



Análisis de Sistemas Dinámicos y Estimación

EL3204-1

Pauta Auxiliar 9 - Detección Bayesiana

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



Resumen

En esta unidad se presenta la perspectiva bayesiana: en lugar de asumir un parámetro desconocido fijo se modela Θ como una variable aleatoria con una distribución a priori que refleja la creencia del agente. Las observaciones actualizan esa creencia mediante la regla de Bayes y, dado un esquema de costos, la decisión óptima minimiza el riesgo posterior. Desde un enfoque subjetivista la prior no tiene que describir un mecanismo físico, sino que recoge información previa del detector; esto permite un tratamiento más flexible de la incertidumbre en problemas de detección y estimación.

En el contexto Bayesiano, Θ se modela como una variable aleatoria con valores en $\mathcal{A} = \{1, \dots, k\}$ y distribución a priori $p_\Theta(\theta)$. Dada una observación X , la probabilidad condicional se expresa mediante la regla de probabilidad condicional:

$$\mathbb{P}(X(w) \in B | \Theta(w) = \theta) = P_{X|\Theta}(B, \{\theta\}) = P_\Theta(\{\theta\}) \cdot P_{X|\Theta}(B | \{\theta\}).$$

Para un vector conjunto (X, Θ) , la distribución conjunta queda determinada por:

$$\mathbb{P}(X(w) \in B, \Theta(w) = \theta) = P_{X,\Theta}(B, \{\theta\}) = p_\Theta(\theta) \cdot \int_B f_{X|\Theta}(x|\theta) dx$$

en el caso continuo, o bien

$$\mathbb{P}(X(w) \in B, \Theta(w) = \theta) = P_{X,\Theta}(B, \{\theta\}) = p_\Theta(\theta) \cdot \sum_{x \in B} p_{X|\Theta}(x|\theta)$$

en el caso discreto. Con esto, un problema de detección Bayesiano consiste en definir un espacio de observación \mathbb{X} , un espacio de decisión \mathcal{A} , las distribuciones condicionales $P_{X|\Theta}(\cdot|\theta)$ o densidades $f_{X|\Theta}(\cdot|\theta)$, una regla de decisión (test) $\pi : \mathbb{X} \rightarrow \mathcal{A}$ y una función de costo $L : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}^+ \cup \{0\}$ que penaliza decisiones incorrectas.

- **Riesgo Promedio Bayesiano:** Para encontrar la regla óptima, se busca minimizar el riesgo promedio Bayesiano, que es el promedio del costo dado por $L(\Theta, \pi(X))$. Primero se define el *riesgo promedio* condicionado a $\Theta = \theta$:

$$R(\theta, \pi) \triangleq \mathbb{E}(L(\theta, \pi(X)) | \Theta = \theta) = \begin{cases} \int_{\mathbb{X}} L(\theta, \pi(x)) f_{X|\Theta}(x|\theta) dx & (\text{caso continuo}) \\ \sum_{x \in \mathbb{X}} L(\theta, \pi(x)) p_{X|\Theta}(x|\theta) & (\text{caso discreto}) \end{cases}$$

El *Riesgo Promedio Bayesiano* se obtiene promediando $R(\theta, \pi)$ respecto a la distribución a priori de Θ :

$$r(\pi) \triangleq \mathbb{E}_\Theta(R(\Theta, \pi)) = \sum_{\theta \in \mathcal{A}} R(\theta, \pi) \cdot p_\Theta(\theta) = \mathbb{E}_{X,\Theta}(L(\Theta, \pi(X))).$$

La *regla óptima Bayesiana* π^* es aquella que minimiza el riesgo promedio Bayesiano:

$$\pi^* = \arg \min_{\pi \in F(\mathbb{X}, \mathcal{A})} r(\pi) = \arg \min_{\pi \in F(\mathbb{X}, \mathcal{A})} \mathbb{E}_{X,\Theta}(L(\Theta, \pi(X))).$$

