



Ingeniería Eléctrica

FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

---

# Análisis de Sistemas Dinámicos y Estimación

EL3204-1

---

*Pauta Auxiliar 8 - Lema de Neyman-Pearson*

Prof Marcos Orchard - Sebastian Espinoza.  
Prof Auxiliar Erik Sáez Aravena.

# 1 Resumen

El **Lema de Neyman-Pearson** es un resultado fundamental en teoría de detección que caracteriza el test óptimo para distinguir entre dos hipótesis simples. Dado un problema de detección binaria:

$$\begin{aligned}\mathcal{H}_0 : X &\sim f_{X_1^n}(x_1, \dots, x_n | \theta = 0) \\ \mathcal{H}_1 : X &\sim f_{X_1^n}(x_1, \dots, x_n | \theta = 1)\end{aligned}$$

donde  $L(x_1^n | \theta)$  representa la **función de verosimilitud**:

- Para variables continuas:

$$L(x_1^n | \theta) = f_{X_1^n}(x_1, \dots, x_n | \theta) \quad (1)$$

- Para variables discretas:

$$L(x_1^n | \theta) = P_{X_1^n}(X_1 = x_1, \dots, X_n = x_n | \theta) \quad (2)$$

Para un  $\nu > 0$  arbitrario y una variable aleatoria binaria  $\rho(w)$ , se tiene que el test aleatorio de la forma:

$$\pi_\nu(w, x) = \begin{cases} 1 & \text{si } L(x | \theta = 1) > \nu L(x | \theta = 0) \\ 0 & \text{si } L(x | \theta = 1) < \nu L(x | \theta = 0) \\ \rho(w) & \text{si } L(x | \theta = 1) = \nu L(x | \theta = 0) \end{cases} \quad (3)$$

es óptimo para su tamaño entre  $[0, 1]$ . Dado que  $X$  es una variable aleatoria, el test se puede reescribir como una función de variable aleatoria notando que:

$$\pi_\nu(\rho, X) = \begin{cases} 1 & \text{si } L(X | \theta = 1) > \nu L(X | \theta = 0) \\ 0 & \text{si } L(X | \theta = 1) < \nu L(X | \theta = 0) \\ \rho & \text{si } L(X | \theta = 1) = \nu L(X | \theta = 0) \end{cases} \quad (4)$$

con  $\rho$  independiente de  $X$ . Además, el test de la forma (haciendo  $\nu \rightarrow \infty$ ):

$$\pi(x) = \begin{cases} 1 & \text{si } L(x | \theta = 1) = 0 \\ 0 & \text{si } L(x | \theta = 1) > 0 \end{cases} \quad (5)$$

es óptimo de tamaño 0 y el test de la forma (haciendo  $\nu \rightarrow 0^+$ ):

$$(\forall x \in \mathbb{X}) \pi(x) = 1 \quad (6)$$

es óptimo de tamaño 1.

Para evaluar un test de detección  $\pi : \mathbb{X} \mapsto \{0, 1\}$ , usamos dos probabilidades fundamentales:

**Tamaño del test** (probabilidad de falsa alarma):

$$\begin{aligned}\alpha_\pi &\triangleq P_{X_1^n}(\pi(X_1^n) = 1 | \theta = 0) \\ &= P_{X_1^n}(X_1^n \in \pi^{-1}(\{1\}) | \theta = 0) \\ &= \int_{\{(x_1, \dots, x_n) \in \mathbb{R}^n : \pi(x_1^n) = 1\}} f_{X_1^n}(x_1, \dots, x_n | \theta = 0) dx_1 \dots dx_n \\ &= \mathbb{E}_{X_1^n}(\pi(X_1^n) | \theta = 0)\end{aligned} \quad (7)$$

**Poder del test** (probabilidad de correcta detección de la hipótesis alternativa):

$$\begin{aligned}
 \beta_\pi &\triangleq P_{X_1^n}(\pi(X_1^n) = 1 | \theta = 1) \\
 &= P_{X_1^n}(X_1^n \in \pi^{-1}(\{1\}) | \theta = 1) \\
 &= \int_{\{(x_1, \dots, x_n) \in \mathbb{R}^n : \pi(x_1^n) = 1\}} f_{X_1^n}(x_1, \dots, x_n | \theta = 1) dx_1 \dots dx_n \\
 &= \mathbb{E}_{X_1^n}(\pi(X_1^n) | \theta = 1)
 \end{aligned} \tag{8}$$

Notar que  $P_{X_1^n}(\pi(X_1^n) = 0 | \theta = 1)$  es la probabilidad de no detección o el error de tipo II que corresponde precisamente a  $1 - \beta_\pi$ . El objetivo del Lema de Neyman-Pearson es **maximizar**  $\beta_\pi$  **sujeto a que**  $\alpha_\pi \leq \alpha_0$ , donde  $\alpha_0$  es un tamaño máximo permitido.

El cociente de verosimilitud mide qué tan “favorable” es la observación para  $\mathcal{H}_1$  en comparación con  $\mathcal{H}_0$ :

- Si  $L(x | \theta = 1) > \nu L(x | \theta = 0)$ : la observación favorece fuertemente a  $\mathcal{H}_1 \rightarrow$  decidimos  $\mathcal{H}_1$  ( $\pi = 1$ )
- Si  $L(x | \theta = 1) < \nu L(x | \theta = 0)$ : la observación favorece a  $\mathcal{H}_0 \rightarrow$  decidimos  $\mathcal{H}_0$  ( $\pi = 0$ )
- Si  $L(x | \theta = 1) = \nu L(x | \theta = 0)$ : zona de incertidumbre  $\rightarrow$  se usa aleatorización con  $\rho(w)$

Los tests aleatorizados (con  $\rho \in (0, 1)$ ) son necesarios cuando las observaciones son discretas y es imposible alcanzar exactamente  $\alpha = \alpha_0$  con un test determinístico.

1. Considere una variable aleatoria  $X$  con distribución Poisson de parámetro  $\lambda$ .

$$P_X(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (9)$$

- (a) Determine la función generadora de momentos de  $X$ , es decir:

$$M_X(t) = \sum_{k \geq 0} P_X(X = k) \cdot e^{tk}, \quad (10)$$

y verifique que es igual a  $e^{\lambda(e^t-1)}$ .

- (b) Considere  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas (i.i.d.) con distribución Poisson de parámetro  $\lambda$ . Verifique que  $X = \sum_{i=1}^n X_i$  es Poisson de parámetro  $n\lambda$ . *Indicación:* Considere los resultados de probabilidades respecto a suma de variables aleatorias y las propiedades de la función generadora de momentos.
- (c) Considere el problema de detección binario en el escenario paramétrico, donde  $\Theta = \{0, 1\}$  y se tiene que:

$$\theta = 0 \Rightarrow X \sim \text{Poisson}(\lambda_0), \quad (11)$$

$$\theta = 1 \Rightarrow X \sim \text{Poisson}(\lambda_1) \quad (12)$$

con  $\lambda_1 > \lambda_0$ . Determine la forma general de la familia de test óptimos dados por el Lema de Neyman-Pearson, y analice la forma de las zonas de decisión considerando que  $\lambda_1 > \lambda_0$ . Comente.

