

UNIVERSIDAD CENTROCCIDENTAL  
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología  
Licenciatura en Ciencias Matemáticas



“CLASIFICACIÓN DE REGISTROS DE POZOS MEDIANTE  
MÉTODOS BAYESIANOS”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR. ALBERTO DORANTE

COMO REQUISITO FINAL

PARA OBTENER EL TÍTULO DE LICENCIADO

EN CIENCIAS MATEMÁTICAS

ÁREA DE CONOCIMIENTO: ESTADÍSTICA Y PROBABILIDAD.

TUTOR: LCDO.MSC. JHONNY ESCALONA

Barquisimeto, Venezuela. Febrero de 2015



Universidad Centroccidental  
 "Lisandro Alvarado"  
 Decanato de Ciencias y Tecnología  
 Licenciatura en Ciencias Matemáticas



ACTA  
 TRABAJO ESPECIAL DE GRADO

Los suscritos miembros del Jurado designado por el Jefe del Departamento de Matemáticas del Decanato de Ciencias y Tecnología de la Universidad Centroccidental "Lisandro Alvarado", para examinar y dictar el veredicto sobre el Trabajo Especial de Grado titulado:

“CLASIFICACIÓN DE REGISTROS DE POZOS MEDIANTE MÉTODOS BAYESIANOS”

presentado por el ciudadano BR. ALBERTO DORANTE titular de la Cédula de Identidad No. 17860147, con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciado en Ciencias Matemáticas.

Luego de realizada la Defensa y en los términos que imponen los Lineamientos para el Trabajo Especial de Grado de la Licenciatura en Ciencias Matemáticas, se procedió a discutirlo con el interesado habiéndose emitido el veredicto que a continuación se expresa:

1 \_\_\_\_\_

Con una calificación de \_\_\_\_\_ puntos.

En fe de lo expuesto firmamos la presente Acta en la Ciudad de Barquisimeto a los \_\_\_\_ días del mes de \_\_\_\_\_ de \_\_\_\_\_.

\_\_\_\_\_  
 TUTOR

\_\_\_\_\_  
 FIRMA

\_\_\_\_\_  
 PRINCIPAL

\_\_\_\_\_  
 FIRMA

\_\_\_\_\_  
 PRINCIPAL

\_\_\_\_\_  
 FIRMA

OBSERVACIONES:

---



---



---

<sup>1</sup> Aprobado ó Reprobado

*A Dios, mis padres y hermana.*

# AGRADECIMIENTOS

Quiero comenzar estas líneas agradeciendo a Dios por concederme la vida y llenarla de grandes bendiciones.

Quién supo guiarme por el buen camino, darme fuerzas para seguir adelante y no desmayar en los problemas que se presentaban, enseñándome a encarar las adversidades sin perder nunca la dignidad ni desfallecer en el intento.

A mis padres, por su apoyo, consejos, comprensión y ayuda en los momentos difíciles. Por brindarme los recursos necesarios para estudiar. Me han dado todo lo que soy como persona, mis valores, mis principios, mi carácter, mi empeño, mi perseverancia, mi coraje para conseguir mis objetivos, por brindarme su amor incondicional, este logro es de ustedes.

A mi hermana, Diana. Por ser tan especial y estar conmigo siempre, eres de las personas más importantes para mí.

A mi Abuela Juana, que con la sabiduría de Dios me has enseñado a ser quien soy hoy. Gracias por tu paciencia, por tus consejos, por el amor que me has dado y por tu apoyo incondicional en mi vida.

A mis amigos Chabela, Steffany, y Victor por apoyarme siempre.

A mis profesores, quienes me formaron y guiaron, ofrecieron sus consejos y apoyo.

A mi tutor y profesor, Jhonny Escalona; por sus enseñanzas, dedicación, confianza.

Gracias a todos aquellos que no están aquí, pero que me ayudaron a que este gran esfuerzo se volviera realidad.

# **CLASIFICACIÓN DE REGISTROS DE POZOS MEDIANTE MÉTODOS BAYESIANOS**

Br. Alberto Rafael Dorante Cordero

## **RESUMEN:**

En este trabajo se estudio el problema de clasificación registros de pozos a partir de muestras etiquetadas, es decir, dado un registro de pozo, se determino su tipo suelo (litología), para ello se utilizo métodos de estadística bayesiana.

# ÍNDICE

Agradecimientos	i
Resumen	iii
Introducción	1
<b>1. DESCRIPCIÓN DEL PROBLEMA</b>	<b>3</b>
1.1. Descripción del Problema . . . . .	3
1.2. Tablas de las Clases de Litologías Disponibles . . . . .	4
1.3. Gráficas de los Registros Litológicos . . . . .	5
<b>2. REGLA DE DECISIÓN DE BAYES</b>	<b>7</b>
2.1. Probabilidad Condicional . . . . .	7
2.2. Teorema de Bayes . . . . .	7
2.3. Regla de Decisión de Bayes . . . . .	8
<b>3. Métodos Monte Carlo con Cadenas de Markov</b>	<b>11</b>
3.1. Cadenas de Markov . . . . .	11
3.2. Algoritmo Metropolis-Hastings . . . . .	16
<b>4. Resultados</b>	<b>19</b>
4.1. Modelo . . . . .	19
Conclusiones	22
Referencias Bibliográficas	23

# ÍNDICE DE TABLAS

1.1. Registros comunes de los pozos MAR402 y MAR203 . . . . .	4
1.2. Litologías Disponibles . . . . .	4
4.1. Resultados para la Estimación de Parámetros . . . . .	20
4.2. Litologías Fusionadas . . . . .	21

