

1 Resumen

Variables Aleatorias

Una **variable aleatoria** X es una función que asigna un número real a cada resultado de un experimento aleatorio: $X : \Omega \rightarrow \mathbb{R}$. En la práctica, trabajamos con las distribuciones de estas variables.

Las variables continuas tienen una **función de densidad de probabilidad** $f(x)$ que describe qué tan “densa” es la probabilidad alrededor de cada punto. No da probabilidades directas, sino que las probabilidades se calculan integrando: $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$. Debe cumplir $f(x) \geq 0$ y $\int_{-\infty}^{\infty} f(x) dx = 1$.

Las variables discretas tienen una **función de probabilidad de masa** $p(x)$ que da directamente la probabilidad de cada valor específico. Debe cumplir $p(x) \geq 0$ y $\sum_x p(x) = 1$. Las probabilidades se calculan sumando: $\mathbb{P}(X \in A) = \sum_{x \in A} p(x)$.

Distribuciones Importantes

Distribución Uniforme $X \sim U(a, b)$: Todos los valores en $[a, b]$ son igualmente probables.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{en otro caso} \end{cases} \quad (1)$$

Distribución Exponencial $X \sim \text{Exp}(\lambda)$: Modela tiempos de espera entre eventos.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (2)$$

Distribución Normal $X \sim \mathcal{N}(\mu, \sigma^2)$: La distribución más importante. Curva en forma de campana.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3)$$

Propiedad clave: cualquier combinación lineal de variables normales independientes también es normal.

Distribución Poisson $X \sim \text{Poisson}(\lambda)$: Modela el número de eventos raros en un intervalo fijo.

$$p(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x \in \{0, 1, 2, \dots\} \\ 0 & \text{en otro caso} \end{cases} \quad (4)$$

Vectores Aleatorios

Un **vector aleatorio** $\mathbf{X} = (X_1, X_2, \dots, X_n)$ agrupa múltiples variables aleatorias. Su **función de densidad conjunta** describe el comportamiento conjunto:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) \quad (5)$$

Para variables **independientes e idénticamente distribuidas (i.i.d.)**:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i) \quad (6)$$

Momentos Estadísticos

Los **momentos** caracterizan las propiedades de una distribución:

Esperanza (valor promedio esperado):

$$\mathbb{E}[X] = \begin{cases} \int_{-\infty}^{\infty} x \cdot f(x) dx & \text{si } X \text{ es continua} \\ \sum_x x \cdot p(x) & \text{si } X \text{ es discreta} \end{cases} \quad (7)$$

Varianza (medida de dispersión):

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \quad (8)$$

Covarianza (dependencia lineal entre dos variables):

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (9)$$

Para la distribución normal $X \sim \mathcal{N}(\mu, \sigma^2)$: $\mu = \mathbb{E}[X]$ y $\sigma^2 = \text{Var}(X)$.

1. Considere una variable aleatoria X con distribución Poisson de parámetro λ .

$$P_X(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (10)$$

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

$$M_X(t) = \sum_{k \geq 0} P_X(X = k) \cdot e^{tk}, \quad (11)$$

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 λ . 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:

$$\theta = 0 \Rightarrow X \sim \text{Poisson}(\lambda_0), \quad (12)$$

$$\theta = 1 \Rightarrow X \sim \text{Poisson}(\lambda_1) \quad (13)$$

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 α sobre los cuales los test determinísticos son óptimos o en su defecto la condición que se debe cumplir para ello.

Solución:

Resolución 1.1

La función generadora de momentos se define como:

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} P_X(X = k) \quad (14)$$

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 λ 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 Poisson(λ), 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 λ es una variable Poisson con parámetro $n\lambda$.

Resolución 1.3

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

$$\begin{aligned}\mathcal{H}_0 : \theta = 0 &\Rightarrow X \sim \text{Poisson}(\lambda_0) \\ \mathcal{H}_1 : \theta = 1 &\Rightarrow X \sim \text{Poisson}(\lambda_1)\end{aligned}$$

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 γ y γ_0 se eligen de manera que la probabilidad de falsa alarma sea exactamente igual a α :

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

donde γ_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 γ_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 γ_0) solo es necesaria cuando el umbral γ cae exactamente en un valor entero y necesitamos ajustar con precisión la probabilidad de falsa alarma a exactamente α .

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 γ y γ_0 tal que la probabilidad de falsa alarma bajo \mathcal{H}_0 sea exactamente $\alpha = 0.01$:

$$\mathbb{P}_{\mathcal{H}_0}(\pi(X) = 1) = \mathbb{P}_{\mathcal{H}_0}(X > \gamma) + \gamma_0 \cdot \mathbb{P}_{\mathcal{H}_0}(X = \gamma) = 0.01 \quad (37)$$

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

$$P(X = k) = \frac{2^k e^{-2}}{k!} \quad (38)$$

$$P(X = 0) = e^{-2} \approx 0.1353 \quad (39)$$

$$P(X = 1) = 2e^{-2} \approx 0.2707 \quad (40)$$

$$P(X = 2) = 2e^{-2} \approx 0.2707 \quad (41)$$

$$P(X = 3) = \frac{4e^{-2}}{3} \approx 0.1804 \quad (42)$$

$$P(X = 4) = \frac{2e^{-2}}{3} \approx 0.0902 \quad (43)$$

$$P(X = 5) = \frac{4e^{-2}}{15} \approx 0.0361 \quad (44)$$

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

$$P(X = 7) = \frac{8e^{-2}}{315} \approx 0.0034 \quad (46)$$

Ahora calculemos las probabilidades acumuladas de cola (tail probabilities):

$$P(X \geq 6) = P(X = 6) + P(X = 7) + \dots \approx 0.0166 \quad (47)$$

$$P(X \geq 7) = P(X = 7) + P(X = 8) + \dots \approx 0.0045 \quad (48)$$

Observamos que $P(X \geq 7) \approx 0.0045 < 0.01 < 0.0166 \approx P(X \geq 6)$. 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 γ_0 tal que:

$$P(X > 6) + \gamma_0 \cdot P(X = 6) = 0.01 \quad (49)$$

Despejando γ_0 :

$$\gamma_0 = \frac{0.01 - P(X > 6)}{P(X = 6)} \quad (50)$$

$$= \frac{0.01 - P(X \geq 7)}{P(X = 6)} \quad (51)$$

$$= \frac{0.01 - 0.0045}{0.0120} \quad (52)$$

$$= \frac{0.0055}{0.0120} \quad (53)$$

$$\approx 0.458 \quad (54)$$

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

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 7 \quad (\text{decidir } \mathcal{H}_1) \\ 0.458 & \text{si } x = 6 \quad (\text{valor esperado de la decisión}) \\ 0 & \text{si } x \leq 5 \quad (\text{decidir } \mathcal{H}_0) \end{cases} \quad (55)$$