- (d) Encuentre el test óptimo para el tamaño  $\alpha = 0.01$ . Considere  $\lambda_0 = 2$  y  $\lambda_1 = 4$ . *Indicación:* Notar que un test aleatorio podría ser necesario.
- (e) Encuentre los valores de tamaño  $\alpha$  sobre los cuales los test determinísticos son óptimos o en su defecto la condición que se debe cumplir para ello.

### Solución:

#### Resolución 1.1

Recordemos que la función generadora de momentos es una herramienta fundamental en probabilidad que permite caracterizar completamente una distribución. Se define como:

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} P_X(X = k) \quad (13)$$

donde  $t$  es un parámetro real. Para variables continuas, la suma se reemplaza por una integral:

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \quad (14)$$

Aplicando esta definición a la distribución de Poisson y sustituyendo la función de probabilidad de Poisson:

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \cdot \frac{\lambda^k e^{-\lambda}}{k!} \quad (15)$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{e^{tk} \lambda^k}{k!} \quad (16)$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} \quad (17)$$

Reconocemos que la suma es la expansión en serie de Taylor de la función exponencial:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (18)$$

Por lo tanto, con  $x = e^t \lambda$ :

$$M_X(t) = e^{-\lambda} \cdot e^{e^t \lambda} \quad (19)$$

$$= e^{-\lambda + e^t \lambda} \quad (20)$$

$$= e^{\lambda(e^t - 1)} \quad (21)$$

Con lo que se demuestra que la función generadora de momentos de una variable Poisson con parámetro  $\lambda$  es  $M_X(t) = e^{\lambda(e^t - 1)}$ .

## Resolución 1.2

Queremos demostrar que si  $X_1, \dots, X_n$  son variables aleatorias i.i.d. con distribución  $\text{Poisson}(\lambda)$ , entonces  $X = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$ .

La función generadora de momentos de la suma de variables independientes es el producto de sus funciones generadoras de momentos individuales:

$$M_X(t) = M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n M_{X_i}(t) \quad (22)$$

Como cada  $X_i \sim \text{Poisson}(\lambda)$ , sabemos por la parte anterior que  $M_{X_i}(t) = e^{\lambda(e^t - 1)}$ . Por lo tanto:

$$M_X(t) = \prod_{i=1}^n e^{\lambda(e^t - 1)} \quad (23)$$

$$= \left[ e^{\lambda(e^t - 1)} \right]^n \quad (24)$$

$$= e^{n\lambda(e^t - 1)} \quad (25)$$

Esta es exactamente la función generadora de momentos de una variable Poisson con parámetro  $n\lambda$ . Como la función generadora de momentos caracteriza únicamente la distribución, concluimos que:

$$X = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda) \quad (26)$$

Con lo que podemos concluir que la suma de  $n$  variables Poisson independientes con el mismo parámetro  $\lambda$  es una variable Poisson con parámetro  $n\lambda$ .

### Resolución 1.3

Consideremos un problema de detección binario donde debemos decidir entre dos hipótesis:

$$\mathcal{H}_0 : \theta = 0 \Rightarrow X \sim \text{Poisson}(\lambda_0)$$

$$\mathcal{H}_1 : \theta = 1 \Rightarrow X \sim \text{Poisson}(\lambda_1)$$

con  $\lambda_1 > \lambda_0$ .

Para encontrar el test óptimo, aplicamos el Lema de Neyman-Pearson, que nos indica que debemos basar nuestra decisión en el cociente de verosimilitud:

$$L(x) = \frac{P_1(X=x)}{P_0(X=x)} = \frac{\frac{\lambda_1^x e^{-\lambda_1}}{x!}}{\frac{\lambda_0^x e^{-\lambda_0}}{x!}} \quad (27)$$

Simplificando esta expresión, los factoriales se cancelan y obtenemos:

$$L(x) = \frac{\lambda_1^x e^{-\lambda_1}}{\lambda_0^x e^{-\lambda_0}} \quad (28)$$

$$= \left(\frac{\lambda_1}{\lambda_0}\right)^x \cdot e^{-(\lambda_1 - \lambda_0)} \quad (29)$$

Según el Lema de Neyman-Pearson, decidimos  $\mathcal{H}_1$  cuando  $L(x) > \eta$ . Esto nos lleva a la condición:

$$\left(\frac{\lambda_1}{\lambda_0}\right)^x \cdot e^{-(\lambda_1 - \lambda_0)} > \eta \quad (30)$$

Para simplificar esta desigualdad, tomamos logaritmo natural en ambos lados (recordando que el logaritmo es una función monótona creciente que preserva la desigualdad):

$$x \ln \left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0) > \ln(\eta) \quad (31)$$

$$x \ln \left(\frac{\lambda_1}{\lambda_0}\right) > \ln(\eta) + (\lambda_1 - \lambda_0) \quad (32)$$

Ahora, dado que  $\lambda_1 > \lambda_0$ , tenemos que  $\frac{\lambda_1}{\lambda_0} > 1$ , lo que implica  $\ln \left(\frac{\lambda_1}{\lambda_0}\right) > 0$ . Por lo tanto, podemos dividir ambos lados por este logaritmo sin invertir la desigualdad:

$$x > \frac{\ln(\eta) + (\lambda_1 - \lambda_0)}{\ln \left(\frac{\lambda_1}{\lambda_0}\right)} = \gamma \quad (33)$$

Con esto, la forma general del test óptimo queda definida como:

$$\pi(x) = \begin{cases} 1 & \text{si } x > \gamma \quad (\text{decidir } \mathcal{H}_1) \\ \gamma_0 & \text{si } x = \gamma \quad (\text{aleatorizar}) \\ 0 & \text{si } x < \gamma \quad (\text{decidir } \mathcal{H}_0) \end{cases} \quad (34)$$

Los valores de  $\gamma$  y  $\gamma_0$  se eligen de manera que la probabilidad de falsa alarma sea exactamente igual a  $\alpha$ :

$$\mathbb{P}_{H_0}(\pi(X) = 1) = \alpha \quad (35)$$

donde  $\gamma_0$  representa el valor esperado (o probabilidad) de decidir  $\mathcal{H}_1$  cuando  $X = \gamma$ , es decir,  $\gamma_0 = \mathbb{E}[\pi(X)|X = \gamma]$ . Notemos que aunque  $\gamma_0$  toma valores en el intervalo  $[0, 1]$ , en el contexto de detección binaria donde la decisión es 0 o 1, este valor representa la probabilidad de aleatorizar, permitiendo alcanzar cualquier tamaño  $\alpha \in [0, 1]$  (intervalo cerrado).

El test óptimo tiene una estructura muy simple para distribuciones Poisson, siendo esencialmente un test de umbral sobre el número de eventos observados. La componente aleatoria (dada por  $\gamma_0$ ) solo es necesaria cuando el umbral  $\gamma$  cae exactamente en un valor entero y necesitamos ajustar con precisión la probabilidad de falsa alarma a exactamente  $\alpha$ .