# Índice de figuras

1.1. Registro GR . . . . .	5
1.2. Registro PEF . . . . .	5
1.3. Registro RHOB . . . . .	5
1.4. Gráfica de dispersión de los datos con 6 litologías. . . . .	6



# INTRODUCCIÓN

En el campo de la geología y de la industria petrolera es de gran importancia identificar la composición litológica del subsuelo lo que permite una mejor explotación del mismo en cuanto a rendimiento y costo. La composición del subsuelo es determinada por los geólogos por medio de sondas que son introducidas en los pozos perforados y que permiten medir las propiedades físicas de las rocas, que son llamadas “registros”, sin embargo, la calidad de esta medida es distorsionada por dificultades técnicas como pérdida del núcleo; el núcleo es la columna de roca por medio del cual se establecen las diferentes clases litológicas presentes en el subsuelo y que a menudo no se recupera en buen estado, esto dificulta la medida de la composición litológica del subsuelo.

Padrón et al. (ver [6]) propusieron una nueva técnica para identificar litologías usando registros de pozo, a saber, los modelos de Markov ocultos. En este trabajo, se pretende estudiar el problema de clasificación de registros utilizando métodos Bayesianos basados en cadenas de Markov siguiendo [5]; con este tipo de método también se puede procesar sucesiones de patrones.

Los métodos Bayesianos son reconocidos como una manera coherente de hacer inferencia, en contraste con los métodos clásicos, donde los datos obtenidos de estudios observacionales y/o experimentales son analizados con modelos que dependen del tipo de datos con procedimientos de inferencia y decisión particulares para cada caso. El análisis Bayesiano trata de una manera unificada la inferencia y la decisión, tomando en consideración la incertidumbre asociada al modelo y a los parámetros, proporcionando de una vez las herramientas para cuantificar esta incertidumbre. Por otra parte, el tratamiento de las cantidades no observadas como variables aleatorias y el análisis condicional, permiten naturalmente considerar modelos jerárquicos o de variables latentes que son difíciles o imposibles de manejar con la estadística clásica.

El presente trabajo de grado está estructurado en capítulos. En el capítulo I se presenta la descripción del problema, el capítulo II contiene la regla de desicisión de Bayes, el capítulo III muestra las cadenas de Markov y el algoritmo Metropolis-Hastings y finalmente en el capítulo IV se explica los resultados.

# CAPÍTULO 1

## DESCRIPCIÓN DEL PROBLEMA

### 1.1. Descripción del Problema

Para investigar la composición del subsuelo, geólogos introducen sondas en los pozos perforados y miden las propiedades físicas de las rocas. A estas medidas se les llama registros (ver [5]). La calidad de estas medidas es distorsionada por dificultades técnicas. A menudo no se recuperan en buen estado la columna de roca, denominada núcleo, que los geólogos quieren estudiar. A partir del núcleo, expertos establecen las diferentes clases litológicas presentes en el subsuelo.

Para este trabajo se tienen datos de dos pozos con nombres MAR203 y MAR402, los cuales fueron perforados en la localidad de Marcoule al sur-este de Francia. En el pozo MAR203 se tienen 5590 medidas de 19 registros a distintas profundidades y en el pozo MAR402, se tienen 9962 medidas de 24 registros también a distintas profundidades. Las tablas 1.1 y 1.2 muestran los registros comunes de los pozos y las clases de litologías disponibles para este trabajo. Estas litologías fueron obtenidas por una red de Kohonen, validadas por un experto, (ver [6]). Como resultado, se dispone de muestras  $(x_i, w_i)$ , donde  $x_i$  es el vector de registros y  $w_i$  la etiqueta de la clase de litología a la cual pertenece. Para otra muestra  $\hat{x}_i$  de registros, el problema de clasificación de registro de pozos consiste en determinar la etiqueta  $\hat{w}_i$  de  $\hat{x}_i$ . Para este fin, se propone el uso de un método estadístico basado en modelos de aprendizaje bayesiano.

## 1.2. Tablas de las Clases de Litologías Disponibles

Registros		Registros	
1	<i>ILD</i>	11	<i>TURA</i>
2	<i>ILM</i>	12	<i>TPRA</i>
3	<i>SFLU</i>	13	<i>UPRA</i>
4	<i>SP</i>	14	<i>RHOB</i>
5	<i>GR</i>	15	<i>PEF</i>
6	<i>SGR</i>	16	<i>CALI</i>
7	<i>CGR</i>	17	<i>NPFI</i>
8	<i>URAN</i>	18	<i>ITT</i>
9	<i>POTA</i>	19	<i>DT4P</i>
10	<i>THOR</i>		

TABLA 1.1: REGISTROS COMUNES DE LOS POZOS MAR402 Y MAR203

Clase	Litología	Clase	Litología
1	Caliza	7	Otros Limos
2	Marga	8	Reniscas Gruesas
3	Arenisca Conglomerado	9	Arenas Puras
4	Arcilla	10	Calizas Gruesas
5	Otras Arcillas	11	Brechas Gruesas
6	Limos	12	Carbón