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 α 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:

$$\alpha = \mathbb{P}_{\mathcal{H}_0}(X \geq k^*) = \sum_{x=k^*}^{\infty} P_{\mathcal{H}_0}(X = x) \quad (56)$$

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

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq k^* \quad (\text{rechazar } \mathcal{H}_0) \\ 0 & \text{si } x < k^* \quad (\text{aceptar } \mathcal{H}_0) \end{cases} \quad (57)$$

y cumple exactamente $\mathbb{P}_{H_0}(\pi(X) = 1) = \alpha$ sin necesidad de aleatorización.

Con $\lambda_0 = 2$ y $\lambda_1 = 4$, utilizando los cálculos de la parte anterior, las probabilidades de cola bajo H_0 son:

$$\begin{aligned}\mathbb{P}_{H_0}(X \geq 4) &\approx 0.1429 \\ \mathbb{P}_{H_0}(X \geq 5) &\approx 0.0527 \\ \mathbb{P}_{H_0}(X \geq 6) &\approx 0.0166 \\ \mathbb{P}_{H_0}(X \geq 7) &\approx 0.0045 \\ \mathbb{P}_{H_0}(X \geq 8) &\approx 0.0011\end{aligned}$$

Por lo tanto, el test determinístico es óptimo para los siguientes valores de α :

$$\alpha \in \{0.1429, 0.0527, 0.0166, 0.0045, 0.0011, \dots\} \quad (58)$$

Específicamente:

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

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 6 \\ 0 & \text{si } x < 6 \end{cases} \quad (59)$$

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

$$\pi(x) = \begin{cases} 1 & \text{si } x \geq 7 \\ 0 & \text{si } x < 7 \end{cases} \quad (60)$$

- 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:

$$\boxed{\alpha \in \{\mathbb{P}_{H_0}(X \geq k) : k \in \mathbb{N}\}} \quad (61)$$

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 α 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:

$$X = S_\theta + N \quad (62)$$

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),

$$X_1, X_2, \dots, X_n \quad (63)$$

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 $\forall x_1^n \in \mathbb{R}^n$

$$\log(L(x_1^n)) = \frac{2}{\sigma^2} \bar{\mu}^T \cdot x_1^n \quad (64)$$

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:

$$A_0 = \pi_\eta^{-1}(\{0\}) = \left\{ x_1^2 \in \mathbb{R}^2 : \ln(L(x_1^2)) \leq \eta \right\}, \quad (65)$$

$$A_1 = \pi_\eta^{-1}(\{1\}) = \left\{ x_1^2 \in \mathbb{R}^2 : \ln(L(x_1^2)) > \eta \right\}. \quad (66)$$

- (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:

$$\alpha_\eta^* = \mathbb{E}(\pi_\eta^*(X_1^n) | \theta = 0) = 0.01 \quad (67)$$

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:

$$X_i = S_\theta + N_i, \quad i = 1, 2, \dots, n \quad (68)$$

donde:

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

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

$$\theta = 0 \Rightarrow S_0 = \mu \quad (69)$$

$$\theta = 1 \Rightarrow S_1 = -\mu \quad (70)$$

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

$$X_i = \mu + N_i \sim \mathcal{N}(\mu, \sigma^2) \quad (71)$$

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 μ 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:

$$X_1^n | \theta = 0 \sim \mathcal{N}(\bar{\mu}, \sigma^2 I_n) \quad (72)$$

donde:

- $\bar{\mu} = (\mu, \mu, \dots, \mu)^T \in \mathbb{R}^n$ es el vector de medias. Cada componente es μ 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 μ precisamente porque cada medición observa la misma señal μ contaminada por ruido independiente.

La función de densidad de probabilidad conjunta es:

$$f_0(x_1^n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \quad (73)$$

o en forma vectorial:

$$f_0(x_1^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n - \bar{\mu}\|^2}{2\sigma^2}\right) \quad (74)$$

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

$$X_i = -\mu + N_i \sim \mathcal{N}(-\mu, \sigma^2) \quad (75)$$

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

$$X_1^n | \theta = 1 \sim \mathcal{N}(-\bar{\mu}, \sigma^2 I_n) \quad (76)$$

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:

$$f_1(x_1^n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i + \mu)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i + \mu)^2\right) \quad (77)$$

o en forma vectorial:

$$f_1(x_1^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right) \quad (78)$$

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

$$\theta = 0 \Rightarrow X_1^n \sim \mathcal{N}\left(\bar{\mu}, \sigma^2 I_n\right) \quad (79)$$

$$\theta = 1 \Rightarrow X_1^n \sim \mathcal{N}\left(-\bar{\mu}, \sigma^2 I_n\right) \quad (80)$$

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 dada por:

$$L(x_1^n) = \frac{f_1(x_1^n)}{f_0(x_1^n)} \quad (81)$$

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

$$L(x_1^n) = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right)}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x_1^n - \bar{\mu}\|^2}{2\sigma^2}\right)} \quad (82)$$

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

$$= \exp\left(\frac{\|x_1^n - \bar{\mu}\|^2 - \|x_1^n + \bar{\mu}\|^2}{2\sigma^2}\right) \quad (84)$$

Desarrollemos el término en el exponente. Tenemos:

$$\|x_1^n - \bar{\mu}\|^2 = \sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \quad (85)$$

$$\|x_1^n + \bar{\mu}\|^2 = \sum_{i=1}^n (x_i + \mu)^2 = \sum_{i=1}^n x_i^2 + 2\mu \sum_{i=1}^n x_i + n\mu^2 \quad (86)$$