#### Resolución 1.4

Ahora debemos encontrar el test óptimo específico para  $\alpha = 0.01$ , con  $\lambda_0 = 2$  y  $\lambda_1 = 4$ . Recordemos de la parte anterior que el test tiene la forma:

$$\pi(x) = \begin{cases} 1 & \text{si } x > \gamma \\ \gamma_0 & \text{si } x = \gamma \\ 0 & \text{si } x < \gamma \end{cases} \quad (36)$$

donde debemos determinar  $\gamma$  y  $\gamma_0$  tal que la probabilidad de falsa alarma bajo  $\mathcal{H}_0$  sea exactamente  $\alpha = 0.01$ :

$$\mathbb{P}(\pi(X) = 1|\theta = 0) = \mathbb{P}(X > \gamma|\theta = 0) + \gamma_0 \cdot \mathbb{P}(X = \gamma|\theta = 0) = 0.01 \quad (37)$$

La probabilidad de falsa alarma es la probabilidad de rechazar  $\mathcal{H}_0$  (decidir 1) cuando  $\mathcal{H}_0$  es verdadera. Dado que el test  $\pi(X)$  toma diferentes valores según la observación  $X$ , debemos calcular la esperanza de  $\pi(X)$  bajo  $\mathcal{H}_0$ :

$$\alpha = \mathbb{P}(\pi(X) = 1|\theta = 0) = \mathbb{E}[\pi(X)|\theta = 0] \quad (38)$$

$$= \sum_{x=0}^{\infty} \pi(x) \cdot P(X = x|\theta = 0) \quad (39)$$

Ahora, sustituyendo la definición del test aleatorizado, separamos la suma en tres regiones:

$$\mathbb{E}[\pi(X)|\theta = 0] = \sum_{x < \gamma} \pi(x) \cdot P(X = x|\theta = 0) + \pi(\gamma) \cdot P(X = \gamma|\theta = 0) + \sum_{x > \gamma} \pi(x) \cdot P(X = x|\theta = 0) \quad (40)$$

$$= \sum_{x < \gamma} 0 \cdot P(X = x|\theta = 0) + \gamma_0 \cdot P(X = \gamma|\theta = 0) + \sum_{x > \gamma} 1 \cdot P(X = x|\theta = 0) \quad (41)$$

$$= 0 + \gamma_0 \cdot \mathbb{P}(X = \gamma|\theta = 0) + \mathbb{P}(X > \gamma|\theta = 0) = \alpha \quad (42)$$

Por lo tanto, obtenemos la expresión que queremos igualar a  $\alpha = 0.01$ .

Bajo la hipótesis nula  $\mathcal{H}_0 : \theta = 0$ , tenemos  $X \sim \text{Poisson}(\lambda_0) = \text{Poisson}(2)$ , por lo que debemos calcular las probabilidades acumuladas. Calculemos las probabilidades para distintos valores de  $k$ :

$$P_0(X = k) = \frac{\lambda_0^k e^{-\lambda_0}}{k!} = \frac{2^k e^{-2}}{k!} \quad (43)$$

$$P_0(X = 0) = e^{-2} \approx 0.1353 \quad (44)$$

$$P_0(X = 1) = 2e^{-2} \approx 0.2707 \quad (45)$$

$$P_0(X = 2) = 2e^{-2} \approx 0.2707 \quad (46)$$

$$P_0(X = 3) = \frac{4e^{-2}}{3} \approx 0.1804 \quad (47)$$

$$P_0(X = 4) = \frac{2e^{-2}}{3} \approx 0.0902 \quad (48)$$

$$P_0(X = 5) = \frac{4e^{-2}}{15} \approx 0.0361 \quad (49)$$

$$P_0(X = 6) = \frac{4e^{-2}}{45} \approx 0.0120 \quad (50)$$

$$P_0(X = 7) = \frac{8e^{-2}}{315} \approx 0.0034 \quad (51)$$

Ahora calculemos las probabilidades acumuladas de cola (tail probabilities) bajo  $\mathcal{H}_0$ :

$$\mathbb{P}(X \geq 6|\theta = 0) = P(X = 6|\theta = 0) + P(X = 7|\theta = 0) + \dots \approx 0.0166 \quad (52)$$

$$\mathbb{P}(X \geq 7|\theta = 0) = P(X = 7|\theta = 0) + P(X = 8|\theta = 0) + \dots \approx 0.0045 \quad (53)$$

Observamos que  $\mathbb{P}(X \geq 7|\theta = 0) \approx 0.0045 < 0.01 < 0.0166 \approx \mathbb{P}(X \geq 6|\theta = 0)$ . Esto significa que un test determinístico puro no puede alcanzar exactamente  $\alpha = 0.01$ : si usamos  $\gamma = 6$ , tendríamos una probabilidad de falsa alarma de aproximadamente  $0.0166 > 0.01$ , mientras que si usamos  $\gamma = 7$ , tendríamos aproximadamente  $0.0045 < 0.01$ .

Por lo tanto, necesitamos un test aleatorizado con  $\gamma = 6$ . Debemos encontrar  $\gamma_0$  tal que:

$$\mathbb{P}(X > 6|\theta = 0) + \gamma_0 \cdot P(X = 6|\theta = 0) = 0.01 \quad (54)$$



Despejando  $\gamma_0$ :

$$\gamma_0 = \frac{0.01 - \mathbb{P}(X > 6|\theta = 0)}{P(X = 6|\theta = 0)} \quad (55)$$

$$= \frac{0.01 - \mathbb{P}(X \geq 7|\theta = 0)}{P(X = 6|\theta = 0)} \quad (56)$$

$$= \frac{0.01 - 0.0045}{0.0120} \quad (57)$$

$$= \frac{0.0055}{0.0120} \quad (58)$$

$$\approx 0.458 \quad (59)$$

Por lo tanto, el test óptimo en su versión aleatorizada es:

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 7 \quad (\text{decidir } \mathcal{H}_1) \\ 0.458 & \text{si } x = 6 \quad (\text{valor esperado de la decisión}) \\ 0 & \text{si } x \leq 5 \quad (\text{decidir } \mathcal{H}_0) \end{cases} \quad (60)$$

En la práctica, para implementar este test en detección binaria (donde solo podemos decidir 0 o 1), cuando observamos exactamente 6 eventos, aleatorizamos: lanzamos una moneda sesgada que con probabilidad 0.458 nos hace rechazar  $\mathcal{H}_0$  (decidir 1) y con probabilidad 0.542 nos hace aceptarla (decidir 0). Si observamos 7 o más eventos, siempre rechazamos  $\mathcal{H}_0$ . Si observamos 5 o menos eventos, siempre aceptamos  $\mathcal{H}_0$ . Esta aleatorización es necesaria para alcanzar exactamente el nivel de significancia deseado de  $\alpha = 0.01$ .

### Resolución 1.5

Para determinar cuándo un test determinístico (sin aleatorización) es óptimo, debemos analizar la naturaleza discreta de la distribución Poisson y cómo esto afecta la construcción del test de Neyman-Pearson.

Un test determinístico es óptimo cuando el tamaño  $\alpha$  deseado coincide exactamente con una probabilidad acumulada de cola de la distribución bajo la hipótesis nula  $\mathcal{H}_0$ . Matemáticamente, esto ocurre cuando existe un entero  $k^*$  tal que:

$$\alpha = \mathbb{P}(X \geq k^*|\theta = 0) = \sum_{x=k^*}^{\infty} P(X = x|\theta = 0) \quad (61)$$

En este caso, el test óptimo determinístico tiene la forma simple:

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq k^* \quad (\text{rechazar } \mathcal{H}_0) \\ 0 & \text{si } x < k^* \quad (\text{aceptar } \mathcal{H}_0) \end{cases} \quad (62)$$

y cumple exactamente  $\mathbb{P}(\pi(X) = 1|\theta = 0) = \alpha$  sin necesidad de aleatorización.

