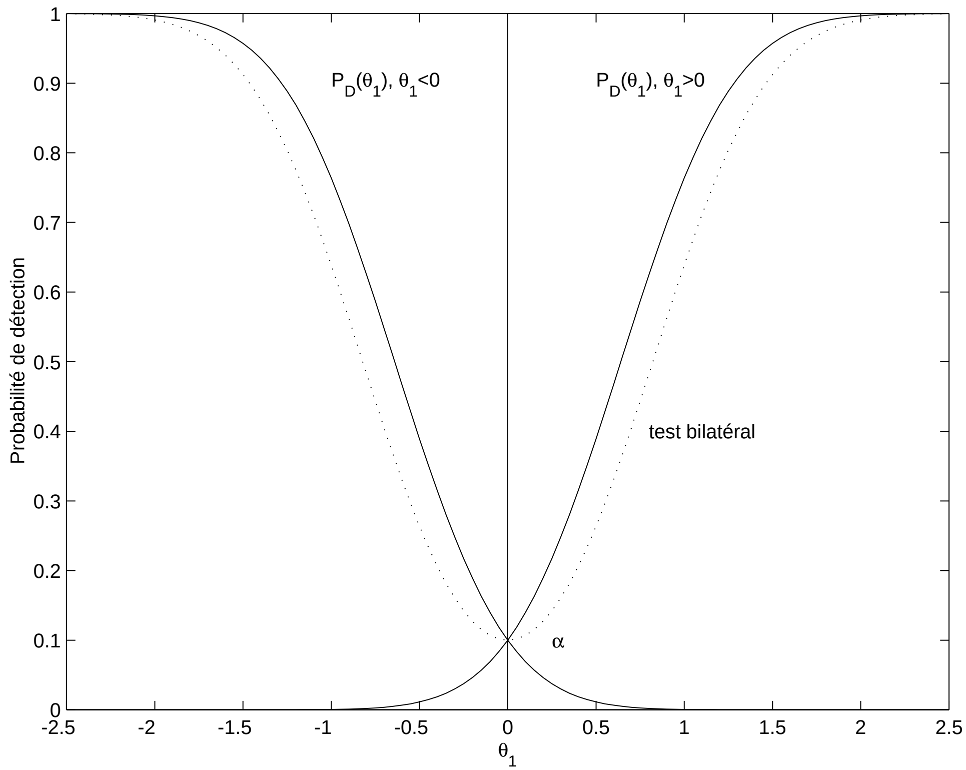
$$d = \frac{\mathbb{E}\left[T|H_1\right] - \mathbb{E}\left[T|H_0\right]}{\sqrt{\mathrm{var}_0(T)}}$$

where $\mathrm{var}_0(T)$ is the variance of T under H_0. Incresed d leads to better test.

- Single sided test under assumption $\theta_1 < 0$ leads to symetrical results :

$$\frac{\sqrt{n}\overline{x}}{\sigma} \underset{H_1}{\overset{H_0}{\gtrless}} F^{-1}(1-\alpha)$$

- The power curve for the SS test ($\theta_1 > 0$) crosses the vertical axis at $P_D(\theta_1 = 0) = P_{FA}$.



Test power for the ”non zero mean” detection problem θ_1 nof a random Gaussian process with known variance . Solid lines are obtained for single sided UMP tests under each hypothesis for the sign of θ_1. Dashed line show the sub-optimal double-sided test power for a fixed P_{FA}. That test does not require any hypothesis on the sign of θ_1 ($\sigma^2 = 2; n = 8$).

Double sided test : $\theta \neq 0$

Motivation : if the assumption e.g. $\theta_1 > 0$ is wrong, the single sided resulting test is BIASED :

A decision test is said to be biaised if for some values of the parameters (say θ), $P_{FA}(\theta) > P_D(\theta)$.