Por lo tanto:

$$\|x_1^n - \bar{\mu}\|^2 - \|x_1^n + \bar{\mu}\|^2 = \left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \right) - \left(\sum_{i=1}^n x_i^2 + 2\mu \sum_{i=1}^n x_i + n\mu^2 \right) \quad (87)$$

$$= -4\mu \sum_{i=1}^n x_i \quad (88)$$

$$= -4\mu \bar{\mu}^T \cdot x_1^n / \mu \quad (89)$$

$$= -4\bar{\mu}^T \cdot x_1^n \quad (90)$$

Sustituyendo en la expresión del cociente de verosimilitud:

$$L(x_1^n) = \exp \left(\frac{-4\bar{\mu}^T \cdot x_1^n}{2\sigma^2} \right) = \exp \left(\frac{-2\bar{\mu}^T \cdot x_1^n}{\sigma^2} \right) \quad (91)$$

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

$$\boxed{\log(L(x_1^n)) = \frac{-2}{\sigma^2} \bar{\mu}^T \cdot x_1^n} \quad (92)$$

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:

$$\log(L(x_1, x_2)) = \frac{-2}{\sigma^2} \mu(x_1 + x_2) \quad (93)$$

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

$$\log(L(x_1, x_2)) > 0 \Rightarrow \frac{-2}{\sigma^2} \mu(x_1 + x_2) > 0 \quad (94)$$

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

$$x_1 + x_2 < 0 \quad (95)$$

Por lo tanto, las zonas de decisión son:

$$A_0 = \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 \geq 0 \right\} \quad (\text{Decidir } \mathcal{H}_0 : S_0 = \mu) \quad (96)$$

$$A_1 = \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 < 0 \right\} \quad (\text{Decidir } \mathcal{H}_1 : S_1 = -\mu) \quad (97)$$

Observación importante: 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 η , 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 $+μ$ (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 $-μ$ (hipótesis \mathcal{H}_1)

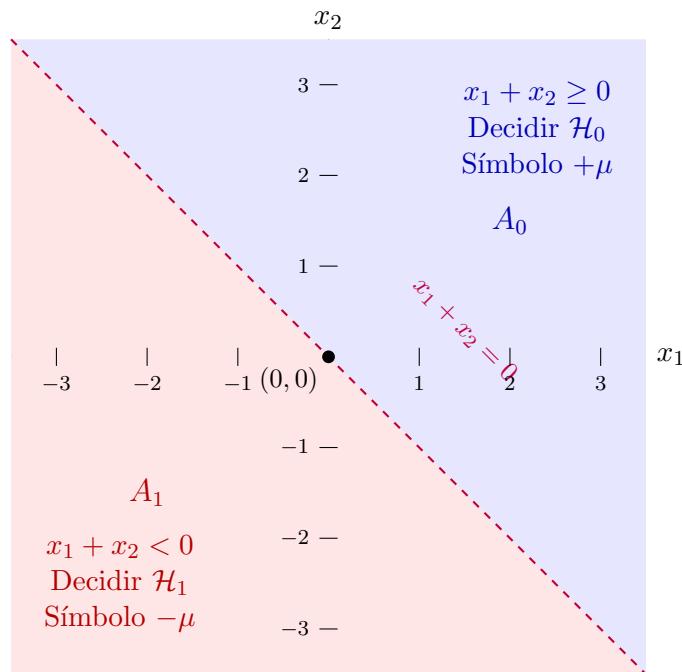


Figura 1: Regiones de decisión en el plano (x_1, x_2) para señalización antipodal con $n = 2$ y $\eta = 0$. La frontera de decisión $x_1 + x_2 = 0$ divide el plano en dos semiplanos: la región A_0 (azul) donde se decide el símbolo $+\mu$, y la región A_1 (roja) donde se decide el símbolo $-\mu$.

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) β_η^* .

Construcción del test óptimo con señalización antipodal:

Del análisis anterior, sabemos que un estadístico suficiente (función de las observaciones que resume toda la información relevante para la decisión) es:

$$T(X_1^n) = \sum_{i=1}^n X_i \quad (98)$$

Bajo la hipótesis nula ($\theta = 0$, símbolo $+\mu$):

$$T(X_1^n)|\theta = 0 = \sum_{i=1}^n (\mu + N_i) = n\mu + \sum_{i=1}^n N_i \sim \mathcal{N}(n\mu, n\sigma^2) \quad (99)$$

Bajo la hipótesis alternativa ($\theta = 1$, símbolo $-\mu$):

$$T(X_1^n)|\theta = 1 = \sum_{i=1}^n (-\mu + N_i) = -n\mu + \sum_{i=1}^n N_i \sim \mathcal{N}(-n\mu, n\sigma^2) \quad (100)$$

La regla de decisión óptima para señalización antipodal es:

$$\text{Decidir } \mathcal{H}_1 \text{ si } T(X_1^n) < \gamma \quad (101)$$

Es decir, decidimos el símbolo $-\mu$ cuando la suma es suficientemente negativa. Notemos que con umbral en cero ($\gamma = 0$), obtenemos una regla de decisión simétrica que es natural para señales antipodales.