TABLA 1.2: LITOLOGÍAS DISPONIBLES

### 1.3. Gráficas de los Registros Litológicos

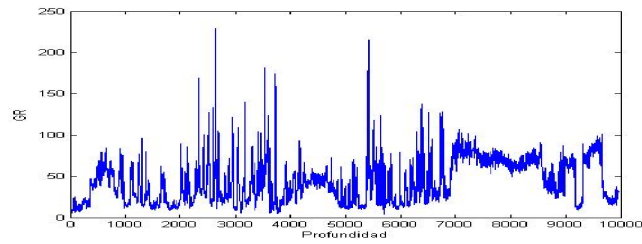


Figura 1.1: Registro GR

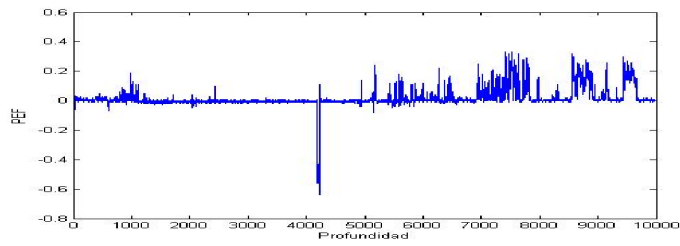


Figura 1.2: Registro PEF

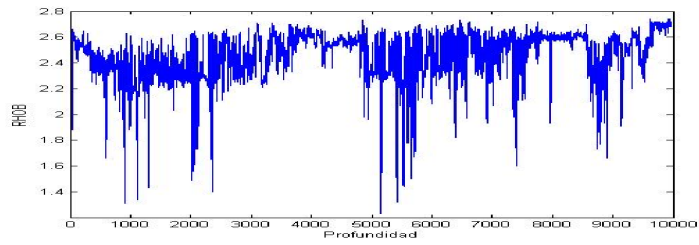
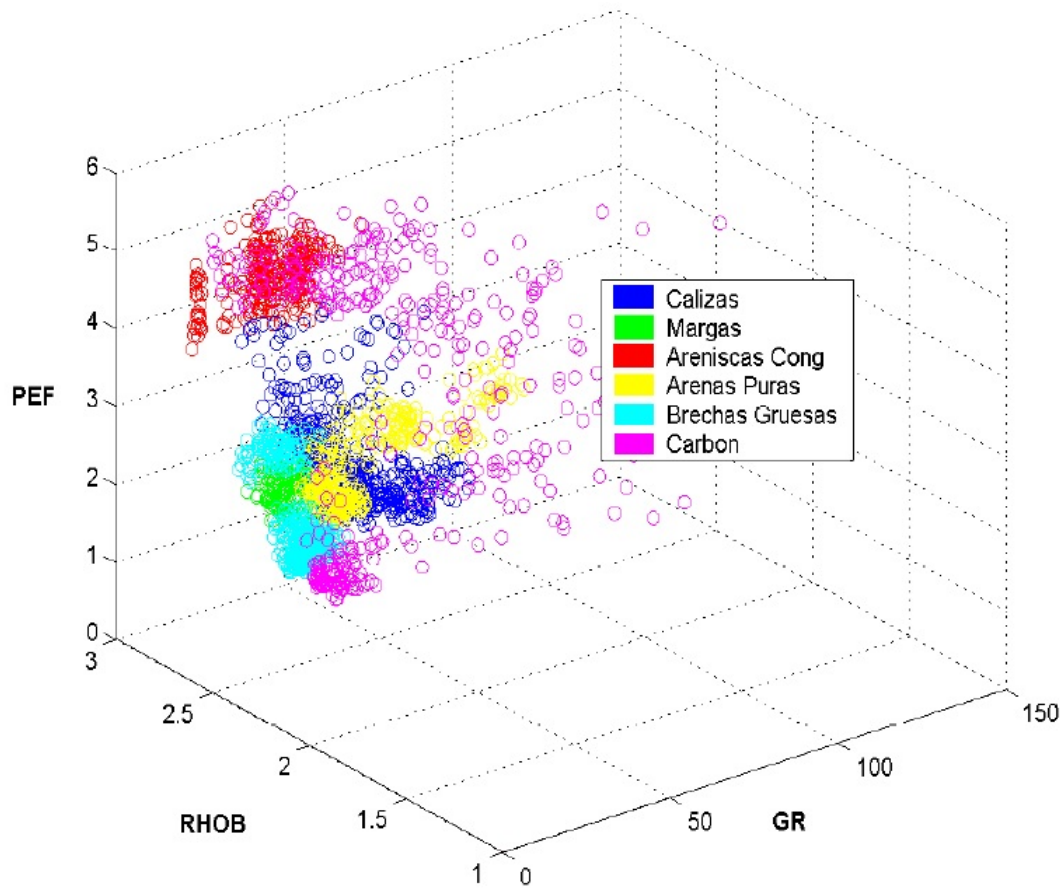


Figura 1.3: Registro RHOB



**Figura 1.4:** Gráfica de dispersión de los datos con 6 litologías.

La figura 1.4 nos da una visión de la complejidad del problema, en la gráfica se muestran solo seis litologías y los registros GR, RHOB y PEF.

# CAPÍTULO 2

## REGLA DE DECISIÓN DE BAYES

Antes de enunciar la regla de decisión de Bayes se dará la definición de probabilidad condicional y el teorema de Bayes, ya que estas serán de gran utilidad en el desarrollo de este trabajo.

