

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“IDENTIFICACIÓN EN MODELOS DINÁMICOS NO
LINEALES.”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

BR. GÉNESIS DE LOS ANGELES SEQUERA GARCÍA

COMO REQUISITO FINAL

PARA OBTENER EL TÍTULO DE LICENCIADA

EN CIENCIAS MATEMÁTICAS

ÁREA DE CONOCIMIENTO: PROBABILIDAD Y ESTADÍSTICA V .

TUTOR: LIC. MSC. JHONNY ESCALONA

Barquisimeto, Venezuela. Mayo de 2014



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:

“IDENTIFICACIÓN EN MODELOS DINÁMICOS NO LINEALES.”

presentado por el ciudadano BR. GÉNESIS DE LOS ANGELES SEQUERA GARCÍA titular de la Cédula de Identidad No. 19856048, con el propósito de cumplir con el requisito académico final para el otorgamiento del título de Licenciada 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:

¹ Aprobado ó Reprobado

*Dedicado a mis padres, hermanos y a todos
mis amigos.*

AGRADECIMIENTOS

En primer lugar agradezco a Dios, por ser mi guía y mi fortaleza, por darme la fuerza y la sabiduría para seguir adelante y ayudarme a alcanzar mis metas.

Con carácter especial, a mi madre, por ser mi amiga, mi gran ejemplo a seguir, por su apoyo incondicional y por siempre creer en mí, a mi padre por creer en mis capacidades, a mi nana Cristina Hernández y mis hermanos Moisés Sequera y María Sequera por brindarme su amor sincero, estar siempre a mi lado y por formar parte de mi maravillosa familia.

A mis amigos, Shaday, Karla, Evelyn, Diana, Celismar, Jessica y Rosa, Gracias por ser parte de esta experiencia, por apoyarme y cuidarme, por los buenos momentos y por brindarme su valiosa amistad. A todos mis amigos de AsoEM, entre ellos Iliana, Ricardo, Joelviz, Uvencio, Rafael y Emely por ayudarme en los malos momentos y hacerme feliz en los buenos, por estar presente cuando mas los necesitaba. Junto a ustedes he vivido momentos muy especiales y gracias a ustedes descubrí el verdadero valor de la amistad.

A mi tutor y profesor Jhonny Escalona por su excelente ayuda y dedicación para la realización de este trabajo, gracias por ser parte de mi formación profesional.

Y finalmente, a mi novio, Juan Bialko por formar parte de mi vida, por su gran apoyo, por estar siempre a mi lado y gracias por que sé que cuento con tu amor y compañía.

RESUMEN

En este trabajo se desarrolla un algoritmo que combina el algoritmo EM, algoritmo de Viterbi Y filtro de partículas, para la identificación de modelos de espacio-estados no lineales, donde la estructura del proceso es desconocida. Se desarrollaron los métodos aplicados, con el fin de estudiar los resultados obtenidos y se implementan en matlab los algoritmos estudiados a un problema económico específico.

ÍNDICE

Agradecimientos	i
Resumen	iii
Introducción	1
1. Preliminares	3
1.1. Probabilidad Condicional	3
1.2. Esperanza y Covarianza de Variables Aleatorias	4
1.3. Estimación por Maxima Verosimilitud	6
1.4. Funciones Radiales	8
1.4.1. Representación matricial	8
1.4.2. Forma de una Función Radial	9
2. Filtro de Partículas	10
2.1. Descripción del Algoritmo	12
3. Algoritmo EM	15
3.1. Descripción del Algoritmo EM	16
3.1.1. El paso E (Esperanza)	16
3.1.2. El paso M (Maximización)	17
3.2. Aplicación al modelo dinámico	20
3.3. Aproximación de la función Q	22
3.4. Maximización de la función Q	24

4. Algoritmo de Viterbi	26
4.1. Algoritmo de Viterbi	26
4.1.1. Cadenas de Markov	26
4.1.2. El Algoritmo de Viterbi	28
4.1.3. Algoritmo	30
4.2. Estimación Paramétrica del Modelo Dinámico no Lineal	30
5. Aplicación del algoritmo	32
5.0.1. El Modelo	32
5.0.2. Resultados	32
Conclusiones	35
Referencias Bibliográficas	36

INTRODUCCIÓN

La identificación de sistemas no lineales trata de la estimación de modelos de sistemas dinámicos a partir de los datos observados, estos han producido un campo de investigación llamado identificación de sistemas, donde se define o describe una estructura modelo que tiene un comportamiento similar al del sistema desconocido. La complejidad del problema de estimación de parámetros considerado en este trabajo, es presentado por [3], [2], donde la verdadera estructura del modelo es desconocida. Las variables del proceso se pueden dividir en variables de estados y de observaciones. Una combinación de estas variables se pueden utilizar para representar el comportamiento dinámico de un proceso no lineal en la siguiente forma,

$$\begin{aligned}x_{t+1} &= f(x_t, u_t, \theta) + w_t, \\y_t &= g(x_t, u_t, \theta) + v_t,\end{aligned}$$

donde $x_t \in \mathbb{R}^n$ es el vector de estados, $y_t \in \mathbb{R}^m$ es el vector de observaciones y w_t, v_t son vectores aleatorios gaussianos independientes e idénticamente distribuidos con media 0 y matriz de covarianza Q y R respectivamente, θ es un vector de parámetros, f y g son funciones no lineales desconocidas que describen la dinámica del proceso. En este trabajo se describen las técnicas utilizadas en [3], donde f y g son aproximadas de la siguiente forma,

$$\begin{aligned}f(x_t) &= \sum_{i=1}^{I_x} h_i \rho_i(x_t, u_t, c_i, \Sigma_x) + Ax_t + Bu_t, \\g(x_t) &= \sum_{i=1}^{I_y} g_i \varphi_i(x_t, u_t, d_i, \Sigma_y) + Cx_t + Du_t,\end{aligned}$$

donde $\rho_i(\cdot, \cdot)$ y $\varphi_i(\cdot, \cdot)$ son las funciones de base radial centradas en c_i y d_i con pesos de salida h_i, g_i y varianzas Σ_x y Σ_y respectivamente, A, B, C, D son matrices apropiadas, usadas para mantener linealidad en el modelo. El algoritmo presentado en el artículo [3] donde la idea central es encontrar el vector de parámetros θ , que maximiza la función de probabilidad de las observaciones, para la estimación de estos parámetros y la aproximación de estas funciones no lineales se usa el algoritmo planteado, el cual combina funciones de base radial para aproximar la dinámica del proceso con el algoritmo EM,

filtros de partículas y algoritmo de Viterbi. El fin de este trabajo es usar este algoritmo para estimar los parámetros que aparezcan en el problema económico planteado.

El trabajo está estructurado de la siguiente manera:

En el primer capítulo mostramos definiciones, propiedades, teoremas que fueron necesarias para la realización de este trabajo, en el segundo capítulo estimaremos los estados del sistema usando Filtro de Partículas, en el tercer capítulo se explican las bases teóricas y la descripción del algoritmo EM y en el quinto capítulo se describe el algoritmo y se aplica a un modelo en particular.

