TRATAMIENTO ESTADÍSTICO DE SEÑALES

• VARIABLES Y PROCESOS ALEATORIOS – CONCEPTOS

• Variable aleatoria X

• fdp. $p(x)$

valor medio : $\mu = E[x]$

pot. media : $P = E[x^2]$

varianza : $\sigma^2 = E[(x - E[x])^2]$

$$\boxed{P = (\mu)^2 + (\sigma^2)}$$

$$E[x] = \int x \cdot p(x)\, dx \quad \text{valor medio de } x$$

$$E[f(x)] = \int f(x)\, p(x)\, dx \quad \text{de } f(x)$$

$$E[x|y] = \int x \cdot p(x|y)\, dx$$

$$E[f(x_1,x_2)] = \int f(x_1,x_2) \cdot p(x_1,x_2)\, dx_1 dx_2$$

• fdp 2D

$p(x,y)$

correlación $E[x \cdot y^*]$

covarianza $E[(x - E[x]) \cdot (y - E[y])^*]$

indep $\leftrightarrow$ $p(x,y) = p_1(x) \cdot p_2(y)$

$E[x \cdot y] = E[x] \cdot E[y]$

• Proceso estocástico X(n)

· estacionario ⇒

media $\mu = E[x(n)]$

pot. media $P = E[x^2(n)]$

varianza $\sigma^2 = E[(x(n) - \mu)^2]$

autocorrelación

$$R(m) = E[x(n+m) \cdot x(n)^*]$$

$$P = R(m=0)$$

Dominio Frecuencial:

$$R(m) \xrightarrow{TF} S(\Omega) \quad \text{densidad espectral de potencia}$$

$$P = \frac{1}{2\pi}\int_0^{2\pi} S(\Omega)\, d\Omega$$

Filtrado de señales

$$x(n) \longrightarrow \boxed{\begin{matrix} h(n) \\ H(\Omega) \end{matrix}} \longrightarrow y(n)$$

$$\mu_y = \mu_x \cdot H(0)$$

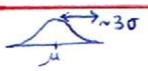
$$S_y(\Omega) = S_x(\Omega)\, |H(\Omega)|^2$$

$$R_y(m) = R_x(m) \underset{\text{conv.}}{*} h(m) * h^*(-m)$$

$$R_{yx}(m) = R_x(m) * h(m)$$

• V.A. Gaussiana

$x : N(\mu, \sigma^2)$



$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

• Ruido blanco media nula

$\mu = 0$

$$R(m) = \begin{cases} P & m = 0 \\ 0 & \text{resto} \end{cases} \qquad S(\Omega) = P$$

• V.A. Chi-cuadrado con N grados de lib

Elevar al cuadrado un vector de N gaussianas $\quad x = \underline{Z} \cdot \underline{Z}^T$

$$\boxed{x = \sum_{i=1}^{N} z_i^2}$$

$$p(x|H_i) = \frac{x^{(\frac{N}{2}-1)}}{2^{\frac{N}{2}}\sigma_i^N \Gamma(\frac{N}{2})} \exp\left(-\frac{x}{2\sigma_i^2}\right) \overbrace{u(x)}^{\text{escalon}}$$

si $N = 2 \rightarrow p(x|H_i) = \frac{1}{2\sigma^2} \exp\left(-\frac{x}{2\sigma^2}\right) u(x)$

· Función de V.A.

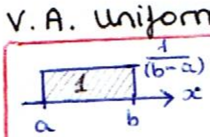
$y = f(x) \longrightarrow p_x(x)$

$x = g(y)$

$$p_y(y) = p_x(x) \cdot \underbrace{\frac{d\,g(y)}{dy}}_{dx/dy} \leftarrow \boxed{dx\, p(x) = dy\, p(y)}$$

Fácil recordar

• V.A. Uniforme



$$E(X) = \frac{b+a}{2}$$

$$var(X) = \frac{(b-a)^2}{12}$$

i

· Función suma de 2 V.A.

$$X = Y + Z$$

$$\boxed{p_x(x) = p_y(y) \underset{\text{convolución}}{*} p_z(z)}$$

V.A. exponencial

$$f_x(x) = a e^{-ax} \qquad E[x] = \frac{1}{a}$$

$$Var[x] = \frac{1}{a^2}$$

· Cálculos típicos

VECTOR de observaciones : $y = [y_0, y_1, \dots, y_{N-1}]^T$

$\rightarrow P_r(y) = \prod_{i=0}^{N-1} P_r(y_i)$

$\rightarrow p(y|\alpha) = \prod_{i=0}^{N-1} p(y_i|\alpha)$

ej. típico $y_i = \alpha + \underset{\hookrightarrow N(0,\sigma_n^2)}{n}$ $\longrightarrow N(\alpha, \sigma_n^2)$

$$p(y|\alpha) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_i-\alpha)^2}{2\sigma_n^2}} = \frac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} \prod_{i=0}^{N-1} \left(e^{-\frac{(y_i-\alpha)^2}{2\sigma_n^2}}\right)$$

$$\ln p(y|\alpha) = K - \sum_{i=0}^{N-1} \frac{(y_i-\alpha)^2}{2\sigma_n^2}$$

VARIANZA (calcular autocorrelación)

ej $y \rightarrow y_i = a + \underset{\hookrightarrow N(0,\sigma_n^2)}{n_i}$

$\rightarrow \hat{a}_{ML} = \frac{1}{N} \sum_{i=0}^{N-1} y_i$ $\longrightarrow$

$$var(\hat{a}_{ML})$$
$$= E[\hat{a}_{ML}^2] - E^2[\hat{a}_{ML}]$$
$$= E\left[\frac{1}{N^2} \sum_i \sum_j y_i y_j\right] - a^2$$
$$= \frac{1}{N^2} \sum_i \sum_j \underbrace{E[y_i y_j]}_{} - a^2$$

i

$$\begin{cases} \boxed{E[y_i^2] = \sigma_y^2 + a^2} \text{ si } i=j \\ \boxed{E[y_i y_j] = \underset{\uparrow \text{ indep}}{E[y_i] \cdot E[y_j]} = a^2} \text{ si } i \neq j \end{cases}$$

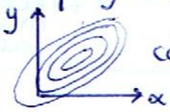
$$= \frac{1}{N^2}\left[\sum_i \underbrace{(\sigma_n^2 + a^2)}_{E[y_i^2]} + \sum_i \sum_{j \neq i} \underset{\uparrow\, E[y_i \cdot y_j]}{a^2}\right] - a^2$$

$$= \frac{1}{N^2}\left[N \cdot (\sigma_n^2 + a^2) + N(N-1)a^2\right] - a^2 = \frac{\sigma_n^2}{N}$$

• ESTIMACIÓN

• Espacios involucrados

parámetros (valor real) → $\alpha \xrightarrow{p(y,\alpha)} y(\alpha)$ observaciones $\xrightarrow[\hat{\alpha}=g(y)]{\text{estimador}} \hat{\alpha}(y)$ estimas



prob. conjunta

lo que max MAP (buscando en α)

incómodo de trabajar → aplicar Bayes

$$\boxed{p(y,\alpha) = p(y|\alpha)\cdot p(\alpha) = p(\alpha|y)\cdot \underbrace{p(y)}_{\text{irrelevante si tenemos y (observación)}}}$$

intersección de sucesos

• Estimadores óptimos

- Maximum a Posteriori $\boxed{\text{MAP}}$ →
$$\boxed{\begin{aligned}\hat{\alpha}_{MAP} &= \max_{\{\alpha\}}(p(\alpha|y)) \\ &= \max(p(y|\alpha)\cdot p(\alpha))\end{aligned}}$$
← útil para entenderlo
← útil para calcular

- Maxima verosimilitud $\boxed{\text{ML}}$ → $\boxed{\hat{\alpha}_{ML} = \max(p(y|\alpha))}$ → no tiene en cuenta la prob del parámetro (es como considerar α determinista → no conocemos su fdp)

- Bayes $\boxed{\text{BAYES}}$

Función error/coste
$$\varepsilon(\alpha,\hat{\alpha}) = \begin{cases} |\alpha-\hat{\alpha}| \\ (\alpha-\hat{\alpha})^2 \\ \vdots \end{cases}$$

$$\hat{\alpha}_{Bayes} = \min_{\hat{\alpha}} E[\varepsilon(\alpha,\hat{\alpha})] \rightarrow \text{minimiza el coste}$$
$$= \min \iint \varepsilon(\alpha,\hat{\alpha})\cdot p(\alpha,y)\,d\alpha\,dy$$
← irrelevante para y conocido

$$\boxed{\hat{\alpha}_{Bayes} = \min_{\hat{\alpha}} \int \varepsilon(\alpha,\hat{\alpha})\cdot p(\alpha|y)\,d\alpha}$$ Coste medio condicionado a y

↳ para $\underset{ECM}{\varepsilon = (\alpha-\hat{\alpha})^2}$: $\boxed{\hat{\alpha}_{Bayes\ ECM} = E[\alpha|y]}$

Demostración

si $\varepsilon(\alpha,\hat{\alpha}) = (\alpha-\hat{\alpha})^2$

entonces

$\min \int (\alpha-\hat{\alpha})^2 p(\alpha|y)\,d\alpha$

$\frac{d}{d\hat{\alpha}}\left(\int (\alpha-\hat{\alpha})^2 p(\alpha|y)\,d\alpha\right) = 0$

$= -2\int(\alpha-\hat{\alpha})\,p(\alpha|y)\,d\alpha$

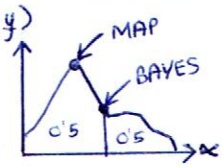
$\underbrace{\int \alpha\, p(\alpha|y)\,d\alpha}_{E[\alpha|y]} = \int \hat{\alpha}\, p(\alpha|y)\,d\alpha$

$\hat{\alpha}\cdot \underbrace{\int p(\alpha|y)\,d\alpha}_{1}$

MAP coge el máximo de $p(\alpha|y)$

BAYES coge la media

BAYES lógico:
- si yo no midiese nada, tomaría $\hat{\alpha} = E[\alpha]$, pero si he medido y, lo condiciono $\hat{\alpha} = E[\alpha|y]$



• Propiedades de los estimadores

- $\text{Sesgo}(\hat{\alpha}) = E[\hat{\alpha}] - E[\alpha]$
- $\text{Varianza}(\hat{\alpha}) = E[(\hat{\alpha}-E[\hat{\alpha}])^2]$
$= E[\hat{\alpha}^2] - E[\hat{\alpha}]^2$

es un lío aprenderlo bien

$$\boxed{\underset{error}{E[\varepsilon]} = \text{sesgo}^2(\hat{\alpha}) + \text{var}(\hat{\alpha})}$$

- insesgado: $E[\hat{\alpha}] = E[\alpha]$
- consistente: $\lim_{N\to\infty} \text{var}(\hat{\alpha}) = 0$ — puedo acercarme todo lo que quiera usando suficientes muestras
- mínima varianza: $\text{var}\,\hat{\alpha}_{MV} < \text{var}\,\hat{\alpha}$ para cualquier $\hat{\alpha}$
- eficiente: $\text{var}\,\hat{\alpha} = \text{CRB}$

- Cota de Cramer-Rao

Suponiendo α determinista
$\hat{\alpha}$ insesgado

Nunca podemos conseguir que $\text{var}(\hat{\alpha}) < \text{CRB}$

$$\boxed{\text{CRB} = \frac{-1}{E_{\{y\}}\left[\frac{\partial^2}{\partial\alpha^2}\ln(p(y|\alpha))\right]}}$$

- **Estimación lineal de señales**

es un escalar
muestra que predecimos

$$\text{señal } s(n) \longrightarrow \oplus \longrightarrow y(n) \longrightarrow \boxed{\text{FIR } \underline{h}} \longrightarrow \hat{s}(n+k) \equiv \hat{s}$$

$w(n)$ ruido

Estimador Lineal

- procesos estacionarios
- media nula
- incorrelados

escalar $s \equiv s(n+k)$

$$\underline{y} = \begin{bmatrix} y(n) \\ y(n-1) \\ y(n-2) \\ \vdots \\ y(n-N+1) \end{bmatrix}$$

$k=0 \rightarrow$ filtrado
$k>0 \rightarrow$ filtrado + predicción
$k<0 \rightarrow$ filtrado + reconstruir muestra que falta

Error Cuadrático Medio

$$ECM = E\left[(s-\hat{s})^2\right]$$
$$= E\left[(s-\underline{h}^T\underline{y})^2\right]$$
$$= E[s^2] - 2\underline{h}^T \underbrace{E[s\underline{y}]}_{N\times 1} + \underline{h}^T \underbrace{E[\underline{y}\,\underline{y}^T]}_{N\times N}\underline{h}$$

$$\qquad\qquad\qquad\qquad \underline{R}_{sy} \qquad\qquad \underline{R}_{yy}$$

Minimizar ECM con $\underline{h}$

$$\frac{\partial ECM}{\partial \underline{h}} = \begin{bmatrix} \partial ECM/\partial h_1 \\ \partial ECM/\partial h_2 \\ \vdots \end{bmatrix} = -2\underline{R}_{sy} + 2\underline{R}_{yy}\underline{h} = 0 \longrightarrow$$

Ecuaciones de Wiener-Hopf

solución lineal óptima

$$\boxed{\underline{h} = \underline{R}_{yy}^{-1}\,\underline{R}_{sy}}$$

$$\boxed{\begin{aligned} \underline{R}_{yy} &= E[\underline{y}\,\underline{y}^T] \\ \underline{R}_{sy} &= E[s(n+k)\cdot\underline{y}] \end{aligned}}$$

Si todo es gaussiano coincide con MAP, ML, ...

