Fundamentos en teoría de detección

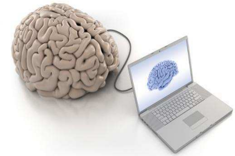
A. M. Alvarez-Meza, Ph.D.
amalvarezme@unal.edu.co

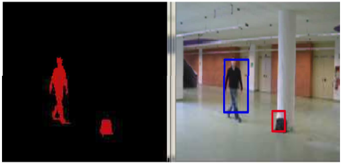
Departamento de ingeniería eléctrica, electrónica y computación
Universidad Nacional de Colombia-sede Manizales

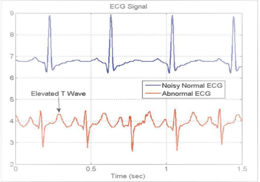
Contenido

Contenido

Teoría de detección y sus aplicaciones



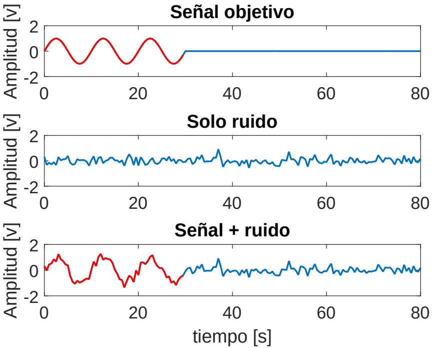




- Comunicaciones [Hsu, 2002].
- Bioseñales [Naït-Ali, 2009].
- Imágenes y video [Tekalp, 2015].
- **Análisis de datos** [Zikopoulos et al., 2011].

La teoría de detección es fundamental para identificar patrones discriminantes desde datos

Teoría de detección desde filtración óptima binaria (FOB)



- Detección de eventos de interés.
- Extraer información del evento.
- Desde estadística: *Test de hipótesis* [Kay, 1998].
- Test de hipótesis binaria: *¿Solo ruido o Señal + ruido?=FOB*.

¡La FOB puede entenderse como un problema de **detección**!

Teoría de detección y su contexto

Campos afines

- Análisis de aleatoriedad [Liptser and Shiryaev, 2013].
- Teoría de estimación [Kay, 1993].
- Habilidades en programación (Matlab, Python, o R).

Objetivos

- Comprender los principios básicos de FOB desde teoría de detección.
- Aplicar la FOB en problemas reales orientados al procesado de señales y al análisis de datos (clasificación $\rightarrow$ Aprendizaje de máquina $\rightarrow$ Analítica de datos).

Contenido

Detección de nivel DC: problema matemático

- Se dispone de una única señal de muestra $x[0] \in \mathbb{R}$.
- Se asume una amplitud de nivel DC $A \in \mathbb{R}$, distorsionada con ruido blanco Gaussiano (RBG) $w[0] \sim \mathcal{N}(0, \sigma^2)$,
  $\sigma^2 \in \mathbb{R}^+$: varianza del ruido.
- **Objetivo de la FOB**: decidir (*discriminar*) entre:
  - Hipótesis solo ruido $\mathcal{H}_o : x[0] = w[0]$.
  - Hipótesis señal + ruido $\mathcal{H}_1 : x[0] = A + w[0]$.
- **Posible solución:** dado que el ruido tiene media cero:
  - Para $x[0] < A/2 \implies \mathcal{H}_o$.
  - Para $x[0] \geq A/2 \implies \mathcal{H}_1$.

  Se espera detectar de forma correcta la mayoría de las veces

Detección de nivel DC: modelado de incertidumbre

- Repetir el experimento varias veces.
- Función de densidad de probabilidad (fdp) del ruido:

$$p(w[0]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} w^2[0]\right)$$

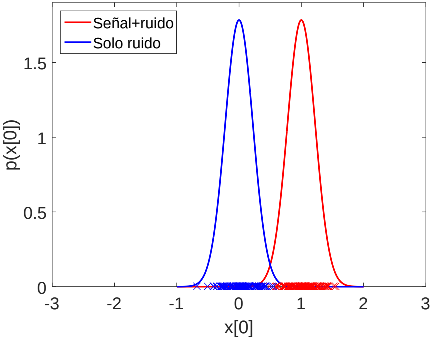
- El desempeño del detector depende de la diferencia entre fpds:

$$p(x[0]; \mathcal{H}_o) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} x^2[0]\right)$$

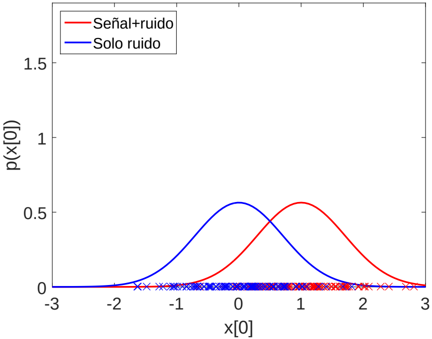
$$p(x[0]; \mathcal{H}_1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[0] - A)^2\right)$$

- La detección de nivel DC es un **test de parámetro** desde la familia de fdps: $p(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (x[0] - A)^2\right)$.

