



Ingeniería Eléctrica

FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

---

# Análisis de Sistemas Dinámicos y Estimación

EL3204-1

---

## *Pauta Auxiliar 9 - Detección Bayesiana*

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



# 1 (Resumen)

## Modelo básico de decisión

Sea  $\{\mathcal{H}_i\}_{i=0}^{M-1}$  el conjunto de hipótesis, una observación  $X \in \mathcal{X}$  con densidades condicionales  $f_{X|\mathcal{H}_i}(x|\mathcal{H}_i)$ , probabilidades a priori  $\pi_i = P(\mathcal{H}_i)$  y matriz de costos  $L(d_j|\mathcal{H}_i)$  (costo de decidir  $d_j$  cuando la hipótesis verdadera es  $\mathcal{H}_i$ ). El **riesgo posterior** para la acción  $d_j$  dado  $x$  es

$$R(d_j|x) = \sum_{i=0}^{M-1} L(d_j|\mathcal{H}_i) P(\mathcal{H}_i|x).$$

La **regla de Bayes** elige la decisión que minimiza el riesgo posterior:

$$\delta^*(x) = \arg \min_{d_j} R(d_j|x).$$

## Caso binario y Regla LRT Bayesiana

Para  $M = 2$  (hipótesis  $\mathcal{H}_0$  vs.  $\mathcal{H}_1$ ), con costos  $L_{10}$  (falso positivo) y  $L_{01}$  (falso negativo), la regla óptima se expresa como una *prueba de razón de verosimilitudes* (LRT):

$$\Lambda(x) = \frac{f_{X|\mathcal{H}_1}(x)}{f_{X|\mathcal{H}_0}(x)} \underset{\mathcal{H}_0}{\geq} \eta, \quad \eta = \frac{L_{10} - L_{00}}{L_{01} - L_{11}} \cdot \frac{\pi_0}{\pi_1}.$$

En el caso de costos 0–1 (errores puntúan 1 y aciertos 0),  $\eta = \pi_0/\pi_1$ .

## Regla MAP y Regla ML

- **MAP (Maximum A Posteriori):**

$$\delta_{\text{MAP}}(x) = \arg \max_i P(\mathcal{H}_i|x) = \arg \max_i \pi_i f_{X|\mathcal{H}_i}(x).$$

Es óptima para costo 0–1: minimiza la probabilidad total de error.

- **ML (Maximum Likelihood):**

$$\delta_{\text{ML}}(x) = \arg \max_i f_{X|\mathcal{H}_i}(x),$$

es un caso especial de MAP cuando las *a priori* son uniformes.

## Curva ROC y desempeño

Al variar el umbral  $\eta$  se obtienen pares  $(P_{\text{FA}}, P_{\text{D}})$  que definen la **curva ROC**. Regiones más próximas a la esquina superior izquierda indican mejores detectores. La elección del umbral se guía por costos y *a priori* (óptimo Bayes) o por una cota en  $P_{\text{FA}}$  (criterio de Neyman–Pearson).

## Detección Gaussiana con media conocida

Si  $X|\mathcal{H}_i \sim \mathcal{N}(\mu_i, \sigma^2)$  con varianza común, la LRT se vuelve umbral en  $x$ :

$$\Lambda(x) = \exp\left(\frac{(\mu_1 - \mu_0)}{\sigma^2} x - \frac{\mu_1^2 - \mu_0^2}{2\sigma^2}\right) \underset{\mathcal{H}_0}{\geq} \eta \iff x \underset{\mathcal{H}_0}{\geq} \gamma,$$

donde el umbral

$$\gamma = \frac{\mu_0 + \mu_1}{2} + \frac{\sigma^2}{\mu_1 - \mu_0} \ln \eta.$$

Con costo 0–1 y *a priori* iguales,  $\ln \eta = 0$  y  $\gamma = \frac{\mu_0 + \mu_1}{2}$ .

### Extensión a $M$ -hipótesis

Para  $M > 2$ :

$$\delta^*(x) = \arg \min_j \sum_i L(d_j | \mathcal{H}_i) P(\mathcal{H}_i | x).$$

Con costo 0–1:

$$\delta_{\text{MAP}}(x) = \arg \max_i P(\mathcal{H}_i | x).$$

Cuando  $f_{X|\mathcal{H}_i}$  y  $\pi_i$  están en la familia exponencial, los *suficientes* lineales permiten implementaciones por *comparadores afines* (fronteras lineales/quadráticas según el caso).

### Estimación Bayesiana (parámetros continuos)

Sea  $\Theta$  un parámetro con *a priori*  $p_\Theta(\theta)$  y verosimilitud  $f_{X|\Theta}(x|\theta)$ .

- **Estimador MMSE** (pérdida cuadrática):

$$\hat{\theta}_{\text{MMSE}}(x) = \mathbb{E}[\Theta | X = x].$$

- **Estimador MAP** (pérdida 0–1 asintótica):

$$\hat{\theta}_{\text{MAP}}(x) = \arg \max_{\theta} p_{\Theta|X}(\theta|x) \propto f_{X|\Theta}(x|\theta) p_{\Theta}(\theta).$$

- **MLE** se recupera con *a priori* plana:  $\hat{\theta}_{\text{ML}}(x) = \arg \max_{\theta} f_{X|\Theta}(x|\theta)$ .