Approach :

$$|\frac{\sqrt{n}\overline{x}}{\sigma}| \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

For this test $P_{FA} = \alpha = 2(1 - F(\gamma))$, and therefore

$$\gamma = F^{-1}(1 - \frac{\alpha}{2})$$

and the power of the test is

$$P_D(\theta_1) = 1 - \left(F\left(\gamma - \frac{\theta_1\sqrt{n}}{\sigma}\right) - F\left(-\gamma - \frac{\theta_1\sqrt{n}}{\sigma}\right)\right)$$

see the resulting powre curve on preceding figure.

C.2. Possible strategies : general approaches and insights

• Define a locally optimal test, i.e. for small range of values of the unknown parameters : $\theta \simeq \theta_0$

• Define a restricted subset of decision functions (e.g. leading to unbiased tests) the within which the solution is derived from some optimality criterion ;

• Determine the least favorable prior distribution $w(\theta)$, to minimize the power of the test :

$$\beta(w) = \int_{\Theta_1} \int_{\mathcal{X}_1} p(x|\theta) w(\theta) d\theta dx$$

... most often very difficult to find.

• Popular alternative : Generalized Likelihood ratio test

The unknown parameters are replaced by some estimated values

LMP single sided Test

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0 \text{ but } \theta \simeq \theta_0$$

Expand the power around θ_0 :

$$P_D(\phi, \theta) = \int_{\mathcal{X}_1} f(x, \theta)\phi(x)dx \simeq P_D(\phi, \theta_0) + (\theta - \theta_0) \left.\frac{\partial P_D(\phi, \theta)}{\partial \theta}\right|_{\theta_0}$$

where $P_D(\phi\theta_0) = P_{FA}(\phi)$; Thus, maximize $Pd(\phi, \theta) \Leftrightarrow$ maximize

$$\left.\frac{\partial P_D(\phi, \theta)}{\partial \theta}\right|_{\theta_0} = \left.\frac{\partial \int_{\mathcal{X}_1} \phi(x) f(x, \theta) dx}{\partial \theta}\right|_{\theta_0} = \int_{\mathcal{X}_1} \phi(x) \left.\frac{\partial f(x, \theta)}{\partial \theta}\right|_{\theta_0} dx$$

Using Lagrange multipliers,

$$\mathcal{L}(\phi) = \int_{\mathcal{X}_1} \phi(x) \left.\frac{\partial f(x,\theta)}{\partial \theta} dx\right|_{\theta_0} - \eta(1 - \int_{\mathcal{X}_1} \phi(x) f_0(x) dx - \alpha)$$
$$= \int_{\mathcal{X}_1} \phi(x) \left[\left.\frac{\partial f(x,\theta)}{\partial \theta} dx\right|_{\theta_0} - \eta f_0(x) \right] dx - \eta\alpha$$

which is max if the braketed term is always > 0 ;

LMP single sided test, cont'nd

the test is therefore given by

$$\frac{\left.\frac{\partial f_1(x,\theta)}{\partial \theta}\right|_{\theta_0}}{f_0(x)} = \left.\frac{\partial \log f(x,\theta)}{\partial \theta}\right|_{\theta_0} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

• H_0 is accepted if the log-likelihood is close to a stationnary point, i.e. the ML estimate of θ_0.

• This test recovers exactly the solution exhibited for UMP single side test in the previous example.

The solution is seeked among the set of unbiaised tests

$\Rightarrow$ add the unbiaseness constraint !

Property : If ϕ leads to un unbiased test, then $P_D(\phi, \theta)$ has a global minimum for $\theta = \theta_0$.

The constraint may then be expressed

$$\left.\frac{\partial \mathsf{E}_1(\phi)}{\partial \theta}\right|_{\theta_0} = \left.\frac{\partial \int_{\mathcal{X}_1} \phi(x) f(x, \theta) dx}{\partial \theta}\right|_{\theta_0} = 0$$

and in order to maximize $P_D(\phi, \theta)$ when θ varies :

$$\left.\frac{\partial^2 \mathsf{E}_1(\phi)}{\partial \theta^2}\right|_{\theta_0} = \left.\frac{\partial^2 \int_{\mathcal{X}_1} \phi(x) f(x, \theta) dx}{\partial \theta^2}\right|_{\theta_0} \quad \text{Max}$$