CAPÍTULO 1

PRELIMINARES

1.1. Probabilidad Condicional

Definición 1.1. Sea (Ω, P) un espacio de probabilidad. Dados dos eventos probabilísticos X y Y , entonces la probabilidad condicional de que ocurra X dado que ocurrió Y , denotado por $P(X|Y)$, se define como:

$$P(X|Y) = \frac{P(XY)}{P(Y)}, \quad \text{si } P(Y) \neq 0. \quad (1.1)$$

Así, si X y Y son variables aleatorias, entonces se define la función de masa de probabilidad condicional de X dado $Y = y$ por:

$$\begin{aligned} P_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{P(x, y)}{P_Y(y)}. \end{aligned}$$

Teorema 1.1. (*Teorema de Bayes.*) Suponga que se observa un variable aleatoria X , y se desea hacer inferencias acerca de otra variable aleatoria θ , donde θ se extrae de alguna distribución $p(\theta)$.

De la definición de probabilidad condicional, se tiene:

$$P(\theta|X) = \frac{P(X, \theta)}{P(X)}.$$

1.2. Esperanza y Covarianza de Variables Aleatorias

Definición 1.2. Sea X una variable aleatoria discreta, con posibles valores X_1, X_2, \dots, X_n y con función de probabilidad $P(X_i)$, $i = 1, \dots, n$. Se define la esperanza como

$$\mathbb{E}(X) = \sum_{i=1}^n X_i P(X_i). \quad (1.2)$$

Y para una variable aleatoria continua, con función de densidad $F(X)$, la esperanza se define como

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} X F(X). \quad (1.3)$$

Definición 1.3. La esperanza condicional de la variable aleatoria X dado el evento $Y = Y_j$ se calcula como sigue

$$\mathbb{E}(X|Y) = \sum_{i=1}^n X_i P(X = X_i | Y = Y_j) = \sum_{i=1}^n X_i \frac{P(X = X_i, Y = Y_j)}{P(Y = Y_j)},$$

que es un valor constante que depende del evento $Y = Y_j$.

Si ahora X es una variable continua, la esperanza condicional viene dada por:

$$\mathbb{E}(X|Y) = \int_X X F_X(X|Y = Y_j) d_X.$$

Propiedad 1.1. Sean X y Y dos variables aleatorias si $\mathbb{E}(X)$ y $\mathbb{E}(Y)$ existen, con a y b constantes, entonces:

- Para todo $\alpha \in \mathbb{R}$, $\mathbb{E}(\alpha X)$ existe y además:

$$\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X).$$

- $\mathbb{E}(X + Y)$ existe, y se cumple:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

- $\mathbb{E}(X + a)$ existe y se tiene que:

$$\mathbb{E}(X + a) = \mathbb{E}(X) + a.$$

- $E(aX + bY)$ existe y se cumple que:

$$E(aX + bY) = aE(X) + bE(Y).$$

- Si X, Y son independientes entonces,

$$E(XY) = E(X)E(Y).$$

Definición 1.4. Sea X un vector aleatorio, $X = X_1, X_2, \dots, X_n$, con valor esperado $E(X) = \mu$, se define la varianza de X como:

$$Var(X) = E[(X - \mu)^2],$$

y si desarrollamos el lado derecho de la igualdad, podemos escribir la varianza de X de la siguiente manera:

$$Var(X) = E[(X)^2] - \mu^2. \quad (1.4)$$

Propiedad 1.2. Sean X e Y dos variables aleatorias, a y b constantes, donde $Var(X)$ y $Var(Y)$ existen, entonces:

- $Var(a) = 0$.
- $Var(aX + b) = a^2Var(X)$.
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$, donde $Cov(X, Y)$ es la covarianza de X e Y .

Definición 1.5. Sean X e Y dos variables aleatorias, la covarianza de X e Y ($Cov(X, Y)$) se define como:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]. \quad (1.5)$$

Propiedad 1.3. Sean X e Y dos variables aleatorias, a y b constantes, entonces:

- $Cov(X, a) = 0$.
- $Cov(X, X) = Var(X)$.

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.
- $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$.
- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

Ahora bien, para los casos n-dimensionales se tienen las siguientes definiciones. Sea $X = (X_1, \dots, X_n)$ un n-vector estocástico, donde el valor esperado $E(X_i)$ existe para todo i , entonces definimos:

Definición 1.6. Sea X un vector aleatorio, $X = X_1, X_2, \dots, X_n$, se define el vector esperanza $E(X)$ por:

$$E(X) := (E(X_1), E(X_2), \dots, E(X_n)).$$

Definición 1.7. Si X, Y son dos vectores estocásticos, n-dimensionales, entonces podemos definir la matriz de Covarianza de X de la siguiente manera:

$$\text{Cov}(X) = \begin{pmatrix} \text{Cov}(X_1; X_1) & \cdots & \text{Cov}(X_1; X_n) \\ \text{Cov}(X_2; X_1) & \cdots & \text{Cov}(X_2; X_n) \\ \vdots & \cdots & \vdots \\ \text{Cov}(X_n; X_1) & \cdots & \text{Cov}(X_n; X_n) \end{pmatrix}$$

donde $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$, para todo $i, j \in 1, 2, \dots, n$.

Observación 1.1. Dado que $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$, para todo $i, j \in 1, 2, \dots, n$, entonces la matriz de covarianza para cualquier n-vector aleatorio X es simétrica.

1.3. Estimación por Maxima Verosimilitud

La función de verosimilitud (o, simplemente, verosimilitud) es una función de los parámetros de un modelo estadístico que permite realizar inferencias acerca de su valor

a partir de un conjunto de observaciones.

La estimación vía Máxima Verosimilitud es una herramienta muy usada para la estimación de parámetros en modelos estadísticos y a continuación, presentaremos una definición de la función Verosimilitud y el Método de Máxima Verosimilitud.

Definición 1.8. Sea X_1, X_2, \dots, X_n una muestra de variables aleatorias, independientes e idénticamente distribuidas, con función de densidad $f_\theta(x_i)$ para cada i , donde θ es el vector de parámetros de la distribución que además es desconocido, se define la función de Verosimilitud como la función $L : (R)^n \rightarrow (R)$ definida por:

$$L(\theta) = L(\theta | x_1, x_2, \dots, x_n) = f_\theta(x_1) \dots f_\theta(x_n),$$

donde x_1, x_2, \dots, x_n son resultados de la muestra aleatoria.

Definición 1.9. Estimador de Máxima Verosimilitud Sea X_1, X_2, \dots, X_n una muestra de variables aleatorias, independientes e idénticamente distribuidas, con función de densidad $f_\theta(x_i)$ para cada i con parámetro θ desconocido, y supongamos que deseamos estimar $\hat{\theta}$ como el valor que está mas próximo al valor verdadero de θ .

En este sentido, definimos $\hat{\theta}$ como el valor (o los valores) que maximizan la función de Verosimilitud $L(\theta | x_1, x_2, \dots, x_n)$. Este se conoce como Estimador de Máxima Verosimilitud.

Matemáticamente es más fácil manipular esta función tomando su logaritmo. Luego la función logarítmica de la verosimilitud se expresa de la siguiente forma:

$$\begin{aligned} \log L(\theta) &= \log f_\theta(x_1) + \dots + \log f_\theta(x_n) \\ &= \sum_{i=1}^n \log f_\theta(x_i). \end{aligned}$$

La estimación consiste en encontrar los parámetros $\hat{\theta}$ óptimos, es decir, los que mas se ajustan a la muestra. Entonces,

$$\hat{\theta} = \arg \max L(\theta).$$

1.4. Funciones Radiales

Una función radial es una función que calcula la distancia Euclídea de un vector de entrada x respecto de un centro d , es decir,

$$\Phi(x) = \|x - d_i\|.$$

Las funciones radiales tienen como modelo matemático la siguiente relación

$$P_f(x) = \sum_{i=1}^N w_i h(\|x - d_i\|), \quad (1.6)$$

donde a cada entrada le corresponde una función radial $h(x)$ y un peso de salida w_i , N es el número de funciones radiales.