Equivalentemente, para cada observación $x \in \mathbb{X}$, la decisión óptima $\pi^*(x)$ minimiza el *riesgo a posteriori*:

$$\pi^*(x) = \arg \min_{y \in \mathcal{A}} \sum_{\theta \in \mathcal{A}} L(\theta, y) P_{\Theta|X}(\Theta = \theta | x) = \arg \min_{y \in \mathcal{A}} \mathbb{E}(L(\Theta, y) | X = x).$$

- **Función de costo 0-1 y regla MAP:** La función de costo 0-1 juega un rol fundamental en problemas de reconocimiento de patrones y comunicaciones digitales, pues su costo promedio equivale a la probabilidad de error de decisión. Se define como:

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

El riesgo promedio condicionado para la función $L_{0,1}$ es:

$$R_{0,1}(\theta, \pi) = \mathbb{E}_X(L_{0,1}(\theta, \pi(X)) | \Theta = \theta) = \int_{\mathbb{X}} L_{0,1}(\theta, \pi(x)) f_{X|\Theta}(x|\theta) dx = P_{X|\Theta}(\pi(X) \neq \theta | \theta).$$

Por lo tanto, el riesgo promedio Bayesiano de la regla π bajo el costo 0-1 es la probabilidad de error:

$$r_{0,1}(\pi) = \mathbb{E}_{X,\Theta}\{L_{0,1}(\Theta, \pi(X))\} = \sum_{\theta=1}^k p_{\Theta}(\theta) \cdot R_{0,1}(\theta, \pi) = P_{X,\Theta}(\pi(X) \neq \Theta).$$

Minimizar el riesgo promedio bajo $L_{0,1}$ equivale a minimizar la probabilidad de error. La regla óptima es:

$$\pi_{0,1}^*(x) = \arg \min_{y \in \mathcal{A}} \sum_{\theta \in \mathcal{A}} L_{0,1}(\theta, y) P_{\Theta|X}(\Theta = \theta | x) = \arg \max_{y \in \mathcal{A}} P_{\Theta|X}(\Theta = y | x),$$

es decir, la regla Bayesiana óptima $\pi_{0,1}^*(x)$ corresponde a **maximizar la probabilidad a posteriori** o regla **MAP** (*maximum a posteriori*). Usando la regla de Bayes:

$$\pi_{0,1}^*(x) = \arg \max_{\theta \in \mathcal{A}} P_{\Theta|X}(\Theta = \theta | x) = \arg \max_{\theta \in \mathcal{A}} \frac{f_{\Theta,X}(\theta, x)}{f_X(x)} = \arg \max_{\theta \in \mathcal{A}} f_{X|\Theta}(x|\theta) \cdot p_{\Theta}(\theta).$$

Un caso particular es cuando la distribución a priori es equiprobable, $p_{\Theta}(\theta) = \frac{1}{|\mathcal{A}|}$, entonces:

$$\pi_{0,1}^*(x) = \arg \max_{\theta \in \mathcal{A}} f_{X|\Theta}(x|\theta),$$

que corresponde al criterio de **máxima verosimilitud** o **ML** (*maximum likelihood*).

- **Función de costo cuadrático y estimador MMSE:** Otra función de costo muy utilizada es la función de costo cuadrático $L_{MSE} : \mathcal{A} \times \Delta \rightarrow \mathbb{R}^+ \cup \{0\}$ (con $\Delta \subset \mathbb{R}$) definida como:

$$L_{MSE}(x, y) = (x - y)^2.$$

Al usar esta función, el espacio de parámetro es distinto al de decisión, lo que permite entregar un valor estimado en lugar de decidir entre hipótesis discretas. El riesgo Bayesiano es:

$$\mathbb{E}_{X,\Theta}(L_{MSE}(\Theta, \pi(X))) = \sum_{\theta \in \mathcal{A}} \int_{\mathbb{X}} (\theta - \pi(x))^2 f_{X,\Theta}(x, \theta) dx.$$