### 2.1. Probabilidad Condicional

**Definición 2.1.** Dado un espacio de probabilidad  $(\Omega, \mathfrak{F}, \mathfrak{P})$  y dos sucesos  $A, B \in \mathfrak{F}$  con  $P(B) > 0$ , la probabilidad condicional de A dado B está dada como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

### 2.2. Teorema de Bayes

En la teoría de la probabilidad el teorema de Bayes es un resultado enunciado por Thomas Bayes en 1763 que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

**Teorema 2.1.** Sean  $A_0, A_1, \dots, A_N$ , eventos mutuamente excluyentes y exhaustivos, tales que la probabilidad de cada uno de ellos es distinta de cero (0). Sea B un suceso cualquiera del que se conocen las probabilidades condicionales  $P(B|A_i)$ . Entonces, la probabilidad  $P(A_i|B)$  viene dada por la expresión:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

donde:  $P(A_i)$  son las probabilidades a priori.

$P(B|A_i)$  es la probabilidad de  $B$  en la hipótesis  $A_i$ .

$P(A_i|B)$  son las probabilidades a posteriori.

### 2.3. Regla de Decisión de Bayes

El aprendizaje bayesiano nos da un enfoque probabilístico de la inferencia, el cual está basado en suponer que los parámetros de interés siguen distribuciones de probabilidad, con la que se puede conseguir una solución óptima. Por medio de estas distribuciones y datos observados, dando así la posibilidad de realizar una ponderación de la probabilidad de ocurrencia de una hipótesis de manera cuantitativa. Los algoritmos de aprendizaje bayesiano pueden calcular probabilidades explícitas para cada hipótesis. También nos proporcionan un marco para estudiar otros algoritmos de aprendizaje.

En general, en el aprendizaje bayesiano se supone que cada clase posee una función de distribución conocida de la forma  $p(x|\theta_i)$  parametrizada por un vector  $\theta_i$ . La distribución a priori de  $\theta_i$  se denota por  $p(\theta_i)$ , esta representa un conocimiento inicial acerca de  $\theta_i$ . Supongamos que se desea estimar el parámetro  $\theta_i$  a partir de un conjunto  $\chi_i = \{x_1^i, \dots, x_{n_i}^i\}$ <sup>1</sup>. Asumamos que el conjunto  $\chi_i$  (correspondiente a la clase  $w_i$ ) no da ninguna información sobre  $p(\theta_j)$ ,  $j \neq i$ . Esta condición denota separabilidad y nos permite trabajar con cada clase por separado. Por último, todas las clases son a priori equiprobables y las muestras  $x_j^i$  de  $\chi_i$  se seleccionan independientemente.

---

<sup>1</sup> El conjunto  $\chi_i = \{x_1^i, \dots, x_{n_i}^i\}$  se le llama muestras de diseño y está disponible para cada clase  $w_i$  con  $i = 1, \dots, c$ .



Al elegir una función de pérdida cero-uno, el clasificador bayesiano clasifica  $x$  en la clase  $w_{i_0}$  tal que:

$$\forall i \neq i_0 \quad P(w_{i_0}|x, \chi_{1,..c}) \geq P(w_i|x, \chi_{1,..c}) \quad (2.1)$$

Para esto se requiere el cálculo de  $P(w_i|x, \chi_{1,..c})$  para cada clase. La regla de Bayes expresa  $P(w_i|x, \chi_{1,..c})$  como una función de las clases a priori y la función de distribución de la siguiente manera :

$$P(w_i|x, \chi_{1,..c}) = \frac{p(x|w_i, \chi_{1,..c})P(w_i|\chi_{1,..c})}{\sum_{i=1}^c p(x|w_i, \chi_{1,..c})P(w_i|\chi_{1,..c})} \quad (2.2)$$

Luego por la condición de separabilidad (ver [3] p. 50) la ecuación (2.2) se reduce a:

$$p(x|w_i, \chi_{1,..c}) = p(x|w_i, \chi_i) = p(x|\chi_i)$$

Para clases equiprobables, la regla de decisión de Bayes puede ser escrita como:

$$\text{Decide } x \in w_{i_0} \iff p(x|\chi_{i_0}) \geq p(x|\chi_i), \quad \forall i \neq i_0 \quad (2.3)$$

La condición de separabilidad asegura la independencia de cada clase, lo que nos permite simplificar las notaciones.

Por otro lado, la función verosimilitud marginal o distribución marginal de los datos, se puede calcular de la siguiente manera:

$$p(x|\chi) = \int p(x|\theta)p(\theta|\chi)d\theta \quad (2.4)$$

donde la distribución a posterior  $p(\theta|\chi)$  se puede escribir como:

$$p(\theta|\chi) \propto \prod_{j=1}^n p(x_j|\theta)p(\theta) \quad (2.5)$$

En general, no existe una expresión cerrada para la integral en la ecuación (2.4). En el siguiente capítulo estudiaremos técnicas que nos permitan aproximar dicha integral.

La ecuación (2.4) muestra que la función de distribución  $p(x|\chi)$  puede ser vista como la media de  $g(\theta) = p(x|\theta)$ , donde  $\theta$  se distribuye de acuerdo a la densidad posteriori  $p(\theta|\chi)$ . Basado en esta observación, este trabajo se propone generar muestras  $\theta_1, \dots, \theta_N$  distribuidos según  $p(\theta|\chi)$  y estimar  $p(x|\chi)$  a través de promedio ergódico como sigue:

$$\hat{p}_N(x|\chi) = \frac{1}{N} \sum_{j=1}^N p(x|\theta_j) \quad (2.6)$$

Las muestras  $\theta_1, \dots, \theta_N$  se obtienen como resultados de una cadena de Markov homogénea ergódica, cuyo objetivo es la distribución  $p(\theta|\chi)$  utilizando el algoritmo de Metropolis-Hastings, el cual se explica más adelante.