1.4.1. Representación matricial

Dados los datos $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, tales que $P_f(x_i) = y_i$, donde P_f es desconocida y puede ser aproximada via funciones radiales de la siguiente manera:

$$\begin{bmatrix} h(\|x_1 - d_1\|) & h(\|x_1 - d_2\|) & \cdots & h(\|x_1 - d_n\|) \\ h(\|x_2 - d_1\|) & h(\|x_2 - d_2\|) & \cdots & h(\|x_2 - d_n\|) \\ \vdots & \vdots & \vdots & \vdots \\ h(\|x_n - d_1\|) & h(\|x_n - d_2\|) & \cdots & h(\|x_n - d_n\|) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{pmatrix} P_{f_1}(x) \\ P_{f_2}(x) \\ \vdots \\ P_{f_n}(x) \end{pmatrix},$$

donde la matriz $\Phi(x)$ es de orden $n \times n$, el peso de salida w_i es de orden $n \times 1$ y la función $P_f(x)$ es de orden $n \times 1$.

1.4.2. Forma de una Función Radial

Para determinar la forma de una Función Radial se suele utilizar algunas funciones, una de las más típicas es la Gaussiana, que en su forma generalizada es:

$$h(x) = \exp[(x - d)^T \Sigma^{-1} (x - d)], \quad (1.7)$$

donde Σ^{-1} es la matriz de radios de la función, ésta es la matriz de varianzas y covarianzas relacionada a la función Normal.

Una propiedad de estas funciones es que forman una base para el espacio de funciones, es decir, cualquier función puede ser aproximada por medio de una combinación lineal de funciones radiales,

$$f(x) = \sum_{i=1}^N w_i h_i(x).$$

CAPÍTULO 2

FILTRO DE PARTÍCULAS

El filtro de partículas fue propuesto por N. Gordon, D. Salmond y A. Smith en 1993, es un método para calcular los estados no observables en el problema de filtrado no lineal, este problema consiste en dos ecuaciones de la forma:

$$\begin{aligned}x_{t+1} &= f(x_t, u_t) + b_t \\y_t &= g(x_t, u_t) + v_t,\end{aligned}\tag{2.1}$$

donde b_t y v_t son vectores aleatorios independientes, u_t y y_t , $t = 1, 2, \dots, n$ son variables observables del sistema, x_t , $t = 1, 2, \dots, n$ es el estado no observable del sistema, f y g son funciones no lineales, en general, f y g tienen una estructura conocida, en este trabajo supondremos que f y g son desconocidas.

Se estima x_t como \hat{x}_t el cual minimiza el error cuadrático medio dadas las observaciones u_t y y_t hasta el tiempo t , es decir,

$$\hat{x}_t = E[x_t | y_1, y_2, \dots, y_t] \equiv E[x_t | y_{1:t}],\tag{2.2}$$

donde $y_{1:t} = (y_1, \dots, y_t)$.

Para obtener \hat{x}_t es necesaria la densidad condicional de x_t dado $y_{1:t}$, la cual denotaremos por:

$$P(x_t | y_{1:t}).$$

El objetivo es encontrar una aproximación discreta de la densidad condicional

$$P(x_{1:t} | y_{1:t}) = p(x_t | y_1, y_2, \dots, y_t).\tag{2.3}$$

La propagación de la densidad a posteriori (2.3) es la principal herramienta para la estimación de \hat{x}_t . Del teorema de Bayes se tiene que:

$$P(x_t|y_{1:t}) = \frac{P(y_t|x_t) P(x_t|y_{1:t-1})}{\int P(y_t|x_t) P(x_t|y_{1:t-1}) \delta x_t}, \quad (2.4)$$

donde

$$P(x_t|y_{1:t-1}) = \int P(x_t|x_{t-1}) P(x_{t-1}|y_{1:t-1}) \delta x_{t-1}. \quad (2.5)$$

Así, la estimación de (2.2) es:

$$\hat{x}_t = \int x_t P(x_t|y_{1:t}) \delta x_t. \quad (2.6)$$

El filtro de partículas nos permite calcular \hat{x}_t ya que este nos da una estimación de $P(x_t|y_{1:t})$. Una forma de obtener $P(x_t|y_{1:t})$ es utilizando muestras \hat{x}_t^i con $i = 1, 2, \dots, n$, y correspondientes pesos \hat{w}_t^i con $i = 1, 2, \dots, n$, de acuerdo a:

$$\hat{p}(x_t|y_{1:t}) = \sum_{i=1}^n \hat{w}_t^i \delta(x_t - \hat{x}_t^i);$$

donde δ es la función delta de Dirac.

Cada muestra \hat{x}_t^i es una realización del estado del sistema con su correspondiente peso \hat{w}_t^i que indica que tan buena es la realización. Otra alternativa es aproximar $p(x_t|y_{1:t})$ y remuestrear de acuerdo a sus pesos, de esta forma obtenemos:

$$p(x_t|y_{1:t}) = \sum_{i=1}^n \frac{1}{n} \delta(x_t - \hat{x}_t^i).$$

En el paso del remuestreo las muestras con alto peso tienden a ser replicadas muchas veces, las muestras con bajo peso tienden a ser rechazadas. Realmente el uso de los filtros de partículas se ha disparado en los últimos años gracias a la introducción de un nuevo paso en el algoritmo que lo mejora, se trata del Resampling que consiste a re-muestrear las partículas: la idea es de dejar propagar las partículas de mayor peso, quitando las que menos información llevan como se muestra en la siguiente figura:

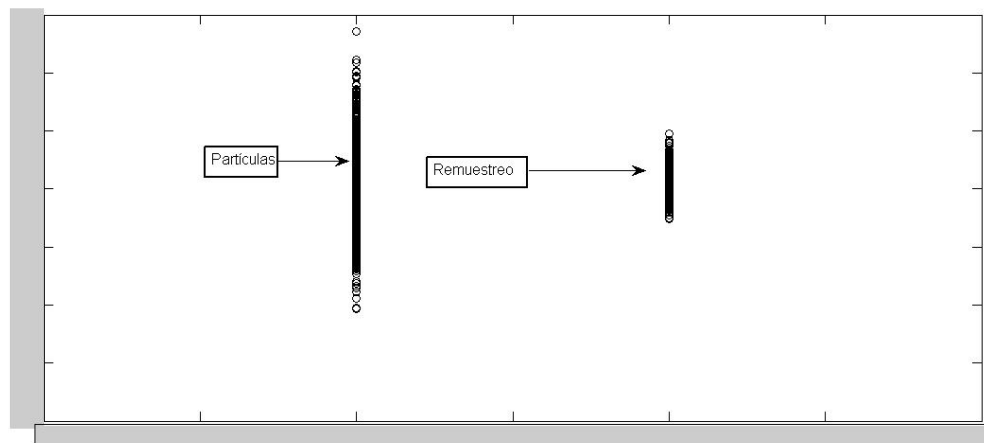


Figura 2.1: Remuestreo en el filtro de partículas.

El remuestreo es importante en la implementación del filtro de partículas porque reduce la varianza en la estimación del estado. La idea básica es que algunas muestras de la distribución de probabilidad son más significantes que otras. De esta forma, reajustando esta distribución a partir de las medidas cada cierto tiempo, la estimación del estado mejorará.

Las muestras \hat{x}_t^i son denominadas partículas, el remuestreo es aplicado como un filtro para obtener las partículas con mayor peso \hat{x}_t^i como se muestra en la figura (2), de aquí el nombre de filtro de partículas.