Para analizar con $\alpha = 0.01$, necesitamos encontrar γ tal que:

$$\mathbb{P}_{H_0}(T(X_1^n) < \gamma) = \alpha = 0.01 \quad (102)$$

Estandarizando bajo \mathcal{H}_0 :

$$\mathbb{P}_{H_0}\left(\frac{T(X_1^n) - n\mu}{\sqrt{n\sigma^2}} < \frac{\gamma - n\mu}{\sqrt{n\sigma^2}}\right) = \alpha \quad (103)$$

Si $Z \sim \mathcal{N}(0, 1)$, entonces:

$$\mathbb{P}(Z < -z_\alpha) = \alpha \Rightarrow \frac{\gamma - n\mu}{\sigma\sqrt{n}} = -z_\alpha \quad (104)$$

Para $\alpha = 0.01$, tenemos $z_{0.01} \approx 2.326$. Por lo tanto:

$$\gamma = n\mu - z_\alpha\sigma\sqrt{n} \quad (105)$$

La probabilidad de detección correcta es:

$$\beta_\eta^* = \mathbb{P}_{H_1}(T(X_1^n) < \gamma) \quad (106)$$

$$= \mathbb{P}_{H_1} \left(\frac{T(X_1^n) + n\mu}{\sqrt{n\sigma^2}} < \frac{\gamma + n\mu}{\sqrt{n\sigma^2}} \right) \quad (107)$$

$$= \mathbb{P} \left(Z < \frac{n\mu - z_\alpha \sigma \sqrt{n} + n\mu}{\sigma \sqrt{n}} \right) \quad (108)$$

$$= \mathbb{P} \left(Z < \frac{2n\mu - z_\alpha \sigma \sqrt{n}}{\sigma \sqrt{n}} \right) \quad (109)$$

$$= \mathbb{P} \left(Z < \frac{2\mu \sqrt{n}}{\sigma} - z_\alpha \right) \quad (110)$$

$$= \Phi \left(\frac{2\mu \sqrt{n}}{\sigma} - z_\alpha \right) \quad (111)$$

donde en el último paso hemos introducido la notación $\Phi(\cdot)$ para la función de distribución acumulada (CDF) de la normal estándar:

$$\Phi(x) = \mathbb{P}(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (112)$$

Esta función Φ nos permite expresar probabilidades de la forma $\mathbb{P}(Z < x)$ de manera compacta. El cambio de notación de $\mathbb{P}(Z < x)$ a $\Phi(x)$ es simplemente una convención: ambas representan exactamente lo mismo, pero Φ es más conveniente cuando trabajamos con la distribución normal estándar repetidamente. Es importante notar que Φ es una función conocida y tabulada (o implementada en software estadístico), lo que nos permite calcular numéricamente la potencia del test para cualquier valor de n .

Cálculo numérico para los valores dados:

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

$$n = 1 : \quad \beta^* = \Phi \left(\frac{2 \cdot 1 \cdot 1}{3.162} - 2.326 \right) = \Phi(-1.694) \approx 0.045 \quad (113)$$

$$n = 10 : \quad \beta^* = \Phi \left(\frac{2 \cdot 1 \cdot \sqrt{10}}{3.162} - 2.326 \right) = \Phi(-0.326) \approx 0.372 \quad (114)$$

$$n = 100 : \quad \beta^* = \Phi \left(\frac{2 \cdot 1 \cdot 10}{3.162} - 2.326 \right) = \Phi(3.999) \approx 1.000 \quad (115)$$

$$n = 1000 : \quad \beta^* = \Phi \left(\frac{2 \cdot 1 \cdot \sqrt{1000}}{3.162} - 2.326 \right) = \Phi(17.674) \approx 1.000 \quad (116)$$

Algunos aspectos importantes a destacar:

- **Efecto del número de mediciones:** La potencia del test β^* 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 μ del origen. La distancia efectiva entre símbolos es 2μ , 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 β (probabilidad de detección) en función del tamaño del test α (probabilidad de falsa alarma) para diferentes valores de n .

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

Para un tamaño de test α dado, el umbral es:

$$\gamma = n\mu - z_\alpha \sigma \sqrt{n} \quad (117)$$

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

$$\beta(\alpha) = \mathbb{P}_{H_1}(T(X_1^n) < \gamma) = \Phi\left(\frac{2\mu\sqrt{n}}{\sigma} - z_\alpha\right) \quad (118)$$

Sustituyendo $z_\alpha = \Phi^{-1}(1 - \alpha)$:

$$\beta(\alpha) = \Phi\left(\frac{2\mu\sqrt{n}}{\sigma} - \Phi^{-1}(1 - \alpha)\right) \quad (119)$$

o equivalentemente:

$$\beta(\alpha) = \Phi\left(\Phi^{-1}(\alpha) + \frac{2\mu\sqrt{n}}{\sigma}\right) \quad (120)$$

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 β 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 α 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.

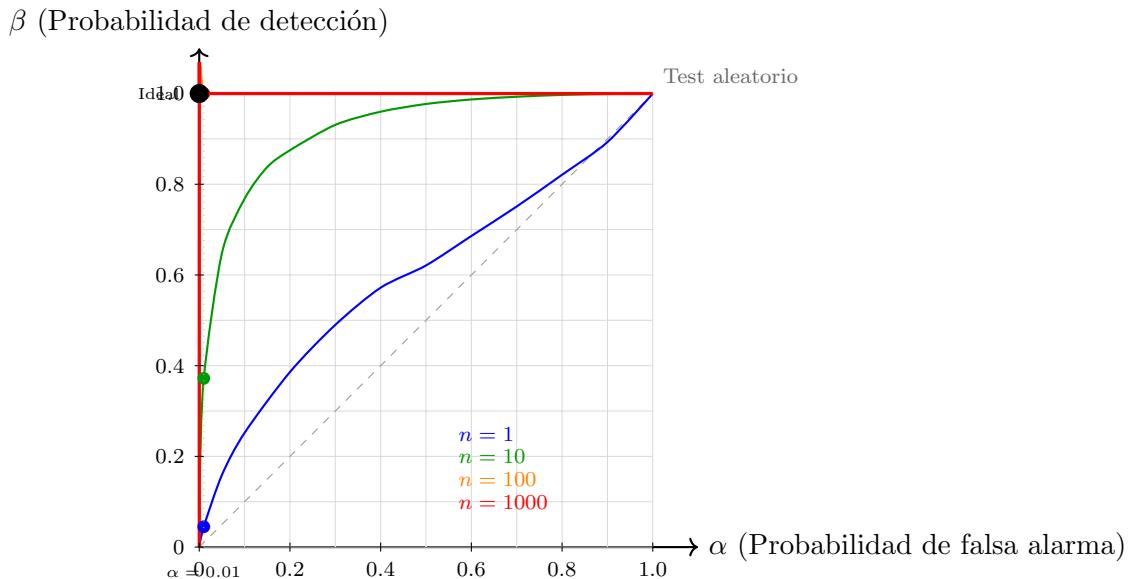


Figura 2: Curvas ROC para detección de señalización antipodal con diferentes números de mediciones n . A medida que n aumenta, las curvas se desplazan hacia el punto ideal (0,1), indicando mejor discriminación entre las hipótesis. La línea punteada vertical marca el punto de operación analizado en la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

Considere un problema de detección binario $\Theta = \{0, 1\}$ donde la variable aleatoria fuente de observación X toma valores en la recta real $X = \mathbb{R}$ y sigue las estadísticas como función del parámetro θ dadas por:

$$\theta = 0 : X \sim \text{Uniforme}[0, 1] \quad (121)$$

$$\theta = 1 : X \sim \text{Uniforme}[0, K] \quad (122)$$

con $K > 1$.