Con  $\lambda_0 = 2$  y  $\lambda_1 = 4$ , utilizando los cálculos de la parte anterior, las probabilidades de cola bajo  $\mathcal{H}_0$  son:

$$\mathbb{P}(X \geq 4|\theta = 0) \approx 0.1429$$

$$\mathbb{P}(X \geq 5|\theta = 0) \approx 0.0527$$

$$\mathbb{P}(X \geq 6|\theta = 0) \approx 0.0166$$

$$\mathbb{P}(X \geq 7|\theta = 0) \approx 0.0045$$

$$\mathbb{P}(X \geq 8|\theta = 0) \approx 0.0011$$

Por lo tanto, el test determinístico es óptimo para los siguientes valores de  $\alpha$ :

$$\alpha \in \{0.1429, 0.0527, 0.0166, 0.0045, 0.0011, \dots\} \quad (63)$$

Específicamente:

- Si  $\alpha = 0.0166$ , el test óptimo es determinístico con  $k^* = 6$ :

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 6 \\ 0 & \text{si } x < 6 \end{cases} \quad (64)$$

- Si  $\alpha = 0.0045$ , el test óptimo es determinístico con  $k^* = 7$ :

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 7 \\ 0 & \text{si } x < 7 \end{cases} \quad (65)$$

- Si  $\alpha = 0.01$  (como en la parte 1.4), que está entre 0.0045 y 0.0166, **no** es posible un test determinístico óptimo, y se requiere aleatorización en  $x = 6$ .

Luego la condición necesaria y suficiente para que el test determinístico sea óptimo es:

$$\alpha \in \{\mathbb{P}(X \geq k|\theta = 0) : k \in \mathbb{N}\} \quad (66)$$

Esta restricción es una consecuencia directa de la naturaleza discreta de la distribución Poisson. En distribuciones continuas, esta limitación no existe y siempre es posible encontrar un test determinístico para cualquier  $\alpha \in (0, 1)$ . Para distribuciones discretas, cuando  $\alpha$  no coincide con ninguna probabilidad de cola, la aleatorización es necesaria para alcanzar exactamente el nivel de significancia deseado, como se demostró en la parte 1.4. Sin embargo, con aleatorización es posible alcanzar cualquier tamaño  $\alpha \in [0, 1]$  (intervalo cerrado), incluyendo los extremos  $\alpha = 0$  y  $\alpha = 1$ .

2. Considere el problema clásico de comunicaciones digitales, de la detección de símbolos binarios contaminadas por ruido aditivo Gaussiano. En este caso  $\Theta = \{0, 1\}$  y la variable aleatoria de observación dado  $\theta \in \Theta$  está dada por:

$$X = S_\theta + N \quad (67)$$

con  $S_0 = \mu$  y  $S_1 = -\mu$ ,  $\mu > 0$  y  $N \sim \mathcal{N}(0, \sigma^2)$ .

Ahora asuma que se tienen múltiples mediciones (o en su defecto transmisiones sucesivas del mismo símbolo),

$$X_1, X_2, \dots, X_n \quad (68)$$

donde  $X_i = S_\theta + N_i$  ( $i = 1, \dots, n$ ), para lo cual  $N_1, \dots, N_n$  son variables aleatorias i.i.d. que siguen una  $\mathcal{N}(0, \sigma^2)$ . Ahora la regla de decisión enfrenta el vector aleatorio Gaussiano  $X_1^n = (X_1, \dots, X_n)$  con valores en  $\mathbb{R}^n$  y el espacio de decisión es  $\Theta = \{0, 1\}$ .

- (a) Condicionado a los valores de  $\theta \in \Theta$ , determine la distribución de  $X_1^n$  y sus parámetros.  
 (b) Analice la familia de test óptimos y verifique que  $\forall x_1^n \in \mathbb{R}^n$

$$\log(L(x_1^n)) = \frac{2}{\sigma^2} \bar{\mu}^T \cdot x_1^n \quad (69)$$

donde  $\bar{\mu} = (\mu, \mu, \dots, \mu) \in \mathbb{R}^n$ . Específicamente para  $n = 2$  y  $\eta = 0$ , determine gráficamente las zonas de decisión, es decir:

$$A_0 = \pi_\eta^{-1}(\{0\}) = \left\{ x_1^2 \in \mathbb{R}^2 : \ln(L(x_1^2)) \leq \eta \right\}, \quad (70)$$

$$A_1 = \pi_\eta^{-1}(\{1\}) = \left\{ x_1^2 \in \mathbb{R}^2 : \ln(L(x_1^2)) > \eta \right\}. \quad (71)$$

- (c) Considere  $\mu = 1$ ,  $\sigma^2 = 10$  y  $n = 1, 10, 10^2, 10^3$ , respectivamente. Para estos distintos escenarios determine el test óptimo  $\pi_\eta^* : \mathbb{R}^n \rightarrow \{0, 1\}$ , tal que:

$$\alpha_\eta^* = \mathbb{E}(\pi_\eta^*(X_1^n) | \theta = 0) = 0.01 \quad (72)$$

y con ello grafique  $\beta_\eta^* = \mathbb{E}(\pi_\eta^*(X_1^n) | \theta = 1)$  como función de  $n$ . Comente que observa en el poder del test y cuál es la influencia en el número de mediciones.

- (d) Complemente el análisis anterior generando la curva ROC completa para los escenarios  $n = 1, 10, 10^2, 10^3$ . Comente en particular qué sucede cuando observado en el punto anterior.

### Solución:

#### Resolución 2.1

Necesitamos determinar la distribución conjunta del vector aleatorio  $X_1^n = (X_1, X_2, \dots, X_n)$  condicionado a cada valor de  $\theta \in \Theta = \{0, 1\}$ .

Recordemos que cada observación tiene la forma:

$$X_i = S_\theta + N_i, \quad i = 1, 2, \dots, n \quad (73)$$

donde:

- $S_\theta$  es la señal transmitida que depende del símbolo  $\theta$
- $N_i \sim \mathcal{N}(0, \sigma^2)$  son ruidos Gaussianos i.i.d.