$\underline{R}_{yy}^{-1}$ = matriz de autocorrelación de y $= E[\underbrace{\underline{y}\cdot\underline{y}^T}_{N\times N}]$

$\underline{R}_{sy}$ = vector $= E[\overset{s(n+k)}{s}\cdot\underline{y}]$

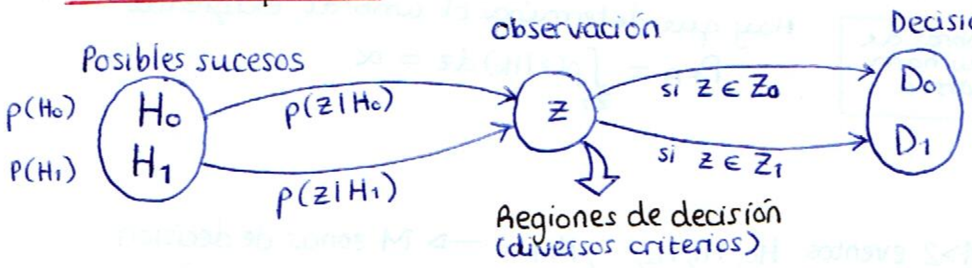
$$R_{sy}(i) = E[s(n+k)y(n-i+1)] = E[s(n+k)s(n-i+1)]$$

ruido incorrelado con s

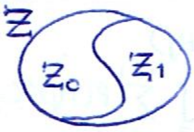
ej $R_{sy}(0) = E[s(n+k)\cdot s(n)]$

DETECCIÓN

• Test de Hipótesis



Regiones de decisión (diversos criterios)



$Z = Z_0 \cup Z_1$
$\emptyset = Z_0 \cap Z_1$

	Coste
• Func normal $H_0 \to D_0$	C_{00}
• Detección $H_1 \to D_1$	C_{11}
• Falsa Alarma $H_0 \to D_1$	C_{10}
• Miss $H_1 \to D_0$	C_{01}

Falsa Alarma

$$P_{FA} = \int_{Z_1} p(z|H_0)\,dz$$

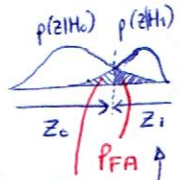
Miss

$$P_M = \int_{Z_0} p(z|H_1)\,dz$$

no tienen en cuenta $P(H_0)$ y $P(H_1)$

Prob. error:

$$P_e = P_{FA}\cdot P(H_0) + P_M \cdot P(H_1)$$



Estas regiones serían las ML. En MAP habría que corregir

Criterios de decisión

• MAP

$$p(H_1|z) \underset{H_0}{\overset{H_1}{\gtrless}} p(H_0|z)$$ ← entenderlo

Bayes

$$p(z|H_1)\cdot P(H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(z|H_0)\cdot p(H_0)$$ ← aplicarlo

$$\boxed{\underbrace{\frac{p(z|H_1)}{p(z|H_0)}}_{\Lambda(z)} \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\frac{p(H_0)}{p(H_1)}}_{\lambda}}$$

$\Lambda(z)$: razón de verosimilitud

λ: umbral (no umbral en Z sino en $\Lambda(z)$)

• ML

$$p(z|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(z|H_0)$$

$$\boxed{\frac{p(z|H_1)}{p(z|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1}$$

• Bayes

Escoge la región Z_0 que minimiza la parte variable del coste medio

$$p(H_1)(C_{01}-C_{11})p(z|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(H_0)(C_{10}-C_{00})p(z|H_0)$$

MISS, DETECT, FA

$$\boxed{\underbrace{\frac{p(z|H_1)}{p(z|H_0)}}_{\Lambda(z)} \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\frac{C_{10}-C_{00}}{C_{01}-C_{11}}\,\frac{p(H_0)}{p(H_1)}}_{\lambda}}$$

FA, MISS

COSTE MEDIO

$$\bar{C} = C_{00}\cdot P(H_0,D_0) + C_{11}\cdot P(H_1,D_1) + C_{10}\cdot P(H_0,D_1) + C_{01}\cdot P(H_1,D_0)$$

$$= C_{00}\cdot P(H_0)\cdot P(D_0|H_0) + C_{11}\cdot P(H_1)\cdot P(D_1|H_1) + C_{10}\cdot P(H_0)\cdot P(D_1|H_0) + C_{01}\cdot P(H_1)\cdot P(D_0|H_1)$$

$$= C_{00}\cdot P(H_0)\cdot \int_{Z_0} p(z|H_0)dz + C_{11}\cdot P(H_1)\cdot\left[1-\int_{Z_0} p(z|H_1)dz\right] + C_{10}\cdot P(H_0)\cdot\left[1-\int_{Z_0} p(z|H_0)dz\right] + C_{01}\cdot P(H_1)\cdot\int_{Z_0} p(z|H_0)dz$$

lógicamente $\int_{Z_1} p(z|H_i)dz = 1 - \int_{Z_0} p(z|H_i)dz$

lo hacemos para integrar siempre en Z_0

$$\bar{C} = C_{11}\,p(H_1) + C_{10}\,p(H_0)$$ ← coste fijo

$$+ \int_{Z_0} p(H_1)\underbrace{(C_{01}-C_{11})}_{>0}p(z|H_1) - p(H_0)\underbrace{(C_{10}-C_{00})}_{>0}p(z|H_0)\,dz$$ ← variable (según Z_0 y Z_1)

$$= C_{11}\,P(H_1) + C_{10}\,p(H_0) + (C_{01}-C_{11})P(H_1)\underbrace{\int_{Z_0} p(z|H_1)dz}_{P_M} - (C_{10}-C_{00})P(H_0)\underbrace{\int_{Z_0} p(z|H_0)dz}_{1-P_{FA}}$$

Y utilizando que $P(H_0) = 1 - P(H_1)$ se llega a

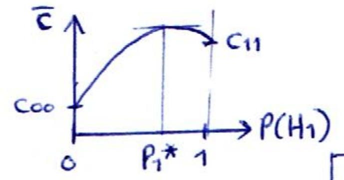
$$\bar{C} = C_{00}(1-P_F) + C_{10}P_F + p(H_1)\left[(C_{11}-C_{00}) + (C_{01}-C_{11})P_M - (C_{10}-C_{00})P_F\right]$$

• Mínima prob. de error

supone $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1$ $\Rightarrow \bar{C} = P_e \Rightarrow$

minimizar coste (Bayes) ||| Minimizar P_e (MAP)

La relación con $P(H_1)$ es:



• Minimax

No conocemos $P(H_1)$ así que tomamos el caso peor

suponemos $\boxed{P(H_1) = P_1^*}$ ⟷

peor caso para $P(H_1)$

$$P_1^* \Leftrightarrow \frac{d\bar{C}}{dP(H_1)} = 0 \Leftrightarrow \begin{array}{l}(C_{11}-C_{00}) \\ +(C_{01}-C_{11})P_M \\ -(C_{10}-C_{00})P_{FA}\end{array} = 0$$

si $C_{00} = C_{11} = 0$, $C_{10} = C_{01} = 1$ $\quad \boxed{P_M = P_{FA}}$

- **Criterio de Neymann-Pearson**

· C_{ij} desconocidas
$P(H_i)$ desconocidas

Se trata de fijar $P_{FA} = \alpha$ (dato) y minimizar P_M (max P_{DET})

Para ello

$$\boxed{\frac{p(z|H_1)}{p(z|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \leftarrow \text{Umbral de Neymann Pearson}}$$

Hay que determinar el umbral exigiendo:

$$P_{FA} = \int_{Z_1} p(z|H_0)\,dz = \alpha$$

- **Hipótesis múltiple**

$M>2$ eventos $H_0, H_1, H_2, \dots, H_{M-1} \longrightarrow M$ zonas de decisión

El coste medio:

$$\bar{C} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} C_{ij}\cdot P(H_j)\cdot P(D_i|H_j)$$

$$= \dots = \underbrace{\sum_{i=0}^{M-1} C_{ii}\,P(H_i)}_{\text{coste fijo}} + \underbrace{\sum_{i=0}^{M-1}\int_{Z_i}\sum_{\substack{j=0\\ j\neq i}}^{M-1} P(H_j)(C_{ij}-C_{jj})\,p(z|H_j)\,dz}_{\text{coste variable que minimizamos}}$$

Escogemos las Z_i que MINIMICEN

$$I_0(z) = \sum_{j\neq 0} P(H_j)\cdot(C_{0j}-C_{jj})\,p(z|H_j) \rightarrow J_0(z) = \sum_{j\neq 0} P(H_j)(C_{0j}-C_{jj})\cdot\frac{p(z|H_j)}{p(z|H_0)}$$

$$I_1(z) = \sum_{j\neq 1} P(H_j)(C_{1j}-C_{jj})\,p(z|H_j) \rightarrow J_1(z) = \sum_{j\neq 1} P(H_j)(C_{1j}-C_{jj})\cdot\frac{p(z|H_j)}{p(z|H_0)}$$

$\vdots$

$$I_i(z) = \sum_{j\neq i} P(H_j)(C_{ij}-C_{jj})\,p(z|H_j) \rightarrow \boxed{J_i(z) = \sum_{j\neq i} P(H_j)\cdot(C_{ij}-C_{jj})\,\underbrace{\frac{p(z|H_j)}{p(z|H_0)}}_{\Lambda_j(z)}}$$

normalizando por $p(z|H_0)$

$$\Lambda_j(z)\begin{cases}\Lambda_0(z) = 1\\ \Lambda_1(z) = p(z|H_1)/p(z|H_0)\\ \vdots\\ \Lambda_{M-1}(z) = p(z|H_{M-1})/p(z|H_0)\end{cases}$$

se escoge $\boxed{z\in Z_i \longleftrightarrow J_i(z) \leq J_j(z)\ \forall j}$

↳ Condición en el dominio $J(z)\rightarrow\Lambda(z)$
Hay que convertirla en condición en el dominio z

CLASIFICACIÓN

Problema de clasificación

$\Rightarrow$ 

- Preproccessing
  - sensing
  - separate fishes from one another and from background (segmentation)
- Feature extraction
  - Reduce data by measuring certain features
  - e.g. lightness, length, width
- Classifier
  - Threshold decision → salmon / → sea bass
  - e.g.

↓ Generalization

- control trade-off between training error and test error

e.g.

would give 0% training-error but bad test-error

→ OVERFITTING

- Features should be:
  - incorrelated with each other
  - do not add "noisy features"
  - too many features is not good (problem of high dimensionality)

  Poner demasiadas dimensiones puede hacer que se estime mal la distribución

  - usar 'discriminative features'
  - saber elegir features es un arte

Tipos

- Statistical learning $x \longrightarrow \boxed{f(\cdot)} \longrightarrow y$
  - Supervised learning: - sample set of x and y during training
    - $f(\cdot)$ is learned
      - → $Y = \{0, 1\}$ → detection
      - → $Y = \{0,1,2,\dots,N\}$ → clasification
      - → $Y = \mathbb{R}$ → regression
  - Unsupervised learning (clustering)
    - only x is known
    - try to cluster data into sets

Bayesian Theory

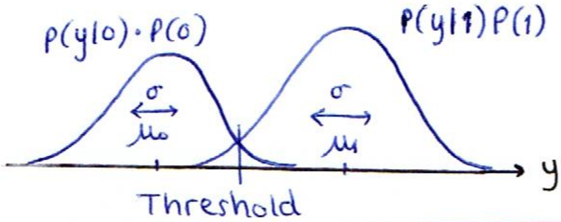
$$P_{X,Y}(x,i) = P_{X|Y}(x|i)\cdot P_Y(i)$$

marginalization (eliminating variables) $$P_{X_1,X_4}(x_1,x_4) = \iint P_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4)\,dx_2\,dx_3$$

Bayes Rule: $$P_{Y|X}(y|x) = \frac{P_{X|Y}(x|y)\cdot P_Y(y)}{P_X(x)}$$ ⟶ no importante para clasificar

MAR Rule Classification $$i^*(x) = \arg\max_i \underbrace{P(x|y)}_{ML}\cdot\underbrace{P(y)}_{Prior}$$

Gaussiana en 1D



$$\boxed{y \gtrless \frac{\mu_1+\mu_0}{2} + \frac{1}{\left(\frac{\mu_1-\mu_0}{\sigma^2}\right)}\cdot \ln\frac{P(0)}{P(1)}}$$

$\frac{1}{\left(\frac{\mu_1-\mu_0}{\sigma^2}\right)}$ weighs the prior

⇐ distance between the means in units of variance

- classes very far apart: "forget the prior"
- classes with same mean "forget the observation, use prior!"

The Gaussian Classifier

Gaussian N-dim.

$$\boxed{p_{X|Y}(\vec{x}|i) = \frac{1}{\sqrt{(2\pi)^d |\underline{\underline{\Sigma}}_i|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \underline{\underline{\Sigma}}_i^{-1}(\vec{x}-\vec{\mu}_i)}}$$

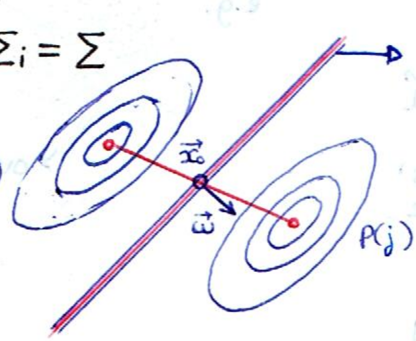
MAP $i^*(\vec{x}) = \arg\max_i \left[\ln p_{X|Y}(\vec{x}|i) + \ln P_Y(i)\right]$

General case $\vec{\mu}_1, \vec{\mu}_2$ $\underline{\underline{\Sigma}}_1, \underline{\underline{\Sigma}}_2$

$$i^*(\vec{x}) = \arg\min_i \Big[\underbrace{(\vec{x}-\vec{\mu}_i)^T \underline{\underline{\Sigma}}_i^{-1}(\vec{x}-\vec{\mu}_i)}_{\substack{\text{Mahalanobis distance} \\ d_i(\vec{x},\vec{\mu}_i)}} + \underbrace{\ln(2\pi)^d |\underline{\underline{\Sigma}}_i| - 2\ln P_Y(i)}_{\substack{\alpha_i \\ \text{class prior}}} \Big]$$

↳ special case: all classes have same covariance

$\Sigma_i = \Sigma$



Hyperplane $\vec{w}^T \cdot (\vec{x} - \vec{x}_0) = 0$

↳ passes through $\vec{x}_0$

↳ normal to $\vec{w}$

changes direction

direction joining the means

$$\vec{w} = \underline{\underline{\Sigma}}^{-1} \cdot (\vec{\mu}_i - \vec{\mu}_j)$$

$$\vec{x}_0 = \underbrace{\frac{\vec{\mu}_i + \vec{\mu}_j}{2}}_{\text{midpoint}} - \underbrace{\frac{(\vec{\mu}_i - \vec{\mu}_j)}{(\vec{\mu}_i-\vec{\mu}_j)^T \underline{\underline{\Sigma}}^{-1} (\vec{\mu}_i - \vec{\mu}_j)}}_{\text{weigh to prior}} \underbrace{\ln \frac{P_Y(i)}{P_Y(j)}}_{\text{prior}}$$

↓ Mahalanobis distance between means

$\underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{I}}$

⇓

Mahalobious distance becomes EUCLIDEAN DISTANCE SQUARED in units of variance

$$d_i(\vec{x}_1, \vec{x}_2) = \frac{\|x_1 - x_2\|^2}{\sigma^2}$$

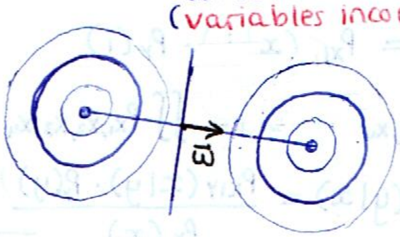
El peso al prior pasa a ser:

$$\frac{\vec{\mu}_i - \vec{\mu}_j}{d_i(\vec{\mu}_i, \vec{\mu}_j)} = \underbrace{\frac{\vec{\mu}_i - \vec{\mu}_j}{\left(\frac{\|\vec{\mu}_i - \vec{\mu}_j\|^2}{\sigma^2}\right)}}_{\text{clara analogía al 1D}}$$

↳ sub-special case

same variance along all dimensions (variables incorreladas)

$$\underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{I}} = \begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \ddots \\ & & & \sigma^2 \end{bmatrix}$$



$\vec{w} = \frac{\vec{\mu}_i - \vec{\mu}_j}{\sigma^2}$	along the line joining the means

$$\vec{x}_0 = \frac{\vec{\mu}_i + \vec{\mu}_j}{2} - \frac{(\vec{\mu}_i - \vec{\mu}_j)}{\left(\frac{\|\vec{\mu}_i - \vec{\mu}_j\|^2}{\sigma^2}\right)} \cdot \ln \frac{P_Y(i)}{P_Y(j)}$$