3. (a) Determine la familia de test óptimos en el sentido del Lema de Neyman-Pearson. Considere solamente las familias de test óptimas de tamaño entre $[0, 1]$.

- (b) Fije un umbral $\tau \in \mathbb{R}$ y considere el siguiente test determinístico:

$$\pi_\tau(x) = 1 \text{ si } \log \frac{f_X(x|\theta=1)}{f_X(x|\theta=0)} \geq \tau \quad (123)$$

y $\pi_\tau(x) = 0$ si la condición en (1.111) no se cumple. Determine las regiones de decisión de π_τ , es decir los conjuntos $A_0^c = \pi_\tau^{-1}(\{0\})$ y $A_1^c = \pi_\tau^{-1}(\{1\})$. Exprese como función además dichas regiones como función de τ . Identifique rangos concretos en el espacio de posibles valores de τ .

- (c) Del punto anterior, determine las expresiones para el poder y tamaño del test como función del valor de τ . Recordar que:

$$\alpha_{\pi_\tau} = P_X(\pi_\tau(X) = 1 | \theta = 0) \quad (124)$$

$$\beta_{\pi_\tau} = P_X(\pi_\tau(X) = 1 | \theta = 1) \quad (125)$$

- (d) Determine la curva ROC. Es posible obtener la curva ROC completa (para todos los tamaños) con test determinísticos? Justifique.

- (e) Vuelva al punto b) y d) y discuta que pasa con las regiones de decisión y la curva ROC si $K \rightarrow \infty$.

Solución:

Resolución 3.1

Para determinar la familia de test óptimos según el Lema de Neyman-Pearson, necesitamos primero especificar las funciones de densidad de probabilidad bajo cada hipótesis.

Bajo la hipótesis $\theta = 0$, tenemos $X \sim \text{Uniforme}[0, 1]$:

$$f_X(x|\theta=0) = \begin{cases} 1 & \text{si } x \in [0, 1] \\ 0 & \text{en otro caso} \end{cases} \quad (126)$$

Bajo la hipótesis $\theta = 1$, tenemos $X \sim \text{Uniforme}[0, K]$ con $K > 1$:

$$f_X(x|\theta=1) = \begin{cases} \frac{1}{K} & \text{si } x \in [0, K] \\ 0 & \text{en otro caso} \end{cases} \quad (127)$$

El Lema de Neyman-Pearson establece que el test óptimo se basa en el cociente de verosimilitud:

$$L(x) = \frac{f_X(x|\theta=1)}{f_X(x|\theta=0)} \quad (128)$$

Analicemos este cociente por regiones:

- **Para** $x \in [0, 1]$: Ambas densidades son distintas de cero, por lo tanto:

$$L(x) = \frac{1/K}{1} = \frac{1}{K} \quad (129)$$

- **Para** $x \in (1, K]$: Solo la densidad bajo $\theta = 1$ es distinta de cero, mientras que bajo $\theta = 0$ es cero. En este caso:

$$L(x) = \frac{1/K}{0} = +\infty \quad (130)$$

- **Para** $x \notin [0, K]$: Ambas densidades son cero, por lo que el cociente no está definido (o se puede considerar que no aporta información).

El test óptimo de Neyman-Pearson tiene la forma:

$$\pi_\eta(x) = \begin{cases} 1 & \text{si } L(x) > \eta \\ \gamma_0 & \text{si } L(x) = \eta \\ 0 & \text{si } L(x) < \eta \end{cases} \quad (131)$$

donde $\eta > 0$ es el umbral y γ_0 representa el valor esperado de la decisión cuando $L(x) = \eta$. Notemos que γ_0 toma valores en $[0, 1]$, y en el contexto de detección binaria, este valor se interpreta como la probabilidad de decidir \mathcal{H}_1 mediante aleatorización.

Dado que $K > 1$, tenemos que $\frac{1}{K} < 1$. Analicemos los diferentes casos según el valor de η :

1. **Si** $\eta < \frac{1}{K}$: Entonces $L(x) > \eta$ para todo $x \in [0, K]$. El test decide \mathcal{H}_1 en todo $[0, K]$:

$$\pi_\eta(x) = \begin{cases} 1 & \text{si } x \in [0, K] \\ 0 & \text{si } x \notin [0, K] \end{cases} \quad (132)$$

2. **Si** $\eta = \frac{1}{K}$: Entonces $L(x) = \eta$ para $x \in [0, 1]$ y $L(x) > \eta$ para $x \in (1, K]$. Se requiere aleatorización en $[0, 1]$:

$$\pi_\eta(x) = \begin{cases} 1 & \text{si } x \in (1, K] \\ \gamma_0 & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, K] \end{cases} \quad (133)$$

3. **Si** $\frac{1}{K} < \eta < +\infty$: Entonces $L(x) < \eta$ para $x \in [0, 1]$ y $L(x) > \eta$ para $x \in (1, K]$. El test decide:

$$\pi_\eta(x) = \begin{cases} 1 & \text{si } x \in (1, K] \\ 0 & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, K] \end{cases} \quad (134)$$

Por lo tanto, la familia de test óptimos de tamaño entre $[0, 1]$ se caracteriza por:

- **Región determinística**: Siempre se decide \mathcal{H}_1 cuando $x \in (1, K]$ (ya que allí $L(x) = +\infty$)
- **Región de decisión variable**: En el intervalo $[0, 1]$, la decisión depende del umbral η y posiblemente de aleatorización

- **Parámetro de control:** El tamaño del test se controla mediante η y γ_0 . Con aleatorización, es posible alcanzar cualquier tamaño $\alpha \in [0, 1]$ (intervalo cerrado).