Del enunciado del problema, sabemos que se trata de señalización antipodal es decir que:

$$\theta = 0 \Rightarrow S_0 = \mu \quad (74)$$

$$\theta = 1 \Rightarrow S_1 = -\mu \quad (75)$$

Luego debemos analizar por caso, por lo tanto cuando  $\theta = 0$ , tenemos  $S_0 = \mu$ , por lo tanto:

$$X_i = \mu + N_i \sim \mathcal{N}(\mu, \sigma^2) \quad (76)$$

Como los  $N_i$  son independientes e idénticamente distribuidos, las variables  $X_1, X_2, \dots, X_n$  son también independientes. Notemos que aunque los ruidos  $N_i$  son i.i.d., cada observación  $X_i$  tiene la misma señal determinística  $\mu$  sumada al ruido. Esto significa que:

- Cada  $X_i$  tiene la misma distribución marginal:  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  (idénticamente distribuidas)
- Las observaciones  $X_1, \dots, X_n$  son independientes entre sí (debido a la independencia de los ruidos  $N_i$ )

Por lo tanto, el vector aleatorio  $X_1^n = (X_1, X_2, \dots, X_n)$  sigue una distribución normal multivariada:

$$X_1^n | \theta = 0 \sim \mathcal{N}(\bar{\mu}, \sigma^2 I_n) \quad (77)$$

donde:

- $\bar{\mu} = (\mu, \mu, \dots, \mu)^T \in \mathbb{R}^n$  es el vector de medias. Cada componente es  $\mu$  porque cada observación  $X_i$  tiene esperanza  $\mathbb{E}[X_i] = \mathbb{E}[\mu + N_i] = \mu + \mathbb{E}[N_i] = \mu$ . Como todas las observaciones tienen la misma media, obtenemos un vector con componentes iguales.
- $\sigma^2 I_n$  es la matriz de covarianza, con  $I_n$  siendo la matriz identidad de dimensión  $n \times n$ . La forma diagonal refleja la independencia de las observaciones:  $\text{Cov}(X_i, X_j) = 0$  para  $i \neq j$ , y  $\text{Var}(X_i) = \sigma^2$  para todo  $i$ .

En otras palabras, trabajamos con un vector porque tenemos  $n$  observaciones simultáneas, y aunque cada una tiene la misma distribución marginal (i.i.d.), necesitamos considerar su distribución conjunta para el test óptimo. El vector de medias tiene todas sus componentes iguales a  $\mu$  precisamente porque cada medición observa la misma señal  $\mu$  contaminada por ruido independiente.

La función de densidad de probabilidad conjunta es:

$$f_0(x_1^n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \quad (78)$$

o en forma vectorial:

$$f_0(x_1^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n - \bar{\mu}\|^2}{2\sigma^2}\right) \quad (79)$$

Por otro lado cuando  $\theta = 1$ , tenemos  $S_1 = -\mu$ , por lo tanto:

$$X_i = -\mu + N_i \sim \mathcal{N}(-\mu, \sigma^2) \quad (80)$$

Nuevamente, por la independencia de los  $N_i$ , el vector aleatorio  $X_1^n$  sigue una distribución normal multivariada:

$$X_1^n | \theta = 1 \sim \mathcal{N}(-\bar{\mu}, \sigma^2 I_n) \quad (81)$$

donde:

- $-\bar{\mu} = (-\mu, -\mu, \dots, -\mu)^T \in \mathbb{R}^n$  es el vector de medias
- $\sigma^2 I_n$  es la matriz de covarianza (la misma que en el caso  $\theta = 0$ )

La función de densidad de probabilidad conjunta es:

$$f_1(x_1^n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i + \mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + \mu)^2\right) \quad (82)$$

o en forma vectorial:

$$f_1(x_1^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right) \quad (83)$$

Luego el vector aleatorio  $X_1^n$  tiene las siguientes distribuciones condicionadas bajo señalización antipodal:

$$\theta = 0 \Rightarrow X_1^n \sim \mathcal{N}(\bar{\mu}, \sigma^2 I_n) \quad (84)$$

$$\theta = 1 \Rightarrow X_1^n \sim \mathcal{N}(-\bar{\mu}, \sigma^2 I_n) \quad (85)$$

donde  $\bar{\mu} = (\mu, \mu, \dots, \mu)^T$  y las matrices de covarianza son idénticas en ambos casos debido a que el ruido tiene la misma varianza independientemente del símbolo transmitido.

## Resolución 2.2

Para aplicar el Lema de Neyman-Pearson, necesitamos calcular el cociente de verosimilitud el cual viene dado por:

$$L(x_1^n) = \frac{f_1(x_1^n)}{f_0(x_1^n)} \quad (86)$$

Sustituyendo las densidades obtenidas en la parte anterior (señalización antipodal):

$$L(x_1^n) = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right)}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n - \bar{\mu}\|^2}{2\sigma^2}\right)} \quad (87)$$

$$= \frac{\exp\left(-\frac{\|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right)}{\exp\left(-\frac{\|x_1^n - \bar{\mu}\|^2}{2\sigma^2}\right)} \quad (88)$$

$$= \exp\left(\frac{\|x_1^n - \bar{\mu}\|^2 - \|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right) \quad (89)$$

Desarrollemos el término en el exponente. Tenemos:

$$\|x_1^n - \bar{\mu}\|^2 = \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \quad (90)$$

$$\|x_1^n + \bar{\mu}\|^2 = \sum_{i=1}^n (x_i + \mu)^2 = \sum_{i=1}^n x_i^2 + 2\mu \sum_{i=1}^n x_i + n\mu^2 \quad (91)$$

Por lo tanto:

$$\|x_1^n - \bar{\mu}\|^2 - \|x_1^n + \bar{\mu}\|^2 = \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right) - \left(\sum_{i=1}^n x_i^2 + 2\mu \sum_{i=1}^n x_i + n\mu^2\right) \quad (92)$$

$$= -4\mu \sum_{i=1}^n x_i \quad (93)$$

Ahora, notemos que el vector  $\bar{\mu} = (\mu, \mu, \dots, \mu)^T$  tiene todas sus componentes iguales a  $\mu$ , por lo que el producto punto  $\bar{\mu}^T \cdot x_1^n$  se puede expresar como:

$$\bar{\mu}^T \cdot x_1^n = \sum_{i=1}^n \mu \cdot x_i = \mu \sum_{i=1}^n x_i \quad (94)$$

Sustituyendo esto en la expresión anterior:

$$\|x_1^n - \bar{\mu}\|^2 - \|x_1^n + \bar{\mu}\|^2 = -4\mu \sum_{i=1}^n x_i = -4\bar{\mu}^T \cdot x_1^n \quad (95)$$

Sustituyendo en la expresión del cociente de verosimilitud:

$$L(x_1^n) = \exp\left(\frac{-4\bar{\mu}^T \cdot x_1^n}{2\sigma^2}\right) = \exp\left(\frac{-2\bar{\mu}^T \cdot x_1^n}{\sigma^2}\right) \quad (96)$$

Tomando logaritmo natural del cociente de verosimilitud  $L(x_1^n)$ :

$$\boxed{\log(L(x_1^n)) = \frac{-2}{\sigma^2} \bar{\mu}^T \cdot x_1^n} \quad (97)$$

donde  $L(x_1^n)$  es el cociente de verosimilitud (likelihood ratio en inglés).

Con lo cual se cumple lo pedido. Ahora determinamos las zonas de decisión para  $n = 2$  y  $\eta = 0$ . Para  $n = 2$ , tenemos  $x_1^2 = (x_1, x_2)$  y  $\bar{\mu} = (\mu, \mu)$ . El logaritmo del cociente de verosimilitud es:

$$\log(L(x_1, x_2)) = \frac{-2}{\sigma^2} \mu(x_1 + x_2) \quad (98)$$

Con  $\eta = 0$ , la regla de decisión es:

$$\log(L(x_1, x_2)) > 0 \quad \Rightarrow \quad \frac{-2}{\sigma^2} \mu(x_1 + x_2) > 0 \quad (99)$$

Dado que  $\mu > 0$  y  $\sigma^2 > 0$ , esto se simplifica a:

$$x_1 + x_2 < 0 \quad (100)$$

Por lo tanto, las zonas de decisión son:

$$A_0 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \geq 0\} \quad (\text{Decidir } \mathcal{H}_0 : S_0 = \mu) \quad (101)$$

$$A_1 = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 < 0\} \quad (\text{Decidir } \mathcal{H}_1 : S_1 = -\mu) \quad (102)$$

En este caso de señalización antipodal con distribuciones Gaussianas, la familia de test óptimos es completamente determinística (sin necesidad de aleatorización). Esto se debe a que la variable aleatoria inducida por el estadístico es absolutamente continua, lo que permite alcanzar cualquier tamaño  $\alpha \in [0, 1]$  mediante la elección apropiada del umbral  $\eta$ , sin requerir aleatorización como en el caso discreto (Poisson).