Construct the Lagrange function :

$$\begin{aligned}\mathcal{L}(\phi) = \; & \int_{\mathcal{X}_1} \phi(x) \left.\frac{\partial^2 f(x,\theta)}{\partial \theta^2}\right|_{\theta_0} dx \\ & -\lambda \left(1 - \int_{\mathcal{X}_1} \phi(x) f(x,\theta) dx - \alpha\right) - \eta \int_{\mathcal{X}_1} \left.\frac{\partial f(x,\theta)}{\partial \theta}\right|_{\theta_0} dx \\ = \; & \int_{\mathcal{X}_1} \phi(x) \left[\left.\frac{\partial^2 f(x,\theta)}{\partial \theta^2}\right|_{\theta_0} - \lambda f_0(x) - \eta \left.\frac{\partial f(x,\theta)}{\partial \theta}\right|_{\theta_0} \right] dx + \lambda\alpha - \lambda\end{aligned}$$

which is max if

$$\frac{\left.\frac{\partial^2 f(x,\theta)}{\partial \theta^2}\right|_{\theta_0}}{\left.\frac{\partial f(x,\theta)}{\partial \theta}\right|_{\theta_0} + \rho f(x,\theta)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

where η et $\rho = \lambda/\eta$ are set to match the constraints.

Form the test on the means of normal random variables (see preceding example),

$$f(x,\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n(\overline{x}-\theta)^2}{2\sigma^2}}$$

then by subtitution

$$\frac{\overline{x}^2 - \sigma^2/n}{\sigma^2/n - \rho\overline{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

and fixing $\rho = 0$, one gets

$$|\overline{x}| \underset{H_0}{\overset{H_1}{\gtrless}} \sigma\sqrt{\eta+1}/\sqrt{n} = \gamma\sigma/\sqrt{n}$$

which was previously obtained...

LMP Double sided Test, cont'nd

These results generalize starightforwardly ;-) to the multiple unknown parameters case :

unbiased test constraint :

$$\| \nabla \mathsf{E}_{\vec{\theta}}[\phi]\Big|_{\vec{\theta}_0} \| = 0$$

Maximal concavity constraint :

$$\mathrm{tr}\, \nabla^2 \mathsf{E}_{\vec{\theta}}[\phi]\Big|_{\vec{\theta}_0} \ \mathsf{Max}$$

Resulting test

$$\frac{\mathrm{tr}\, \nabla^2 f(x,\theta)\Big|_{\vec{\theta}_0}}{f_0(x) + \rho \, \| \nabla f(x,\theta)|_{\vec{\theta}_0}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

Neyman-Pearson MinMax test

IDEA : Maximize the power of the test for the least favorable distribution of the unknown parameters

Find $p_0^*(\theta)$ and $p_1^*(\theta)$ in order to

$$\text{Maximize} \int_{\Theta_1} \int_{\mathcal{X}_1} f_1(x|\theta) p_1^*(\theta) d\theta$$

with the constraint on the level of the test

$$\int_{\Theta_0} \int_{\mathcal{X}_1} f_0(x|\theta) p_0^*(\theta) d\theta \leq \alpha$$

Problems

• if $p_0^*(\theta)$ concentrates upon very unlikely (atypical) values, the test may be very poor (low power).

• $p_0^*(\theta)$ and $p_1^*(\theta)$ may extremely difficult to find.

Classical example : detection of a sinusoidal signal with unknown phase ψ ; $\psi \in [0, 2\pi]$ is the least favorable dist.

Generalized LRT

- "ad-hoc" detector
- "plug-in" receiver
- in general, no optimilaty is asserted

most commonly : the unknown parameters are replaced by their "maximum likelihood" estimates

$$L_{GLR} = \frac{\max_{\theta \in \Theta_1} p(x,\theta)}{p_{\theta_0}(x)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$$

(notice that in our case H_0 is a simple hypothesis, and no plug in estimate is necessary under H_0).

GLRT asymptotics

— As the nb. of indep. obs. $n \to \infty$, the ML estimate θ_{1MLE} is a <u>consistent</u> estimator of θ_1, the GLRT is asymptotically UMP.