**Ejemplo (Gaussiano conjugado):** Si  $X|\theta \sim \mathcal{N}(\theta, \sigma^2)$  y  $\Theta \sim \mathcal{N}(\mu_0, \tau^2)$ , entonces

$$\Theta | X = x \sim \mathcal{N}\left(\frac{\frac{1}{\sigma^2}x + \frac{1}{\tau^2}\mu_0}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right),$$

y por tanto  $\hat{\theta}_{\text{MMSE}} = \hat{\theta}_{\text{MAP}}$  es un promedio ponderado de  $x$  y  $\mu_0$ .

### Riesgo Bayesiano y probabilidad de error

El **riesgo Bayesiano** es

$$R_B = \mathbb{E}[L(\delta(X)|\mathcal{H})] = \sum_i \pi_i \int L(\delta(x)|\mathcal{H}_i) f_{X|\mathcal{H}_i}(x) dx,$$

y bajo costo 0–1 coincide con la *probabilidad total de error*. En el caso binario con LRT óptima:

$$P_e = \pi_0 P(\text{decidir } \mathcal{H}_1 | \mathcal{H}_0) + \pi_1 P(\text{decidir } \mathcal{H}_0 | \mathcal{H}_1).$$

1. Se pide que implemente un sistema de decisión que detecte la presencia de una señal  $s_t \triangleq s(t)$ . Para eso suponga que se tiene un sistema que observa  $n$  muestras ruidosas de la señal  $(s_k)_{k=1,\dots,n}$ .

En concreto se distinguen dos escenarios posibles de observación.

**Presencia de señal  $\Theta = 1$ :**

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} + \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \end{pmatrix} \quad (2.87)$$

**Ausencia de señal  $\Theta = 0$ :**

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \end{pmatrix} \quad (2.88)$$

donde  $N_1, \dots, N_n$  son variables aleatorias independientes que distribuyen  $\mathcal{N}(0, \sigma^2)$ .

- (a) Notar que dado el valor de  $\Theta = \theta$ , el vector  $\mathbf{X}_1^n = (X_1, \dots, X_n)$  es un vector Gaussiano. Determine su vector de media y matriz de covarianza en ambos escenarios (presencia y ausencia de señal). *Indicación:* Notar que  $X_1, \dots, X_n$  son variables aleatorias independientes.
- (b) Del punto anterior determine la función de log-verosimilitud

$$L(x_1, \dots, x_n | \theta) = \ln f_{X_1, \dots, X_n | \Theta}(x_1, \dots, x_n | \theta)$$

y la solución del problema:

$$\tau_{ML}(x_1, \dots, x_n) = \arg \max_{\theta \in \{0,1\}} L(x_1, \dots, x_n | \theta). \quad (2.89)$$

*Indicación:* Se debe llegar a una expresión cerrada para  $\tau_{ML}(x_1, \dots, x_n)$ , función de  $x_1, \dots, x_n$  y los parámetros conocidos del problema.

- (c) Determine la probabilidad de error del test del punto anterior cuando  $p_{\Theta}(1) = p_{\Theta}(0) = \frac{1}{2}$ .
- (d) Determine que pasa con la probabilidad de error del test óptimo en (2.89), si la potencia de la señal dada por  $\|s\|^2 = \sum_{i=1}^n s_i^2$  tiende a infinito, es decir,

$$\lim_{n \rightarrow \infty} \|s\|^2 = \infty$$

**Solución:****Resolución 1.1**

Analizamos el vector aleatorio  $\mathbf{X}_1^n = (X_1, \dots, X_n)^T$  bajo ambas hipótesis.

- **Caso  $\Theta = 1$  (Presencia de señal):**

De la ecuación (2.87), tenemos que  $X_i = s_i + N_i$  para  $i = 1, \dots, n$ . Como los  $N_i$  son independientes con  $N_i \sim \mathcal{N}(0, \sigma^2)$  y la señal  $s_i$  es determinista:

Primero calculamos el vector de media y la matriz de covarianza. Para lo primero tenemos que:

$$\mathbb{E}[X_i | \Theta = 1] = \mathbb{E}[s_i + N_i] = s_i + \mathbb{E}[N_i] = s_i$$

Por lo tanto:

$$\boldsymbol{\mu}_1 = \mathbb{E}[\mathbf{X}_1^n | \Theta = 1] = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \mathbf{s}$$

Ahora debemos caracterizar cómo varían las observaciones alrededor de su media. Para un vector aleatorio  $\mathbf{X} = (X_1, \dots, X_n)^T$ , la matriz de covarianza  $\boldsymbol{\Sigma}$  captura tanto:

- La variabilidad de cada variable individual (elementos diagonales)
- La relación lineal entre pares de variables (elementos fuera de la diagonal)

La matriz de covarianza se define como:

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

donde cada elemento  $(i, j)$  de esta matriz es:

$$\Sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \text{Cov}(X_i, X_j)$$

Esta matriz es fundamental en el análisis de vectores Gaussianos porque especifica completamente la estructura de correlación entre las componentes del vector.