Esta ecuación se conoce como el *error cuadrático medio* o *Mean Square Error*. Para encontrar la regla óptima, consideramos el operador de esperanza condicional:

$$\mathbb{E}(\Theta | X = x) = \sum_{\theta \in \mathcal{A}} \theta P_{\Theta|X}(\Theta = \theta | x),$$

que corresponde a la esperanza condicional de Θ dado $X = x$. Descomponiendo la suma:

$$\sum_{\theta \in \mathcal{A}} (\theta - y)^2 P_{\Theta|X}(\Theta = \theta|x) = \sum_{\theta \in \mathcal{A}} (\theta - \mathbb{E}(\Theta|X = x))^2 P_{\Theta|X}(\Theta = \theta|x) + (\mathbb{E}(\Theta|X = x) - y)^2,$$

se observa que:

$$\sum_{\theta \in \mathcal{A}} (\theta - \mathbb{E}(\Theta|X = x))^2 P_{\Theta|X}(\Theta = \theta|x) = \text{Var}(\Theta|X = x)$$

es la varianza condicional de Θ dado $X = x$. Por lo tanto, el argumento que minimiza la expresión es:

$$\pi^*(x) = \arg \min_{y \in \mathcal{A}} \text{Var}(\Theta|X = x) + (\mathbb{E}(\Theta|X = x) - y)^2 = \arg \min_{y \in \mathcal{A}} (\mathbb{E}(\Theta|X = x) - y)^2 = \mathbb{E}(\Theta|X = x).$$

El detector óptimo que minimiza el error cuadrático medio es:

$$\pi_{MMSE}(x) = \mathbb{E}(\Theta|X = x) = \sum_{\theta \in \mathcal{A}} \theta P_{\Theta|X}(\Theta = \theta|x),$$

que es la **esperanza condicional** o la **esperanza de la distribución a posteriori** de Θ dado $X = x$. El riesgo Bayesiano mínimo o error cuadrático medio mínimo (MMSE) está dado por:

$$MMSE = \min_{\phi: \mathbb{X} \rightarrow \mathcal{A}} \mathbb{E}_{X, \Theta}(L(\Theta, \phi(X))) = \mathbb{E}_{\Theta, X}((\Theta - \phi(X))^2) = \int_{\mathbb{X}} \text{Var}(\Theta|X = x) f_X(x) dx = \mathbb{E}(\text{Var}(\Theta|X)),$$

que corresponde al promedio de la varianza condicional.

- **Relación con el Lema de Neyman-Pearson:** En el caso binario $\mathcal{A} = \{0, 1\}$ con una función de costo arbitraria de la forma:

$\mathcal{A} \backslash \mathcal{A}$	0	1
0	l_{00}	l_{01}
1	l_{10}	l_{11}

la regla óptima es, para un $x \in \mathbb{X}$ y aplicando la expresión de decisión punto a punto:

$$\pi^*(x) = \arg \min_{y \in \mathcal{A}} \sum_{\theta \in \mathcal{A}} L(\theta, y) P_{\Theta|X}(\Theta = \theta|x) = \begin{cases} 1 & \text{si } \sum_{\theta \in \mathcal{A}} L(\theta, 0) P_{\Theta|X}(\Theta = \theta|x) > \sum_{\theta \in \mathcal{A}} L(\theta, 1) P_{\Theta|X}(\Theta = \theta|x) \\ 0 & \text{si } \sum_{\theta \in \mathcal{A}} L(\theta, 0) P_{\Theta|X}(\Theta = \theta|x) < \sum_{\theta \in \mathcal{A}} L(\theta, 1) P_{\Theta|X}(\Theta = \theta|x) \\ I & \text{si } \sum_{\theta \in \mathcal{A}} L(\theta, 0) P_{\Theta|X}(\Theta = \theta|x) = \sum_{\theta \in \mathcal{A}} L(\theta, 1) P_{\Theta|X}(\Theta = \theta|x) \end{cases}$$