2.1. Descripción del Algoritmo

El algoritmo tiene una serie de etapas en las cuales se usan las operaciones explicadas anteriormente:

1. Inicialización

Existe una etapa de inicialización del algoritmo que únicamente se ejecuta en el

instante $t = 1$,

$$x_1^i \quad i = 1 : n.$$

En ella se obtiene un conjunto de n partículas o muestras independientes e idénticamente distribuidas que son n realizaciones de $p(x_1)$, a partir de esta distribución a priori se lanzan n partículas de forma aleatoria en el espacio de búsqueda.

2. Predicción

Las partículas se propagan a través del modelo de estado del sistema. Obtener nuevas partículas \hat{x}_t^i utilizando (2.1)

$$\hat{x}_t^i = f(x_{t-1}, u_{t-1}) + b_{t-1}^i.$$

Se actualizan las N partículas en función de la acción anterior a $t-1$ y la observación actual z_t . Para ello, se aplica el modelo de movimiento

$$p(x_t | x_{t-1}^i).$$

A cada una de las N partículas, generando un nuevo conjunto de partículas. Las partículas nuevas representan la predicción de la variable de estado, sin considerar la observación.

3. Actualización

Consiste en encontrar los pesos de la densidad del filtro w_t^i usando

$$w_t^i = p(y_t | \hat{x}_t^i),$$

y normalizados:

$$w_t^i = \frac{w_t^i}{\sum_{i=1}^N w_t^i}. \quad (2.7)$$

El conjunto de pesos de cada partícula es proporcional a la probabilidad de su estado y a la suma normalizada de sus pesos.

4. Resampling o Remuestreo

El cálculo computacional se hace más liviano si la mayoría de las partículas tienen pesos iguales. Este es el propósito de la etapa de resampling. Volver a muestrear n partículas con reemplazamiento del conjunto $\{x_t^1, x_t^2, \dots, x_t^n\}$ con la probabilidad de escoger x_t^i con peso w_t^i .

$$p(x_t^i = \hat{x}_t^j) = \hat{w}_t^j \quad \text{con } j = 1, \dots, n.$$

5. $t = t + 1$ y repetir desde el paso 2 hasta $t = n$.

En el próximo capítulo se desarrollará el Algoritmo EM, un método para calcular el estimador de máxima verosimilitud donde dicha función es muy compleja y no se pueden usar métodos estándar.

CAPÍTULO 3

ALGORITMO EM

El algoritmo EM (Expectation-Maximization) es un procedimiento matemático iterativo para calcular el estimador de máxima verosimilitud de los parámetros de una distribución de probabilidad en problemas donde la función de verosimilitud es muy compleja y no se pueden utilizar métodos estándar para maximizarla. Tal situación se presenta cuando parte de la información esta oculta (no es observada).

Este algoritmo fue desarrollado por Dempster en 1977, aunque el método ya había sido propuesto y estudiado, Dempster consiguió formalizar, justificar y generalizar el método y supone un punto de partida para el desarrollo del mismo. Se trata de un algoritmo de optimización elegante que calcula el valor esperado de los datos completos (variables observables y no observables), y luego se maximiza dicha función.

La idea de este algoritmo es simple, dada una primera estimación de los parámetros, se predicen los datos faltantes, se re-estiman los parámetros y se continúa iterando hasta obtener convergencia.

Su principal aplicación constituye una herramienta iterativa para obtener la estimación por maxima verosimilitud a partir de los datos desconocidos. Se considera entonces la muestra dividida en un vector de datos incompletos (vector de observaciones) y un vector de datos no observable (vector de estados). Se requiere una transformación apropiada P que relacione el vector de datos observados con el vector de datos completos.

3.1. Descripción del Algoritmo EM

Sea el conjunto de datos completos $Z = (x_{1:T}, y_{1:T})$ con función de densidad de probabilidad $P_Z = (Z|\Theta)$ asociada a los datos $y_{1:T} = y_1, \dots, y_t, \dots, y_T$ que son observados y a los datos $x_{1:T} = x_1, \dots, x_t, \dots, x_T$ que no son observados y Θ es el conjunto de parámetros de la representación de espacio de estados. Sea P_y la función de densidad de $y_{1:T}$ determinada por el vector de parámetros Θ .

La función de densidad conjunta de las observaciones viene dada por:

$$P(y|\Theta) = P_y(y_{1:T}|\Theta). \quad (3.1)$$

La función de densidad de $y_{1:T}$ esta asociada con la función de densidad conjunta de los datos completos Z y esta dada por:

$$P_Z((y, x)|\Theta) = P(x_{1:T}|y_{1:T}, \Theta) (P(y_{1:T}|\Theta)). \quad (3.2)$$

La distribución de la densidad de probabilidad conjunta, sobre los datos completos e incompletos, es definida mediante la construcción de relaciones entre los valores $x_{1:T}, y_{1:T}$. Con base en esto, se define la función de verosimilitud sobre los parámetros Θ , de la distribución conjunta de información completa:

$$L(\Theta|Z) = L(\Theta|y_{1:T}, x_{1:T}) = P((y_{1:T}, x_{1:T})|\Theta). \quad (3.3)$$

3.1.1. El paso E (Esperanza)

En este paso se calcula el valor esperado del logaritmo de la función de verosimilitud de los datos completos $L(\Theta|Z)$ con respecto a la distribución condicional de los datos ocultos dados los datos observados y el parámetro Θ en su valor actual Θ_T .

Se considera $x_{1:T}$ como variable aleatoria, por lo que la función de verosimilitud $L(\Theta|y_{1:T}, x_{1:T})$ es variable aleatoria, por lo tanto se puede aplicar la esperanza condi-

cional a $L(\Theta|y_{1:T}, x_{1:T})$ dado $y_{1:T}$ y Θ_i , lo cual se denota por:

$$Q(\Theta_i, \Theta) = E[\log L(\Theta_i|y_{1:T}, x_{1:T}) | y_{1:T}, \Theta]. \quad (3.4)$$

La esperanza es tomada con respecto a la función de densidad de $x_{1:T}$ dado $y_{1:T}$, $P(x_{1:T}|y_{1:T}, \Theta_i)$, es decir,

$$\begin{aligned} Q(\Theta_i, \Theta) &= \int \log(L(\Theta_i|y_{1:T}, x_{1:T}) | y_{1:T}, \Theta) (P(x_{1:T}|y_{1:T}), \Theta_i dx_{1:T}) \\ &= \int \log(P(x_{1:T}, y_{1:T}|y_{1:T}, \Theta))(P(x_{1:T}|y_{1:T}, \Theta_i) dx_{1:T}). \end{aligned} \quad (3.5)$$

El paso E consiste en:

- Elegir un valor inicial para Θ, Θ_0 .
- Dados los datos observados $y_{1:T}$ y pretendiendo de la suposición actual Θ_0 , se determina la distribución condicional $P(x_{1:T}|y_{1:T}, \Theta_i)$.
- Quisiéramos maximizar $\log P(x_{1:T}|\Theta_i)$, pero como no conocemos realmente $x_{1:T}$, vamos a maximizar su valor esperado a lo cual llamamos "función Q" (3.5).

3.1.2. El paso M (Maximización)

El paso M busca determinar el conjunto de parámetros que maximice la función Q, esto es, actualiza la estimación del parámetro maximizando la función calculada en el paso E con respecto a

$$\Theta_{i+1} = \arg \max Q(\Theta_i, \Theta).$$

Estos dos pasos anteriores se desarrollan de manera iterativa tantas veces como sea necesario hasta alcanzar la convergencia del máximo de la función de verosimilitud de $(x_{1:T}, y_{1:T})$.