Detección de nivel DC: modelado de incertidumbre



$A = 1,\ \sigma^2 = 0.05$



$A = 1,\ \sigma^2 = 0.5$

Mayor distancia entre fdps = Mayor relación señal ruido (RSR): A^2/σ^2
¡Mejor desempeño en FOB!

Consideraciones alrededor de la FOB

- En la práctica se trabaja con RSR bajas.
- Se debe **restringir** el valor de RSR requerido.
- El éxito de la FOB dependerá entonces de la **cantidad** de datos procesados.
- Detección nivel DC desde la señal $\{x[n] \in \mathbb{R}\}_{n=0}^{N-1}$:
  - Hipótesis solo ruido $\mathcal{H}_o : x[n] = w[n]$
  - Hipótesis señal + ruido $\mathcal{H}_1 : x[n] = A + w[n]$

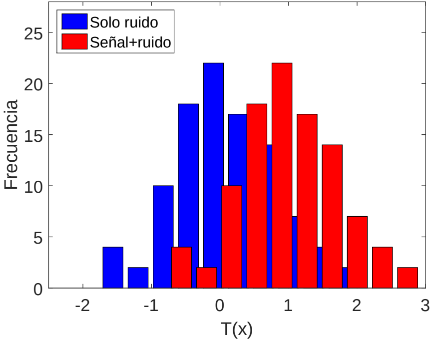
  donde $w[n] \sim \mathcal{N}\left(0, \sigma^2\right)$.

FOB desde estadístico $T(x)$

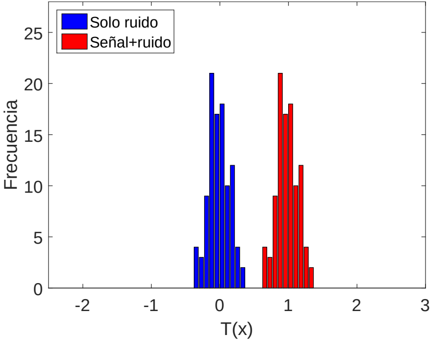
- Para aprovechar las N muestras $x[n]$ se detecta a partir de algún estadístico estimado desde los datos: $T : \mathbb{R}^N \to \mathbb{R}$.
- Ej: Media muestral $T(x) = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$.
- La FOB se resume entonces:
  - Para $T(x) < \zeta \implies \mathcal{H}_o$.
  - Para $T(x) \geq \zeta \implies \mathcal{H}_1$.

  $\zeta \in \mathbb{R}$: umbral.
- Desde teoría de estimación: a **mayor** número de **muestras**, mejor será el **desempeño** en FOB.

Estadístico $T(x)$ en FOB: detección nivel DC



$A = 1,\ \sigma^2 = 0.5,\ N = 1,\ d^2 = 2$



$A = 1,\ \sigma^2 = 0.5,\ N = 20,\ d^2 = 40$

Experimento con 100 repeticiones

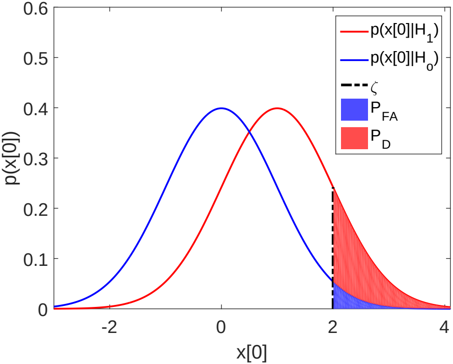
- **Coeficiente de desvío.** $d^2 = \frac{(\mathbb{E}\{T(x);\mathcal{H}_1\} - \mathbb{E}\{T(x);\mathcal{H}_o\})^2}{\text{var}\{T(x);\mathcal{H}_o\}}$
- En nivel DC $d^2 = NA^2/\sigma^2$; Si $N \to \infty$, $d \to \infty$

Contenido

Detección por Neyman-Pearson

- La FOB se resuelve desde decisión estadística como un **test de hipótesis** simple.
- Se asume que las fdps $p(x; \mathcal{H}_o)$ y $p(x; \mathcal{H}_1)$ son conocidas y se impone un **nivel de significancia**.
- **Neyman-Pearson** (NP) se puede entender como la solución clásica (frecuentista) en detección desde teoría de estimación.
- NP: **maximización** de la probabilidad de **detección** P_D sujeta a un valor **deseado** de probabilidad de **falsa alarma** P_{FA}.