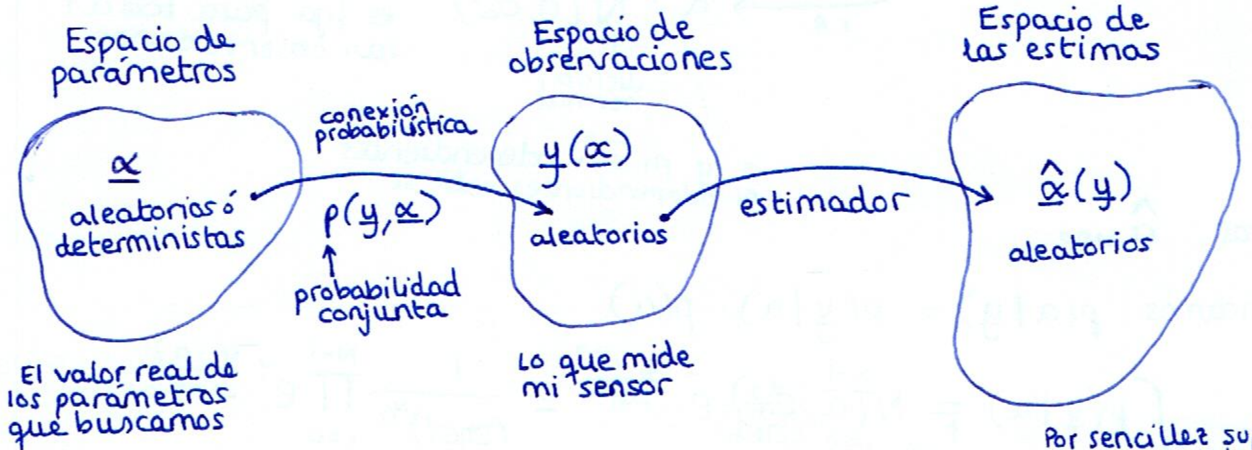
↳ sub-special case

$P_Y(i) = P_Y(j)$ equiprobables

$\vec{x}_0 = \frac{\vec{\mu}_i + \vec{\mu}_j}{2}$	midpoint of the means

ESTIMACIÓN

Espacios involucrados y modelos de probabilidad



Por sencillez supondremos un valor escalar en lugar de un vector
$\underline{\alpha} \equiv \alpha$

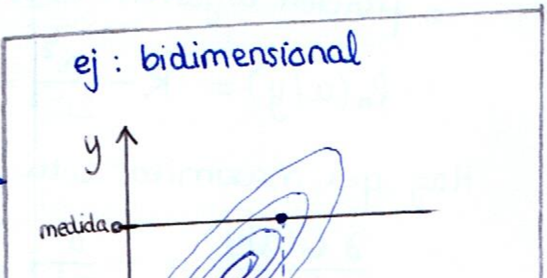
Estimadores óptimos

1. Estimador "Maximum a posteriori" (MAP)

Supondremos α aleatorio con $p(\alpha)$ conocida (si no conociésemos $p(\alpha)$ podríamos decir que es determinista desconocido)

$$\hat{\alpha}_{MAP} = \hat{\alpha}(y) = \max_{\{\alpha\}} P(y, \alpha)$$

he medido en y — me muevo en α para buscar el max.



Es engorroso trabajar con $P(y,\alpha)$
Pero aplicando Teorema de Bayes

$$P(y,\alpha) = \frac{P(y|\alpha)\cdot P(\alpha)}{P(y)}$$ ← no depende de α

$$\boxed{\hat{\alpha}_{MAP} = \max_{\{\alpha\}} \Big\{ \underbrace{P(y|\alpha)\cdot P(\alpha)}_{= P(\alpha|y)} \Big\}}$$

← en la práctica ésta es la función fácil de conseguir

← es equivalente a maximizar $P(\alpha|y)$ lo cual le da nombre al método

A la función que maximizamos ($p(y,\alpha) \rightarrow p(y|\alpha)\cdot p(\alpha) \rightarrow p(\alpha|y)$) se le llama función de verosimilitud (likelyness)

$$\ell(y/\alpha)$$

$$\ln[\ell(y/\alpha)] = L(y/\alpha)$$ función logarítmica de verosimilitud

ejemplo 1:

suponemos $y = [y_0, y_1, \dots, y_{N-1}]^T$

donde $y_i = a + n_i$

$n_i : N(0, \sigma_n^2)$ V.A.

$a : N(0, \sigma_a^2)$ V.A. — distrib. normal

'a' es una V.A. pero es fija para todas las observaciones

a y n_i son independientes
n_i independientes entre si

Encontrar $\hat{a}_{MAP}$

Buscamos $p(a|y) = p(y|a) \cdot p(a)$

$$p(y|a) \underset{y=a+n}{=} \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y_i-a)^2}{2\sigma_n^2}} = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \prod_{i=0}^{N-1} e^{-\frac{(y_i-a)^2}{2\sigma_n^2}}$$

la prob. del vector será la prob de que se cumplan todos sus elementos (que como son independientes es el producto de gaussianas)

$$p(a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} e^{-\frac{a^2}{2\sigma_a^2}}$$

La función logarítmica de verosimilitud será

$$l_n(a/y) = K - \frac{a^2}{2\sigma_a^2} - \sum_{i=0}^{N-1} \frac{(y_i-a)^2}{2\sigma_n^2}$$

Hay que maximizar está función con respecto a 'a'

$$\frac{\partial l_n(a|y)}{\partial a} = -\frac{a}{\sigma_n^2} + \sum_{i=0}^{N-1} \frac{(y_i - a)}{\sigma_n^2}$$

$$= -a\left(\frac{1}{\sigma_a^2} + \frac{N}{\sigma_n^2}\right) + \sum_{i=0}^{N-1} \frac{y_i}{\sigma_n^2} = 0$$

despejamos $\hat{a}_{MAP}$ que cumple esta ecuación

$$\Rightarrow \boxed{\hat{a}_{MAP} = \frac{1}{\frac{\sigma_n^2}{\sigma_a^2} + N} \sum_{i=0}^{N-1} y_i}$$

Esta fórmula es muy lógica sobretodo si la vemos en casos extremos

si $\frac{\sigma_n^2}{\sigma_a^2} \ll N \longrightarrow \hat{a}_{MAP} \simeq \frac{1}{N} \sum_{i=0}^{N-1} y_i$ la media de las observaciones

$\sigma_a \uparrow\uparrow\uparrow$

si $\frac{\sigma_n^2}{\sigma_a^2} \gg N \longrightarrow \hat{a}_{MAP} \simeq \underbrace{\frac{\sigma_a^2}{\sigma_n^2}}_{\downarrow\downarrow\downarrow} \sum_{i=0}^{N-1} y_i$ ¡esta ya no es tan intuitivo!

sabiendo que $\sigma_a \downarrow\downarrow \rightarrow$ el valor estimado lo reducimos acercándolo a 0

2. Estimador de máxima verosimilitud (ML)

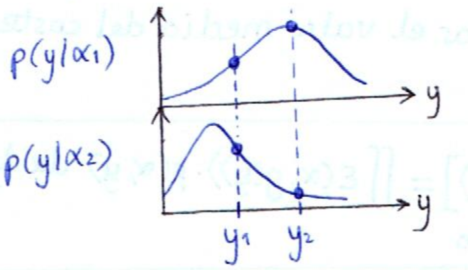
Suponemos α determinista

$$\boxed{\hat{\alpha}_{ML} = \max_{\{\alpha\}} p(y|\alpha)}$$

Equivale a MAP
en el caso de
$p(\alpha) = K$ uniforme

Es decir, no introducimos información a priori acerca del parámetro

ejemplo: si α tiene dos posibilidades α_1, α_2



con y_1 elijo α_2
con y_2 elijo α_1

ejemplo 2: igual que 1 pero 'a' determinista

$$\ln\{p(y|a)\} = K - \sum_{i=0}^{N-1} \frac{(y_i - a)^2}{2\sigma_n^2}$$

función logarítmica de verosimilitud

maximizando respecto a 'a'

$$\frac{\partial \ln p(y|a)}{\partial a} = -\sum_{i=0}^{N-1} \frac{(y_i - a)}{\sigma_n^2} = 0 \longrightarrow \boxed{\hat{a}_{ML} = \frac{1}{N} \sum_{i=0}^{N-1} y_i}$$

coincide con el MAP cuando no hay info de 'a' a priori $\sigma_a^2 \to \infty$

En resumen:

$\hat{\alpha}_{ML} = \max_{\alpha} p(y|\alpha)$ escoge el parámetro que da la máxima probabilidad de lo que he medido

$\hat{\alpha}_{MAP} = \max_{\alpha} p(\alpha|y) = \max_{\alpha} p(y|\alpha) \cdot p(\alpha)$ escoge el parámetro que con máxima probabilidad ha generado lo que he medido

no estoy seguro

3. Estimador de Bayes

Cada pareja $\alpha, \hat{\alpha}$ tiene un coste (típicamente grande si he errado mucho)
a través de una función de coste

$$C(\alpha, \hat{\alpha})$$

típicamente

$$C(\alpha,\hat{\alpha}) = \left\{ \begin{matrix} (\alpha - \hat{\alpha})^2 \\ |\alpha - \hat{\alpha}| \\ \vdots \end{matrix} \right\} = \mathcal{E}(\alpha, \hat{\alpha})$$

será una V.A.
- valor medio
- varianza
- ⋮

El estimador de Bayes trata de minimizar el valor medio del coste (i.e. minimiza el 'error' medio)

$$\boxed{\hat{\alpha}_{BAYES} = g(y) \text{ tal que } E[\mathcal{E}(\alpha,\hat{\alpha})] = \iint \mathcal{E}(\alpha, g(y)) \cdot p(\alpha, y)\, d\alpha\, dy \text{ sea mínimo}}$$

Recuerda

$$E\{X\} = \int x \cdot p(x)\, dx$$
$$E\{f(X)\} = \int f(x) \cdot p(x)\, dx$$
$$E\{f(X_1, X_2)\} = \int f(x_1, x_2) \cdot p(x_1, x_2)\, dx_1\, dx_2$$

se puede expresar la integral:

$$E[\mathcal{E}(\alpha,\hat{\alpha})] = \iint \mathcal{E}(\alpha, g(y)) \cdot \underbrace{p(\alpha, y)}_{p(\alpha|y)\cdot p(y)}\, d\alpha\, dy$$

$$= \int p(y) \underbrace{\left[\int \mathcal{E}(\alpha, g(y))\, p(\alpha|y)\, d\alpha\right]}_{\text{esto dará lugar a un valor que depende de } y} dy$$

siempre es positivo

puedo limitarme a minimizar esto (llamado: Coste medio condicionado a y)

si lo minimizo para todo y estaré minimizando todo el coste medio

- caso particular:

  si escogemos $\mathcal{E}(\alpha, g(y)) = (\alpha - g(y))^2 = (\alpha - \hat{\alpha})^2$

  i.e. error cuadrático como función de coste

  - Sustituyendo en la integral anterior
  - Derivando respecto a g (aunque dependa de y)
  - Igualando a cero

$$\frac{\delta}{\delta g} \int (\alpha - g(y))^2 p(\alpha|y)\, d\alpha = -2\int (\alpha - g(y)) \cdot p(\alpha|y)\, d\alpha = 0$$

$$\Rightarrow \underbrace{\int \alpha\, p(\alpha|y)\, d\alpha}_{E\{\alpha|y\}} = \int g(y) \cdot p(\alpha|y) = g(y) \cdot \underbrace{\int p(\alpha|y)}_{=1}$$

La integral de CUALQUIER densidad de probabilidad (aunque sea condicionada, etc...) debe ser 1

$$\Rightarrow \boxed{\hat{\alpha}_{BAYES} = E\{\alpha|y\}}$$

- En general es un procesado no lineal en y
- si todo es gaussiano, sí es un operador lineal (es decir $\alpha = A \cdot y$) ← matriz

Es lógico:

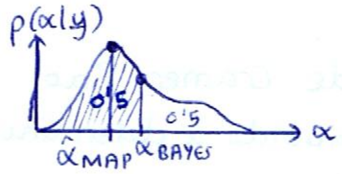
si yo no midiese nada, tomaría $\hat{\alpha} = E[\alpha]$

pero si he medido y, tomo el valor medio CONDICIONADO a y

Es decir:

MAP cogía el MÁXIMO de $p(\alpha|y)$

BAYES coge la MEDIA de $\alpha|y$



---

ejemplo 3

$y = [y_0, \dots, y_{N-1}]^T$

$y_i = a + n_i$

$a : N(0, \sigma_a^2)$

$n_i : N(0, \sigma_n^2)$

minimiz. error cuadrático medio

$$\hat{a}_{BAYES} \overset{\downarrow}{=} E[a|y] = \int_{-\infty}^{\infty} a \cdot p(a|y)\, da$$

recuerda:

$$p(a|y) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \left[\prod_{i=0}^{N-1} e^{-\frac{(y_i - a)^2}{2\sigma_n^2}} \right] \frac{1}{\sqrt{2\pi\sigma_a^2}}\, e^{-\frac{a^2}{2\sigma_a^2}}$$

Propiedades de los estimadores

- Insesgado: $E[\hat{\alpha}] = E[\alpha]$ i.e. si hago ∞ estimaciones, la media será el valor correcto

  $\text{sesgo}(\alpha) = E[\hat{\alpha}] - E[\alpha]$

- Consistente: $\lim_{N\to\infty} P_r(\hat{\alpha} - \alpha < \varepsilon) = 0$ i.e. puedo conseguir que el valor estimado sea tan cercano al real como se desee sin más que repetir la estimación N veces

- Mínima varianza: $\text{var}\,\hat{\alpha}_{MV} \leq \text{var}\,\hat{\alpha}$, para cualquier $\hat{\alpha}$

  recuerda, se cumple:

  $$\text{var}\,\hat{\alpha} = E[\hat{\alpha}^2] - E^2[\hat{\alpha}]$$

  $$\boxed{\underset{\text{error}}{E[\varepsilon^2]} = \text{var}\,\hat{\alpha} + \text{sesgo}^2(\hat{\alpha})}$$

  Normalmente preocupa más la varianza que el sesgo (a veces el sesgo se puede compensar)

- Cota de Cramer-Rao

  suponiendo α determinista, $\hat{\alpha}$ insesgado

  $\text{var}(\hat{\alpha}) \geq CRB$ nunca podemos conseguir que la varianza baje por debajo de la cota

  La distancia a la cota indica lo mejorable que es el estimador

  $$CRB = \frac{-1}{E_y\left[\frac{\partial^2}{\partial \alpha^2} \ln(p(y|\alpha))\right]}$$

- Estimador eficiente $\text{var}(\hat{\alpha}) = CRB$

ejemplo 4 : ML

Hipótesis de ejemplos 1,2,3

$$\hat{a}_{ML} = \frac{1}{N}\sum_{i=0}^{N-1} y_i$$

$$E[\hat{a}_{ML}] = \frac{1}{N}\sum_{i=0}^{N-1} \underbrace{E[y_i]}_{a} = a \longrightarrow$$ estimador insesgado

$$\text{var}[\hat{a}_{ML}] = E[\hat{a}_{ML}^2] - E^2[\hat{a}_{ML}]$$

$$= E\left[\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} y_i y_j\right] - a^2$$

$$= \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \underbrace{E[y_i y_j]}_{} - a^2$$

$$= \begin{cases} E[y_i^2] = \sigma_y^2 + a^2 & \text{si } i=j \\ E[y_i y_j] = E[y_i]\cdot E[y_j] = a^2 & \text{si } i\neq j \end{cases}$$

(ⓘ) independientes

$$= \frac{1}{N^2}\left[\sum_{i=0}^{N-1} E[y_i^2] + \sum_{i=0}^{N-1}\sum_{\substack{j=0 \\ j\neq i}}^{N-1} E[y_i]\cdot E[y_j]\right] - a^2$$

$$= \frac{1}{N^2}\left[N\cdot(\sigma_n^2 + a^2) + (N)(N-1)\cdot a^2\right] - a^2$$

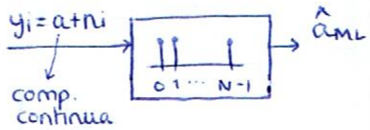
$$\boxed{\text{var}\,\hat{a}_{ML} = \frac{\sigma_n^2}{N}}$$

Estamos reduciendo el ruido

$$\begin{cases} y_i = a + n_i \longrightarrow \sigma_n^2 \\ \hat{x}_{ML} = a + n_i' \longrightarrow \frac{\sigma_n^2}{N} \end{cases}$$

$$\text{var}\,\hat{a}_{ML} \xrightarrow[N\to\infty]{} 0 \longrightarrow$$ estimador consistente

si te fijas bien, $\hat{a}_{ML}$ en este caso no es mas que un filtro FIR que elimina ruido



Calculemos la Cota de Kramer-Rau

recuerda que ya teníamos

$$\frac{\partial \ln p(y|a)}{\partial a} = \frac{\partial}{\partial a}\left\{K - \sum_{i=0}^{N-1}\frac{(y_i - a)^2}{2\sigma_n^2}\right\}$$

$$= \frac{1}{2\sigma_n^2}(-2)\sum_{i=0}^{N-1}(y_i - a)(-1) = \frac{1}{\sigma_n^2}\sum_{i=0}^{N-1}(y_i - a)$$

$$\frac{\partial^2 \ln(p(y|a))}{\partial^2 a} = \frac{1}{\sigma_n^2}\sum_{i=0}^{N-1}(-1) = \frac{-N}{\sigma_n^2}$$

$$CRB = -\frac{1}{E\left[\frac{\partial^2 \ln p(y|a)}{\partial^2 a}\right]} = \frac{\sigma_n^2}{N}$$

$$\text{var}\,\hat{a}_{ML} = CRB \longrightarrow$$ estimador eficiente

Si el ruido no fuese gaussiano, habrían otras formas mejores de estimar el valor medio

Ejemplo 5.