El algoritmo EM, converge cuando la siguiente condición se cumple:

$$\|\Theta_{i+1} - \Theta_i\| < \epsilon_\Theta,$$

donde $\epsilon_\Theta > 0$ es un nivel de tolerancia especificado.

Ejemplo 3.1. A 197 niños les dan a escoger un juguete de entre cuatro opciones, las elecciones (los datos observados) son:

$$Y = (Y_1, Y_2, Y_3, Y_4) = (125, 18, 20, 34)$$

donde y_i es el numero de niños que escogieron el juguete i .

Podemos modelar los datos observados Y mediante la distribución multinomial, la cual tiene dos parámetros, el numero de niños y la probabilidad de que los niños elijan cada uno de los cuatro juguetes.

El problema consiste en estimar el valor de θ que maximiza la función de verosimilitud de las observaciones.

$$P(Y|\theta) = \frac{(Y_1 + Y_2 + Y_3 + Y_4)!}{Y_1!Y_2!Y_3!Y_4!} P_1^{Y_1} \cdot P_2^{Y_2} \cdot P_3^{Y_3} \cdot P_4^{Y_4}.$$

Supongamos que las probabilidades de que una observación corresponda a alguno de los juguetes es:

$$(P_1, P_2, P_3, P_4) = \left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1 - \theta), \frac{1}{4}(1 - \theta), \frac{\theta}{4} \right),$$

para algún $\theta \in (0, 1)$, donde $P_1 + P_2 + P_3 + P_4 = 1$.

La probabilidad de escoger el juguete $Y = (Y_1, Y_2, Y_3, Y_4)$ es, entonces:

$$P(Y|\theta) = \frac{(Y_1 + Y_2 + Y_3 + Y_4)!}{Y_1!Y_2!Y_3!Y_4!} \left(\frac{1}{2} + \frac{\theta}{4} \right)^{Y_1} \left(\frac{1}{4}(1 - \theta) \right)^{Y_2} \left(\frac{1}{4}(1 - \theta) \right)^{Y_3} \left(\frac{\theta}{4} \right)^{Y_4}.$$

Para utilizar EM necesitamos especificar a que llamamos X (los datos completos) y hacerlo de manera que su distribución también se pueda expresar en términos del

mismo parámetro.

Primero debemos considerar el modelo completo de la siguiente manera: la primera categoría que tiene probabilidad de ocurrencia $\frac{1}{2} + \frac{\theta}{4}$, puede ser considerada como que esta formada por dos subcategorías, una de ellas con probabilidad de ocurrencia $\frac{1}{2}$ y la otra con probabilidad de ocurrencia $\frac{\theta}{4}$.

El modelo completo consistirá de X_1, X_2, X_3, X_4 y X_5 , donde $Y_1 = X_1 + X_2$, $Y_2 = X_3$, $Y_3 = X_4$, $Y_4 = X_5$. La función de probabilidad del modelo completo estaría dado por:

$$P(X|\theta) = \frac{(X_1 + X_2 + X_3 + X_4 + X_5)!}{X_1!X_2!X_3!X_4!X_5!} \left(\frac{1}{2}\right)^{X_1} \left(\frac{\theta}{4}\right)^{X_2} \left(\frac{1}{4}(1-\theta)\right)^{X_3} \left(\frac{1}{4}(1-\theta)\right)^{X_4} \left(\frac{\theta}{4}\right)^{X_5}$$

aplicando logaritmo tenemos:

$$\log P(X|\theta) = K + (X_2 + X_5) \log\left(\frac{\theta}{4}\right) + (X_3 + X_4) \log\left(\frac{1}{4}(1-\theta)\right).$$

Así,

$$E[\log P(X|\theta) | Y, \theta^p] = K + E(X_2 | Y_1, \theta^p) \log\left(\frac{\theta}{4}\right) + Y_4 \log\left(\frac{\theta}{4}\right) + (Y_3 + Y_2) \log\left(\frac{1}{4}(1-\theta)\right),$$

donde K no depende de θ .

Ahora, utilizando el hecho de que X_2 y Y_1 tienen distribución binomial obtenemos

$$E(X_2 | Y_1, \theta^p) = Y_1 \frac{\frac{\theta^p}{4}}{\frac{1}{2} + \frac{\theta^p}{4}}.$$

- Paso 1 (Esperanza)

Aplicando el paso E tenemos:

$$\begin{aligned} Q(\theta | \theta^p) &= E[\log P(X|\theta) | Y, \theta^p] \\ &= K + Y_1 \frac{\frac{\theta^p}{4}}{\frac{1}{2} + \frac{\theta^p}{4}} \log\left(\frac{\theta}{4}\right) + Y_4 \log\left(\frac{\theta}{4}\right) + (Y_3 + Y_2) \log\left(\frac{1}{4}(1-\theta)\right). \end{aligned}$$

- Paso 2 (Maximización)

Para maximizar Q usamos la misma técnica de cálculo diferencial, derivamos e igualamos a cero. Entonces derivando con respecto a θ se tiene:

$$Q'(\theta|\theta^p) = 0 + Y_1 \frac{\frac{\theta^p}{4}}{\frac{1}{2} + \frac{\theta^p}{4}} \left(\frac{1}{\theta}\right) + Y_4 \left(\frac{1}{\theta}\right) + (Y_3 + Y_2) \left(\frac{1}{1 - \theta}\right) = 0,$$

despejando θ se obtiene:

$$\theta = \frac{Y_1 + Y_4\theta^p + 2Y_4}{(Y_1 + Y_2 + Y_3 + Y_4)\theta^p + 2(Y_2 + Y_3 + Y_4)}.$$

Implementando el algoritmo EM en matlab resulta que el parámetro aproximado es $\theta^p = 0,6268$ con 8 iteraciones, con valor inicial $\theta^0 = 0,5$ y con una tolerancia de 10^{-6} .

3.2. Aplicación al modelo dinámico

Se supone que la dinámica del proceso es desconocida, donde y_1, y_2, \dots, y_t son las observaciones y los datos x_1, x_2, \dots, x_t que son desconocidos, representaremos el comportamiento dinámico de un proceso no lineal de la siguiente forma,

$$\begin{aligned} x_{t+1} &= f(x_t, u_t) + b_t \\ y_t &= g(x_t, u_t) + v_t, \end{aligned} \tag{3.6}$$

donde $x_t \in \mathbb{R}^n$ es el vector de estados, $y_t \in \mathbb{R}^m$ es el vector de observaciones y u_t, v_t son vectores aleatorios gaussianos independientes e idénticamente distribuidos con media 0 y matriz de covarianza Q y R respectivamente, $f(\cdot)$ y $g(\cdot)$ son funciones no lineales que describen la dinámica del proceso. En este trabajo las funciones no lineales $f(\cdot)$ y $g(\cdot)$ son desconocidas.