Neyman-Pearson: falsa alarma y detección



Regla de detección

$$L(\boldsymbol{x}) = \frac{p(\boldsymbol{x};\mathcal{H}_1)}{p(\boldsymbol{x};\mathcal{H}_o)} > \zeta$$

$\zeta \in \mathbb{R}$ se estima a partir de un nivel de significancia $\alpha \in \mathbb{R}^+$:

$$P_{FA} = P(\mathcal{H}_1;\mathcal{H}_o) = P(x[0] > \zeta;\mathcal{H}_o)$$

$$P_{FA} = \int_{\{\boldsymbol{x}:L(\boldsymbol{x})>\zeta\}} p(\boldsymbol{x};\mathcal{H}_o)d\boldsymbol{x} = \alpha,$$

Se utilizan funciones **cumulativas complementarias**. Además:

$$P_D = P(\mathcal{H}_1;\mathcal{H}_1) = P(x[0] > \zeta;\mathcal{H}_1)$$

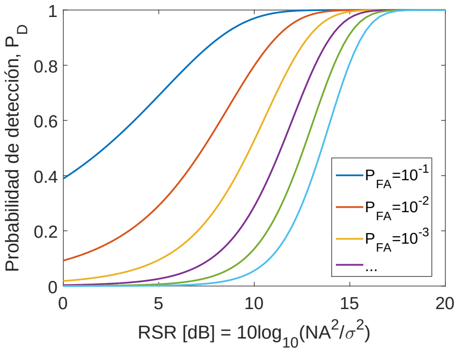
$$P_D = \int_{\{\boldsymbol{x}:L(\boldsymbol{x})>\zeta\}} p(\boldsymbol{x};\mathcal{H}_1)d\boldsymbol{x}$$

Neyman-Pearson: detección de nivel DC

- Para una señal $\boldsymbol{x} = \{x[n]\}_{n=0}^{N-1}$, las hipótesis se fijan como:
  - Hipótesis solo ruido $\mathcal{H}_o : \boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$
  - Hipótesis señal + ruido $\mathcal{H}_1 : \boldsymbol{x} \sim \mathcal{N}(A\mathbf{1}, \sigma^2 \boldsymbol{I})$
- $L(\boldsymbol{x}) = \frac{-1}{2\sigma^2}\left(-2A\sum\limits_{n=0}^{N-1} x[n] + NA^2\right) > \log(\zeta)$
- $L(\boldsymbol{x}) = T(\boldsymbol{x}) = \frac{1}{N}\sum\limits_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\log(\zeta) + \frac{A}{2} = \zeta'$ (Tarea: Demostrar)
- La detección de nivel DC se basa en la estimación de A desde:

$$T(\boldsymbol{x}) \sim \begin{cases} \mathcal{N}(0, \sigma^2/N) & \mathcal{H}_o \\ \mathcal{N}(A, \sigma^2/N) & \mathcal{H}_1 \end{cases}$$

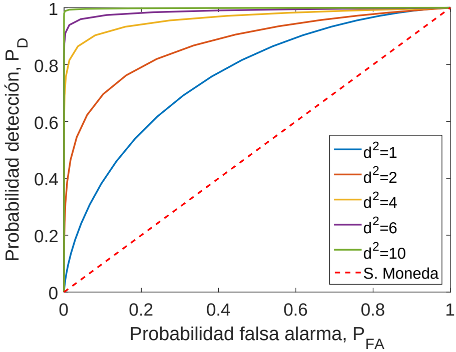
Neyman-Pearson: detección de nivel DC



$$P_{FA} \downarrow RSR \uparrow \implies P_D \uparrow$$
$$d^2 = \frac{N(\mu_o - \mu_1)^2}{\sigma^2}$$

- $P_{FA} = Q(\zeta'/\sqrt{\sigma^2/N})$
- $P_D = Q\left((\zeta' - A)/\sqrt{\sigma^2/N}\right)$
- $\zeta' = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(P_{FA})$
- Q: función de distribución cumulativa complementaria para $\mathcal{N}(0,1)$.
  (Tarea: Consultar y demostrar para el ejemplo)
- P_{FA} y P_D pueden aproximarse mediante simulación de **Monte Carlo**.