Demostrar que cualquier estimador de Bayes es insesgado

(lo demostramos para el caso del error cuadrático como función de coste)

$$E\left[\hat{\alpha}_{BAYES}\right] = E\left[E[\alpha|y]\right]$$

valor medio de la estima para varias observaciones (i.e. la aleatoriedad es en y)

valor medio de α en la función $p(\alpha|y)$ (i.e. aleatoriedad en α)

v.A.

$$= E\left[\int \alpha \cdot p(\alpha|y)\, d\alpha\right] = \int \alpha \cdot E\left[p(\alpha|y)\right] \cdot d\alpha$$

$$= \int \alpha \cdot \left[\int \underbrace{p(\alpha|y) \cdot p(y)}_{p(y|\alpha)\cdot p(\alpha)} dy\right] d\alpha$$

Bayes

$$= \int \alpha \cdot \left[\int p(y|\alpha) \cdot p(\alpha)\, dy\right] d\alpha$$

$$= \int \alpha \cdot p(\alpha) \underbrace{\left[\int p(y|\alpha)\, dy\right]}_{1} d\alpha$$

$$= \int \alpha \cdot p(\alpha)\, d\alpha = E[\alpha]$$

> recuerda:
>
> valor medio de x
>
> $E[x] = \int x\, p(x)\, dx$
>
> valor medio de $f(x)$
>
> $E[f(x)] = \int f(x)\, p(x)\, dx$

Estimación lineal de señales

Vamos a trabajar con → Procesos estocásticos (¡señales!)
→ Estimadores lineales

$$y(n) = \underbrace{s(n)}_{\text{señal}} + \underbrace{w(n)}_{\text{ruido}}$$

- $\{\tilde{s}(n)\}$ proceso estacionario en sentido amplio
  $E[\tilde{s}(n+m)\cdot s(n)] = R_{ss}(m)$
  y con media nula $E[\tilde{s}(n)] = 0$
  (si no lo fuera, se le resta y ya está)

- $\{\tilde{w}(n)\}$ proceso estacionario en sentido amplio
  $E[\tilde{w}(n+m)\cdot w(n)] = R_{ww}(m)$
  y con media nula $E[\tilde{w}(n)] = 0$

- La señal y el ruido son incorrelados
  $E[\tilde{w}(n+m)\cdot \tilde{s}(n)] = R_{ws}(m) = 0$

Problema: Estimar $s(n+k) \equiv s$
a partir de $y = [y(n), y(n-1), \dots, y(n-N+1)]$

si $k=0$ → filtrado → eliminar el ruido
$k>0$ → filtrado + predicción
$k<0$ → filtrado + interpolación → intento reconstruir muestras que me faltan en el pasado

y quiero que sea un procesado LINEAL

$$\hat{s}(n+k) = \underline{h}^T \underline{y}$$ → Es un filtro FIR

$\underline{y} \rightarrow \boxed{\underline{h}} \rightarrow \hat{\underline{s}}$

hay que encontrar esta h

para cada instante n es un vector que va cambiando (con las N muestras anteriores)

y(n)

y

Error cuadrático medio

$$ECM = E\left[(s-\underline{h}^T y)^2\right]$$

$$= E\left[s^2 - 2s\underline{h}^T y + \underline{h}^T y y^T \underline{h}\right]$$

NOTA:

$\underline{h}^T y$ no es más que un número

$\underline{h}^T y y^T \underline{h}$ no es mas que ese número al cuadrado. He cambiado los traspuestos porque me conviene y el producto escalar es conmutativo

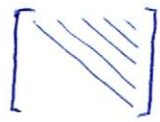
$$= E[s^2] - 2\underline{h}^T \underbrace{E[s\underline{y}]}_{\substack{\text{un vector}\\ \text{de valores}\\ \text{medios}\\ N\times 1}} + \underline{h}^T \underbrace{E[\underline{y}\,\underline{y}^T]}_{\substack{\text{matriz}\\ N\times N}} \underline{h}$$

indaguemos en ambos

$$E[\underline{y}\,\underline{y}^T] = \underline{\underline{R}}_{yy} \longrightarrow \underbrace{R_{yy}(i,j)}_{\substack{\text{la comp.}\\ (i,j)\text{ de la}\\ \text{matriz}}} = E[y(i)\cdot y(j)] = E[y(n-i+1)\,y(n-j+1)]$$

$$= R_{yy}(j-i)$$

La matriz de autocorrelación puede construirse conociendo la función de autocorrelación del proceso estocástico $y(n)$



$$E[s\cdot \underline{y}] = \underline{R}_{sy} \longrightarrow \underline{R}_{sy}(i) = E[s(n+k)\,y(n-i+1)]$$

podemos quitar la parte del ruido de la y ya que es incorrelado con s

$$= E[s(n+k)\cdot s(n-i+1)]$$

$$= R_{ss}(k+i-1)$$

a la correlación sólo le importa la diferencia de índices

$$ECM = E[s^2] - 2\underline{h}^T \underline{R}_{sy} + \underline{h}^T \underline{\underline{R}}_{yy}\,\underline{h}$$

Nos interesa minimizar el ECM, por tanto derivamos
(con Reglas de derivación con matrices y vectores se resuelve muy fácilmente)

$$\frac{\partial\, ECM}{\partial \underline{h}} = \begin{bmatrix} \frac{\partial ECM}{\partial \underline{h}(1)} \\ \vdots \\ \frac{\partial ECM}{\partial \underline{h}(n)} \end{bmatrix} = -2 \underline{R}_{sy} + 2 \underline{\underline{R}}_{yy} \underline{h} = 0$$

Igualando $\partial ECM / \partial \underline{h}$ a cero se obtiene un sistema de ecuaciones lineales que se resuelve con:

$$\boxed{\underline{h} = \underline{\underline{R}}_{yy}^{-1} \underline{R}_{sy}}$$

Ecuaciones de Wiener-Hopf

Ésta es la solución lineal óptima
Además si todo es gaussiano resulta que coincide esta solución con las soluciones MAP, ML, … $p(\alpha/y)$

Ejemplo: Predecir el valor de un proceso estacionario (sin ruido)
(por ej. predecir la bolsa)

Ejemplo 6: construir el predictor lineal óptimo (ECM mínimo) de $s(n+1)$ a partir de $s(n)$

$w(n) = 0 \rightarrow R_{ww}(m) = 0$ $\qquad s \equiv s(n+1)$ ($k=1$)

$\underline{y} = [s(n)]^T = s(n)$ ($N=1$)

nota de campo: algunos que trabajan en esto dicen que una matriz es mínimo 1000x1000. Cualquier matriz menor que eso lo consideran un escalar .... ☺

La matriz de autocorrelación

$\underline{\underline{R}}_{yy} = \underline{\underline{R}}_{ss} = R_{ss}(0)$
matriz tamaño 1 $\qquad$ escalar

vector de correlación señal

$\underline{R}_{sy} = E[s(n+1)s(n)] = R_{ss}(1)$

$$\underline{h} = h = \frac{R_{ss}(1)}{R_{ss}(0)} = \rho$$

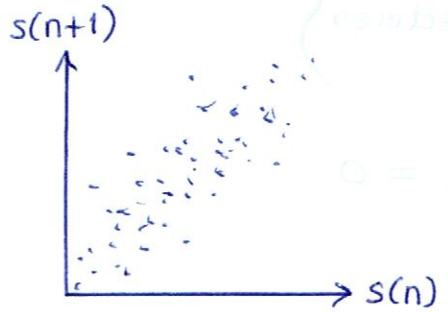
es decir

$$s(n) \longrightarrow \bigotimes \longrightarrow \hat{s}(n+1)$$

$\rho \qquad |\rho| \leq 1$

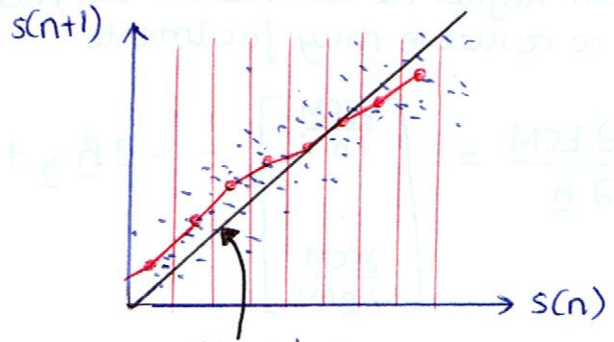
si $\rho = 1 \rightarrow$ mucha correlación entre n y $n+1$ (tomo el valor anterior)

si $\rho = 0 \rightarrow$ no tengo ni idea así que digo 0 que es la media si emparejo muestras contiguas.



Forma sencilla experimental de hacer un predictor con N=1

Puedo calcular la media alrededor de ventanas de s(n)



será la media de los puntos si el proceso es gaussiano

Problema 1

Tengo un conjunto de números binarios $x_0, x_1, \dots, x_{N-1}$ independientes

x_0: 0, 1

MUY IMPORTANTE

$$\text{Prob}\{x_i = 1\} = \theta \quad \forall i$$

Hallar $\hat{\theta}_{ML} \rightarrow \hat{\theta}_{ML} = \max_{\theta} \; p(\underline{x}|\theta)$

calculemos:

$$p(\underline{x}|\theta) = \prod_{i=0}^{N-1} \underbrace{p(x_i|\theta)}$$

$$p(x_i|\theta) = \begin{cases} 1-\theta & x_i = 0 \\ \theta & x_i = 1 \end{cases} = \theta^{x_i} \cdot (1-\theta)^{1-x_i}$$

compactando (i)

Por tanto:

$$p(\underline{x}|\theta) = \prod_{i=0}^{N-1} \theta^{x_i} \cdot (1-\theta)^{1-x_i}$$

$$= \theta^{k} \cdot (1-\theta)^{N-k}$$

siendo k el número de unos $k = \sum_{i=0}^{N-1} x_i$

$\downarrow$ ln (i) muy aconsejable aplicar el ln para simplificar

$$\ln p(\underline{x}|\theta) = k \ln\theta + (N-k)\ln(1-\theta)$$

Maximizándolo:

$$\frac{\partial \ln p(\underline{x}|\theta)}{\partial\theta} = \frac{k}{\theta} - (N-k)\frac{1}{1-\theta}$$

$$= \frac{k(1-\theta) - (N-k)\theta}{\theta(1-\theta)} = 0$$

$$\Rightarrow k(1-\theta) - (N-k)\theta = 0$$

$$\Rightarrow \boxed{\hat{\theta}_{ML} = \frac{k}{N}}$$

Esto es cierto y sencillo gracias a que los x_i son incorrelados
(ej: intención de voto: PP o PSOE?
... si siempre preguntas al mismo.....)

Problema 3

$x_0, \dots, x_{M-1}$
independientes

$$p(x_i/\theta) = \theta e^{-\theta x_i} \qquad E[x_i] = \frac{1}{\theta}$$
$(x_i \geq 0, \theta \geq 0)$

- Calcular $\hat{\theta}_{ML}$
- Calcular $\hat{\theta}_{MAP}$ suponiendo $p(\theta) = \lambda e^{-\lambda\theta}$ $\begin{matrix}\theta \geq 0 \\ \lambda \geq 0\end{matrix}$

• ¿$\hat{\theta}_{ML}$?

$$\hat{\theta}_{ML} = \max_{\theta} \; p(\underline{x} \mid \theta)$$

!

$$p(\underline{x}/\theta) = \prod_{i=0}^{M-1} p(x_i/\theta) = \prod_{i=0}^{M-1} \theta e^{-\theta x_i} = \theta^M e^{-k\theta} \quad \text{siendo } k = \sum_i x_i$$

Hay que buscar el máximo.
Puede ser más sencillo tomando la función logarítmica

$$\ln p(\underline{x}/\theta) = M \ln\theta - k\theta$$

$$\frac{\partial \ln p(\underline{x}|\theta)}{\partial\theta} = \frac{M}{\theta} - k = 0$$

$$\Rightarrow \hat{\theta}_{ML} = \frac{M}{k} = \frac{1}{\underbrace{\frac{1}{M}\sum_{i=0}^{M-1} x_i}_{\text{la inversa de la media}}}$$

lo cual es lógico ya que teníamos $E[x_i] = \frac{1}{\theta}$

• ¿$\hat{\theta}_{MAP}$?

$$\hat{\theta}_{MAP} = \max_{\theta} \; p(\theta/\underline{x})$$
$$= \max_{\theta} \; p(\underline{x}/\theta) \cdot p(\theta)$$
$$= \max_{\theta} \left[\prod_{i=0}^{M-1} p(x_i/\theta)\right] \cdot (\lambda e^{-\lambda\theta}) = \max_{\theta} \left[\theta^M e^{-k\theta}\right] \cdot (\lambda e^{-\lambda\theta})$$

tomando logaritmo neperiano para maximizar más fácil

$$\ln\left[p(\underline{x}|\theta) \cdot p(\theta)\right] = M\ln\theta - k\theta + \ln\lambda - \lambda\theta$$

$$\frac{\partial \ln\left[p(\underline{x}|\theta) \cdot p(\theta)\right]}{\partial\theta} = \frac{M}{\theta} - k - \lambda$$

$$\Rightarrow \hat{\theta}_{MAP} = \frac{M}{k+\lambda}$$

se ha corregido el valor anterior a partir del conocimiento del parámetro

si $\lambda \to 0$ se tiende a una varianza muy grande en la V.A. θ

$p(\theta) = \lambda e^{-\lambda\theta} \to E[\theta] = \frac{1}{\lambda} \to var[\theta] = \frac{1}{\lambda^2}$

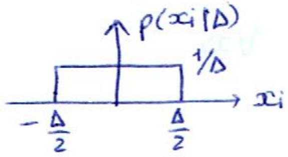
una varianza muy grande no aporta información

Problema 4 $\quad x_0, \dots, x_{M-1}$ independientes

x_i uniforme entre $-\frac{\Delta}{2}, \frac{\Delta}{2}$

- Hallar $\hat{\Delta}_{ML}$
- Hallar $\hat{\Delta}_{MAP}$ sabiendo $p(\Delta) = \lambda e^{-\lambda\Delta}$

- ¿$\hat{\Delta}_{ML}$? Vamos a poner la distribución de cada x_i



$$\longrightarrow \quad p(x_i/\Delta) = \frac{1}{\Delta}\,\text{rect}\left(\frac{x_i}{\Delta}\right)$$

$$\hat{\Delta}_{ML} = \max_{\Delta} \; p(\underline{x}|\Delta)$$
$$= \max_{\Delta} \prod_{i=0}^{M-1} p(x_i|\Delta)$$
$$= \max_{\Delta} \frac{1}{\Delta^M} \cdot \prod \text{rect}\left(\frac{x_i}{\Delta}\right)$$

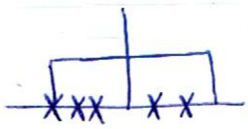
En lugar de derivar vamos a ver si encontramos el máximo razonando

$$p(\underline{x}|\Delta) = \begin{cases} 0 & \text{si } |x_i| > \frac{\Delta}{2} \text{ en algún } x_i \text{ (basta con uno!)} \\ \frac{1}{\Delta^M} & \text{si } |x_i| < \frac{\Delta}{2} \; \forall x_i \leftarrow \text{i.e. todo el vector} \end{cases}$$

me interesaría Δ muy pequeño para tener valor muy alto pero no lo suficiente como para anular algún x_i !!!