Es bien conocido que cualquier función se puede aproximar utilizando funciones de base tales como funciones de base radial. luego,

$$f(x_t) = \sum_{i=1}^{I_x} h_i \rho_i(x_t, c_i, \Sigma_x)$$

$$g(x_t) = \sum_{i=1}^{I_y} g_i \varphi_i(x_t, d_i, \Sigma_y).$$

Por lo tanto, el modelo en (3.6) es aproximado utilizando funciones de base radial de la siguiente manera:

$$x_{t+1} = \sum_{i=1}^{I_x} h_i \rho_i(x_t, u_t, c_i, \Sigma_x) + Ax_t + Bu_t + b_t$$

$$y_t = \sum_{i=1}^{I_y} g_i \varphi_i(x_t, u_t, d_i, \Sigma_y) + Cx_t + Du_t + v_t, \quad (3.7)$$

donde $\rho_i(\cdot, \cdot)$ y $\varphi_i(\cdot, \cdot)$ son las funciones de base radial centradas en c_i y d_i , con una varianza Σ_x y Σ_y respectivamente, h_i y g_i son el peso de salida, I_x, I_y son el número de funciones de base radial utilizadas en las ecuaciones de estado y de observación.

En este trabajo usaremos como aproximación de funciones de base radial la función Gaussiana, que se utiliza de la siguiente manera:

$$\rho_i(x_t, u_t, c_i, \Sigma_x) = e^{[-(x_t - c_i)^T \Sigma_x^{-1} (x_t - c_i)]}$$

$$\varphi_i(x_t, u_t, d_i, \Sigma_y) = e^{[-(x_t - d_i)^T \Sigma_y^{-1} (x_t - d_i)]}.$$

Así, el modelo en (3.7) es aproximado de la siguiente manera:

$$x_{t+1} = \sum_{i=1}^{I_x} h_i e^{[-(x_t - d_i)^T \Sigma_x^{-1} (x_t - c_i)]} + b_t$$

$$y_t = \sum_{i=1}^{I_y} g_i e^{[-(x_t - d_i)^T \Sigma_y^{-1} (x_t - d_i)]} + v_t. \quad (3.8)$$

3.3. Aproximación de la función Q

La aproximación de las no linealidades fallaran si estas son prominentes, aunque ya existen una serie de aproximaciones de algoritmo EM que han sido desarrolladas anteriormente con la participación de diferentes aproximaciones de la función Q, pero los métodos involucrados para la aproximación del valor esperado son generalmente intensivos, por esto, aproximaremos la función Q mediante una combinación de filtros de partículas estudiado en el capítulo 2 para el modelo de espacio-estado que desarrollaremos en este trabajo debido a que en el modelo aplicado las no linealidades son prominentes.

Recordemos que la función Q del algoritmo EM viene dada por:

$$Q(\Theta_i, \Theta) = \int \log (P(x_{1:T}, y_{1:T} | y_{1:T}, \Theta)) P(x_{1:T} | y_{1:T}, \Theta_i) dx_{1:T}.$$

Utilizando la propiedad de Markov del modelo de espacio de estado, la función de densidad conjunta de los datos ocultos y las observaciones se pueden escribir como:

$$P(x_{1:T}, y_{1:T} | y_{1:T}, \Theta) = P(x_1 | y_{1:T}, \Theta) \prod_{t=2}^T P(x_t | x_{t-1}, \Theta) \prod_{t=1}^T P(y_t | x_t, \Theta),$$

Luego, sustituyendo esta expresión en la función Q se tiene

$$\begin{aligned} Q(\Theta_i, \Theta) &= \int \log [P(x_1 | y_{1:T}, \Theta)] P(x_1 | y_{1:T}, \Theta_i) dx_1 \\ &+ \sum_{t=2}^T \int \log [P(x_t | x_{t-1}, \Theta)] P(x_{t-1:t} | y_{1:T}, \Theta_i) dx_{t-1:T} \\ &+ \sum_{t=1}^T \int \log [P(y_t | x_t, \Theta)] P(x_t | y_{1:T}, \Theta_i) dx_t. \end{aligned}$$

A partir de esta expresión, se puede ver que para obtener una aproximación de la función Q se necesitan las siguientes funciones de densidad:

1. $P(x_1|y_{1:T}, \Theta_i)$.
2. $P(x_t|y_{1:T}, \Theta_i)$.
3. $P(x_{t-1:t}|y_{1:T}, \Theta_i)$.

Usando filtro de partículas, es posible obtener aproximaciones de las funciones de densidad que se necesitan para la aproximación de la función Q, esto es,

$$P(x_1|y_{1:T}, \Theta_i) = \sum_{i=1}^N w_{1|1}^{(i)} \delta(x_1 - x_1^{(i)}).$$

$$P(x_t|y_{1:T}, \Theta_i) = \sum_{i=1}^N w_{t|T}^{(i)} \delta(x_t - x_t^{(i)}).$$

$$P(x_{t-1:t}|y_{1:T}, \Theta_i) = \sum_{i=1}^N w_{t-1,t}^{(i)} \delta(x_{t-1} - x_{t-1}^{(i)}) \delta(x_t - x_t^{(i)}).$$

Donde $w_{1|1}^{(i)}$, $w_{t|T}^{(i)}$, y $w_{t-1,t}^{(i)}$ son los pesos apropiados, correspondientes a las funciones de densidad y calculados usando la regla de bayes, $\delta(\cdot)$ representara la función delta Kronecker. Los pesos dependen de las funciones y de sus respectivas aproximaciones.

Usando las aproximaciones anteriores de las funciones de densidad de la función Q, se puede escribir la siguiente expresión:

$$Q(\Theta_i, \Theta) = \sum_{i=1}^N w_{1|1}^{(i)} \log [P(x_1^{(i)}|y_{1:T}, \Theta)]$$

$$+ \sum_{t=2}^T \sum_{i=1}^N w_{t-1,t}^{(i)} \log [P(x_t^{(i)}|x_{t-1}^{(i)}, \Theta)]$$

$$+ \sum_{t=1}^T \sum_{i=1}^N w_{t|T}^{(i)} \log [P(y_t|x_t^{(i)}, \Theta)]. \quad (3.9)$$

En la anterior aproximación, las funciones de densidad:

$$P(x_t^{(i)}|x_{t-1}^{(i)}, \Theta)$$

$$P(y_t|x_t^{(i)}, \Theta),$$

se pueden escribir en términos de funciones de densidad gaussianas, por lo tanto

$$\log \left[P \left(x_t^{(i)} | x_{t-1}^{(i)}, \Theta \right) \right] = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\det(Q)) - \frac{1}{2} (x_t^{(i)} - x_{t-1}^{(i)})^T Q^{-1} (x_t^{(i)} - x_{t-1}^{(i)}).$$

$$\log \left[P \left(y_t | x_t^{(i)}, \Theta \right) \right] = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\det(R)) - \frac{1}{2} (y_t^{(i)} - x_t^{(i)})^T R^{-1} (y_t^{(i)} - x_t^{(i)}).$$

Donde

$$x_t^i = \sum_{i=1}^{I_x} h_i \rho_i(x_t^i, c_i, \Sigma_x) + b_t$$

$$y_t^i = \sum_{i=1}^{I_y} g_i \varphi_i(x_t^i, d_i, \Sigma_y) + v_t.$$

3.4. Maximización de la función Q

La maximización de la función Q se lleva a cabo en dos pasos utilizando mínimos cuadrados, es fácil darse cuenta de que los parámetros en θ_l entran en el modelo lineal, mientras que los de θ_{nl} entran en el modelo no lineal.

Usaremos un procedimiento donde los parámetros lineales se estiman en el primer paso utilizando los mínimos cuadrados lineales y los parámetros no lineales se estiman en el segundo paso a través de mínimos cuadrados no lineales, el procedimiento es:

- Paso 1

A partir de una aproximación inicial para el vector de parámetros no lineal, θ_{nl} , la función Q se maximiza con respecto a θ_l . Esta maximización puede lograrse a través de mínimos cuadrados lineales. Se define:

$$\Omega_x = [w_1 w_2 \dots w_{I_x}]$$

$$S_t = I_1 [\rho_1(x_t^i, c_i, \Sigma_x) I_1 \rho_2(x_t^i, c_i, \Sigma_x) \dots I_1 \rho_{I_x}(x_t^i, c_i, \Sigma_x) x_t],$$

donde I_1 es un vector de unos con dimensiones apropiadas.