Luego, para la matriz de covarianza en nuestro problema, recordemos que como las variables  $N_i$  son independientes, entonces:

$$\text{Cov}(X_i, X_j | \Theta = 1) = \text{Cov}(s_i + N_i, s_j + N_j) = \text{Cov}(N_i, N_j) = \begin{cases} \sigma^2 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

La matriz de covarianza  $\boldsymbol{\Sigma}_1$  es una matriz de  $n \times n$  donde el elemento  $(i, j)$  es  $\text{Cov}(X_i, X_j | \Theta = 1)$ . Explícitamente, esta matriz tiene la forma:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

Sustituyendo los valores de las covarianzas (las varianzas  $\sigma^2$  en la diagonal y ceros fuera de ella):

$$\Sigma_1 = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{I}_n$$

donde  $\mathbf{I}_n$  es la matriz identidad de  $n \times n$ . Esta estructura diagonal refleja la independencia entre las observaciones  $X_i$ , ya que todas tienen la misma varianza  $\sigma^2$  y covarianza cero entre ellas.

• **Caso  $\Theta = 0$  (Ausencia de señal):**

De la ecuación, tenemos que  $X_i = N_i$  para  $i = 1, \dots, n$ , por lo que análogamente tenemos que:

$$\mu_0 = \mathbb{E}[\mathbf{X}_1^n | \Theta = 0] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

y por otro lado:

$$\Sigma_0 = \sigma^2 \mathbf{I}_n$$

## Resolución 1.2

La regla de Máxima Verosimilitud (ML, por sus siglas en inglés *Maximum Likelihood*) es un caso particular de la regla MAP cuando las hipótesis son equiprobables o no se tienen priors informativos. Como  $\Sigma_0 = \Sigma_1$  y no se especifican costos ni priors distintos, con  $p_\Theta(0) = p_\Theta(1)$ , la regla MAP coincide con ML.

Matemáticamente, la regla ML decide la hipótesis  $\theta^*$  que maximiza la función de verosimilitud:

$$\theta^* = \arg \max_{\theta \in \Theta} f_{X|\Theta}(x|\theta)$$

Equivalentemente, podemos maximizar la log-verosimilitud (que es más conveniente computacionalmente):

$$\theta^* = \arg \max_{\theta \in \Theta} L(x|\theta) = \arg \max_{\theta \in \Theta} \ln f_{X|\Theta}(x|\theta)$$

Sabemos que la función de densidad para un vector Gaussiano multivariado es:

$$f_{\mathbf{X}_1^n|\Theta}(\mathbf{x}|\theta) = \frac{1}{(2\pi)^{n/2} |\Sigma_\theta|^{1/2}} \exp \left( -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_\theta)^T \Sigma_\theta^{-1} (\mathbf{x} - \boldsymbol{\mu}_\theta) \right)$$

Luego al aplicar el logaritmo, obtenemos la log-verosimilitud:

$$\begin{aligned} L(\mathbf{x}|\theta) &= \ln f_{\mathbf{X}_1^n|\Theta}(\mathbf{x}|\theta) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_\theta| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_\theta)^T \Sigma_\theta^{-1} (\mathbf{x} - \boldsymbol{\mu}_\theta) \end{aligned}$$

Para  $\theta = 1$ :

$$\begin{aligned} L(\mathbf{x}|1) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - s_i)^2 \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - s_i)^2 \end{aligned}$$

Para  $\theta = 0$ :

$$L(\mathbf{x}|0) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2$$

Para determinar  $\tau_{ML}(\mathbf{x})$ , comparamos las log-verosimilitudes:

$$\begin{aligned} \tau_{ML}(\mathbf{x}) = 1 &\iff L(\mathbf{x}|1) > L(\mathbf{x}|0) \\ &\iff -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - s_i)^2 > -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \\ &\iff \sum_{i=1}^n (x_i - s_i)^2 < \sum_{i=1}^n x_i^2 \\ &\iff \sum_{i=1}^n (x_i^2 - 2x_i s_i + s_i^2) < \sum_{i=1}^n x_i^2 \\ &\iff -2 \sum_{i=1}^n x_i s_i + \sum_{i=1}^n s_i^2 < 0 \\ &\iff \sum_{i=1}^n x_i s_i > \frac{1}{2} \sum_{i=1}^n s_i^2 = \frac{\|s\|^2}{2} \end{aligned}$$

Por lo que se llega a la siguiente regla de decisión:

$$\tau_{ML}(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } T = \sum_{i=1}^n x_i s_i > \frac{\|s\|^2}{2} \\ 0 & \text{en caso contrario} \end{cases}$$

donde el umbral  $\frac{\|s\|^2}{2}$  corresponde al punto medio entre las medias de  $T$  bajo ambas hipótesis.

### Resolución 1.3

Con  $p_{\Theta}(1) = p_{\Theta}(0) = \frac{1}{2}$ , la probabilidad de error es:

$$\begin{aligned} P_e &= P(\tau_{ML} = 0, \Theta = 1) + P(\tau_{ML} = 1, \Theta = 0) \\ &= P(\Theta = 1)P(\tau_{ML} = 0|\Theta = 1) + P(\Theta = 0)P(\tau_{ML} = 1|\Theta = 0) \\ &= \frac{1}{2}P\left(\sum_{i=1}^n X_i s_i \leq \frac{\|s\|^2}{2} \middle| \Theta = 1\right) + \frac{1}{2}P\left(\sum_{i=1}^n X_i s_i > \frac{\|s\|^2}{2} \middle| \Theta = 0\right) \end{aligned}$$

Analizamos la distribución de  $\sum_{i=1}^n X_i s_i$  bajo cada hipótesis.