Cogeremos el mínimo posible siempre y cuando $\Delta/2$ esté por encima de todos los x_i

$$\Rightarrow \quad \hat{\Delta}_{ML} = \max_i \; 2|x_i|$$



lógicamente coger un Δ menor sería estúpido (tratamos de estimar la fdp de x_i viendo sus muestras)

Nota: y si no supiéramos que la f.d.p. es simétrica, cogeríamos entre el mínimo y el máximo

- $\hat{\Delta}_{MAP} = \max\limits_{\Delta} \; p(\underline{x}/\Delta)\, p(\Delta)$

$$= \max_{\Delta} \left[\frac{1}{\Delta^M} \prod_{i=0}^{M-1} \text{rect}\left(\frac{x_i}{\Delta}\right) \right] \cdot \lambda e^{-\lambda\Delta}$$

$$= \max_{\Delta} \begin{cases} 0 & \text{si } |x_i| \geq \frac{\Delta}{2} \text{ en algún } x_i \\ \underbrace{\lambda e^{-\lambda\Delta}}\cdot \frac{1}{\Delta^M} & \text{si } |x_i| < \frac{\Delta}{2} \; \forall x_i \end{cases}$$

este conocimiento no
nos ha aportado nada
ya que nos sigue pidiendo
que busquemos el valor
más pequeño posible de Δ
(siempre y cuando entren dentro
todas las observaciones)

$$\Rightarrow \hat{\Delta}_{MAP} = \max_i \; 2|x_i|$$

Problema 2 $x_0, \dots, x_{M-1}$ independientes

$$p(x_i|\theta) = e^{-\theta} \cdot \frac{1}{x_i!} \theta^{x_i}$$ Distribución de Poisson

$x_i = 0, 1, 2, 3, \dots$

- Hallar $\hat{\theta}_{ML}$
- Indicar si $\hat{\theta}_{ML}$ es eficiente

- $$p(\underline{x}|\theta) = \prod_{i=0}^{M-1} p(x_i|\theta) = \frac{e^{-M\theta}}{\prod_{i=0}^{M-1} x_i!} \theta^k \qquad k = \sum_{i=0}^{M-1} x_i$$

$$\ln p(\underline{x}|\theta) = -M\theta - \ln\left(\prod_{i=0}^{M-1} x_i!\right) + k \ln\theta$$

Hay que halla el máximo

$$\frac{\partial \ln p(\underline{x}|\theta)}{\partial\theta} = -M + \frac{k}{\theta}$$

$$\Rightarrow \boxed{\hat{\theta}_{ML} = \frac{k}{M} = \frac{1}{M}\sum x_i}$$

La estimación de θ es la media muestral.

Tiene sentido ya que el valor medio de una distrib de Poisson

$$p(x_i) = e^{-\theta} \cdot \frac{1}{x_i!} \theta^{x_i} \rightarrow E[x_i] = \theta$$

- Para hallar si es eficiente, hay que hallar la cota de Kramer-Raus, la cual sólo existe si el estimador es insesgado

  - ¿es insesgado? $$E[\hat{\theta}_{ML}] = E\left[\frac{1}{M}\sum_{i=0}^{M-1} x_i\right] = \frac{1}{M}\sum_{i=0}^{M-1} E[x_i] = \frac{1}{M} \cdot M\theta = \theta$$

- ¿varianza del estimador?

$$var[\hat{\theta}_{ML}] = E[(\hat{\theta}_{ML} - \theta)^2] = E[\hat{\theta}_{ML}^2] - 2E[\hat{\theta}_{ML} \cdot \theta] + E[\theta^2]$$

ya que no tenemos info de θ, para nosotros es una constante

$$= E[\hat{\theta}_{ML}^2] - 2\theta \underbrace{E[\hat{\theta}_{ML}]}_{\theta} + \theta^2$$

$$= E[\hat{\theta}_{ML}^2] - \theta^2$$

Podríamos haber llegado usando $\boxed{E[x^2] = E^2[x] + var[x]}$ siempre

falta hallar este término (valor cuadrático medio)

$$E[\hat{\theta}_{ML}^2] = \frac{1}{M^2} E\left[\left(\sum x_i\right)^2\right]$$ se puede hacer con sumatorios e integrales

$$= \frac{1}{M^2} E\left[\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} x_n \cdot x_m\right]$$

$$= \frac{1}{M^2}\sum_n\sum_m \underbrace{E[x_n \cdot x_m]}$$

$$\begin{cases} E[x_n^2] \text{ si } m=n \leftarrow M \text{ casos} \\ E[x_n x_m] \text{ si } m\neq n \leftarrow M^2-M \text{ casos}\end{cases}$$

i al ser independientes, $E[x_i x_j] = E[x_i]\cdot E[x_j]$

$$= \frac{1}{M^2}\left(\sum_{n=0}^{M-1} E[x_n^2] + \sum_{n=0}^{M-1}\sum_{m\neq n} E[x_n]\cdot E[x_m]\right)$$

$\sum E[x_i] = M\cdot E[x_i]$

$\sum\sum E[x_i]\cdot E[x_j] = \sum\sum E^2[x_i] = (M^2-M)E^2[x_i]$

$$= \frac{1}{M^2}\left\{M\cdot E[x_n^2] + (M^2-M)E^2[x_n]\right\}$$

Sabiendo que; para x_i de Poisson se tiene

$$E[x_i] = \theta$$
$$E[x_i^2] = \theta$$

$$= \frac{1}{M^2}\cdot M\cdot\theta + \frac{1}{M^2}M(M-1)\cdot\theta^2 = \frac{\theta}{M} + \frac{M-1}{M}\theta^2$$

Por tanto ya sabemos:

$$E[\hat{\theta}_{ML}] = \frac{\theta}{M} + \frac{M-1}{M}\theta^2$$
$$\text{var}[\hat{\theta}_{ML}] = \frac{\theta}{M} + \frac{M-1}{M}\theta^2 - \theta^2$$

Por tanto:

$$CCR = \frac{-1}{E\left[\frac{\partial^2 p(\mathbf{x}/\theta)}{\partial\theta^2}\right]} = \frac{1}{E\left[\frac{\delta(-M+\frac{k}{\theta})}{\partial\theta}\right]} = \frac{1}{E\left[\frac{k}{\theta^2}\right]}$$

i no olvidar aplicar $E[\cdot]$

$$= \frac{1}{E\left[\frac{\sum x_i}{\theta^2}\right]} = \frac{\theta^2}{M\theta} = \frac{\theta}{M}$$

Por lo tanto el estimador es asintóticamente eficiente (cuando $M\to\infty$)

DETECCIÓN

TEORÍA DE LA DETECCIÓN: BIBLIOGRAFÍA

[1] M.D. Srinath, P.K. Rajasekaran, ... "Introduction to statistical signal processing with applications" (Capítulo 3) Ed. Prentice Hall

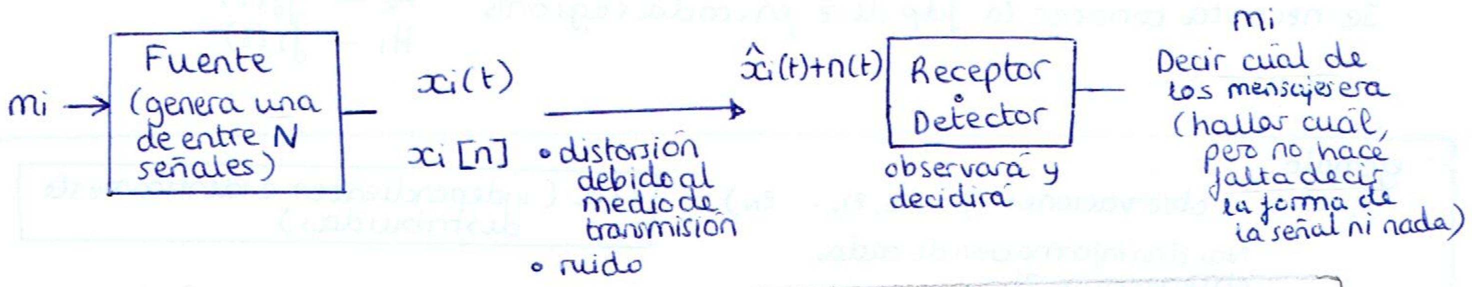
[2] Hippeustiel, Ralph Dieter: "Detection theory: applications and digital signal processing"

Índice

1. Introducción

- Deseamos elegir entre varias situaciones de manera automática

→ Seleccionar la opción correcta



queremos saber qué mensaje era el qué generó la señal
↳ Test de hipótesis

La observación es una V A.
(ya que la distorsión y el ruido no lo conocemos)

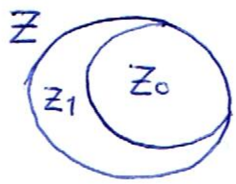
2. Test de Hipótesis

La detección y la clasificación son problemas que resolvemos aplicando Test de Hipótesis

Test de Hipótesis
- Binario (H_0 ó H_1)
  - sobre la V.A. donde tomamos la decisión 'z'
    - simple (unidimensional) $z = z_1$
    - múltiple (varias observaciones) $\underline{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$ → normalmente se reduce a una dimensión
- Múltiple ($H_0, H_1, \ldots, H_N$)
  - V.A. de la decisión 'z'
    - simple
    - múltiple

Z: región de los valores posibles de z

- En Test de Hipótesis binario se divide en dos zonas $\underbrace{Z_0 \text{ y } Z_1}_{\text{regiones de decisión}}$



$$Z = Z_0 \cup Z_1$$
$$Z_0 \cap Z_1 = \emptyset$$

De forma que decidimos la hipótesis según a qué región pertenezca z

$$z \in Z_0 \rightarrow H_0$$
$$z \in Z_1 \rightarrow H_1$$

Se necesita conocer la fdp de z para cada hipótesis
$H_0 \rightarrow f_0(z)$
$H_1 \rightarrow f_1(z)$

---

ejemplo:

vector de observaciones $\underline{z} = (z_0, z_1, \cdots z_N)$

Nos dan información de cada observación z_i

iid (independientes e idénticamente distribuidas)

no hay juego: $H_0 \longrightarrow$ z_i tiene fdp gaussiana $f_G(z_i) \rightarrow p(z_0, z_1, \ldots, z_N | H_0) = \prod_{i=1}^{n} f_G(z_i)$

sí hay juego: $H_1 \longrightarrow$ z_i tiene fdp laplaciana $f_L(z_i) \rightarrow p(\underbrace{z_0, z_1, \ldots, z_n}_{\underline{z}} | H_1) = \prod f_L(z_i)$

¿con qué criterio escojo las regiones?

Suele reducirse el problema a escoger un umbral ejemplo: Test de Hipótesis Compuesto

$z_1, z_2, \ldots, z_N$ iid gaussianas con media θ y varianza 1

$$H_1 \longrightarrow \theta = 0 \longrightarrow p(\underline{z} \mid H_0) = \prod_{i=1}^{N} N(0,1) = \frac{1}{(\sqrt{2\pi})^N} e^{\sum_{i=1}^{N} -\frac{z_i^2}{2}}$$

recuerda

$$N(m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-m)^2}{2\sigma^2}}$$

compruebe

$$H_0 \longrightarrow \theta \neq 0 \quad \text{(Test de Hipótesis compuesto)}$$

¿Cuál será la fdp en ese caso?

$\downarrow$

¡¡Yo no sé cuál será la media de los z_i en ese caso!!

$$p(\underline{z} \mid H_1) = \prod_{i=1}^{N} N(\theta, 1)$$

todas las z_i tendrán la misma θ, pero yo no sé cuál es.

Previamente tendré que estimar θ (teoría de la estima)

3. Criterios de decisión

3.1 criterio MAP (el que nos dice el sentido común)

si $p(H_0 \mid z) > p(H_1 \mid z) \longrightarrow H_0$

si $p(H_1 \mid z) > p(H_0 \mid z) \longrightarrow H_1$

$$\Longleftrightarrow \boxed{\frac{p(H_1 \mid z)}{p(H_0 \mid z)} \underset{H_0}{\overset{H_1}{\gtrless}} 1}$$

criterio MAP

¿cómo podemos conocer esto?

Por Bayes: $p(H_i \mid z) = \dfrac{p(z \mid H_i) \cdot p(H_i)}{p(z)}$

Por tanto

$$\boxed{\frac{p(z \mid H_1)}{p(z \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)}}$$

$$\boxed{\Lambda(z) \gtrless \lambda}$$

$\Lambda(z)$: razón de verosimilitud

λ: umbral

ejemplo:

Canal de comunicaciones:

Fuente: T segundos $\begin{cases} \nearrow 1 \\ \searrow 0 \end{cases}$

Canal: Añade ruido $v(t) \to$ gaussiana media nula varianza uno

Recibido: $r(t) = \begin{cases} 1 + v(t) \\ \text{ó} \\ v(t) \end{cases}$ en T segundos

Con una sola observación decidir H_0 ó H_1

$$H_0: \; z = v$$
$$H_1: \; z = 1 + v$$

Podemos deducir

- $p(z \mid H_0) = p(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

- $p(z \mid H_1)$: será también gaussiana pero con la media desplazada

$$E[z] = E[1+v] = E[1] + E[v] = 1$$
$$var[z] = E[(z-1)^2] = E[v^2] = 1$$

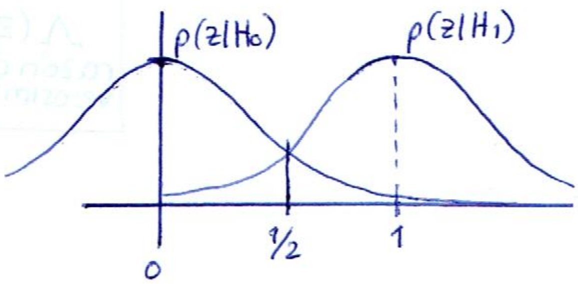
$$p(z \mid H_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-1)^2}{2}}$$

Por lo tanto podemos calcular

$$\left.\begin{aligned} \Lambda(z) &= \frac{p(z \mid H_1)}{p(z \mid H_0)} = \frac{e^{-\frac{(z-1)^2}{2}}}{e^{-\frac{z^2}{2}}} = e^{z-1/2} \\ \lambda &= \frac{p(H_0)}{p(H_1)} \end{aligned}\right\} \quad \boxed{e^{z-\frac{1}{2}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p(H_0)}{p(H_1)}}$$

$\downarrow$ tomando ln

$$\boxed{z \gtrless \frac{1}{2} + \ln\left(\frac{p(H_0)}{p(H_1)}\right)}$$



El principal problema es obtener la fdp. bajo cada hipótesis

- Criterio de Bayes

Vamos a tener en cuenta el coste que supone equivocarse, y vamos a minimizar el coste.