— The GLR statistic has a Chi-square limiting distribution : if $p_{\theta_0}(x)$ is smooth (under H_0), it can be shown that for large n

$$2 \log L_{GLR} \sim \mathcal{X}_p^2$$

p= nb. of components of the unknown parameter θ_1, which are relevant in for the test under consideration

GLRT asymptotics : proof of convergence toward the $\mathcal{X}^2$ law

(H_0 is simple and θ_0 is a scalar)

•Let $\widehat{\theta}_n$ an ML estimate from a set x of n i.i.d. observations.

$$\begin{aligned}\log \mathcal{L}(x,\theta_0) &= \log \mathcal{L}(x,\widehat{\theta}_n) \\ &+(\theta_0-\widehat{\theta}_n)\frac{\partial}{\partial\theta}\log \mathcal{L}(x,\widehat{\theta}_n)+\frac{1}{2}(\theta_0-\widehat{\theta}_n)^2\frac{\partial^2}{\partial\theta^2}\log \mathcal{L}(x,\widehat{\theta}^\star)\end{aligned}$$

where $\theta^\star \in [\theta_0, \widehat{\theta}_n]$.

• $\widehat{\theta}_n$ is the ML estimate $\Rightarrow \frac{\partial}{\partial\theta}\log \mathcal{L}(x,\widehat{\theta}_n)=0$

•By the law of large numbers, noticing that MLE is consistent, for $n\to\infty$,

$$\begin{aligned}&\widehat{\theta}_n \to \theta_0 \\ &\widehat{\theta}^\star \to \theta_0 \\ &\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log f(x_i,\theta^\star) \to \mathrm{E}_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}\log f_0(x)\right] = -\mathrm{I}_0\end{aligned}$$

where I_0 is the Fisher information of the sample for estimating θ_0.

GLRT asymptotics : proof of convergence, continued.

From estimation theory :

$$\sqrt{n\mathrm{I}_0}(\widehat{\theta}_n - \theta_0) \to \mathcal{N}(0, 1)$$

thus

$$n\mathrm{I}_0(\widehat{\theta}_n - \theta_0)^2 \to \mathcal{X}_1^2$$

and by inserting this in the previous expansion of the log-likelihood

$$2 \log \mathcal{L}(x, \theta_0) \to \mathcal{X}_1^2$$

if $\mathcal{L}$ is not a smooth function of θ_0.... :-(

Approximate statistics with e.g. Gram Charlier or Edgeworth expansions

This course : test $H0$ agains $H1$, in a single (or multiple, with not that many) hypothesis framework

What happens for hundreds or thousands of possible instances, possibly departing from $H0$?

Next Lecture : multiple hypothesis testing and false discovery rate control

References, textbooks

A.D. Whalen, ”Detection of signal in noise”, Academic Press,1971.

N.L .Johnson, S. Kotz, A.W.Kemp, ”Univariate discrete distribution”, Wiloey and Sons, 1992

P.J.Bickel, K.A.Docksum, ”Mathematical Statistics” vol.1, Prentice Hall, 2001

T. Cover, J.A.Thomas, ”Information Theory”, Wiley Series in Telecommunications, 1991.

C.Fourgeaud, A. Fuchs, ”Statistique”, Collection Universitaire de Mathematiques Dunod, vol.24, 1967.

C.W.Helstrom, ”Elements of Signal Detection and Estimation”,Prentice Hall, Englewood Cliffs, 1995.

S.M.Kay, ”Fundamentals of Statistical Signal Processing”, vol.2, Prentice Hall, 1998.

A.Papoulis,”Probability, random variables and stochastic processes”, McGraw Hill International, 1991.

H.V.Poor, ”An introduction to Signal Detection and Estimation”, Springer, 1994.

L.L.Scharf, ”Statistical Signal Processing : Detection, estimation and time series”, Addison Wesley, 1991

H.L.Van Trees, ”Detection, Estimation and Modulation Theory”, Wiley and Sons, 1968.