**Bajo  $\Theta = 1$ :** Como  $X_i = s_i + N_i$ , tenemos:

$$\sum_{i=1}^n X_i s_i = \sum_{i=1}^n (s_i + N_i) s_i = \sum_{i=1}^n s_i^2 + \sum_{i=1}^n N_i s_i = \|s\|^2 + \sum_{i=1}^n N_i s_i$$

Como  $N_i \sim \mathcal{N}(0, \sigma^2)$  son independientes, la suma  $\sum_{i=1}^n N_i s_i$  es Gaussiana con media cero y varianza  $\sigma^2 \sum_{i=1}^n s_i^2 = \sigma^2 \|s\|^2$ . Por lo tanto:

$$\sum_{i=1}^n X_i s_i \Big| \Theta = 1 \sim \mathcal{N}(\|s\|^2, \sigma^2 \|s\|^2)$$

**Bajo  $\Theta = 0$ :** Como  $X_i = N_i$ , entonces:

$$\sum_{i=1}^n X_i s_i \Big| \Theta = 0 = \sum_{i=1}^n N_i s_i \sim \mathcal{N}(0, \sigma^2 \|s\|^2)$$

Ahora calculamos las probabilidades. Para  $\Theta = 1$ , estandarizamos:

$$\begin{aligned} P\left(\sum_{i=1}^n X_i s_i \leq \frac{\|s\|^2}{2} \Big| \Theta = 1\right) &= P\left(\frac{\sum_{i=1}^n X_i s_i - \|s\|^2}{\sigma \|s\|} \leq \frac{-\|s\|^2/2}{\sigma \|s\|} \Big| \Theta = 1\right) \\ &= P\left(Z \leq -\frac{\|s\|}{2\sigma}\right) = Q\left(\frac{\|s\|}{2\sigma}\right) \end{aligned}$$

donde  $Z \sim \mathcal{N}(0, 1)$  y  $Q(x) = P(Z > x)$  es la función Q.

Para  $\Theta = 0$ :

$$\begin{aligned} P\left(\sum_{i=1}^n X_i s_i > \frac{\|s\|^2}{2} \Big| \Theta = 0\right) &= P\left(\frac{\sum_{i=1}^n X_i s_i}{\sigma \|s\|} > \frac{\|s\|^2/2}{\sigma \|s\|} \Big| \Theta = 0\right) \\ &= P\left(Z > \frac{\|s\|}{2\sigma}\right) = Q\left(\frac{\|s\|}{2\sigma}\right) \end{aligned}$$

Por simetría, ambos términos son iguales, por lo tanto:

$$P_e = \frac{1}{2}Q\left(\frac{\|s\|}{2\sigma}\right) + \frac{1}{2}Q\left(\frac{\|s\|}{2\sigma}\right) = Q\left(\frac{\|s\|}{2\sigma}\right)$$

### Resolución 1.4

Del punto anterior, tenemos que la probabilidad de error es:

$$P_e = Q\left(\frac{\|s\|}{2\sigma}\right)$$



Cuando  $\|s\|^2 \rightarrow \infty$ , entonces  $\|s\| \rightarrow \infty$ , y por lo tanto:

$$\lim_{\|s\| \rightarrow \infty} P_e = \lim_{\|s\| \rightarrow \infty} Q\left(\frac{\|s\|}{2\sigma}\right) = Q(\infty) = 0$$

Podemos entenderlo como cuando la potencia de la señal tiende a infinito, el test de máxima verosimilitud puede distinguir perfectamente entre presencia y ausencia de señal, ya que la probabilidad de error tiende a cero. Esto tiene sentido físicamente: una señal con potencia infinita es fácilmente distinguible del ruido, sin importar cuán grande sea la varianza del ruido.

2. El estudiante Juanito fue a una degustación a ciegas en el Paseo Eléctrico. Lamentablemente Juanito estaba muy congestionado por lo que le costaba identificar los sabores y además por ser una degustación a ciegas no sabía cuál bebida estaba tomando. Se sabe que este local durante la degustación ofreció 3 tipos de bebidas: vino, pisco y baltica, además, la frecuencia con la cuál salía cada bebida del bar era del 50%, 20% y 30% respectivamente. Las bebidas pueden gustarle o no gustarle a Juanito y, por experiencia, él sabe que si la bebida es vino le gusta el 40% de las veces, si es pisco le gusta el 20% de las veces y si es baltica le gusta el 70% de las veces. Juanito probó la bebida recibida y le gustó. El objetivo de esta pregunta es que usted ayude a Juanito a detectar cuál fue la bebida que tomó. Para esto siga los siguientes pasos:
- Plantee un espacio de observación y un espacio de decisión adecuado, indique la distribución de la variable aleatoria  $\Theta$  asociado al espacio de decisión (distribución a priori) y las densidades y/o probabilidades de masa condicionales de  $X$  dado  $\Theta = \theta$ .
  - Suponga que la función de costo del problema Bayesiano es la función  $L_{0,1}$ , es decir,

$$L_{0,1}(x, y) = \begin{cases} 1 & \text{si } x \neq y \\ 0 & \text{si } x = y. \end{cases} \quad (2.90)$$

Plantee la regla óptima de decisión y a partir de esto decida cuál fue la bebida que tomó.