Llamamos D_i a la decisión tomada $i = 0, 1$
D_0 ó D_1 según creamos que la hipótesis es H_0 ó H_1

prob. conjunta intersección

Pueden pasar 4 cosas:

1. H_0 hipótesis verdadera y decidimos $D_0 \rightarrow p(H_0, D_0)$ } aciertos
2. H_1 hipótesis verdadera y decidimos $D_1 \rightarrow p(H_1, D_1)$ }
3. H_0 hipótesis verdadera y decidimos $D_1 \rightarrow p(H_0, D_1) \rightarrow$ falsa alarma C_{10}
4. H_1 hipótesis verdadera y decidimos $D_0 \rightarrow p(H_1, D_0) \rightarrow$ pérdida (miss) C_{01}

Cada decisión tendrá un coste según el contexto.

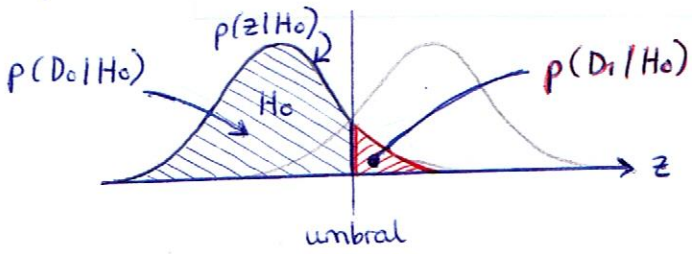
C_{ij} : coste de ser la hipótesis "j" verdadera y nosotros haber decidido "i"

Vamos a intentar minimizar el coste medio $\bar{C}$

$$\bar{C} = C_{00} \cdot p(H_0, D_0) + C_{11} \cdot p(H_1, D_1) + C_{10} \cdot p(H_0, D_1) + C_{01} \cdot p(H_1, D_0)$$

Por Bayes: $$p(D_i \mid H_j) = \frac{p(D_i, H_j)}{p(H_j)}$$

$$\bar{C} = C_{00} \cdot p(H_0) \cdot p(D_0 \mid H_0) + C_{11}\, p(H_1) p(D_1 \mid H_1) + C_{10}\, p(H_0) p(D_1 \mid H_0) + C_{01}\, p(H_1) p(D_0 \mid H_1)$$



Es fácil ver que:

$$p(D_i \mid H_j) = \int_{Z_i} p(z \mid H_j)\, dz$$

Además, puesto que $Z = Z_0 \cup Z_1$ y $Z_0 \cap Z_1 = \emptyset$

$$\int_Z p(z \mid H_j)\, dz = 1 = \int_{Z_0} p(z \mid H_j)\, dz + \int_{Z_1} p(z \mid H_j)\, dz$$

$$\Rightarrow \int_{Z_1} p(z \mid H_j) = 1 - \int_{Z_0} p(z \mid H_j)\, dz$$

Por lo tanto se deduce que podemos expresarlo todo en función de la integral en una única zona:

$$\bar{C} = C_{00} \cdot p(H_0) \cdot \int_{Z_0} p(z|H_0)\,dz$$
$$+ C_{11}\, p(H_1) \cdot \left[1 - \int_{Z_0} p(z|H_1)\,dz\right]$$
$$+ C_{10} \cdot p(H_0) \cdot \left[1 - \int_{Z_0} p(z|H_0)\,dz\right]$$
$$+ C_{01}\, p(H_1) \cdot \int_{Z_0} p(z|H_1)\,dz$$

el coste de equivocarnos es mayor que el de acertar

$$\bar{C} = \underbrace{C_{11}\,p(H_1) + C_{10}\,p(H_0)}_{\substack{\text{cantidad fija} \\ \text{(viene de} \\ \text{los unos)}}} + \underbrace{\int_{Z_0} p(H_1)\cdot \overbrace{(C_{01} - C_{11})}^{\text{positivo}}\, p(z|H_1) - p(H_0)\overbrace{(C_{10} - C_{00})}^{\text{positivo}}\, p(z|H_0)\; dz}_{\substack{\text{minimizar esto} \\ \text{(puede ser negativo)}}}$$

Nuestro criterio será:

$$p(H_1)\cdot(C_{01} - C_{11})\, p(z|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(H_0)\cdot(C_{10} - C_{00})\, p(z|H_0)$$

que queda como:

$$\boxed{\begin{array}{l} \dfrac{p(z|H_1)}{p(z|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \dfrac{C_{10} - C_{00}}{C_{01} - C_{11}} \cdot \dfrac{p(H_0)}{p(H_1)} \\[2ex] \Lambda(z) \gtrless \lambda \end{array}}$$

Criterio de Bayes

ejercicio: Una fuente emite señales con 2 posibles varianzas

$\boxed{F}$ — señal fdp gaussiana media $\mu = 0$ $\quad H_1$: varianza σ_1^2 $\quad \sigma_1^2 > \sigma_0^2$

H_0 : varianza σ_0^2

Hay que hallar:

$$p(z|H_0) = N(0;\sigma_0^2) = \frac{1}{\sqrt{2\pi}\,\sigma_0}\exp\left(-\frac{z^2}{2\sigma_0^2}\right)$$

$$p(z|H_1) = N(0:\sigma_1^2) = \frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{z^2}{2\sigma_1^2}\right)$$

Por tanto: $$\Lambda(z) = \frac{p(z|H_1)}{p(z|H_0)} = \frac{\sigma_0}{\sigma_1}\exp\left(-\frac{1}{2}z^2\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right)\right)$$

Aplicando la regla de decisión (con el ln) se obtiene

$$\ln\Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\lambda \qquad \lambda = \frac{C_{10}-C_{00}}{C_{01}-C_{11}}\cdot\frac{P(H_0)}{P(H_1)}$$

$$\ln\left(\frac{\sigma_0}{\sigma_1}\right) - \frac{1}{2}z^2\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_0^2}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\lambda$$

$$\frac{1}{2}z^2\left(\frac{\sigma_1^2-\sigma_0^2}{\sigma_1^2\cdot\sigma_0^2}\right) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\lambda - \ln\left(\frac{\sigma_0}{\sigma_1}\right)$$

si z era gaussiana, esta será chi-cuadrado

$$z^2 \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\frac{2\sigma_1^2\sigma_0^2}{\sigma_1^2\sigma_0^2}\ln\left(\frac{\sigma_1}{\sigma_0}\lambda\right)}_{\text{umbral } \gamma}$$

$$\boxed{z^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma}$$

¿cómo es de bueno el sistema?

P_F Probabilidad de falsa alarma $\quad P_I = P_r(D_1/H_0) = \int_{Z_1} p(z|H_0)\,dz \quad = \int_\lambda^\infty p(\Lambda(z)|H_0)\,dz$

P_M Probabilidad de pérdida (miss) $\quad P_{II} = P_r(D_0/H_1) = \int_{Z_0} p(z|H_1)\,dz \quad = \int_{-\infty}^{\lambda} p(\Lambda(z)|H_1)\,dz$

Probabilidad de error $\quad P_e = P_I\cdot p(H_0) + P_{II}\cdot p(H_1)$

⚠ P_F, P_M no tienen en cuenta $pr(H_0)$ frente a $p(H_1)$. P_e sí

Probabilidad de detección $\quad P_D = P_r(D_1|H_1)$

Sabemos que el criterio de Bayes es óptimo, pero se requiere conocer ciertos datos.

Hay otros criterios (ya no óptimos) que necesitan menos datos

- ## Criterio de mínima probabilidad de error

supone $C_{00} = C_{11} = 0$ (observador ideal)
$C_{10} = C_{01} = 1$

$\hookrightarrow \quad \lambda = \dfrac{P(H_0)}{P(H_1)} \rightarrow$ Equivale al MAP
Requiere conocer las prob. a priori

El coste medio

$$\bar{C} = C_{00}(1-P_F) + C_{10}P_F + p(H_1)\cdot[(C_{11}-C_{00}) + (C_{01}-C_{11})P_M - (C_{10}-C_{00})P_F]$$
$$= P_F + p(H_1)\cdot[P_M - P_F]$$
$$= P_F(1-p(H_1)) + p(H_1)P_M$$
$$= P_F\,p(H_0) + P_M\,p(H_1)$$
$$= P_e$$

El coste medio es igual a la prob. de error

---

ejemplo: $H_1 : m + n$
$H_0 : n$ ← ruido $N(0,\sigma^2)$

Determinar la regla de decisión aplicando el criterio de mínima probabilidad de error, sabiendo $p(H_0) = p(H_1) = 0'5$

$$\Lambda(z) = \frac{p(z|H_1)}{p(z|H_0)} = \frac{N(m,\sigma^2)}{N(0,\sigma^2)} = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(z-m)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{z^2}{2\sigma^2}\right)} = \exp\left(-\frac{(z-m)^2}{2\sigma^2} + \frac{z^2}{2\sigma^2}\right)$$

Por tanto el criterio queda $\quad \Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \dfrac{p(H_0)}{p(H_1)} = 1$

ln ↓

$$\frac{-\cancel{z^2} + 2mz - m^2 + \cancel{z^2}}{2\sigma^2} \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

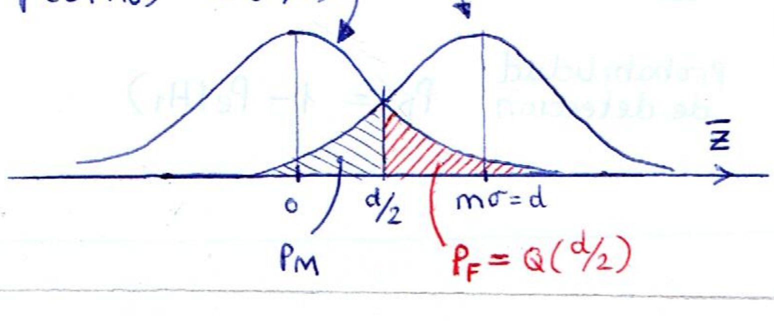
$$\boxed{z \underset{H_0}{\overset{H_1}{\gtrless}} \frac{m}{2}}$$

Típicamente se normaliza z

$\boxed{\bar{z} = z/\sigma}$ para que $p(\bar{z}/H_1) = N(d,1)$
$p(\bar{z}/H_0) = N(0,1)$

$$\boxed{\bar{z} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{m}{2\sigma} = \frac{d}{2}}$$

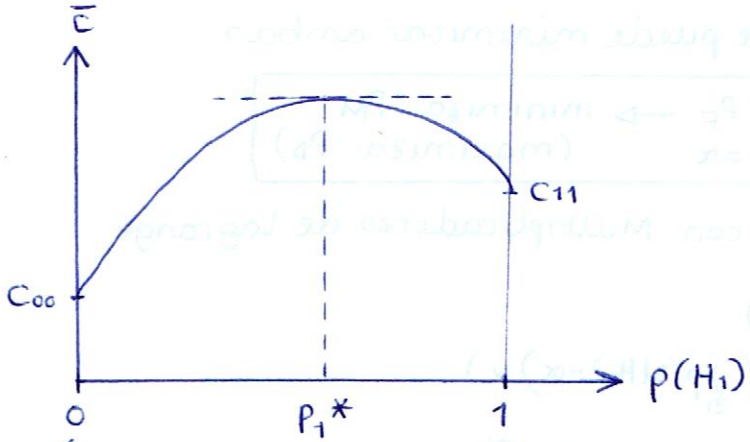
$d = \dfrac{m}{\sigma}$



- <u>Criterio minimax</u>

Se aplica cuando no conocemos las probabilidades a priori

Recuerda: $\bar{C} = C_{00}(1-P_F) + C_{10} P_F + p(H_1)\left[(C_{11}-C_{00}) + (C_{01}-C_{11})P_M - (C_{10}-C_{00})P_F\right]$



si $p(H_1)=0 \rightarrow P_F=0 \rightarrow \bar{c}=C_{00}$

caso peor para $p(H_1)$

si $p(H_1)=1 \rightarrow P_M=0 \rightarrow \bar{c}=C_{11}$

El criterio minimax supone que $p(H_1) = P_1^*$

P_1^* se obtiene maximizando $\bar{c}$

$$\frac{d\bar{c}}{dp(H_1)} = 0$$

$$= (C_{11}-C_{00}) + (C_{01}-C_{11})P_M - (C_{10}-C_{00})P_F$$

Si además no conocemos los costes y suponemos $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1$

$$\frac{d\bar{c}}{dp(H_1)} = P_M - P_F = 0$$

$$\Rightarrow \boxed{P_M = P_F}$$

- Criterio de Neyman-Pearson

→ C_{ij} desconocidos
→ $p(H_i)$ desconocidas

¿Puedo minimizar a la vez $P_F=\alpha$ y $P_M=\beta$?

Neumann vió que no se puede minimizar ambas.

se le ocurrió: $\boxed{\text{fijar } P_F=\alpha \longrightarrow \text{minimizar } P_M \text{ (maximizar } P_D)}$

Pudo hacerlo con Multiplicadores de Lagrange

$$J = P_M + \lambda(P_F - \alpha)$$

$$J = \int_{Z_0} p(z|H_1)\,dz + \lambda\left(\int_{Z_1} p(z|H_0)\,dz - \alpha\right)$$

$$= \int_{Z_0} p(z|H_1)\,dz + \lambda\left(\left[1-\int_{Z_0} p(z|H_0)\right] - \alpha\right)$$

$$= \lambda(1-\alpha) \int_{Z_0} \left[p(z/H_1) - \lambda\, p(z/H_0)\right] dz$$

Puedo escoger el λ que minimice J condicionada a la decisión sin más que hacer:

$$p(z/H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(z|H_0)\lambda \Longrightarrow \frac{p(z/H_1)}{p(z/H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$$ ← el umbral de Neyman-Pearson

Es decir, hay que resolver la ecuación:

$$\int_{\lambda}^{\infty} p(z|H_0)\,dz = \alpha$$

λ ↑ umbral que debemos determinar

α ↑ prob. de falsa alarma (dato) (exijimos un valor)

Hipótesis múltiple

- Supongamos $M > 2$ eventos $H_0, H_1, \ldots, H_{M-1}$
- Asumimos:

  a) $p(H_0), p(H_1), \ldots, p(H_{M-1})$ conocidos $\left(\sum P_i = 1\right)$

  b) C_{ij} es el coste de D_i/H_j

  c) $p(z|H_i)$

- El objetivo de decisión es minimizar el coste medio $\bar{C}$

$$\bar{C} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} C_{ij} \cdot P(H_j) \cdot P(D_i/H_j)$$

Sabiendo que $P(D_i/H_j) = \int_{Z_i} p(z|H_j)\, dz$

siendo Z_i la región de decisión de H_i $\quad Z = \bigcup_{i=0}^{M-1} Z_i$

$Z_i \cap Z_j = \emptyset \quad \forall i \neq j$

$$\bar{C} = \sum_{i=0}^{M-1} C_{ii}\, P(H_i) \int_{Z_i} p(z|H_i)\, dz + \sum_{\substack{j=0 \\ j\neq i}}^{M-1} C_{ij}\, P(H_j) \int_{Z_j} p(z|H_j)\, dz$$

Y puesto que: $\int_{Z_i} p(z|H_i)\, dz = 1 - \sum_{\substack{j=0 \\ j\neq i}}^{M-1} \int_{Z_j} p(z|H_j)\, dz$

$$\bar{C} = \underbrace{\sum_{i=0}^{M-1} C_{ii}\, P(H_i)}_{\substack{\text{coste fijo} \\ \text{(no depende} \\ \text{de } Z_i)}} + \underbrace{\sum_{i=0}^{M-1} \int_{Z_i} \sum_{\substack{j=0 \\ j\neq i}}^{M-1} P(H_j)(C_{ij} - C_{jj})\, P(z/H_j)\, dz}_{\text{coste variable}}$$

Escogeremos las regiones Z_i que minimicen

$$I_i(z) = \sum_{\substack{j=0 \\ j\neq i}}^{M-1} P(H_j)\,(C_{ij} - C_{jj})\, P(z|H_j) \qquad \text{(sin referencia)}$$

tomando referencia $p(z|H_0)$

minimizar: $$\boxed{J_i(z) = \frac{I_i(z)}{p(z|H_0)} = \sum_{\substack{j=0 \\ j\neq i}}^{M-1} P(H_j) \cdot (C_{ij} - C_{jj}) \underbrace{\frac{p(z|H_j)}{p(z|H_0)}}_{\Lambda_j(z)}}$$

siendo $\Lambda_0(z) = 1$ (referencia!)