En el plano  $(x_1, x_2)$ , la frontera de decisión es la recta  $x_1 + x_2 = 0$  (la diagonal que pasa por el origen). Esta recta divide el plano en dos semiplanos:

- **Región  $A_0$  (azul):** Semiplano donde  $x_1 + x_2 \geq 0$  (primer y cuarto cuadrante más el eje positivo), decidimos  $\mathcal{H}_0$ , es decir, símbolo  $+\mu$
- **Región  $A_1$  (roja):** Semiplano donde  $x_1 + x_2 < 0$  (segundo y tercer cuadrante más el eje negativo), decidimos  $\mathcal{H}_1$ , es decir, símbolo  $-\mu$

La frontera de decisión es una recta con pendiente  $-1$  que pasa por el origen. Esto tiene perfecto sentido para señalización antipodal:

- Si las observaciones  $x_1$  y  $x_2$  son en promedio positivas ( $x_1 + x_2 > 0$ ), es más probable que el símbolo transmitido haya sido  $+\mu$  (hipótesis  $\mathcal{H}_0$ )
- Si las observaciones son en promedio negativas ( $x_1 + x_2 < 0$ ), es más probable que el símbolo transmitido haya sido  $-\mu$  (hipótesis  $\mathcal{H}_1$ )

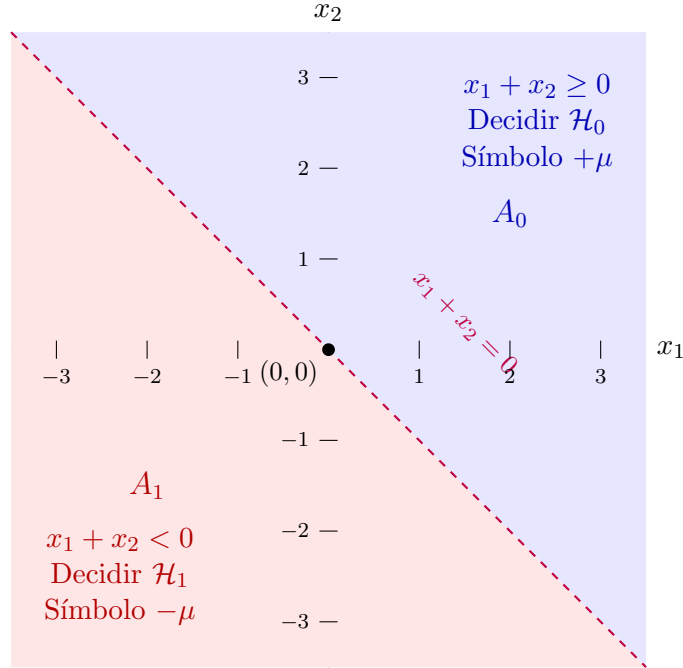


Figura 1: Regiones de decisión en el plano  $(x_1, x_2)$  para señalización antipodal con  $n = 2$  y  $\eta = 0$ . La frontera de decisión  $x_1 + x_2 = 0$  divide el plano en dos semiplanos: la región  $A_0$  (azul) donde se decide el símbolo  $+\mu$ , y la región  $A_1$  (roja) donde se decide el símbolo  $-\mu$ .

### Resolución 2.3

Para resolver esta parte, necesitamos determinar el test óptimo para cada valor de  $n$  tal que la probabilidad de falsa alarma sea  $\alpha = 0.01$ , y luego calcular la potencia del test (probabilidad de detección correcta)  $\beta_\eta^*$ .

En la resolución 2.2, encontramos que el logaritmo del cociente de verosimilitud es:

$$\log(L(x_1^n)) = \frac{-2}{\sigma^2} \bar{\mu}^T \cdot x_1^n = \frac{-2\mu}{\sigma^2} \sum_{i=1}^n x_i \quad (103)$$

Esto nos dice que la regla de decisión óptima solo depende de la suma  $\sum_{i=1}^n x_i$ , no de los valores individuales de cada  $x_i$ .

Para el caso particular de  $n = 2$  y  $\eta = 0$ , obtuvimos la regla:

$$\text{Decidir } \mathcal{H}_1 \text{ si } x_1 + x_2 < 0 \quad (104)$$

Esto se generaliza para cualquier  $n$  y cualquier umbral  $\eta$ . La regla de decisión de Neyman-Pearson es:

$$\log(L(x_1^n)) > \eta \quad \Leftrightarrow \quad \frac{-2\mu}{\sigma^2} \sum_{i=1}^n x_i > \eta \quad (105)$$



Multiplicando ambos lados por  $\frac{-\sigma^2}{2\mu}$  (y cambiando la dirección de la desigualdad porque multiplicamos por un número negativo):

$$\sum_{i=1}^n x_i < \frac{-\eta\sigma^2}{2\mu} \equiv \gamma \quad (106)$$

donde hemos definido  $\gamma = \frac{-\eta\sigma^2}{2\mu}$  como el umbral transformado. Por lo tanto, la regla de decisión óptima es:

$$\boxed{\text{Decidir } \mathcal{H}_1 \text{ si } \sum_{i=1}^n X_i < \gamma} \quad (107)$$

La notación  $X_1^n$  es una forma compacta de escribir la secuencia de observaciones  $(X_1, X_2, \dots, X_n)$ . Es decir:

$$X_1^n \equiv (X_1, X_2, X_3, \dots, X_n) \quad (108)$$

Cada  $X_i$  representa una medición individual. Por ejemplo, si  $n = 3$ , entonces  $X_1^3 = (X_1, X_2, X_3)$ .

El estadístico  $T(X_1^n)$  es simplemente la **suma de todas las observaciones**:

$$T(X_1^n) = \sum_{i=1}^n X_i = X_1 + X_2 + X_3 + \dots + X_n \quad (109)$$

Este estadístico es suficiente porque resume toda la información relevante de las  $n$  mediciones para tomar la decisión óptima. Bajo la hipótesis nula ( $\theta = 0$ , símbolo  $+\mu$ ), cada observación es  $X_i = \mu + N_i$ , donde  $N_i \sim \mathcal{N}(0, \sigma^2)$ . Entonces:

$$T(X_1^n)|\theta = 0 = \sum_{i=1}^n X_i = \sum_{i=1}^n (\mu + N_i) = n\mu + \sum_{i=1}^n N_i \quad (110)$$

Como la suma de variables normales independientes es también normal, tenemos:

$$T(X_1^n)|\theta = 0 \sim \mathcal{N}(n\mu, n\sigma^2) \quad (111)$$

- *Media:*  $n\mu$  (porque sumamos  $n$  términos iguales a  $\mu$ , más ruidos que promedian a cero)
- *Varianza:*  $n\sigma^2$  (porque sumamos  $n$  ruidos independientes, cada uno con varianza  $\sigma^2$ )