donde I indica indiferencia. Simplificando las sumas y usando la regla de Bayes, se obtiene:

$$\pi^*(x) = \begin{cases} 1 & \text{si } f_{X|\Theta}(x|\Theta = 1) > \frac{(l_{01} - l_{00})p_{\Theta}(0)}{(l_{10} - l_{11})p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \\ 0 & \text{si } f_{X|\Theta}(x|\Theta = 1) < \frac{(l_{01} - l_{00})p_{\Theta}(0)}{(l_{10} - l_{11})p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \\ I & \text{si } f_{X|\Theta}(x|\Theta = 1) = \frac{(l_{01} - l_{00})p_{\Theta}(0)}{(l_{10} - l_{11})p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \end{cases}$$

Observamos que la regla óptima es una instancia del **test de Neyman-Pearson** y el umbral es $\nu = \frac{(l_{01} - l_{00})p_{\Theta}(0)}{(l_{10} - l_{11})p_{\Theta}(1)}$. La distribución a priori y los costos juegan un rol de ajuste del umbral y establecen un compromiso entre tamaño y poder del test. Para la función de costo $L_{0,1}$ tenemos:

$$\pi_{0,1}^*(x) = \begin{cases} 1 & \text{si } f_{X|\Theta}(x|\Theta = 1) > \frac{p_{\Theta}(0)}{p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \\ 0 & \text{si } f_{X|\Theta}(x|\Theta = 1) < \frac{p_{\Theta}(0)}{p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \\ I & \text{si } f_{X|\Theta}(x|\Theta = 1) = \frac{p_{\Theta}(0)}{p_{\Theta}(1)} \cdot f_{X|\Theta}(x|\Theta = 0) \end{cases}$$

La función de riesgo promedio Bayesiano en este caso equivale a la probabilidad de error:

$$r_{0,1}(\pi) = P_{X,\Theta}(\pi(X) \neq \Theta) = P_{X|\Theta}(\pi(X) = 1|\Theta = 0)p_{\Theta}(0) + P_{X|\Theta}(\pi(X) = 0|\Theta = 1)p_{\Theta}(1).$$

Se observa que la probabilidad de error no es más que la combinación convexa del error de tipo I (α_{π}) y el error de tipo II ($1 - \beta_{\pi}$) donde sus ponderadores están dados por la distribución a priori. La distribución a priori indica cuánto peso o importancia se le dará a cada tipo de error, lo que nuevamente entrega un compromiso entre ambos tipos de errores, similar al test de Neyman-Pearson.

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 σ^2 en la diagonal y ceros fuera de ella):

$$\mathbf{\Sigma}_1 = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{I}_n$$

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 σ^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:

$$\boldsymbol{\mu}_0 = \mathbb{E}[\mathbf{X}_1^n | \Theta = 0] = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

y por otro lado:

$$\mathbf{\Sigma}_0 = \sigma^2 \mathbf{I}_n$$

Podemos verificar que este resultado es idéntico al de $\text{Cov}(\mathbf{X}_1^n, \mathbf{X}_1^n | \Theta = 1)$. Esto se debe a que la matriz de covarianza es invariante frente a traslaciones de la media, ya que mide la dispersión alrededor de la media y no depende del valor absoluto de ésta.

Resolución 1.2

La regla de Máxima Verosimilitud (ML, por sus siglas en inglés *Maximum Likelihood*) decide la hipótesis θ^* que maximiza la función de verosimilitud:

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

Equivalentemente, podemos maximizar la log-verosimilitud. Esto es válido porque la función logaritmo es estrictamente creciente, por lo que preserva el orden: si $f_{X|\Theta}(x|\theta_1) > f_{X|\Theta}(x|\theta_2)$, entonces $\ln f_{X|\Theta}(x|\theta_1) > \ln f_{X|\Theta}(x|\theta_2)$. Por lo tanto, el $\arg \max$ no cambia:

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

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

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

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. Note que en el caso de dos hipótesis, el problema $\arg \max_{\theta \in \{0,1\}} L(\mathbf{x}|\theta)$ es equivalente a determinar cuál log-verosimilitud es mayor mediante una desigualdad: si $L(\mathbf{x}|1) > L(\mathbf{x}|0)$, entonces $\theta^* = 1$; si $L(\mathbf{x}|1) < L(\mathbf{x}|0)$, entonces $\theta^* = 0$. Por lo tanto:

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

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

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

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

Resolución 1.3

Con $p_\Theta(1) = p_\Theta(0) = \frac{1}{2}$, calculamos la probabilidad de error. Nótese que utilizamos la *regla de la probabilidad total* para descomponer la probabilidad de error en función de las hipótesis, combinando las probabilidades condicionales bajo cada escenario (presencia o ausencia de señal) ponderadas por las probabilidades a priori equiprobables:

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

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:

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

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

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

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

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

donde $Z \sim \mathcal{N}(0, 1)$. Usando la simetría de la distribución normal estándar, sabemos que $P(Z \leq -a) = P(Z \geq a) = P(Z > a)$ para cualquier $a > 0$. Por lo tanto:

$$P\left(Z \leq -\frac{\|s\|}{2\sigma}\right) = P\left(Z > \frac{\|s\|}{2\sigma}\right) = Q\left(\frac{\|s\|}{2\sigma}\right)$$

donde $Q(x) = P(Z > x)$ es la función Q.

Para $\Theta = 0$:

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

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 que 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:
- (a) Plantee un espacio de observación y un espacio de decisión adecuado, indique la distribución de la variable aleatoria Θ asociado al espacio de decisión (distribución a priori) y las densidades y/o probabilidades de masa condicionales de X dado $\Theta = \theta$.
 - (b) 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 $\pi(x)$ es:

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

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

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

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:

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

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 δ 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 :

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

Note que en el caso de dos opciones, el problema $\arg \min_{\delta \in \{0,1\}} r(\delta|x)$ es equivalente a determinar cuál riesgo es menor mediante una desigualdad: si $r(1|x) < r(0|x)$, entonces $r^*(x) = 1$; si $r(1|x) > r(0|x)$, entonces $r^*(x) = 0$. Por lo tanto, 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 π 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:

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

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 $r^*(x)$ minimiza $P_{\Theta|X}(\pi(X) \neq \Theta|X = x)$ para cada $x \in \mathbb{X}$, cualquier otra regla $\pi(x)$ que difiera de $r^*(x)$ en algún punto x tendrá un valor mayor o igual de $P_{\Theta|X}(\pi(X) \neq \Theta|X = x)$ en ese punto. Por lo tanto, r^* minimiza la probabilidad de error punto a punto.

Ahora, integrando sobre todo el espacio de observación usando la medida de probabilidad, tenemos que la probabilidad de error total es:

$$P_{X,\Theta}(r^*(X) \neq \Theta) = \int_{\mathbb{X}} P_{\Theta|X}(r^*(x) \neq \Theta|X = x) f_X(x) dx$$

donde usamos la notación $f_X(x) dx$ para denotar la medida de probabilidad (que puede ser una densidad o una masa de probabilidad, según el caso). Como para cada x se tiene que:

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

para cualquier otra regla π , entonces integrando sobre todos los $x \in \mathbb{X}$:

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

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

4. Considere un cuerpo radiactivo que emite θ 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 θ conocidos, X distribuye binomial de parámetros p y θ , 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 θ asumiendo conocido p , pero en un contexto Bayesiano, donde la cantidad de partículas emitidas distribuye Poisson de parámetro λ 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 λp , 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 Θ .

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

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

Dado que ya tenemos la distribución conjunta $P_{X,\Theta}(k, \theta)$ del punto anterior, podemos calcular la distribución condicional:

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

Para $\theta \geq k$:

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

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

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

La regla 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):**

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

La regla 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):**

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

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