# CAPÍTULO 3

## MÉTODOS MONTE CARLO CON CADENAS DE MARKOV

Antes de presentar el algoritmo Metropolis-Hastings y los métodos de Monte Carlo con Cadenas de Markov (MCMC), exponemos algunos resultados necesarios relativos a las cadenas de Markov.

### 3.1. Cadenas de Markov

**Definición 3.1.** Un proceso estocástico es una familia de variables aleatorias,  $X_t$ , donde  $t \in T$ , y  $T$  es un conjunto adecuado de índices.

**Definición 3.2.** Un proceso de Markov  $\{X_t\}$  es un proceso estocástico con la siguiente propiedad: dado el valor de  $X_t$ , los valores de  $X_s$  con  $s > t$  no están influenciados por los valores de  $X_u$  con  $u < t$ . Es decir, la probabilidad de cualquier comportamiento futuro del proceso, cuando el estado actual se conoce con exactitud, no se ve alterada por el conocimiento de su comportamiento en el pasado. **Una cadena de Markov de tiempo discreto**, es un proceso de Markov cuyo espacio-estado (rango de posibles valores de las variables aleatorias) es finito o un conjunto numerable, y donde el conjunto de índices es  $T = (0, 1, 2, 3, \dots)$ .

En términos formales, la propiedad de Markov es la siguiente:

$$Pr \{X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} = Pr \{X_{n+1} = j | X_n = i\}, \quad (3.1)$$

para todo tiempo  $n$  y todos los estados  $i_0, i_1, \dots, i_{n-1}, i, j$ .

**Observación 3.1.** En una cadena de Markov, interpretamos a  $X_n$  como el “estado del sistema en el instante  $n$ ”.

La probabilidad de que  $X_{n+1}$  este en el estado  $j$  dado que  $X_n$  está en el estado  $i$  es llamada **la probabilidad de transición de un paso** y es denotado por  $P_{ij}^{n,n+1}$ , esto es:

$$P_{ij}^{n,n+1} = Pr \{X_{n+1} = j | X_n = i\}.$$

Esta notación hace hincapié en que en general la probabilidad de transición no solo es función del estado inicial y final, si no también del tiempo de transición. Cuando la probabilidad de transición de los pasos son independientes de la variable de tiempo  $n$ , se dice que la cadena de Markov es **homogénea**. Como la mayoría de las cadenas de Markov que se encuentran poseen esta propiedad, pondremos nuestro interés en este caso. Así, tenemos que  $P_{ij}^{n,n+1} = P_{ij}$ , donde  $P_{ij}$  es la probabilidad condicional de que el valor del estado  $i$  se mueva al estado  $j$  en un paso.

**Observación 3.2.** Muchas veces es conveniente etiquetar el espacio-estado de una cadena de Markov con los enteros no-negativos  $\{0, 1, 2, \dots\}$ .

**Definición 3.3.** Se define la matriz de probabilidades transición o matriz de Markov, como la matriz  $\mathbf{P}$  cuyo elemento  $(i, j)$  es  $P_{ij}$ , es decir, la matriz en cuyo elemento  $(i, j)$  se encuentra la probabilidad de pasar del estado  $i$  al estado  $j$ .

Así, la matriz  $\mathbf{P}$  es como sigue:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Si el número de estados es finito, entonces  $\mathbf{P}$  es una matriz cuadrada finita, cuyo orden (el número de filas) es igual al número de estados. Claramente, las cantidades  $P_{ij}$  satisfacen las condiciones:

$$P_{ij} \geq 0, \quad \forall i, j = 0, 1, 2, \dots \quad (3.2)$$

$$\sum_{j=0}^{\infty} P_{ij} = 1, \quad \forall i, j = 0, 1, 2, \dots \quad (3.3)$$

La condición (3.3) simplemente indica que alguna transición se produce en cada ensayo (por conveniencia se dice que la transición se ha producido incluso si el estado se mantiene sin cambios).

**Ejemplo 1.** Después de algunos estudios sobre el clima en determinada ciudad, se ha observado que si un día está soleado en el 70 % de los casos el día siguiente continúa soleado y en el 30 % nublado. Por el contrario, si un día está nublado, la probabilidad de que esté soleado el día siguiente es de 60 % y la probabilidad de que se siga nublado es de 40 %. Además, según estos estudios el clima de un determinado día solo depende del clima del día anterior.

Para estudiar este problema primeramente encontramos las probabilidades de transición, es decir las probabilidades de que teniendo cierto clima un día, al día siguiente se tenga otro o el mismo clima. Así, si indicamos con  $s$  a un día soleado y con  $n$  a un día nublado, tendremos:

$$P_{ss} = 0,7,$$

$$P_{sn} = 0,3,$$

$$P_{ns} = 0,6,$$

$$P_{nn} = 0,4.$$

Si acomodamos estos datos en una matriz, obtenemos la matriz de probabilidades de transición:

$$\mathbf{P} = \begin{bmatrix} 0,7 & 0,3 \\ 0,6 & 0,4 \end{bmatrix},$$

donde las filas indican el clima en el día, fila 1 soleado y fila 2 nublado, y las columnas el clima en el día siguiente, columna 1 soleado y columna 2 nublado.

Observamos que en esta matriz no sólo los coeficientes son no-negativos, sino que al sumarlos por filas obtenemos 1. Este es un ejemplo de cadena de Markov ya que:

1. Tenemos ciertos estados, en este caso  $s$  y  $n$ .
2. En cada momento estamos en uno de estos estados.
3. En el próximo momento volveremos a estar en ese u otro estado.
4. Pasamos de un estado a otro con cierta probabilidad, que sólo puede depender del estado inmediatamente anterior y esa probabilidad no cambia con el transcurso del tiempo.