Análogamente, bajo la hipótesis alternativa ( $\theta = 1$ , símbolo  $-\mu$ ), cada observación es  $X_i = -\mu + N_i$ :

$$T(X_1^n)|\theta = 1 = \sum_{i=1}^n (-\mu + N_i) = -n\mu + \sum_{i=1}^n N_i \sim \mathcal{N}(-n\mu, n\sigma^2) \quad (112)$$

Para analizar con  $\alpha = 0.01$ , necesitamos encontrar el umbral  $\gamma$  tal que:

$$\mathbb{P}_{H_0}(T(X_1^n) < \gamma) = \alpha = 0.01 \quad (113)$$

Sabemos que bajo  $\mathcal{H}_0$ , la suma  $T(X_1^n) \sim \mathcal{N}(n\mu, n\sigma^2)$ . El problema es que trabajar con distintas medias y varianzas es complicado. La **estandarización** es un truco matemático que convierte cualquier variable normal en una *normal estándar*  $\mathcal{N}(0, 1)$  (media 0 y varianza 1), que está tabulada.

Si  $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ , entonces la variable estandarizada:

$$Z = \frac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}(0, 1) \quad (114)$$

En nuestro caso,  $Y = T(X_1^n)$  con media  $\mu_Y = n\mu$  y desviación estándar  $\sigma_Y = \sqrt{n\sigma^2} = \sigma\sqrt{n}$ . Entonces:

$$Z = \frac{T(X_1^n) - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0, 1) \quad (115)$$

Aplicando esta transformación a la desigualdad  $T(X_1^n) < \gamma$ :

$$\mathbb{P}_{H_0}(T(X_1^n) < \gamma) = \mathbb{P}_{H_0}\left(\frac{T(X_1^n) - n\mu}{\sigma\sqrt{n}} < \frac{\gamma - n\mu}{\sigma\sqrt{n}}\right) \quad (116)$$

$$= \mathbb{P}(Z < \frac{\gamma - n\mu}{\sigma\sqrt{n}}) = \alpha \quad (117)$$

donde  $Z \sim \mathcal{N}(0, 1)$  es una normal estándar.

Sabemos que  $\mathbb{P}(Z < -z_\alpha) = \alpha$  donde  $z_\alpha$  es el cuantil de la normal estándar. Para  $\alpha = 0.01$ , tenemos  $z_{0.01} \approx 2.326$  (este valor se obtiene de tablas o software estadístico).

Entonces:

$$\frac{\gamma - n\mu}{\sigma\sqrt{n}} = -z_\alpha \Rightarrow \gamma = n\mu - z_\alpha\sigma\sqrt{n} \quad (118)$$

El umbral  $\gamma$  está  $z_\alpha = 2.326$  desviaciones estándar por debajo de la media bajo  $\mathcal{H}_0$ . Esto garantiza que la probabilidad de error tipo I (falsa alarma) sea exactamente 0.01.

La potencia  $\beta_\eta^*$  es la probabilidad de decidir correctamente  $\mathcal{H}_1$  cuando  $\mathcal{H}_1$  es verdadera:

$$\beta_\eta^* = \mathbb{P}_{H_1}(T(X_1^n) < \gamma) \quad (119)$$

Bajo  $\mathcal{H}_1$ , sabemos que  $T(X_1^n) \sim \mathcal{N}(-n\mu, n\sigma^2)$ . Estandarizamos nuevamente, pero ahora usando la media y varianza bajo  $\mathcal{H}_1$ :

$$\beta_\eta^* = \mathbb{P}_{H_1}(T(X_1^n) < \gamma) \quad (120)$$

$$= \mathbb{P}_{H_1}\left(\frac{T(X_1^n) - (-n\mu)}{\sigma\sqrt{n}} < \frac{\gamma - (-n\mu)}{\sigma\sqrt{n}}\right) \quad (121)$$

$$= \mathbb{P}_{H_1}\left(\frac{T(X_1^n) + n\mu}{\sigma\sqrt{n}} < \frac{\gamma + n\mu}{\sigma\sqrt{n}}\right) \quad (122)$$

$$= \mathbb{P}\left(Z < \frac{\gamma + n\mu}{\sigma\sqrt{n}}\right) \quad (123)$$

donde  $Z \sim \mathcal{N}(0, 1)$ . Ahora sustituimos  $\gamma = n\mu - z_\alpha \sigma \sqrt{n}$ :

$$\beta_\eta^* = \mathbb{P} \left( Z < \frac{(n\mu - z_\alpha \sigma \sqrt{n}) + n\mu}{\sigma \sqrt{n}} \right) \quad (124)$$

$$= \mathbb{P} \left( Z < \frac{2n\mu - z_\alpha \sigma \sqrt{n}}{\sigma \sqrt{n}} \right) \quad (125)$$

$$= \mathbb{P} \left( Z < \frac{2n\mu}{\sigma \sqrt{n}} - z_\alpha \right) \quad (126)$$

$$= \mathbb{P} \left( Z < \frac{2\mu \sqrt{n}}{\sigma} - z_\alpha \right) \quad (127)$$

donde  $Z \sim \mathcal{N}(0, 1)$  es una variable aleatoria normal estándar. Esta probabilidad se puede calcular usando tablas de la distribución normal o software estadístico.

Con  $\mu = 1$ ,  $\sigma^2 = 10$  (por lo tanto  $\sigma = \sqrt{10} \approx 3.162$ ), y  $z_{0.01} \approx 2.326$ :

$$n = 1 : \quad \beta^* = \mathbb{P} \left( Z < \frac{2 \cdot 1 \cdot 1}{3.162} - 2.326 \right) = \mathbb{P}(Z < -1.694) \approx 0.045 \quad (128)$$

$$n = 10 : \quad \beta^* = \mathbb{P} \left( Z < \frac{2 \cdot 1 \cdot \sqrt{10}}{3.162} - 2.326 \right) = \mathbb{P}(Z < -0.326) \approx 0.372 \quad (129)$$

$$n = 100 : \quad \beta^* = \mathbb{P} \left( Z < \frac{2 \cdot 1 \cdot 10}{3.162} - 2.326 \right) = \mathbb{P}(Z < 3.999) \approx 1.000 \quad (130)$$

$$n = 1000 : \quad \beta^* = \mathbb{P} \left( Z < \frac{2 \cdot 1 \cdot \sqrt{1000}}{3.162} - 2.326 \right) = \mathbb{P}(Z < 17.674) \approx 1.000 \quad (131)$$

Algunos aspectos importantes a destacar:

- **Efecto del número de mediciones:** La potencia del test  $\beta^*$  aumenta significativamente con el número de mediciones  $n$ . Con una sola medición ( $n = 1$ ), la potencia es apenas 4.5%. Con  $n = 10$  mediciones, la potencia aumenta a 37.2%. Con  $n = 100$  mediciones, la potencia es prácticamente 100%.
- **Ventaja de la señalización antipodal:** Comparado con señalización ON-OFF (donde un símbolo es cero), la señalización antipodal aprovecha mejor la energía disponible, ya que ambos símbolos están a distancia  $\mu$  del origen. La distancia efectiva entre símbolos es  $2\mu$ , lo que duplica la capacidad de discriminación.
- **Relación con  $\sqrt{n}$ :** La potencia mejora con  $\sqrt{n}$  debido a que la varianza del promedio decrece como  $1/n$ , mientras que la diferencia entre medias escala con  $n$ .