La familia de test óptimos queda completamente caracterizada por el umbral $\eta \in (0, +\infty]$ y el valor esperado $\gamma_0 \in [0, 1]$ cuando $\eta = \frac{1}{K}$, donde γ_0 representa la probabilidad de decidir \mathcal{H}_1 al aleatorizar en el intervalo $[0, 1]$.

Resolución 3.2

Consideremos el test determinístico dado por:

$$\pi_\tau(x) = 1 \text{ si } \log L(x) \geq \tau \quad (135)$$

y $\pi_\tau(x) = 0$ en caso contrario.

Primero, calculemos el logaritmo del cociente de verosimilitud en cada región:

- **Para** $x \in [0, 1]$:

$$\log L(x) = \log \left(\frac{1}{K} \right) = -\log K \quad (136)$$

- **Para** $x \in (1, K]$:

$$\log L(x) = \log (+\infty) = +\infty \quad (137)$$

- **Para** $x \notin [0, K]$: El cociente no está definido o ambas densidades son cero.

La regla de decisión es: decidir \mathcal{H}_1 (es decir, $\pi_\tau(x) = 1$) si $\log L(x) \geq \tau$.

Analicemos las regiones de decisión según el valor de τ :

Caso 1: $\tau < -\log K$

En este caso:

- Para $x \in [0, 1]$: $\log L(x) = -\log K$ y como $-\log K \geq \tau$, se decide \mathcal{H}_1
- Para $x \in (1, K]$: $\log L(x) = +\infty > \tau$, se decide \mathcal{H}_1
- Para $x \notin [0, K]$: No hay información, se decide \mathcal{H}_0 por defecto

Las regiones de decisión son:

$$A_1^c = \pi_\tau^{-1}(\{1\}) = [0, K] \quad (138)$$

$$A_0^c = \pi_\tau^{-1}(\{0\}) = \mathbb{R} \setminus [0, K] \quad (139)$$

Caso 2: $\tau = -\log K$

En este caso:

- Para $x \in [0, 1]$: $\log L(x) = -\log K = \tau$, se decide \mathcal{H}_1 (por la condición \geq)
- Para $x \in (1, K]$: $\log L(x) = +\infty > \tau$, se decide \mathcal{H}_1
- Para $x \notin [0, K]$: Se decide \mathcal{H}_0

Las regiones de decisión son:

$$A_1^c = \pi_\tau^{-1}(\{1\}) = [0, K] \quad (140)$$

$$A_0^c = \pi_\tau^{-1}(\{0\}) = \mathbb{R} \setminus [0, K] \quad (141)$$

Caso 3: $-\log K < \tau < +\infty$

En este caso:

- Para $x \in [0, 1]$: $\log L(x) = -\log K < \tau$, se decide \mathcal{H}_0
- Para $x \in (1, K]$: $\log L(x) = +\infty > \tau$, se decide \mathcal{H}_1
- Para $x \notin [0, K]$: Se decide \mathcal{H}_0

Las regiones de decisión son:

$$A_1^c = \pi_\tau^{-1}(\{1\}) = (1, K] \quad (142)$$

$$A_0^c = \pi_\tau^{-1}(\{0\}) = [0, 1] \cup (\mathbb{R} \setminus [0, K]) \quad (143)$$

Caso 4: $\tau \geq +\infty$ (o τ muy grande)

En este caso, ninguna observación finita satisface $\log L(x) \geq \tau$, por lo que:

$$A_1^c = \pi_\tau^{-1}(\{1\}) = \emptyset \quad (144)$$

$$A_0^c = \pi_\tau^{-1}(\{0\}) = \mathbb{R} \quad (145)$$

En resumen, las regiones de decisión como función de τ son:

$$A_1^c = \begin{cases} [0, K] & \text{si } \tau \leq -\log K \\ (1, K] & \text{si } -\log K < \tau < +\infty \\ \emptyset & \text{si } \tau \rightarrow +\infty \end{cases} \quad (146)$$

$$A_0^c = \begin{cases} \mathbb{R} \setminus [0, K] & \text{si } \tau \leq -\log K \\ \mathbb{R} \setminus (1, K] & \text{si } -\log K < \tau < +\infty \\ \mathbb{R} & \text{si } \tau \rightarrow +\infty \end{cases} \quad (147)$$

Los rangos concretos de τ que definen comportamientos distintos son:

- $\tau \in (-\infty, -\log K]$: Test que decide \mathcal{H}_1 en todo $[0, K]$
- $\tau \in (-\log K, +\infty)$: Test que decide \mathcal{H}_1 solo en $(1, K]$

Resolución 3.3

Ahora determinaremos las expresiones para el tamaño (α_{π_τ}) y la potencia (β_{π_τ}) del test como función de τ .

Cálculo del tamaño del test:

El tamaño del test es la probabilidad de rechazar \mathcal{H}_0 cuando es verdadera:

$$\alpha_{\pi_\tau} = P_X(\pi_\tau(X) = 1 | \theta = 0) = P_X(X \in A_1^c | \theta = 0) \quad (148)$$

Bajo $\theta = 0$, tenemos $X \sim \text{Uniforme}[0, 1]$.

Caso 1: $\tau \leq -\log K$

En este caso, $A_1^c = [0, K]$. La probabilidad es:

$$\alpha_{\pi_\tau} = P_X(X \in [0, K] | \theta = 0) = P_X(X \in [0, 1] | \theta = 0) = 1 \quad (149)$$

ya que bajo $\theta = 0$, X solo toma valores en $[0, 1] \subset [0, K]$.

Caso 2: $-\log K < \tau < +\infty$

En este caso, $A_1^c = (1, K]$. La probabilidad es:

$$\alpha_{\pi_\tau} = P_X(X \in (1, K] | \theta = 0) = 0 \quad (150)$$

ya que bajo $\theta = 0$, $X \in [0, 1]$ y nunca puede estar en $(1, K]$.