Decidimos

$z \to H_0$ si $J_0 < J_1$, $J_0 < J_2$, $J_0 < J_3$, $\ldots$, $J_0 < J_{M-1}$

$z \to H_1$ si $J_1 < J_0$, $J_1 < J_3$, $\ldots$, $J_1 < J_{M-1}$

$\vdots$

$z \to H_{M-1}$ si $J_{M-1} < J_0$, $J_{M-1} < J_1$, $\ldots$, $J_{M-1} < J_{M-2}$

Ejemplo: Detectar entre 3 señales en fondo de ruido $n \sim N(0, \sigma_n^2)$

se decidirá con un vector de N muestras

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}$$

$S0 =$ DC m_0
$S1 =$ DC m_1
$S2 =$ DC m_2

$P(H_i) = \frac{1}{3}$

Los costes
$C_{ii} = 0$ (acertar)
$C_{ij} = 1$ (equivocarse)

Hay que calcular:

$$J_i(z) = \sum_{\substack{j=0 \\ j \neq i}}^{2} \frac{1}{3} \cdot \Lambda_j(\underline{z})$$

$$\begin{cases} J_0(\underline{z}) = \frac{1}{3}\Lambda_1(\underline{z}) + \frac{1}{3}\Lambda_2(\underline{z}) \\ J_1(\underline{z}) = \frac{1}{3}\underbrace{\Lambda_0(\underline{z})}_{1} + \frac{1}{3}\Lambda_2(\underline{z}) \\ J_2(\underline{z}) = \frac{1}{3}\underbrace{\Lambda_0(\underline{z})}_{1} + \frac{1}{3}\Lambda_1(\underline{z}) \end{cases}$$

Decidiremos H_0 cuando $J_0(\underline{z}) < J_1(\underline{z})$ y $J_0(\underline{z}) < J_2(\underline{z})$

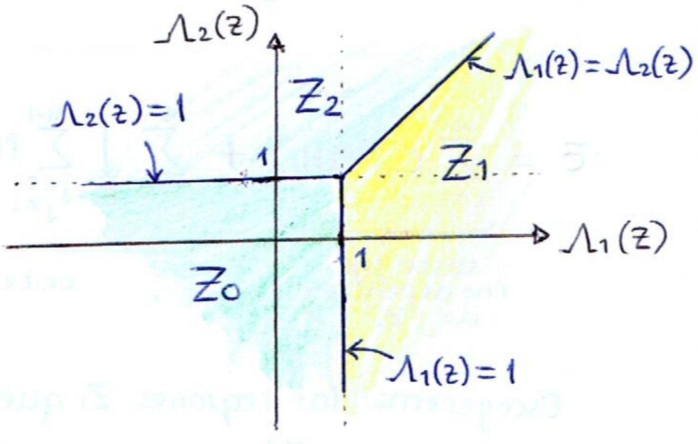
- $\frac{1}{3}\Lambda_1(\underline{z}) + \cancel{\frac{1}{3}\Lambda_2(\underline{z})} < \frac{1}{3} + \cancel{\frac{1}{3}\Lambda_2(\underline{z})} \rightarrow \Lambda_1(\underline{z}) < 1$
- $\cancel{\frac{1}{3}\Lambda_1(\underline{z})} + \frac{1}{3}\Lambda_2(\underline{z}) < \frac{1}{3} + \cancel{\frac{1}{3}\Lambda_1(\underline{z})} \rightarrow \Lambda_2(\underline{z}) < 1$

Z_0 : $\Lambda_1(\underline{z}) < 1$
$\quad\;\;\;$ $\Lambda_2(\underline{z}) < 1$

Procediendo igual

Z_1 : $\Lambda_1(\underline{z}) > 1$
$\quad\;\;\;$ $\Lambda_2(\underline{z}) < \Lambda_1(\underline{z})$

Z_2 : $\Lambda_2(\underline{z}) > 1$
$\quad\;\;\;$ $\Lambda_2(\underline{z}) > \Lambda_1(\underline{z})$



Habrá que expresar estas regiones como condición en $\underline{z}$ (no como condición en $\Lambda(\underline{z})$)

$$\Lambda_1(\underline{z}) = \frac{p(\underline{z}/H_1)}{p(\underline{z}/H_0)} \qquad \Lambda_2(\underline{z}) = \frac{p(\underline{z}/H_2)}{p(\underline{z}/H_0)}$$

con $p(\underline{z}/H_i) = \prod_{j=1}^{N} \underbrace{p(z_j/H_i)}_{N(0,\sigma_n^2)} = \prod_{j=1}^{N} \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot \exp\left(-\frac{(z_j - m_i)^2}{2\sigma^2}\right)$

Por tanto

$$p(\underline{z} \mid H_i) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\underbrace{\sum_{j=1}^{N}}_{\substack{\text{multiplicación} \\ \text{de exponenciales} \\ \text{es el sumatorio}}} \frac{(z_j - m_i)^2}{2\sigma^2}\right)$$

$$\Lambda_1(\underline{z}) = \frac{p(\underline{z} \mid H_1)}{p(\underline{z} \mid H_0)} = \exp\left[-\sum_{i=1}^{N} \frac{(z_i - m_1)}{2\sigma^2} + \sum_{i=1}^{N} \frac{(z_i - m_0)}{2\sigma^2}\right]$$

$$= \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (\cancel{z_i^2} - 2m_1 z_i + m_1^2 - \cancel{z_i^2} + 2m_0 z_i - m_0^2)\right]$$

entonces $\Lambda_1(\underline{z}) = 1$

$$\ln \Lambda_1(\underline{z}) = 0$$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \left[2(m_0 - m_1) z_i + m_1^2 - m_0^2\right] = 0$$

$$2(m_0 - m_1) \sum z_i = (m_0^2 - m_1^2) N$$

$$\underbrace{\frac{1}{N} \sum_{i=1}^{N} z_i}_{\text{la media } \bar{z}} = \frac{m_0 + m_1}{2}$$

igualmente

$$\Lambda_2(\underline{z}) = 1 \rightarrow \frac{1}{N} \sum_{i=1}^{N} z_i = \frac{m_0 + m_2}{2}$$

$$\Lambda_2(\underline{z}) = \Lambda_1(\underline{z}) \rightarrow \frac{1}{N} \sum_{i=1}^{N} z_i = \frac{m_1 + m_2}{2}$$

Por lo tanto podemos escribir las regiones como:

$$Z_0 \rightarrow \begin{bmatrix} \Lambda_1(\underline{z}) < 1 \\ \Lambda_2(\underline{z}) < 1 \end{bmatrix} \rightarrow \begin{bmatrix} \bar{z} < \frac{m_0 + m_1}{2} \\ \bar{z} < \frac{m_0 + m_2}{2} \end{bmatrix}$$

$$Z_1 \rightarrow \begin{bmatrix} \Lambda_1(\underline{z}) > 1 \\ \Lambda_2(\underline{z}) < \Lambda_1(\underline{z}) \end{bmatrix} \rightarrow \begin{bmatrix} \bar{z} > \frac{m_0 + m_1}{2} \\ \bar{z} < \frac{m_1 + m_2}{2} \end{bmatrix}$$

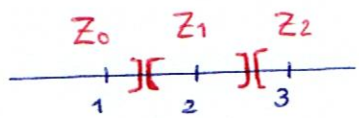
$$Z_2 \rightarrow \begin{bmatrix} \Lambda_2(\underline{z}) > 1 \\ \Lambda_2(\underline{z}) > \Lambda_1(\underline{z}) \end{bmatrix} \rightarrow \begin{bmatrix} \bar{z} > \frac{m_0 + m_1}{2} \\ \bar{z} > \frac{m_1 + m_2}{2} \end{bmatrix}$$

ejemplo: $m_0 = 1$, $m_1 = 2$, $m_2 = 3$ $\rightarrow$

$H_0 \rightarrow \begin{matrix} \bar{z} < 3/2 \\ \bar{z} < 2 \end{matrix} \rightarrow \bar{z} < 3/2$

$H_1 \rightarrow \begin{matrix} \bar{z} > 3/2 \\ \bar{z} < 5/2 \end{matrix} \rightarrow \frac{3}{2} < \bar{z} < \frac{5}{2}$

$H_2 \rightarrow \begin{matrix} \bar{z} > 5/2 \\ \bar{z} > 2 \end{matrix} \rightarrow \bar{z} > \frac{5}{2}$



Anexo: Distribución chi-cuadrado (producto de gaussianas)

Supongamos

$$l = \underline{z}^T \cdot \underline{z} = \sum_{i=1}^{N} z_i^2 \longrightarrow$$ Distribución chi-cuadrado con N grados de libertad

$\uparrow$

z_i gaussianas

$$p(l/H_i) = \begin{cases} \dfrac{l^{\left(\frac{N}{2}-1\right)}}{\left(2^{\frac{N}{2}}\right)\sigma_i^N \, \Gamma\left(\frac{N}{2}\right)} \exp\left(-\dfrac{l}{2\sigma_i^2}\right) & l \geq 0 \\ 0 & \text{resto} \end{cases}$$

Se supone sabido

al calcular $\Lambda_i(z)$ se va todo menos esto

Es muy util cuando hay que detectar CAMBIOS en la varianza de la observación

Problema 1. $H_0 \rightarrow$ transmitida continua DC de 0V
$H_1 \rightarrow$ transmitida continua DC de 0'75 V } equiprobables $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1$

· Se suma ruido uniforme de varianza 0'5 y media nula
· sólo una muestra en detección

a) Deduzca el detector óptimo y qué criterio está usando
Obtenga el umbral adecuado

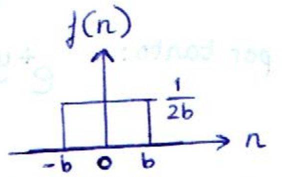
Dados los costes $C_{00} = C_{11} = 0$, $C_{01} = C_{10} = 1$ se deduce que el coste mínimo (criterio de Bayes) coincide con la mínima probabilidad de error (criterio MAP). Además, por ser las hipótesis equiprobables este es a su vez equivalente al criterio ML.

Por tanto: $\Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda = 1$

$$\frac{p(z|H_1)}{p(z|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

Se tiene:
$H_0: \quad y = n$
$H_1: \quad y = 0'75 + n$

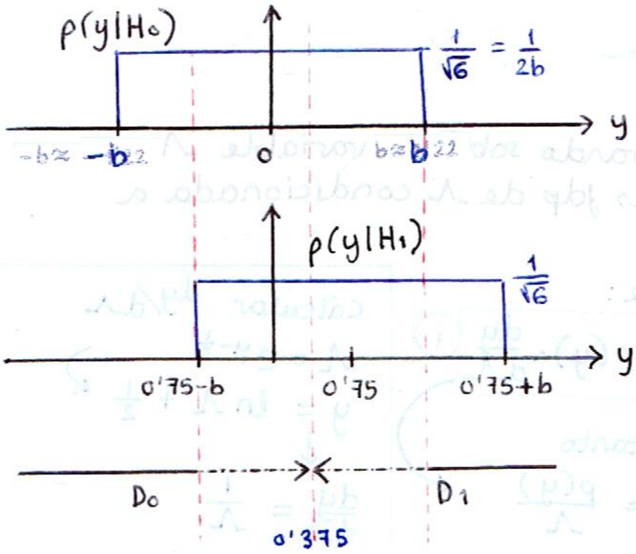
fdp de n



sabiendo que para una V.A. uniforme se tiene $\sigma_n^2 = \frac{(b-a)^2}{12} = \frac{(2b)^2}{12}$

podemos obtener b:
$$\sigma_n^2 = \frac{b^2}{3} = 0'5 \longrightarrow b = \sqrt{1'5} = \sqrt{\frac{3}{2}}$$

Parece claro que el valor DC sólo desplazará las fdp de n, por lo tanto:



Escogeremos H_0 si $y < 0'75 - b$
Escogeremos H_1 si $y > b$
¿Qué hacemos en la zona intermedia donde $\Lambda(z) = 1$?
Podemos tomar simplemente la mitad como umbral

$$y_{umb} = \frac{0'75 - b + b}{2} = 0'375$$

Por tanto: $D_0 \equiv y < 0'375$
$D_1 \equiv y > 0'375$

b) Probabilidad de error

$$P_e = P(H_0) \cdot P_{FA} + P(H_1) \cdot P_M$$
$$= 0'347$$

$P(H_0) = P(H_1) = 0'5$

$$P_{FA} = \int_{D_1} p(y|H_0)\,dy = \int_{0'375}^{b} \frac{1}{2b}\,dy = 0'347$$

$$P_M = \int_{D_0} p(y|H_1)\,dy = \int_{0'75-b}^{0'375} \frac{1}{2b}\,dy = 0'347$$

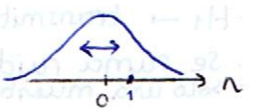
Problema 2 :

· Sólo una muestra en detección

$$H_0 : \; y = n \qquad H_1 : \; y = 1 + n \qquad n = N(0,1)$$

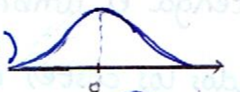
distribución Normal/Gaussiana

media varianza

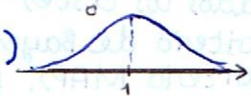


a) Detector ML y umbral

Parece claro que: $p(y|H_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = N(0,1)$



$p(y|H_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}} = N(1,1)$



$$ML \rightarrow \quad \Lambda(y) = \frac{p(y|H_1)}{p(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda = 1$$

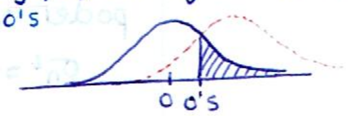
$$\Lambda(y) = \frac{e^{-\frac{(y-1)^2}{2}}}{e^{-\frac{(y-0)^2}{2}}} = e^{+\frac{y^2}{2} - \frac{(y-1)^2}{2}} = e^{\frac{\cancel{-y^2} + \cancel{y^2} + 2y - 1}{2}} = e^{+y-\frac{1}{2}}$$

por tanto: $e^{+y-\frac{1}{2}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$