**Resolución 2.4**

Para generar la curva ROC completa, debemos analizar cómo varía la potencia del test  $\beta$  (probabilidad de detección) en función del tamaño del test  $\alpha$  (probabilidad de falsa alarma) para diferentes valores de  $n$ .

La curva ROC se construye graficando  $\beta(\alpha)$  versus  $\alpha$  para  $\alpha \in [0, 1]$ . Del análisis anterior, sabemos que:

Para un tamaño de test  $\alpha$  dado, el umbral es:

$$\gamma = n\mu - z_\alpha\sigma\sqrt{n} \quad (132)$$

donde  $z_\alpha$  es el cuantil de orden  $1 - \alpha$  de la distribución normal estándar tal que  $\mathbb{P}(Z \leq z_\alpha) = 1 - \alpha$  (equivalentemente,  $\mathbb{P}(Z < -z_\alpha) = \alpha$ ). La potencia del test correspondiente es:

$$\beta(\alpha) = \mathbb{P}_{H_1}(T(X_1^n) < \gamma) = \mathbb{P}\left(Z < \frac{2\mu\sqrt{n}}{\sigma} - z_\alpha\right) \quad (133)$$

De esta forma, obtenemos la curva ROC que relaciona  $\beta$  y  $\alpha$ :

$$\beta(\alpha) = \mathbb{P}\left(Z < \frac{2\mu\sqrt{n}}{\sigma} - z_{1-\alpha}\right) \quad (134)$$

donde  $z_{1-\alpha}$  es el cuantil de orden  $1 - \alpha$  de la distribución normal estándar  $Z \sim \mathcal{N}(0, 1)$  (es decir,  $\mathbb{P}(Z \leq z_{1-\alpha}) = 1 - \alpha$ ).

Equivalentemente:

$$\beta(\alpha) = \mathbb{P}\left(Z < z_\alpha + \frac{2\mu\sqrt{n}}{\sigma}\right) \quad (135)$$

donde  $z_\alpha$  es el cuantil de orden  $\alpha$  de la distribución normal estándar.

La curva ROC tiene las siguientes propiedades importantes:

- **Punto (0,0):** Corresponde a un test que nunca rechaza  $\mathcal{H}_0$  (decisión trivial: siempre aceptar  $\mathcal{H}_0$ ). Se tiene  $\alpha = 0$  y  $\beta = 0$ .
- **Punto (1,1):** Corresponde a un test que siempre rechaza  $\mathcal{H}_0$  (decisión trivial: siempre decidir  $\mathcal{H}_1$ ). Se tiene  $\alpha = 1$  y  $\beta = 1$ .
- **Diagonal  $\beta = \alpha$ :** Representa un test aleatorio sin información (equivalente a lanzar una moneda). Un buen test debe estar por encima de esta diagonal.
- **Punto de operación anterior ( $\alpha = 0.01$ ):** Este es el punto específico analizado en la parte 2.3, donde fijamos  $\alpha = 0.01$  y calculamos los valores de  $\beta$  correspondientes.

A continuación, se presentan las curvas ROC para los diferentes valores de  $n$  (1, 10, 100, 1000) con  $\mu = 1$  y  $\sigma^2 = 10$ . Cada curva se genera variando  $\alpha$  desde 0 hasta 1 y calculando el correspondiente  $\beta(\alpha)$ .

1.  $n = 1$ :

Con una sola medición en señalización antipodal, la curva ROC muestra una capacidad de discriminación limitada pero mejor que el caso unipolar. Para  $\alpha = 0.01$ , obtuvimos  $\beta \approx 0.045$ , lo que significa que la probabilidad de detectar correctamente el símbolo  $-\mu$  es del 4.5%.

El área bajo la curva (AUC) para este caso es aproximadamente 0.58, indicando rendimiento modesto pero mejor que el azar.

2.  $n = 10$ :

Con 10 mediciones, la curva ROC se aleja notablemente de la diagonal. Para  $\alpha = 0.01$ , tenemos  $\beta \approx 0.372$ , mejorando significativamente la capacidad de detección. El AUC aumenta considerablemente a aproximadamente 0.79.

3.  $n = 100$ :

Con 100 mediciones, la curva ROC se acerca mucho a la esquina superior izquierda (punto ideal (0,1)). Para  $\alpha = 0.01$ , tenemos  $\beta \approx 1.000$ , lo que significa que podemos detectar la señal con probabilidad casi perfecta manteniendo una tasa de falsa alarma muy baja.

El AUC es mayor a 0.99, indicando un excelente rendimiento.

4.  $n = 1000$ :

Con 1000 mediciones, la curva ROC es prácticamente una línea que va de (0,0) a (0,1) y luego a (1,1), siguiendo el eje vertical y luego el horizontal. Para  $\alpha = 0.01$ , tenemos  $\beta \approx 1.0$ , lo que significa detección perfecta.

El AUC es prácticamente 1, indicando rendimiento perfecto.

$\beta$  (Probabilidad de detección)

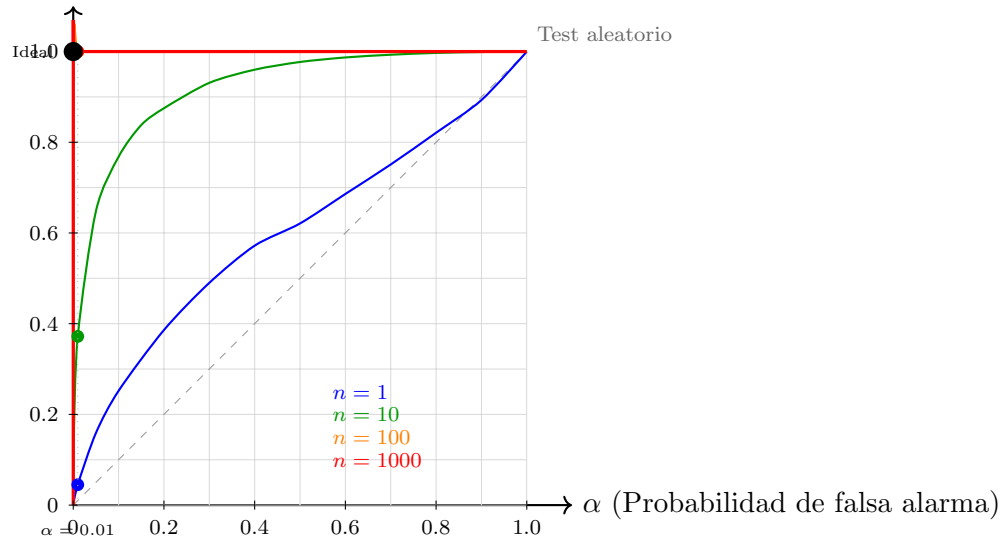


Figura 2: Curvas ROC para detección de señalización antipodal con diferentes números de mediciones  $n$ . A medida que  $n$  aumenta, las curvas se desplazan hacia el punto ideal (0,1), indicando mejor discriminación entre las hipótesis. La línea punteada vertical marca el punto de operación analizado en la parte 2.3 con  $\alpha = 0.01$ .