Desarrollando la aproximación de la función Q (3.9) se tiene que Ω_x se puede escribir como:

$$\Omega_x = \left[\sum_{t=1}^T \langle x_t, S_t^T \rangle_{xx} \right] \left[\sum_{t=1}^T \langle S_t, S_t^T \rangle_{xx} \right]^{-1}, \quad (3.10)$$

donde $\langle . \rangle_{xx}$ se utiliza para denotar la integración con respecto a la función de densidad $P(x_{t-1:t} | y_{1:T}, \Theta)$. Esta integración se puede aproximar mediante la aproximación de las partículas de $P(x_{t-1:t} | y_{1:T}, \Theta)$.

Similarmente se definen:

$$\begin{aligned} \Omega_y &= [p_1 p_2 \dots p_{I_x}] \\ R_t &= I_1 [\varphi_1(x_t^i, d_i, \Sigma_y) I_1 \varphi_1(x_t^i, d_i, \Sigma_y) \dots I_1 \varphi_1(x_t^i, d_i, \Sigma_y) x_t], \end{aligned}$$

donde I_1 es un vector de unos con dimensiones apropiadas.

Desarrollando la aproximación de la función Q (3.9) se tiene que Ω_y se puede escribir como:

$$\Omega_y = \left[\sum_{t=1}^T \langle y_t, R_t^T \rangle_x \right] \left[\sum_{t=1}^T \langle R_t, R_t^T \rangle_{yx} \right]^{-1}, \quad (3.11)$$

donde $\langle . \rangle_x$ se utiliza para denotar la integración con respecto a la función de densidad $P(x_t | y_{1:T}, \Theta)$ y $\langle . \rangle_{yx}$ denota la integración con respecto a la función de densidad $P(y_t | x_t, \Theta)$.

■ Paso 2

En el paso uno se presentó el enfoque para estimar los centros. la idea es obtener una estimación máximo a posteriori de la trayectoria del estado y de los centros que proporcionan las mejores predicciones posibles de estimación de los estados y las observaciones. La estimación máximo a posteriori del estado es obtenido utilizando el algoritmo de Viterbi que se describe continuación.

CAPÍTULO 4

ALGORITMO DE VITERBI

4.1. Algoritmo de Viterbi

4.1.1. Cadenas de Markov

Durante los últimos 30 años, los Modelos de cadenas de Markov, se han transformado en una herramienta de amplio uso en la comunidad científica y como un modelo probabilístico, utilizado para representar la probabilidad conjunta de un conjunto de variables aleatorias. Dado lo anterior, se han desarrollado valiosos aportes a diferentes problemáticas como reconocimiento de voz, Bioinformática, Finanzas y control estocástico.

Dichos modelos también nos permiten modelar la dinámica de un sistema (oculto), al cual no podemos acceder (observar) de forma directa; por el contrario de forma indirecta mediante la observación de eventos externos, suponemos que están correlacionados con dicho sistema y su estado.

Uno de los modelos mas utilizados que describen la dependencia de los estados es la regla de la cadena de Markov. Sean $X = x_1, x_2, \dots, x_n$ una secuencia de estados (vector de estados) y sea $Y = y_1, \dots, y_n$ una secuencia de n vectores de observaciones (vector de observaciones), entonces el modelo de Markov cumpliendo la propiedad de que en una cadena de Markov el conocimiento de los valores de x_1, x_2, \dots, x_{n-1} no agrega información a lo que puede esperarse como valor de x_{n+1} si se sabe el valor de x_n .

Dado esto se supone que:

$$P(x_k | x_{k-1}, x_{k-2}, \dots, x_1) = P(x_k | x_{k-1}). \quad (4.1)$$

El significado de esto es que la dependencia de un estado es limitado solo dentro de dos estados sucesivos. En otras palabras, dado que las observaciones y_1, \dots, y_{k-1} pertenecen a los estados x_1, \dots, x_{k-1} respectivamente, la probabilidad de que la observación y_k pertenezca a el estado x_k solo depende de el estado en el cual la observación y_{k-1} , se ha producido.

Así, usando la regla de la cadena de probabilidad tenemos,

$$\begin{aligned} P(X) &= P(x_1, x_2, \dots, x_n) \\ &= P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \dots P(x_1), \end{aligned}$$

obtenemos

$$P(X) = P(x_1) \prod_{k=1}^n P(x_k | x_{k-1}), \quad (4.2)$$

donde $P(x_1)$ es la probabilidad a priori para el estado x_1 que ocurra. Además,

- Dada la secuencia de los estados, las observaciones son independientes.
- La función de probabilidad en un estado no depende de las otros estados.

$P(Y|X)$ viene dada por:

$$P(Y|X) = \prod_{k=1}^n P(y_k | x_k). \quad (4.3)$$

Luego, combinando (4.2) y (4.3) tenemos:

$$P(Y|X) P(X) = P(x_1) P(y_1 | x_1) \prod_{k=2}^n P(x_k | x_{k-1}) P(y_k | x_k). \quad (4.4)$$

4.1.2. El Algoritmo de Viterbi

El algoritmo de Viterbi es uno de los algoritmos de programación dinámica más utilizados en reconocimiento automático del habla. La idea central de este algoritmo es recorrer el diagrama de transiciones de estados a través del tiempo, almacenando para cada estado la máxima probabilidad acumulada y el estado anterior desde el que se llega con esta probabilidad.

Este algoritmo realiza la detección por máxima similitud, para esto utiliza las propiedades del diagrama de trellis. También intenta reducir la complejidad del cálculo evitando tener en cuenta la totalidad de las secuencias posibles.

Se considera un algoritmo de máxima verosimilitud y altamente paralelizable, consiste en determinar el camino más óptimo expresado como la escogencia de la cadena con la métrica de máxima verosimilitud o la escogencia de la cadena con la métrica de mínima distancia, en la medida que este criterio se va aplicando se evalúa la secuencia que tiene la menor distancia con respecto de la recibida, de forma que la secuencia de máxima verosimilitud finalmente aparece.

El algoritmo de Viterbi permite encontrar la secuencia de estados más probable en un modelo de Markov, x_1, x_2, \dots, x_k a partir de las observaciones y_1, \dots, y_n , es decir, obtiene una secuencia óptima.

En la siguiente figura se muestra por medio de un diagrama de Trellis como el algoritmo de Viterbi encuentra la secuencia más probable.