### Solución:

#### Resolución 2.1

Para plantear el problema de decisión Bayesiana, debemos identificar los espacios y distribuciones involucradas.

El espacio de decisión corresponde al tipo de bebida que Juanito debe identificar:

$$\Theta = \{\text{Vino, Pisco, Baltica}\} = \{V, P, B\}$$

La distribución a priori  $p_{\Theta}(\theta)$  corresponde a la frecuencia con que sale cada bebida del bar:

$$p_{\Theta}(V) = 0.50$$

$$p_{\Theta}(P) = 0.20$$

$$p_{\Theta}(B) = 0.30$$

El espacio de observación corresponde a si la bebida le gustó o no a Juanito:

$$\mathcal{X} = \{\text{Gusta, No Gusta}\} = \{1, 0\}$$

Las probabilidades de masa condicionales  $P_{X|\Theta}(x|\theta)$  representan la probabilidad de que a Juanito le guste la bebida dado el tipo de bebida:

Para  $X = 1$  (le gusta):

$$P_{X|\Theta}(1|V) = 0.40$$

$$P_{X|\Theta}(1|P) = 0.20$$

$$P_{X|\Theta}(1|B) = 0.70$$

Para  $X = 0$  (no le gusta):

$$P_{X|\Theta}(0|V) = 1 - 0.40 = 0.60$$

$$P_{X|\Theta}(0|P) = 1 - 0.20 = 0.80$$

$$P_{X|\Theta}(0|B) = 1 - 0.70 = 0.30$$

## Resolución 2.2

Con la función de costo  $L_{0,1}$  dada en (2.90), el riesgo Bayesiano para una regla de decisión  $\delta(x)$  es:

$$\begin{aligned} r(\delta) &= \mathbb{E}_{X,\Theta}[L_{0,1}(\Theta, \delta(X))] \\ &= P(\Theta \neq \delta(X)) \\ &= P(\text{error}) \end{aligned}$$

Es decir, minimizar el riesgo equivale a minimizar la probabilidad de error. La regla de decisión óptima es el estimador MAP:

$$\delta^*(x) = \arg \max_{\theta \in \{V,P,B\}} p_{\Theta|X}(\theta|x)$$

Usando el teorema de Bayes:

$$p_{\Theta|X}(\theta|x) = \frac{P_{X|\Theta}(x|\theta) \cdot p_{\Theta}(\theta)}{p_X(x)}$$

Como  $p_X(x)$  es constante para todas las hipótesis, podemos comparar directamente:

$$\delta^*(x) = \arg \max_{\theta \in \{V,P,B\}} P_{X|\Theta}(x|\theta) \cdot p_{\Theta}(\theta)$$

Juanito observó que le gustó la bebida, es decir,  $X = 1$ . Calculamos las probabilidades a posteriori (sin normalizar):

$$P_{X|\Theta}(1|V) \cdot p_{\Theta}(V) = 0.40 \times 0.50 = 0.20$$

$$P_{X|\Theta}(1|P) \cdot p_{\Theta}(P) = 0.20 \times 0.20 = 0.04$$

$$P_{X|\Theta}(1|B) \cdot p_{\Theta}(B) = 0.70 \times 0.30 = 0.21$$

Comparando los valores:

$$\max\{0.20, 0.04, 0.21\} = 0.21$$

La regla de decisión óptima indica que Juanito tomó **baltica**, ya que es la hipótesis con mayor probabilidad a posteriori dado que le gustó la bebida.

Primero calculamos  $p_X(1)$ :

$$p_X(1) = 0.20 + 0.04 + 0.21 = 0.45$$

Las probabilidades a posteriori normalizadas son:

$$\begin{aligned}p_{\Theta|X}(V|1) &= \frac{0.20}{0.45} = 0.444 \approx 44.4\% \\p_{\Theta|X}(P|1) &= \frac{0.04}{0.45} = 0.089 \approx 8.9\% \\p_{\Theta|X}(B|1) &= \frac{0.21}{0.45} = 0.467 \approx 46.7\%\end{aligned}$$

La baltica tiene la mayor probabilidad a posteriori (46.7%), confirmando nuestra decisión óptima.

3. Considere un problema de detección Bayesiano con un espacio de observación  $\mathbb{X}$  arbitrario tal que  $\Theta = \{0, 1\}$ , es decir, un problema de decisión binario. Además considere la función de costo  $L_{0,1}$ .

(a) Demuestre que en este caso la regla óptima puede escribirse como:

$$r^*(x) = \begin{cases} 1 & \text{si } \eta(x) > 1/2 \\ 0 & \text{si } \eta(x) < 1/2 \\ I & \text{si } \eta(x) = 1/2 \end{cases} \quad (2.91)$$

donde  $\eta(x) = P_{\Theta|X}(\Theta = 1|X = x)$  e  $I$  indica indiferencia, es decir, 0 o 1.