Un proceso de Markov es completamente definido una vez que la matriz de probabilidades de transición y el estado inicial  $X_0$  (de forma más general, la distribución de probabilidad de  $X_0$ ) están especificados. Probemos este hecho:

Sea  $Pr \{X_0 = i_0\} = p_{i_0}$ . Es suficiente mostrar como calcular las cantidades:

$$Pr \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\}, \quad (3.4)$$

ya que cualquier probabilidad que involucra  $X_{j_1}, \dots, X_{j_k}$ , para  $j_1 < \dots < j_k$ , puede ser obtenida, acorde al axioma de probabilidad total, por la suma de términos de la forma (3.4).

Por definición de probabilidad condicional:

$$\begin{aligned} & Pr \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\} \\ &= Pr \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_{n-1} = i_{n-1}\} \\ &\times Pr \{X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\}. \end{aligned} \quad (3.5)$$

Ahora, por definición de cadena de Markov,

$$\begin{aligned} Pr \{X_n = i_n | X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_{n-1} = i_{n-1}\} &= Pr \{X_n = i_n | X_{n-1} = i_{n-1}\} \\ &= P_{i_{n-1}i_n}. \end{aligned} \quad (3.6)$$

Sustituyendo (3.6) en (3.5):

$$\begin{aligned} & Pr \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\} \\ &= Pr \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_{n-1} = i_{n-1}\} P_{i_{n-1}i_n}. \end{aligned} \quad (3.7)$$

Al repetir el proceso  $n - 1$  veces, (3.4) se convierte en:

$$Pr \{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = p_{i_0} P_{i_0 i_1} \cdots P_{i_{n-2} i_{n-1}} P_{i_{n-1} i_n}.$$

Esto demuestra que toda probabilidad finita dimensional es especificada una vez que la matriz de probabilidades de transición y la distribución del estado inicial son dados, y el proceso es definido por estas cantidades.

Por otro lado, sea  $F$  la distribución condicional de  $X_k$  dado  $X_0$ , para algún  $k$ ; se puede demostrar que bajo ciertas condiciones esta distribución converge a una única distribución  $\Pi$ , la cual llamamos **distribución estacionaria o invariante**.

Además, bajo algunas condiciones sobre la distribución condicional  $F$ , se puede mostrar que una cadena de Markov con distribución estacionaria  $\pi$  es **ergódica**, esto es:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(X_k) = \mathbb{E}_{\Pi}[h(x)],$$

donde  $h$  es alguna función que se desea estimar.

Esto nos permite estimar valores esperados a partir de los resultados de una realización de la cadena. De esta manera, construyendo una cadena de Markov cuya distribución estacionaria corresponda a la distribución posterior, se puede encontrar la esperanza posterior de cualquier función, simulando la cadena y tomando el promedio de la función sobre los valores simulados. Por supuesto, este procedimiento descansa en los resultados asintóticos (versión vectorial de la ley fuerte de los grandes números), por lo que se hace necesario realizar un número grande de simulaciones y estudiar la convergencia de los valores obtenidos.

Así, el objetivo de los métodos Monte Carlo con cadenas de Markov es encontrar una cadena de Markov en el espacio de parámetros, de manera tal que la distribución estacionaria de la cadena coincida con la distribución posterior.

### 3.2. Algoritmo Metropolis-Hastings

La referencia al algoritmo de Metropolis-Hastings corresponde a un término general que se utiliza para una familia de métodos de simulación, compuesta de algoritmos universales que generan cadenas de Markov con distribución estacionaria que corresponde al posterior de interés, estos se derivan del algoritmo de Metrópolis.

El algoritmo de Metropolis, es una modificación de un paseo al azar que utiliza una regla de aceptación-rechazo para obtener convergencia de la cadena a una distribución específica. Este algoritmo consiste de los siguientes pasos:

1. Simular un punto inicial para el cual  $\pi(\theta^0|Z) > 0$ , a partir de una distribución inicial  $p_0(\theta)$ .

Para  $s = 1, \dots, n_{sim}$

2. Obtener una realización candidata  $\vartheta$  a partir de una distribución de salto en el tiempo  $s$ ,  $J_s(\vartheta|\theta^{(s-1)})$ . Esta distribución debe ser simétrica en el sentido de que  $J_s(\theta^{(a)}|\theta^{(b)}) = J_s(\theta^{(b)}|\theta^{(a)})$ , para todo  $\theta^{(a)}, \theta^{(b)}, s$ .

3. Calcular:

$$r = \frac{\pi(\vartheta|Z)}{\pi(\theta^{(s-1)}|Z)}.$$

4. Hacer:

$$\theta^{(s)} = \begin{cases} \vartheta, & \text{con probabilidad } \min\{1, r\} \\ \theta^{(s-1)}, & \text{en otro caso} \end{cases} \quad (3.8)$$

La regla de aceptación y rechazo del algoritmo anterior se puede interpretar como sigue: si el “salto” produce un valor para el que se aumenta la densidad posterior, hacer  $\theta^{(s)} = \vartheta$ ; si el “salto” no aumenta la densidad a posteriori, hacer  $\theta^{(s)} = \vartheta$  con probabilidad  $r$  y  $\theta^{(s)} = \theta^{(s-1)}$ , si no. Esto puede ser visto como una versión estocástica de un



algoritmo de búsqueda de moda por pasos.