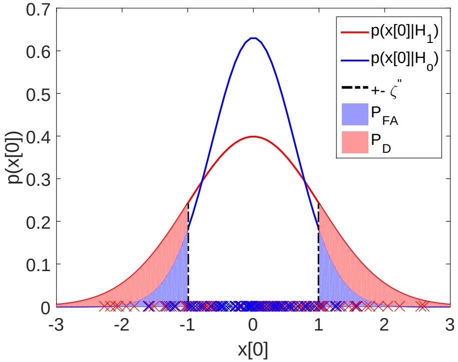
Neyman-Pearson: Curva ROC detección nivel DC



$$P_{FA} \downarrow RSR \uparrow \Longrightarrow P_D \uparrow$$
$$d^2 = \frac{N(\mu_o - \mu_1)^2}{\sigma^2}$$

- P_{FA} = 1-especificidad = FPR
- P_D = sensibilidad = TPR
- FPR y TPR comunmente usados en **reconocimiento de patrones!**
- ROC: Receiver Operating Characteristic (Característica Operativa del Receptor) - gráfica de sensibilidad vs especificidad en clasificador binario al variar umbral.

Neyman-Pearson: ¿Qué información extraer desde $T(x)$?



$$\mathcal{H}_o : x[n] \sim \mathcal{N}(0, \sigma_o^2)$$
$$\mathcal{H}_1 : x[n] \sim \mathcal{N}(0, \sigma_1^2)$$
$$\sigma_o^2 > \sigma_1^2$$

- $L(\boldsymbol{x}) = T(\boldsymbol{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]^2 > \zeta''$
- $\zeta'' = \dfrac{\frac{2}{N}\log(\zeta) + \log\left(\frac{\sigma_o^2}{\sigma_1^2}\right)}{\frac{1}{\sigma_o^2} - \frac{1}{\sigma_1^2}}$
  (Tarea: Demostrar)
- NP: estimación de $\theta \in \mathbb{R}$ desde una **fdp parametrizada** $p(\boldsymbol{x}; \theta)$.
- $T(\boldsymbol{x})$ codifica la **información relevante** en $\boldsymbol{x}$ para FOB (clasificación binaria).

Contenido

FOB desde Bayes: Mínima probabilidad de error

- **Información a priori** = asignar probabilidades a $\mathcal{H}_c$.
- **Probabilidad de error** desde Bayes (FOB $c \in \{0,1\}$):

$$P_e = P(\mathcal{H}_o|\mathcal{H}_1)P(\mathcal{H}_1) + P(\mathcal{H}_1|\mathcal{H}_o)P(\mathcal{H}_o)$$

- $P(\mathcal{H}_c|\mathcal{H}_{c'})$: probabilidad de decidir $\mathcal{H}_c$ dado que $\mathcal{H}_{c'}$ es cierto (condicional).
- $P(\mathcal{H}_c)$: prior = **suposiciones** sobre ocurrencia de $\mathcal{H}_c$.
- **Regla de detección**

$$L(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{H}_1)}{p(\boldsymbol{x}|\mathcal{H}_o)} > \frac{P(\mathcal{H}_o)}{P(\mathcal{H}_1)} = \zeta$$

FOB desde Bayes: Máxima verosimilitud

- Si $P(\mathcal{H}_o) = P(\mathcal{H}_1)$, se escoge la hipótesis con **máxima verosimilitud** (MV) condicional $p(\boldsymbol{x}|\mathcal{H}_c)$.
- En MV $L(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{H}_1)}{p(\boldsymbol{x}|\mathcal{H}_o)} > 1$
- Para detección nivel DC:
  - En NP: $L(\boldsymbol{x}) = T(\boldsymbol{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\log(\zeta) + \frac{A}{2} = \zeta'$;
    $P_{FA} = Q\left(\sqrt{\frac{N}{\sigma^2}}(\zeta' - A)\right)$
  - En MV se escoge $\mathcal{H}_1$ si: $L(\boldsymbol{x}) = T(\boldsymbol{x}) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{A}{2}$;
    $P_e = Q\left(\sqrt{\frac{NA^2}{4\sigma^2}}\right)$