(b) Demuestre que para cualquier otra regla de decisión  $\pi : \mathbb{X} \rightarrow \{0, 1\}$  se tiene que:

$$P_{X,\Theta}(r^*(X) \neq \Theta) \leq P_{X,\Theta}(\pi(X) \neq \Theta).$$

Es decir, la regla  $r^*$  es aquella que minimiza la probabilidad de error. Para esto use el hecho que, dado  $x \in \mathbb{X}$  y una regla  $\pi : \mathbb{X} \rightarrow \{0, 1\}$  arbitraria entonces:

$$P_{\Theta|X}(\pi(X) \neq \Theta|X = x) = 1 - \left[ \mathbf{1}_{\pi(x)=1}(x) \cdot \eta(x) + \mathbf{1}_{\pi(x)=0}(x) \cdot (1 - \eta(x)) \right], \quad (2.92)$$

donde  $\mathbf{1}_B(x)$  corresponde a la indicatriz, es decir, vale 1 si  $x \in B$  y 0 en caso contrario.

### Solución:

#### Resolución 3.1

Para encontrar la regla de decisión óptima bajo la función de costo  $L_{0,1}$ , debemos minimizar el riesgo Bayesiano.

Con la función de costo  $L_{0,1}$ , tenemos:

$$L_{0,1}(\theta, \delta) = \begin{cases} 0 & \text{si } \theta = \delta \\ 1 & \text{si } \theta \neq \delta \end{cases}$$

El riesgo Bayesiano condicional para una decisión  $\delta$  dado  $X = x$  es:

$$\begin{aligned} r(\delta|x) &= \mathbb{E}_{\Theta|X}[L_{0,1}(\Theta, \delta)|X = x] \\ &= \sum_{\theta \in \{0,1\}} L_{0,1}(\theta, \delta) \cdot P_{\Theta|X}(\theta|x) \end{aligned}$$

Para  $\delta = 1$ :

$$\begin{aligned} r(1|x) &= L_{0,1}(0, 1) \cdot P_{\Theta|X}(0|x) + L_{0,1}(1, 1) \cdot P_{\Theta|X}(1|x) \\ &= 1 \cdot P_{\Theta|X}(0|x) + 0 \cdot P_{\Theta|X}(1|x) \\ &= P_{\Theta|X}(0|x) = 1 - \eta(x) \end{aligned}$$

Para  $\delta = 0$ :

$$\begin{aligned} r(0|x) &= L_{0,1}(0,0) \cdot P_{\Theta|X}(0|x) + L_{0,1}(1,0) \cdot P_{\Theta|X}(1|x) \\ &= 0 \cdot P_{\Theta|X}(0|x) + 1 \cdot P_{\Theta|X}(1|x) \\ &= P_{\Theta|X}(1|x) = \eta(x) \end{aligned}$$

La regla óptima minimiza el riesgo condicional para cada  $x$ :

$$r^*(x) = \arg \min_{\delta \in \{0,1\}} r(\delta|x)$$

Comparando los riesgos:

$$\begin{aligned} r^*(x) = 1 &\iff r(1|x) < r(0|x) \\ &\iff 1 - \eta(x) < \eta(x) \\ &\iff 1 < 2\eta(x) \\ &\iff \eta(x) > \frac{1}{2} \end{aligned}$$

Similarmente:

$$r^*(x) = 0 \iff \eta(x) < \frac{1}{2}$$

Cuando  $\eta(x) = 1/2$ , ambos riesgos son iguales ( $r(0|x) = r(1|x) = 1/2$ ), por lo que hay indiferencia.

Por lo tanto, la regla óptima es:

$$r^*(x) = \begin{cases} 1 & \text{si } \eta(x) > 1/2 \\ 0 & \text{si } \eta(x) < 1/2 \\ I & \text{si } \eta(x) = 1/2 \end{cases}$$

### Resolución 3.2

Para demostrar que  $r^*$  minimiza la probabilidad de error, usamos la expresión (2.92) y el teorema de la probabilidad total. La probabilidad de error para una regla  $\pi$  es:

$$\begin{aligned} P_{X,\Theta}(\pi(X) \neq \Theta) &= \mathbb{E}_X[P_{\Theta|X}(\pi(X) \neq \Theta|X)] \\ &= \int_{\mathbb{X}} P_{\Theta|X}(\pi(X) \neq \Theta|X=x) dP_X(x) \end{aligned}$$

Usando la ecuación (2.92):

$$P_{\Theta|X}(\pi(X) \neq \Theta|X=x) = 1 - \left[ \mathbf{1}_{\pi(x)=1}(x) \cdot \eta(x) + \mathbf{1}_{\pi(x)=0}(x) \cdot (1 - \eta(x)) \right]$$

Notemos que para cualquier  $x$ , se cumple que  $\mathbf{1}_{\pi(x)=1}(x) + \mathbf{1}_{\pi(x)=0}(x) = 1$ , ya que  $\pi(x) \in \{0,1\}$ .