Por lo tanto:

$$\boxed{\alpha_{\pi_\tau} = \begin{cases} 1 & \text{si } \tau \leq -\log K \\ 0 & \text{si } \tau > -\log K \end{cases}} \quad (151)$$

Cálculo de la potencia del test:

La potencia del test es la probabilidad de rechazar \mathcal{H}_0 cuando \mathcal{H}_1 es verdadera:

$$\beta_{\pi_\tau} = P_X(\pi_\tau(X) = 1 | \theta = 1) = P_X(X \in A_1^c | \theta = 1) \quad (152)$$

Bajo $\theta = 1$, tenemos $X \sim \text{Uniforme}[0, K]$.

Caso 1: $\tau \leq -\log K$

En este caso, $A_1^c = [0, K]$. La probabilidad es:

$$\beta_{\pi_\tau} = P_X(X \in [0, K] | \theta = 1) = 1 \quad (153)$$

Caso 2: $-\log K < \tau < +\infty$

En este caso, $A_1^c = (1, K]$. La probabilidad es:

$$\beta_{\pi_\tau} = P_X(X \in (1, K] | \theta = 1) \quad (154)$$

$$= \int_1^K \frac{1}{K} dx \quad (155)$$

$$= \frac{1}{K}(K - 1) \quad (156)$$

$$= \frac{K - 1}{K} \quad (157)$$

$$= 1 - \frac{1}{K} \quad (158)$$

Por lo tanto:

$$\beta_{\pi_\tau} = \begin{cases} 1 & \text{si } \tau \leq -\log K \\ 1 - \frac{1}{K} & \text{si } \tau > -\log K \end{cases} \quad (159)$$

Observaciones importantes:

- El test solo tiene dos posibles comportamientos, determinados por si τ es mayor o menor/igual que $-\log K$.
- Cuando $\tau \leq -\log K$, el test es trivial: siempre decide \mathcal{H}_1 para cualquier observación en $[0, K]$, resultando en $\alpha = \beta = 1$.
- Cuando $\tau > -\log K$, el test decide \mathcal{H}_1 solo cuando $X \in (1, K]$. Este es el test no trivial interesante, con $\alpha = 0$ y $\beta = 1 - \frac{1}{K}$.
- Note que el test con $\tau > -\log K$ tiene tamaño cero ($\alpha = 0$), lo cual es óptimo desde el punto de vista de no cometer falsa alarma, pero tiene una potencia limitada de $1 - \frac{1}{K}$.
- A medida que K aumenta, la potencia $\beta = 1 - \frac{1}{K}$ se acerca a 1, lo cual tiene sentido intuitivo: si K es muy grande, la mayoría de las observaciones bajo \mathcal{H}_1 caerán en $(1, K]$.

Resolución 3.4

Para construir la curva ROC, debemos graficar β versus α para todos los posibles tests.

Del análisis anterior, vimos que los tests determinísticos solo pueden producir dos puntos en el espacio (α, β) :

1. **Punto 1:** Cuando $\tau \leq -\log K$:

$$(\alpha, \beta) = (1, 1) \quad (160)$$

Este es el test trivial que siempre decide \mathcal{H}_1 .

2. **Punto 2:** Cuando $\tau > -\log K$:

$$(\alpha, \beta) = \left(0, 1 - \frac{1}{K}\right) \quad (161)$$

Este es el test que decide \mathcal{H}_1 solo cuando $X > 1$.

Adicionalmente, siempre existe el test trivial que nunca decide \mathcal{H}_1 :

3. **Punto 3:** Test que siempre decide \mathcal{H}_0 :

$$(\alpha, \beta) = (0, 0) \quad (162)$$

¿Es posible obtener la curva ROC completa con tests determinísticos?

No, no es posible obtener la curva ROC completa usando únicamente tests determinísticos. Los tests determinísticos solo producen tres puntos discretos:

- $(0, 0)$ - Test que nunca rechaza \mathcal{H}_0

- $\left(0, 1 - \frac{1}{K}\right)$ - Test óptimo no trivial
- $(1, 1)$ - Test que siempre rechaza \mathcal{H}_0

Para obtener la curva ROC completa, necesitamos tests aleatorizados. Consideremos un test aleatorizado general:

$$\pi_{\gamma_1}(x) = \begin{cases} 1 & \text{si } x \in (1, K] \\ \gamma_1 & \text{si } x \in [0, 1] \\ 0 & \text{si } x \notin [0, K] \end{cases} \quad (163)$$

donde γ_1 representa el valor esperado de la decisión (o probabilidad de decidir \mathcal{H}_1) cuando $x \in [0, 1]$, con $\gamma_1 \in [0, 1]$. En la implementación práctica de detección binaria, cuando $x \in [0, 1]$ se aleatoriza: se decide \mathcal{H}_1 con probabilidad γ_1 y \mathcal{H}_0 con probabilidad $1 - \gamma_1$.

Para este test aleatorizado:

$$\alpha = P(X \in (1, K] | \theta = 0) + \gamma_1 \cdot P(X \in [0, 1] | \theta = 0) \quad (164)$$

$$= 0 + \gamma_1 \cdot 1 \quad (165)$$

$$= \gamma_1 \quad (166)$$

$$\beta = P(X \in (1, K] | \theta = 1) + \gamma_1 \cdot P(X \in [0, 1] | \theta = 1) \quad (167)$$

$$= \frac{K-1}{K} + \gamma_1 \cdot \frac{1}{K} \quad (168)$$

$$= 1 - \frac{1}{K} + \frac{\gamma_1}{K} \quad (169)$$

Eliminando $\gamma_1 = \alpha$ de la segunda ecuación:

$$\beta = 1 - \frac{1}{K} + \frac{\alpha}{K} = 1 - \frac{1-\alpha}{K} \quad (170)$$

Esta es la ecuación de la curva ROC completa para $\alpha \in [0, 1]$:

$$\boxed{\beta(\alpha) = 1 - \frac{1-\alpha}{K} = \frac{K-1+\alpha}{K}} \quad (171)$$

Esta es una recta que conecta los puntos $(0, 1 - \frac{1}{K})$ y $(1, 1)$.