Esta primera versión del algoritmo fue construida por Metropolis (1949) y más tarde, es generalizada por Hastings (1970), cuya generalización consiste esencialmente en que las reglas de salto, dadas por  $J_s$  no necesitan ser simétricas y el radio  $r$  es reemplazado por,

$$r = \frac{\pi(\vartheta|Z) / J_s(\vartheta|\theta^{(s-1)})}{\pi(\theta^{(s-1)}|Z) / J_s(\theta^{(s-1)}|\vartheta)}.$$

Al ser este algoritmo de aceptación-rechazo, la eficiencia en la generación de la cadena dependerá de las propiedades de la distribución de salto; una buena distribución de salto debería cumplir con las siguientes propiedades:

- Para cualquier  $\theta$ , es fácil muestrear de  $J(\vartheta|\theta)$ .
- Es fácil calcular los cocientes de importancia  $r$ .
- Cada salto produce resultados a una distancia razonable en el espacio de parámetros.
- Los saltos no son rechazados muy frecuentemente.

En esta sección presentamos el algoritmo RWM (Random Walk Metrópolis) el cual pertenece a la clase más general de los algoritmos Metropolis-Hastings, en la cual es utilizada como distribución de salto, la distribución normal y donde  $h$  es la función que se desea estimar.

### Algoritmo Random Walk Metrópolis

#### Inicialización:

- Denote la moda del posterior por  $\tilde{\theta}$ .
- Sea  $\tilde{\Sigma}$  la estimación de la matriz de covarianza de  $\tilde{\theta}$ .

1. Generar  $\theta^{(0)}$  desde  $N(\tilde{\theta}, c^2\tilde{\Sigma})$  o dar un valor inicial específico.

Para  $s = 1, \dots, n_{sim}$ .

2. Generar  $\vartheta$  desde  $N(\theta^{(s-1)}, c^2\tilde{\Sigma})$ .

3. Calcular:

$$r(\theta^{(s-1)}, \vartheta | Z) = \frac{\mathcal{L}_{\vartheta}(Z) \pi(\vartheta)}{\mathcal{L}_{\theta^{(s-1)}}(Z) \pi(\theta^{(s-1)})}.$$

4. Con probabilidad  $\min\{1, r(\theta^{(s-1)}, \vartheta | Z)\}$  aceptar  $\vartheta$  y hacer  $\theta^{(s)} = \vartheta$ , en otro caso, rechazar  $\vartheta$  y hacer  $\theta^{(s)} = \theta^{(s-1)}$ .

5. Aproximar el valor esperado del posterior de la función  $h(\theta)$  por:

$$\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} h(\theta^{(s)}).$$

**Observación 3.3.** El escalar  $c$  utilizado en el algoritmo, permite re-escalar la matriz  $\tilde{\Sigma}$  de tal modo que los vectores generados  $\vartheta$ , sean aceptados en la mayoría de los casos, lo cual es esperado en este algoritmo.

# CAPÍTULO 4

## RESULTADOS

En este capítulo, mostramos los resultados obtenidos para la selección de los mejores registros y también mostramos la clasificación obtenida usando mezclas de gaussianas.

### 4.1. Modelo

Para la implementación del modelo se usó el software matemático MatLab versión (7.12.0), este es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M) y servicio de especie. Fue creado por el matemático y programador de computadoras Cleve Moler en 1984.

Nosotros modelamos los datos usando mezclas de gaussianas para aproximar la distribución de las observaciones, el cual está representada por los registros, esto es, cada observación  $\chi_t$  a tiempo  $t$  es emitida por la litología  $j$  con distribución:

$$p(\chi_t|\theta_j) = \sum_{m=1}^M c_{jm} \mathcal{N}(\chi_t, \mu_{jm}, \Sigma_{jm})$$

donde,  $\theta_j = \{c_{jm}, \mu_{jm}, \Sigma_{jm}\}$ ,  $c_{jm}$  son los coeficientes de la mezcla,  $\mu_{jm}$  los vectores de las medias y  $\Sigma_{jm}$  las matrices de covarianzas. Los coeficientes  $c_{jm}$  satisfacen las siguientes restricciones estocásticas:

$$c_{jm} \geq 0$$

y

$$\sum_{m=1}^M c_{jm} = 1$$

Inicialmente se considero  $\theta_0$ , utilizamos  $M = 1$ , es decir, una Gaussina y realizamos una corrida del Metropolis-Hastings para cada clase de los datos del pozo MAR402, con la cadena obtenida calculamos la ecuación (2.6).

A continuación se muestran los resultados de la clasificación para algunos de estos registros y los resultados para los registros GR, RHOB y PEF (los registros comúnmente usados por los geólogos).

$C$	1	2	3	4	5	6	7	8	9	10	11	12	$T$	$I$	%
1	291	1	0	0	0	0	0	0	0	11	0	2	305	36	95.41
2	15	144	0	16	1	0	0	0	0	2	0	0	178	70	80.90
3	0	0	199	3	0	37	36	28	2	4	29	0	338	22	58.88
4	0	49	0	1155	58	57	35	0	0	0	0	0	1354	24	85.30
5	0	8	0	164	556	47	19	0	0	0	0	3	797	25	69.76
6	0	4	16	6	1	237	84	18	0	1	0	2	369	45	64.23
7	0	0	2	0	47	95	205	0	0	0	0	0	349	41	58.73
8	1	0	16	0	0	1	0	298	28	24	13	10	391	44	76.21
9	0	0	1	0	0	0	0	6	431	15	20	76	549	46	78.51
10	7	3	16	1	0	5	0	66	33	290	2	8	431	39	67.29
11	0	0	0	0	0	0	0	6	25	0	99	0	130	52	76.15
12	35	6	0	0	14	19	0	25	2	0	0	298	399	26	74.69

TABLA 4.1: RESULTADOS PARA LA ESTIMACIÓN DE PARÁMETROS

La tabla 4.1 muestra los resultados usando el pozo MAR203 y los registros GR, RHOB y PEF para la clasificación, en ella se encuentran las clases, el número total de secuencias en cada litología, el porcentaje de correcta clasificación y (I) el es el número de iteraciones.