Podemos reescribir:

$$\begin{aligned}
 & \mathbf{1}_{\pi(x)=1}(x) \cdot \eta(x) + \mathbf{1}_{\pi(x)=0}(x) \cdot (1 - \eta(x)) \\
 &= \mathbf{1}_{\pi(x)=1}(x) \cdot \eta(x) + [1 - \mathbf{1}_{\pi(x)=1}(x)] \cdot (1 - \eta(x)) \\
 &= \mathbf{1}_{\pi(x)=1}(x) \cdot \eta(x) + (1 - \eta(x)) - \mathbf{1}_{\pi(x)=1}(x) \cdot (1 - \eta(x)) \\
 &= (1 - \eta(x)) + \mathbf{1}_{\pi(x)=1}(x) \cdot [2\eta(x) - 1]
 \end{aligned}$$

Por lo tanto:

$$P_{\Theta|X}(\pi(X) \neq \Theta | X = x) = \eta(x) - \mathbf{1}_{\pi(x)=1}(x) \cdot [2\eta(x) - 1]$$

Para minimizar esta expresión dado  $x$ , debemos elegir  $\mathbf{1}_{\pi(x)=1}(x)$  óptimamente:

- Si  $2\eta(x) - 1 > 0$  (es decir,  $\eta(x) > 1/2$ ): elegimos  $\mathbf{1}_{\pi(x)=1}(x) = 1$ , o sea  $\pi(x) = 1$ .
- Si  $2\eta(x) - 1 < 0$  (es decir,  $\eta(x) < 1/2$ ): elegimos  $\mathbf{1}_{\pi(x)=1}(x) = 0$ , o sea  $\pi(x) = 0$ .
- Si  $2\eta(x) - 1 = 0$  (es decir,  $\eta(x) = 1/2$ ): cualquier decisión da el mismo resultado.

Esta es exactamente la regla  $r^*(x)$  definida en (2.91).

Como  $r^*$  minimiza  $P_{\Theta|X}(\pi(X) \neq \Theta | X = x)$  para cada  $x \in \mathbb{X}$ , también minimiza la probabilidad de error total:

$$P_{X,\Theta}(r^*(X) \neq \Theta) = \int_{\mathbb{X}} \min_{\delta \in \{0,1\}} P_{\Theta|X}(\delta \neq \Theta | X = x) dP_X(x)$$

Por lo tanto, para cualquier otra regla  $\pi$ :

$$P_{X,\Theta}(r^*(X) \neq \Theta) \leq P_{X,\Theta}(\pi(X) \neq \Theta)$$

Esto demuestra que  $r^*$  es la regla óptima que minimiza la probabilidad de error.

4. Considere un cuerpo radiactivo que emite  $\theta$  partículas, con  $\theta \in \mathbb{N}$ . Para detectar las partículas emitidas, se cuenta con un detector imperfecto, el cual detecta cada partícula emitida de forma independiente. Para modelar el proceso de detección, consideremos la variable aleatoria  $B_i$  que toma el valor 1 si la partícula  $i$ -ésima fue detectada y 0 si no, donde  $B_i$  distribuye Bernoulli de parámetro  $p$  ( $P_{B_i}(B_i = 1) = p$ ).

Finalmente, la variable de observación  $X$  es el número de partículas totales detectadas dada por

$$X = \sum_{i=1}^{\theta} B_i \in \{0, \dots, \theta\}$$

Notar que dados  $p$  y  $\theta$  conocidos,  $X$  distribuye binomial de parámetros  $p$  y  $\theta$ , es decir:

$$P_X(X = k|p, \theta) = \binom{\theta}{k} p^k (1-p)^{\theta-k}$$

Considere el problema de estimar la cantidad de partículas emitidas  $\theta$  asumiendo conocido  $p$ , pero en un contexto Bayesiano, donde la cantidad de partículas emitidas distribuye Poisson de parámetro  $\lambda$  conocido, es decir:

$$p_{\Theta}(\theta) = \frac{\lambda^{\theta}}{\theta!} e^{-\lambda}, \quad \forall \theta \in \{0, 1, 2, \dots\}$$

Luego, se busca el estimador que minimice el error cuadrático medio  $\phi_{MMSE}(X)$ , dada una observación de  $X$ . Para ello, siga los siguientes pasos:

- (a) Determine la probabilidad conjunta  $P_{X,\Theta}(X = k, \Theta = \theta)$  y con ello muestre que la variable aleatoria  $X$  (número de partículas detectadas) distribuye Poisson de parámetro  $\lambda p$ , es decir:

$$p_X(k) = \frac{(\lambda p)^k}{k!} e^{-\lambda p}, \quad \forall k \in \{0, 1, 2, \dots\}$$

- (b) Muestre que:

$$P_{\Theta|X}(\Theta = \theta|X = k) = \frac{(\lambda(1-p))^{\theta-k}}{(\theta-k)!} e^{-\lambda(1-p)}, \quad \text{si } \theta \geq k$$

y

$$P_{\Theta|X}(\theta|k) = 0 \quad \text{si } \theta < k$$

y con ello obtenga  $\pi_{MMSE}(X)$ . Comente sobre los regímenes  $p \approx 1$  y  $p \approx 0$ .