$\big) \ln$

$$y - \frac{1}{2} \underset{H_0}{\overset{H_1}{\gtrless}} 0 \longrightarrow \boxed{y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2}}$$

calcular $P(D_1|H_0) = \int_{0'5}^{\infty} p(y|H_0)\,dy = Q(0'5) = 0'3085$



b) Volver a calcular $P(D_1|H_0)$ pero integrando sobre la variable Λ (una exponencial). Deberá obtener las fdp de Λ condicionada a cada hipótesis.

$$\Lambda(y) = e^{y-\frac{1}{2}}$$

$$\Lambda \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

Sabemos que: $\boxed{p(\Lambda) = p(y) \cdot \frac{dy}{d\Lambda}}$ i

por tanto $p(\Lambda) = \frac{p(y)}{\Lambda}$

Cálcular $dy/d\Lambda$

$\Lambda = e^{y-\frac{1}{2}}$

$y = \ln \Lambda + \frac{1}{2}$

$\frac{dy}{d\Lambda} = \frac{1}{\Lambda}$

entonces:

$$p(\Lambda|H_0) = \frac{p(y|H_0)}{\Lambda} = \frac{1}{\Lambda\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{\Lambda\sqrt{2\pi}} e^{-\frac{(\ln\Lambda + \frac{1}{2})^2}{2}}$$

expresas 'y' en función de Λ

$$p(\Lambda|H_1) = \frac{p(y|H_1)}{\Lambda} = \; \ldots \; = \frac{1}{\Lambda\sqrt{2\pi}} e^{-\frac{(\ln\Lambda - \frac{1}{2})^2}{2}}$$

$$P(D_1|H_0) = \int_{D_1} p(\Lambda|H_0)\,d\Lambda = \int_1^{\infty} \frac{1}{\Lambda\sqrt{2\pi}} e^{-\frac{(\ln\Lambda + \frac{1}{2})^2}{2}} d\Lambda = Q(\ln 1 + \tfrac{1}{2}) = Q(0'5)$$

cambio de variable

$\mu = \ln\Lambda + \frac{1}{2}$

$d\mu = \frac{1}{\Lambda} d\Lambda$

0'5

Problema 3. $H_0: f(y) = e^{-y} \cdot u(y)$ u(y) es el escalón

$H_1: f(y) = 2e^{-2y}\, u(y)$

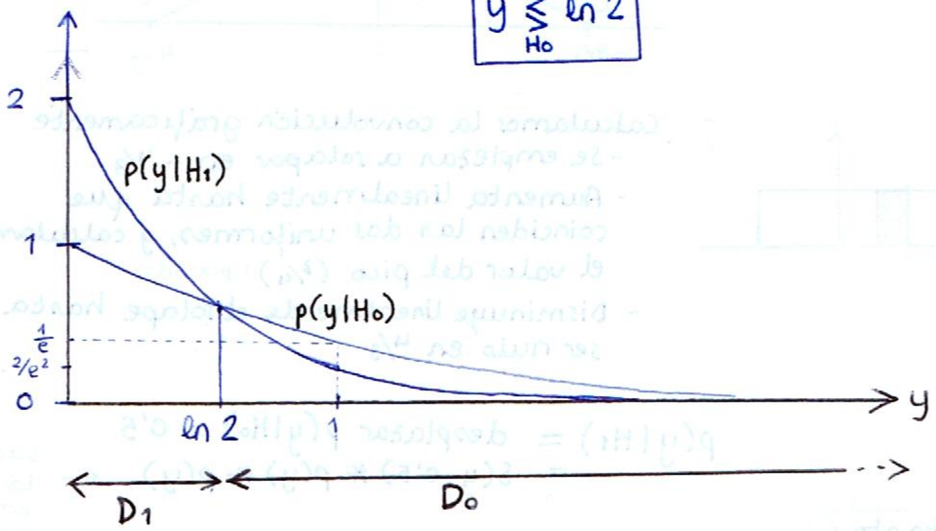
a) Detector y umbral, representar gráficamente

$$\Lambda(z) = \frac{p(y|H_1)}{p(y|H_0)} = \frac{2e^{-2y}}{e^{-y}} = 2e^{-y} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda = \frac{(C_{10}-C_{00})P(H_0)}{(C_{01}-C_{11})P(H_1)} = 1$$

$\downarrow \ln$

$$\ln 2 - y \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

$$\boxed{y \underset{H_0}{\overset{H_1}{\lessgtr}} \ln 2}$$



b) P_{FA}

$$P_D = Pr(D_0|H_0) = \int_{D_0} p(y|H_0)\,dy = \int_{\ln 2}^{\infty} e^{-y}\,dy = -\left[e^{-y}\right]_{\ln 2}^{\infty} = 1/2$$

$$P_{FA} = Pr(D_1|H_0) = \int_{D_1} p(y|H_0)\,dy = \int_0^{\ln 2} e^{-y}\,dy = -\left[e^{-y}\right]_0^{\ln 2} = 1 - 1/2 = 1/2$$

Problema 4.

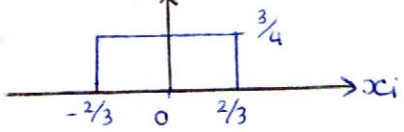
Sean x_1 y x_2 V.A. indep. uniformemente distribuidas $x_i \sim U(-\frac{2}{3}, \frac{2}{3})$

$H_0 : \quad y = x_1 + x_2$

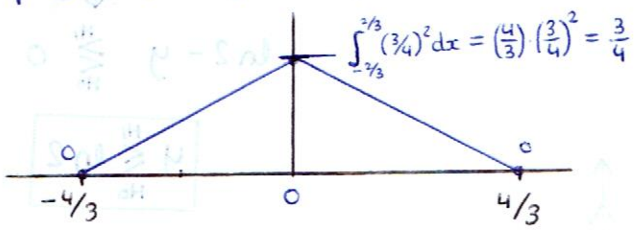
$H_1 : \quad y = 0'5 + x_1 + x_2$

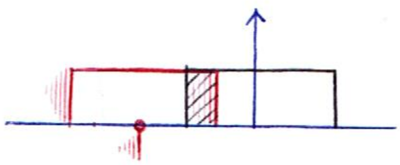
a) Detector óptimo y el umbral

$p(x_1) = p(x_2)$



$p(y|H_0) = p(x_1) * p(x_2)$





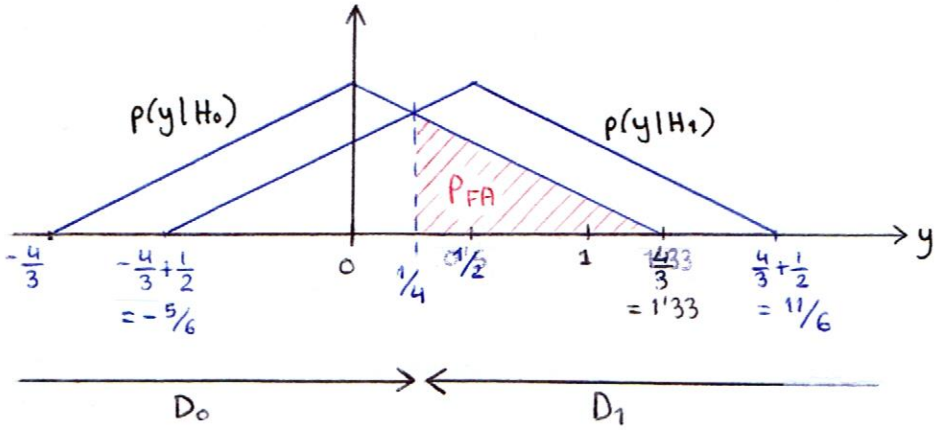
Calculamos la convolución gráficamente
- Se empiezan a solapar en $-4/3$
- Aumenta linealmente hasta que coinciden las dos uniformes, y calculamos el valor del pico $\int_\infty p(x_1) \cdot p(x_2) dx$
- Disminuye linealmente el solape hasta ser nulo en $4/3$

$p(y|H_1) =$ desplazar $p(y|H_0)$ a $0'5$
$= \delta(y-0'5) * p(y) * p(y)$ ← suma de 3 V.A. La 1ª de ellas es una constante no aleatoria (su $p(y)$ puede verse una delta en el valor de la cte)

Aplicando:

$$\Lambda(z) = \frac{p(y|H_1)}{p(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda = \frac{(C_{10}-C_{00})P(H_0)}{(C_{01}-C_{11})P(H_1)} = 1$$

$$ML \longleftrightarrow p(y|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} p(y|H_0)$$



Problema 5.

$$H_0 : f_x(x) = \frac{1}{2} e^{-|x|}$$
$$H_1 : f_x(x) = N(0,4) = \frac{1}{\sqrt{2\pi 4}} e^{-\frac{x^2}{8}}$$

Determinar la LRT y el umbral $\qquad \Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \lambda$

sin información de los costes ni de las $p(H_0)$, $p(H_1)$ asumimos $\lambda = 1$

$$\Lambda(x) = \frac{p(x|H_1)}{p(x|H_0)} = \frac{\frac{1}{\sqrt{2\pi 4}} e^{-\frac{x^2}{8}}}{\frac{1}{2} e^{-|x|}} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{8} + |x|} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

$$\ln \Lambda(x) = -\ln\sqrt{2\pi} - \frac{x^2}{8} + |x| \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

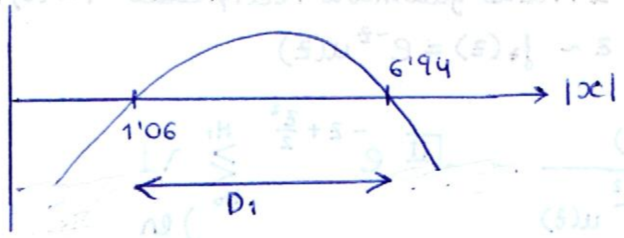
$$\boxed{|x| - \frac{x^2}{8} \underset{H_0}{\overset{H_1}{\gtrless}} \ln\sqrt{2\pi}} = \frac{1}{2}\ln 2\pi$$

Para obtener los umbrales habrá que resolver la ecuación cuadrática

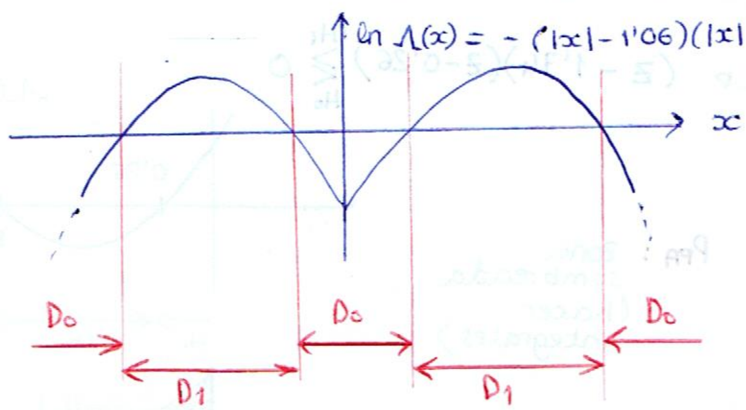
$$-|x|^2 + 8|x| - 4\ln 2\pi \gtrless 0$$
$$[\, a = -1 \quad b = 8 \quad c = -4\ln 2\pi \,] \qquad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

soluciones $\quad x = \begin{cases} 1'06 \\ 6'94 \end{cases}$

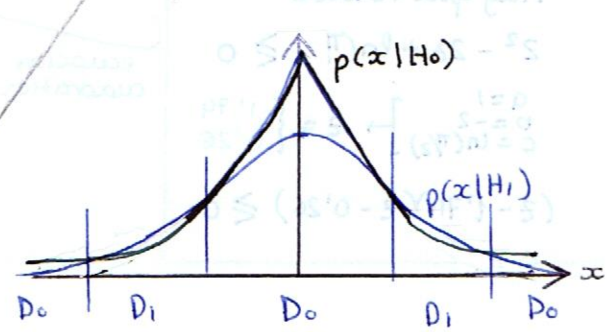
$$-(|x| - 1'06)(|x| - 6'94) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$



Nota: Cuidado. Las regiones en x serán más!



Gráficamente tiene sentido:



Problema 6 : $H_0 : p(y) = a e^{-ay}$ $a > b > 0$
$H_1 : p(y) = b e^{-by}$

en el problema original había que deducir esta constante $\Rightarrow \int_0^\infty k e^{-ay} dy = \left[-\frac{k}{a} e^{-ay}\right]_0^\infty = \frac{k}{a} = 1$

a) Criterio ML

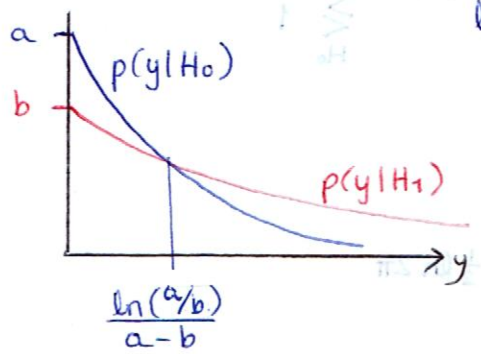
$$\Lambda(y) = \frac{p(y|H_1)}{p(y|H_0)} = \frac{b e^{-by}}{a e^{-ay}} = \frac{b}{a} e^{+(a-b)y} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda = 1$$

$a > b$

$\Big) \ln$

$$\ln\left(\frac{b}{a}\right) + (a-b)y \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

$$\boxed{y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\ln(a/b)}{a-b}}$$



b) Probabilidad de falsa alarma

$$P_{FA} = P_r(D_1, H_0) = \int_{D_1} p(y|H_0) dy = \int_{\frac{\ln(a/b)}{a-b}}^{\infty} e^{-ay} dy = \left[-e^{-ay}\right]_{\frac{\ln(a/b)}{a-b}}^{\infty} = e^{-\frac{a}{a-b} \cdot \ln(a/b)}$$

$$= e^{-\frac{a}{a-b} \ln\left(\frac{a}{b}\right)} = e^{\ln\left[\left(\frac{a}{b}\right)^{-\frac{a}{a-b}}\right]} = \left(\frac{b}{a}\right)^{\frac{a}{a-b}}$$

Problema 7. $H_0 : y = x$ x: ruido gaussiano rectificado $\sim |N(0,1)|$
$H_1 : y = z$ $z \sim f_z(z) = e^{-z} u(z)$

$$\Lambda(z) = \frac{p(y|H_1)}{p(y|H_0)} = \frac{e^{-z} u(z)}{\frac{2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} u(z)} = \sqrt{\frac{\pi}{2}} e^{-z + \frac{z^2}{2}} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

gaussiana rectificada

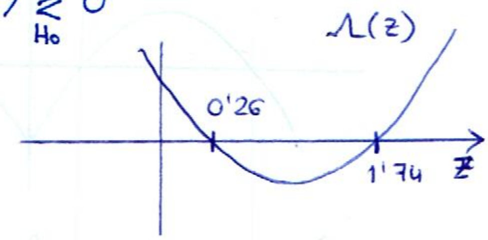
$\Big) \ln$

$$\frac{1}{2} \ln\left(\frac{\pi}{2}\right) - z + \frac{z^2}{2} \gtrless 0$$

$$(z - 1'74)(z - 0'26) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

Hay que resolver
$z^2 - 2z + \ln(\pi/2) \gtrless 0$
$\left.\begin{array}{l} a = 1 \\ b = -2 \\ c = \ln(\pi/2) \end{array}\right] \rightarrow z = \begin{cases} 1'74 \\ 0'26 \end{cases}$
$(z - 1'74)(z - 0'26) \gtrless 0$

ecuación cuadrática



P_{FA} : zona sombreada (hacer integrales)