FBO desde Bayes: Máximo a posteriori

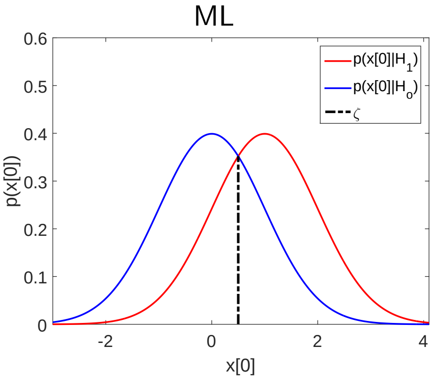
- Desde teorema de Bayes:

$$p(\boldsymbol{x}, \mathcal{H}_c) = p(\mathcal{H}_c|\boldsymbol{x})p(\boldsymbol{x}) = p(\boldsymbol{x}|\mathcal{H}_c)p(\mathcal{H}_c)$$

- Se escoge la hipótesis con **máximo posterior** (MAP) $p(\mathcal{H}_c|\boldsymbol{x})$.
- En MAP se escoge $\mathcal{H}_1$ si: $L(\boldsymbol{x}) = p(\mathcal{H}_1|\boldsymbol{x}) > p(\mathcal{H}_o|\boldsymbol{x})$
- MAP con respecto a las verosimilitudes:

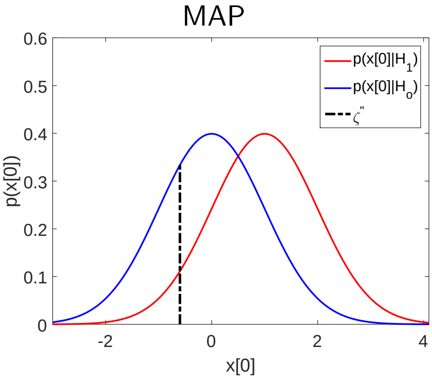
$$L(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{H}_1)}{p(\boldsymbol{x}|\mathcal{H}_o)} > \frac{P(\mathcal{H}_o)}{P(\mathcal{H}_1)} = \zeta$$

Detección de nivel DC: ML vs MAP



$A=1,\ \sigma^2=1,\ N=1\ P(\mathcal{H}_o)=P(\mathcal{H}_1)$

$T(\boldsymbol{x})>\zeta=A/2$



$A=1,\ \sigma^2=1,\ N=1$

$P(\mathcal{H}_o)=1/2\neq P(\mathcal{H}_1)=3/4$

$T(\boldsymbol{x})>\zeta''=\frac{\sigma^2}{NA}\log\left(\frac{P(\mathcal{H}_o)}{P(\mathcal{H}_1)}\right)+A/2$

Tarea: Demostrar las expresiones anteriores.

Riesgo Bayesiano generalizado: múltiples hipótesis

- Múltiples hipótesis = **clasificación multi-clase**.
- Incluye **penalización** (costo) $\xi_{cc'} \in \mathbb{R}^+$, con $c, c' \in \{1, 2, \ldots, C\}$ y $\xi_{cc} = 0$.
- **Riesgo de Bayes generalizado**:

$$\mathcal{R} = \sum_{c=1}^{C} \sum_{c'=1}^{C} \xi_{cc'} P(\mathcal{H}_c | \mathcal{H}_{c'}) P(\mathcal{H}_{c'})$$

- De FOB a múltiples hipótesis (clases) = clasificación = **reconocimiento de patrones**

Contenido

Conclusiones

- La **FOB** es la base de los métodos de **teoría de detección**.
- Neyman-Person y Bayes (ML y MAP) facilitan la **inclusión** de **restricciones** (FA o priors) de la aplicación.
- **Reconocimiento de patrones** $=$ test múltiples hipótesis.

Ejercicios Laboratorio

En un cuaderno (notebook) implemente el problema de detección de nivel DC desde Neyman Person, ML y MAP variando los niveles de confianza y para diferentes configuraciones de los priors. Envie/comparta su notebook al correo `amalvarezme@unal.edu.co`.

Referencias I

Hsu, H. P. (2002).
Analog and Digital Communications (Schaum's Outlines).
New York: McGraw Hill.

Kay, S. M. (1993).
Fundamentals of statistical signal processing, volume i: estimation theory.

Kay, S. M. (1998).
Fundamentals of statistical signal processing, vol. ii: Detection theory.
Signal Processing. Upper Saddle River, NJ: Prentice Hall.

Liptser, R. S. and Shiryaev, A. N. (2013).
Statistics of random Processes: I. general Theory, volume 5.
Springer Science & Business Media.

Naït-Ali, A. (2009).
Advanced biosignal processing.
Springer Science & Business Media.

Tekalp, A. M. (2015).
Digital video processing.
Prentice Hall Press.

Zikopoulos, P., Eaton, C., et al. (2011).
Understanding big data: Analytics for enterprise class hadoop and streaming data.
McGraw-Hill Osborne Media.