**Solución:****Resolución 4.1**

Para encontrar la distribución marginal de  $X$ , primero calculamos la distribución conjunta y luego marginalizamos sobre  $\Theta$ .

$$\begin{aligned} P_{X,\Theta}(X = k, \Theta = \theta) &= P_{X|\Theta}(X = k|\Theta = \theta) \cdot P_{\Theta}(\theta) \\ &= \binom{\theta}{k} p^k (1-p)^{\theta-k} \cdot \frac{\lambda^\theta}{\theta!} e^{-\lambda} \end{aligned}$$

Para  $\theta \geq k$ :

$$\begin{aligned} P_{X,\Theta}(k, \theta) &= \frac{\theta!}{k!(\theta-k)!} p^k (1-p)^{\theta-k} \cdot \frac{\lambda^\theta}{\theta!} e^{-\lambda} \\ &= \frac{p^k (1-p)^{\theta-k} \lambda^\theta}{k!(\theta-k)!} e^{-\lambda} \end{aligned}$$

Para  $\theta < k$ :  $P_{X,\Theta}(k, \theta) = 0$  (no se pueden detectar más partículas de las emitidas).

$$\begin{aligned} p_X(k) &= \sum_{\theta=0}^{\infty} P_{X,\Theta}(k, \theta) \\ &= \sum_{\theta=k}^{\infty} \frac{p^k (1-p)^{\theta-k} \lambda^\theta}{k!(\theta-k)!} e^{-\lambda} \\ &= \frac{p^k e^{-\lambda}}{k!} \sum_{\theta=k}^{\infty} \frac{(1-p)^{\theta-k} \lambda^\theta}{(\theta-k)!} \end{aligned}$$

Hacemos el cambio de variable  $m = \theta - k$ :

$$\begin{aligned} p_X(k) &= \frac{p^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \frac{(1-p)^m \lambda^{m+k}}{m!} \\ &= \frac{p^k \lambda^k e^{-\lambda}}{k!} \sum_{m=0}^{\infty} \frac{[\lambda(1-p)]^m}{m!} \\ &= \frac{(p\lambda)^k e^{-\lambda}}{k!} \cdot e^{\lambda(1-p)} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda + \lambda(1-p)} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned}$$

Por lo tanto,  $X \sim \text{Poisson}(\lambda p)$ .

**Resolución 4.2**

Usando el teorema de Bayes:

$$P_{\Theta|X}(\theta|k) = \frac{P_{X,\Theta}(k, \theta)}{p_X(k)}$$

Para  $\theta \geq k$ :

$$\begin{aligned} P_{\Theta|X}(\theta|k) &= \frac{\frac{p^k(1-p)^{\theta-k}\lambda^\theta}{k!(\theta-k)!}e^{-\lambda}}{\frac{(\lambda p)^k}{k!}e^{-\lambda p}} \\ &= \frac{p^k(1-p)^{\theta-k}\lambda^\theta e^{-\lambda}}{(\theta-k)!} \cdot \frac{k!}{(\lambda p)^k e^{-\lambda p}} \\ &= \frac{(1-p)^{\theta-k}\lambda^\theta e^{-\lambda}}{(\theta-k)!} \cdot \frac{1}{\lambda^k e^{-\lambda p}} \\ &= \frac{\lambda^{\theta-k}(1-p)^{\theta-k}}{(\theta-k)!} e^{-\lambda+\lambda p} \\ &= \frac{[\lambda(1-p)]^{\theta-k}}{(\theta-k)!} e^{-\lambda(1-p)} \end{aligned}$$

Para  $\theta < k$ :  $P_{\Theta|X}(\theta|k) = 0$ .

Observamos que dado  $X = k$ , la variable  $\Theta$  tiene una distribución Poisson desplazada:  $\Theta - k \sim \text{Poisson}(\lambda(1-p))$ .

El estimador MMSE es la esperanza condicional:

$$\begin{aligned} \pi_{MMSE}(k) &= \mathbb{E}[\Theta|X = k] \\ &= \sum_{\theta=k}^{\infty} \theta \cdot P_{\Theta|X}(\theta|k) \\ &= \sum_{\theta=k}^{\infty} \theta \cdot \frac{[\lambda(1-p)]^{\theta-k}}{(\theta-k)!} e^{-\lambda(1-p)} \end{aligned}$$

Usando el cambio de variable  $m = \theta - k$ :

$$\begin{aligned} \pi_{MMSE}(k) &= \sum_{m=0}^{\infty} (m+k) \cdot \frac{[\lambda(1-p)]^m}{m!} e^{-\lambda(1-p)} \\ &= k \sum_{m=0}^{\infty} \frac{[\lambda(1-p)]^m}{m!} e^{-\lambda(1-p)} + \sum_{m=0}^{\infty} m \cdot \frac{[\lambda(1-p)]^m}{m!} e^{-\lambda(1-p)} \\ &= k \cdot 1 + \lambda(1-p) \\ &= k + \lambda(1-p) \end{aligned}$$

Por lo tanto:

$$\boxed{\pi_{MMSE}(X) = X + \lambda(1-p)}$$

**Análisis de regímenes:**

- **Régimen  $p \approx 1$  (detector casi perfecto):**

$$\pi_{MMSE}(X) \approx X + \lambda(1 - 1) = X$$

El estimador es aproximadamente igual al número de partículas detectadas, lo cual tiene sentido ya que el detector detecta casi todas las partículas emitidas.

- **Régimen  $p \approx 0$  (detector muy malo):**

$$\pi_{MMSE}(X) \approx X + \lambda(1 - 0) = X + \lambda$$

El estimador se sesga fuertemente hacia el prior  $\lambda$ . Como el detector detecta muy pocas partículas, la observación  $X$  es poco informativa y el estimador se apoya más en la información a priori. El término  $X$  es pequeño (pocas detecciones) y el término  $\lambda$  domina, estimando cerca del valor esperado a priori.