Los registros utilizados para la tabla 4.1 fueron GR, RHOB y PEF. En estos resultados observamos que los modelos de clasificación confunden cierto tipos de litologías, por ejemplo las clases 4 y 5 correspondientes a Arcillas y otras Arcillas son confundidas, las clases 6 y 7 (Limos y otros Limos) son confundidas y también las clases 8 y 9 (Areniscas gruesas y arenas puras) son confundidas. El porcentaje promedio de correcta clasificación fue de 73.84. Por otra parte, debido a que estos modelos de clasificación tienden a confundir ciertos tipos de litologías ya que se observo un gran solapamiento

entre estas, se fusiono estas litologías en un solo tipo, es decir ahora tendremos 10 clases, las cuales se muestran en la tabla 4.2.

Clase	Litología	Clase	Litología
1	Caliza	6	Areniscas Gruesas
2	Marga	7	Arenas Puras
3	Arenisca Conglomerado	8	Calizas Gruesas
4	Arcilla y otras Arcillas (Arcillas)	9	Brechas Gruesas
5	Limos y Otros Limos(Limos)	10	Carbón

TABLA 4.2: LITOLOGÍAS FUSIONADAS

Los resultados de la clasificación para 10 clases se encuentran en la siguientes tablas:

<i>Clase</i>	1	2	3	4	5	6	7	8	9	10	<i>Total</i>	%
1	297	1	0	0	0	1	0	4	0	2	305	97.38
2	15	151	0	9	0	0	0	2	0	1	178	84.83
3	0	0	217	4	48	29	2	1	36	1	338	64.20
4	0	85	3	1970	91	0	0	0	0	2	2151	91.59
5	0	6	71	74	544	19	0	2	0	2	718	75.77
6	1	0	20	0	13	307	12	19	8	11	391	78.52
7	0	0	2	0	0	19	447	22	20	39	549	81.42
8	14	3	16	0	1	83	21	279	3	11	431	64.73
9	0	0	0	0	0	8	18	0	104	0	130	80.00
10	26	11	0	14	28	23	1	0	0	296	399	74.19

Resultados obtenidos luego de 100 corridas del algoritmo Metropolis-Hastings, los porcentajes promedios de clasificación en cada clase se muestran a continuación:

<i>Clase</i>	1	2	3	4	5	6	7	8	9	10
%	98.36	90.44	79.88	68.34	75.77	78.52	81.42	64.73	80.00	65.16

# Conclusiones

En conclusión, los resultados confirman la efectividad de aplicar los métodos de aprendizaje bayesiano a problemas del campo de la litología y de la industria petrolera. El clasificador bayesiano es óptimo respecto al criterio de error bayesiano, es decir, minimiza el error de clasificar un patrón  $x$  a una clase que no es la suya (siempre y cuando las funciones de verosimilitud estimadas sean las que efectivamente describen a la muestra).

Por otro lado, el clasificador bayesiano es óptimo aún cuando las probabilidades a priori sean distintas para cada clase. La idea básica de usar estos métodos para la identificación de litologías, fue modelar cada litología como un evento oculto que posteriormente será reconocido usando el algoritmo de Metropolis-Hastings. Los mejores resultados se obtuvieron con los tres registros usados, es decir, con los registros GR, RHOB y PEF. Los resultados obtenidos usando los métodos bayesianos son una herramienta estadística que además de arrojar buenos resultados, lo hace a un bajo costo computacional.

# REFERENCIAS BIBLIOGRÁFICAS

- [1] P. Bremaud: *“Markov Chains, Gibbs field, Monte Carlo Simulation, and Queues”*, Springer, (1998).
- [2] M. Davy, C. Doncarli, and J. Tournet: *Supervised Classification using MCMC Methods*, vol. 1, pp.33-36, June (2000).
- [3] R. O. Duda, and P. E. Hart: *Pattern Classification and Scene Analysis*, Wiley Interscience, (1973).
- [4] Escalona. J. *“Modelos Estadísticos en Selección y Clasificación de Registro de Pozos”*, Tesis de Maestría. Instituto Venezolano de Investigaciones Científicas, (2004).
- [5] D. Frayssinet, S. Thiria, F. Badran, and L. Briqueu: *Use of Neural Network in Log’s Data Processing-Prediction and Rebuilding of Lithologic Facies”*, *Petrophysics meets Geophysics*, Paris-France, 6-8 November, (2000), 305-317.
- [6] M. Padrón, S. Garcia, D. Barráez, B. Dorizzi, and S. Thiria: *A Hidden Markov Model Approach for Lithology Identification from Logs”*, *3rd Conference on Artificial Intelligence Applications to the Environmental Science*, Long Beach, CA. Feb. (2003).
- [7] S. Theodoridis, and K. Koutroumbas: *“Pattern Recognition”*, Academic Press. (1999).