Características de la curva ROC:

- Es una función lineal de α
- Tiene pendiente $\frac{1}{K}$
- Comienza en el punto $(0, 1 - \frac{1}{K})$

- Termina en el punto $(1, 1)$
- No pasa por el origen $(0, 0)$ a menos que se considere el test trivial que nunca rechaza

La curva ROC completa consiste en dos segmentos:

1. Segmento vertical desde $(0, 0)$ hasta $(0, 1 - \frac{1}{K})$: Se obtiene variando entre el test que nunca rechaza y el test determinístico óptimo
2. Segmento lineal desde $(0, 1 - \frac{1}{K})$ hasta $(1, 1)$: Se obtiene con tests aleatorizados según la ecuación

$$\beta = 1 - \frac{1-\alpha}{K}$$

Por lo tanto, la respuesta es: **No es posible obtener la curva ROC completa con tests determinísticos.** Se requieren tests aleatorizados para llenar el segmento entre $(0, 1 - \frac{1}{K})$ y $(1, 1)$. Sin embargo, el test determinístico con $\tau > -\log K$ es óptimo en el sentido de que alcanza el punto con máxima potencia para tamaño cero.

Resolución 3.5

Ahora analizaremos qué sucede con las regiones de decisión y la curva ROC cuando $K \rightarrow \infty$.

Análisis de las regiones de decisión (parte b):

Recordemos que las regiones de decisión son:

Para $\tau \leq -\log K$:

$$A_1^c = [0, K] \quad (172)$$

$$A_0^c = \mathbb{R} \setminus [0, K] \quad (173)$$

Para $\tau > -\log K$:

$$A_1^c = (1, K] \quad (174)$$

$$A_0^c = \mathbb{R} \setminus (1, K] \quad (175)$$

Cuando $K \rightarrow \infty$:

- El umbral crítico $-\log K \rightarrow -\infty$. Esto significa que para cualquier valor finito de τ , eventualmente tendremos $\tau > -\log K$ para K suficientemente grande.
- Las regiones $[0, K]$ y $(1, K]$ se expanden hacia el infinito:

$$\lim_{K \rightarrow \infty} [0, K] = [0, \infty) \quad (176)$$

$$\lim_{K \rightarrow \infty} (1, K] = (1, \infty) \quad (177)$$

- Para cualquier τ finito, el test con $K \rightarrow \infty$ decide:

$$A_1^c \rightarrow (1, \infty) \quad (178)$$

$$A_0^c \rightarrow (-\infty, 1] \quad (179)$$

- La frontera de decisión se estabiliza en $x = 1$, independientemente del valor de K .

Interpretación: En el límite $K \rightarrow \infty$, el test óptimo determinístico tiene una forma simple: decide \mathcal{H}_1 si y solo si $x > 1$. Esta regla de decisión es intuitiva: bajo \mathcal{H}_0 , $X \in [0, 1]$ siempre, mientras que bajo \mathcal{H}_1 con K muy grande, hay una probabilidad alta de observar $X > 1$.

Análisis de la curva ROC (parte d):

La curva ROC tiene la forma:

$$\beta(\alpha) = 1 - \frac{1 - \alpha}{K} \quad (180)$$

Cuando $K \rightarrow \infty$:

$$\lim_{K \rightarrow \infty} \beta(\alpha) = \lim_{K \rightarrow \infty} \left(1 - \frac{1 - \alpha}{K}\right) = 1 - 0 = 1 \quad (181)$$

para todo $\alpha \in (0, 1]$.

Esto significa que la curva ROC converge a:

$$\beta(\alpha) = \begin{cases} 0 & \text{si } \alpha = 0 \\ 1 & \text{si } \alpha \in (0, 1] \end{cases} \quad (182)$$

En otras palabras, la curva ROC converge a una forma de "escalón" o función de Heaviside:

- Parte verticalmente desde $(0, 0)$ hasta $(0, 1)$
- Luego se extiende horizontalmente desde $(0, 1)$ hasta $(1, 1)$

Interpretación y significado:

- **Test casi perfecto:** Cuando $K \rightarrow \infty$, el problema de detección se vuelve "casi perfecto". Es posible alcanzar una potencia arbitrariamente cercana a 1 con un tamaño de test arbitrariamente pequeño.
- **Separación perfecta:** La razón es que, en el límite, las dos distribuciones se vuelven casi completamente separadas. Bajo \mathcal{H}_0 , $X \in [0, 1]$ con probabilidad 1. Bajo \mathcal{H}_1 con $K \rightarrow \infty$, la probabilidad de que $X > 1$ tiende a 1.
- **Región de soporte disjunto:** En el límite, tenemos:

$$\text{Soporte de } f_0 = [0, 1] \quad (183)$$

$$\lim_{K \rightarrow \infty} \text{Soporte de } f_1 = [0, \infty) \quad (184)$$

La fracción del soporte de f_1 que se solapa con el soporte de f_0 es:

$$\frac{1}{K} \rightarrow 0 \text{ cuando } K \rightarrow \infty \quad (185)$$

- **Test óptimo simple:** El test óptimo en el límite es simplemente: decidir \mathcal{H}_1 si $X > 1$. Este test tiene:

$$\alpha = P(X > 1 | \theta = 0) = 0 \quad (186)$$

$$\beta = P(X > 1 | \theta = 1) = \lim_{K \rightarrow \infty} \frac{K - 1}{K} = 1 \quad (187)$$

- **AUC perfecta:** El área bajo la curva ROC converge a:

$$\text{AUC} = \lim_{K \rightarrow \infty} \int_0^1 \beta(\alpha) d\alpha = \lim_{K \rightarrow \infty} \int_0^1 \left(1 - \frac{1-\alpha}{K}\right) d\alpha = 1 \quad (188)$$

Un AUC de 1 indica discriminación perfecta entre las dos hipótesis.

Conclusión: Cuando $K \rightarrow \infty$, el problema de detección se simplifica enormemente. Las dos distribuciones se separan casi completamente, haciendo posible una discriminación casi perfecta con un test determinístico simple (decidir \mathcal{H}_1 si $X > 1$). La curva ROC converge a la forma ideal de una función escalón, indicando rendimiento perfecto del detector.