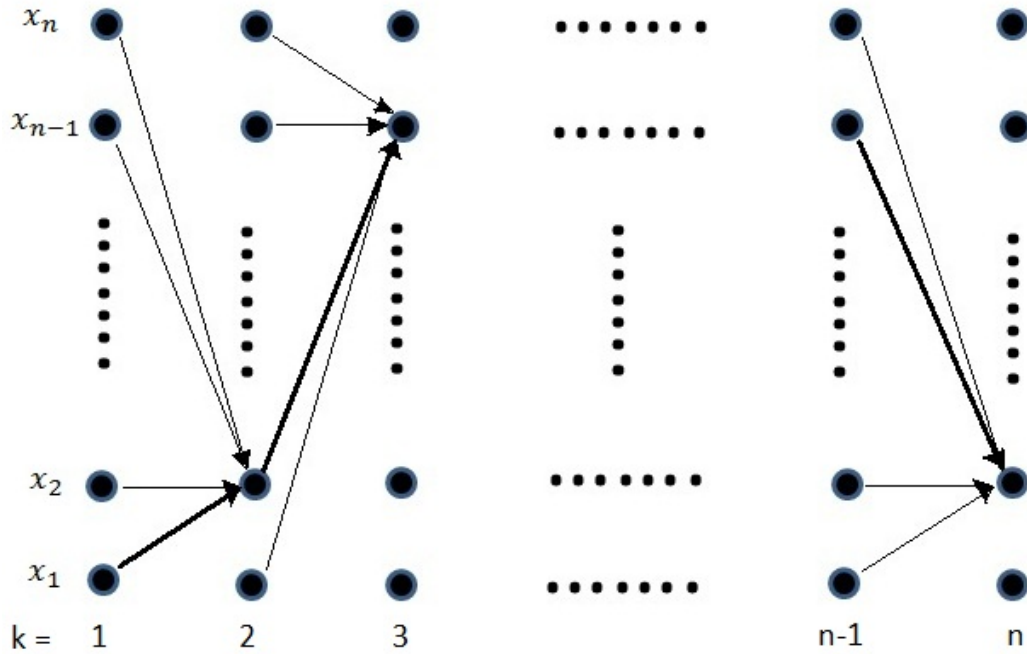


Figura 4.1: Diagrama de Viterbi

Donde las flechas mas resaltadas son las que indican el camino mas probable.

Denotemos como la distancia de un estado x_{k-1} al estado x_k como $\hat{\delta}(x_k|x_{k-1})$, y usando (4.4) se define de la siguiente forma

$$\hat{\delta}(x_k|x_{k-1}) = P(x_k|x_{k-1}) P(y_k|x_k). \tag{4.5}$$

Para la condición inicial $k=1$ se tiene,

$$\hat{\delta}(x_1|x_0) = P(x_1) P(y_1|x_1),$$

así, la distancia de los estados la denotamos por \hat{D} y esta dada por:

$$\hat{D} = \prod_{k=1}^n \hat{\delta}(x_k|x_{k-1}), \tag{4.6}$$

Para que los cálculos sean más sencillos, en lugar de calcular (4.6) se calculará

$$\begin{aligned} \log(\widehat{D}) &= \sum_{k=1}^n \log(\widehat{\delta}(x_k|x_{k-1})) \\ &= \sum_{k=1}^n \delta(x_k|x_{k-1}) \\ &= D. \end{aligned}$$

4.1.3. Algoritmo

Algoritmo 4.1. Algoritmo de viterbi

1. *Inicialización*

Para $1 \leq i \leq n$,

$$\delta_1(i) = \log(P(x_1^{(i)})) + \log(P(y_1|x_1^{(i)})).$$

2. *Recursion*

Para $2 \leq t \leq T$ y $1 \leq j \leq n$,

$$\delta_t(j) = \log(P(y_t|x_t^{(j)})) + \max[\delta_{t-1}(i) + \log(P(x_t^j|x_{t-1}^{(i)}))].$$

$$\psi_t(j) = \arg \max[\delta_{t-1}(i) + \log(P(x_t^j|x_{t-1}^{(i)}))].$$

3. *Terminación*

$$i_T = \arg \max \delta_T^{(i)} \text{ y } x_{MAP}(T) = x_T^{(i_T)}.$$

Donde $x_{MAP}(T)$ = máximo a posteriori.

4. *Backtracking (vuelta atrás)*

$$\text{Para } t = T - 1, T - 2, \dots, 1, \quad i_t = \psi_{t+1}(i_{t+1}) \text{ y } x_{MAP}(t) = x_t(i_t).$$

4.2. Estimación Paramétrica del Modelo Dinámico no Lineal

Los pasos del algoritmo para la estimación de los parámetros del modelo no lineal usando los algoritmos antes desarrollados son:

1. Inicializar.

$$\theta_0 = (\theta_{l_0}, \theta_{nl_0})$$

2. Dados y_1, y_2, \dots, y_t (observaciones) y x_1 inicial, Usamos filtro de Partículas, planteado en el capítulo 2, para obtener los estados x_2, x_3, \dots, x_t del modelo.
3. Luego de tener el vector de estados x_1, x_2, \dots, x_t , usamos el algoritmo de Viterbi para obtener la trayectoria que escoge los mejores centros en la aproximación del modelo.
4. Aplicamos el algoritmo EM desarrollado en el capítulo 3, para obtener θ_l y θ_{nl} .
5. hacemos $\theta_k = (\theta_l, \theta_{nl})$, volver al paso 2 y repetir hasta que $\|\theta_k - \theta_{k-1}\| < tol$.

CAPÍTULO 5

APLICACIÓN DEL ALGORITMO

5.0.1. El Modelo

En esta sección se describe el modelo utilizado para probar el algoritmo antes descrito, para ello, tomamos un modelo donde las funciones $f(\cdot)$ y $g(\cdot)$ del modelo (3.6) son conocidas ya que de esta manera podremos realizar una comparación entre la verdadera solución del modelo con la dada por el pasos del algoritmo planteado.

El modelo económico es el siguiente (Para detalles de este modelo ver ([6])):

$$\begin{aligned}x_{t+1} &= (1 - \rho)\bar{x} + \rho x_t + \sigma\eta\varepsilon_{t+1} \\y_t &= \sum_{i=1}^{100} \beta^i e^{a_i + b_i} (x_t - \bar{x}).\end{aligned}$$

donde

$$a_i = \theta\bar{x}i + \frac{\theta^2\sigma^2\eta^2}{2(1-\rho)^2} \left[i - \frac{2\rho(1-\rho^i)}{1-\rho} + \frac{\rho^{2(1-\rho^{2i})}}{1-\rho^2} \right]$$

y

$$b_i = \frac{\theta\rho(1-\rho^i)}{1-\rho}.$$

5.0.2. Resultados

Para estudiar las bondades del algoritmo propuesto utilizamos el modelo económico mostrado anteriormente, donde su representación espacio estado no-lineal es conocida, de esta forma podemos medir el error cometido al aplicar el algoritmo propuesto.

Iter	0	1	2	3	4	5
h_1	0.0123	0.0106	0.0103	0.0101	0.0091	0.0090
h_2	-0.0123	-0.0106	-0.0101	-0.0102	-0.0103	-0.0103
A	0.9994	0.9995	0.9996	0.9997	0.9997	0.9997
g_1	-7.8082	-6.3056	-6.7056	-6.0816	-6.0810	-6.0811
g_2	7.8082	6.3155	6.4885	6.4811	6.4915	6.4967
C	0.194	0.9995	0.9998	1.001	1.003	1.005
c_1	0.0174	0.0179	0.0118	0.0099	0.0098	0.0099
c_2	0.0051	0.0073	0.0107	0.0109	0.0111	0.0114
LL	105.3374	118.6551	119.8986	120.0345	120.4706	120.4891

Luego de aplicar los pasos del algoritmo, para 5 iteraciones tenemos los siguientes resultados: donde, Iter es el numero de iteraciones, h_1 , h_2 , g_1 y g_2 son el peso de salida de las funciones radiales, c_i y d_i son los centros de las funciones de base radial y LL es el logaritmo de la función de verosimilitud.

Desde la iteración 4 a la iteración 5 se puede observar la convergencia del algoritmo usado, los valores de los parámetros se están acercando mucho y el logaritmo de la función de verosimilitud se maximiza pero muy poco en la ultima iteración.

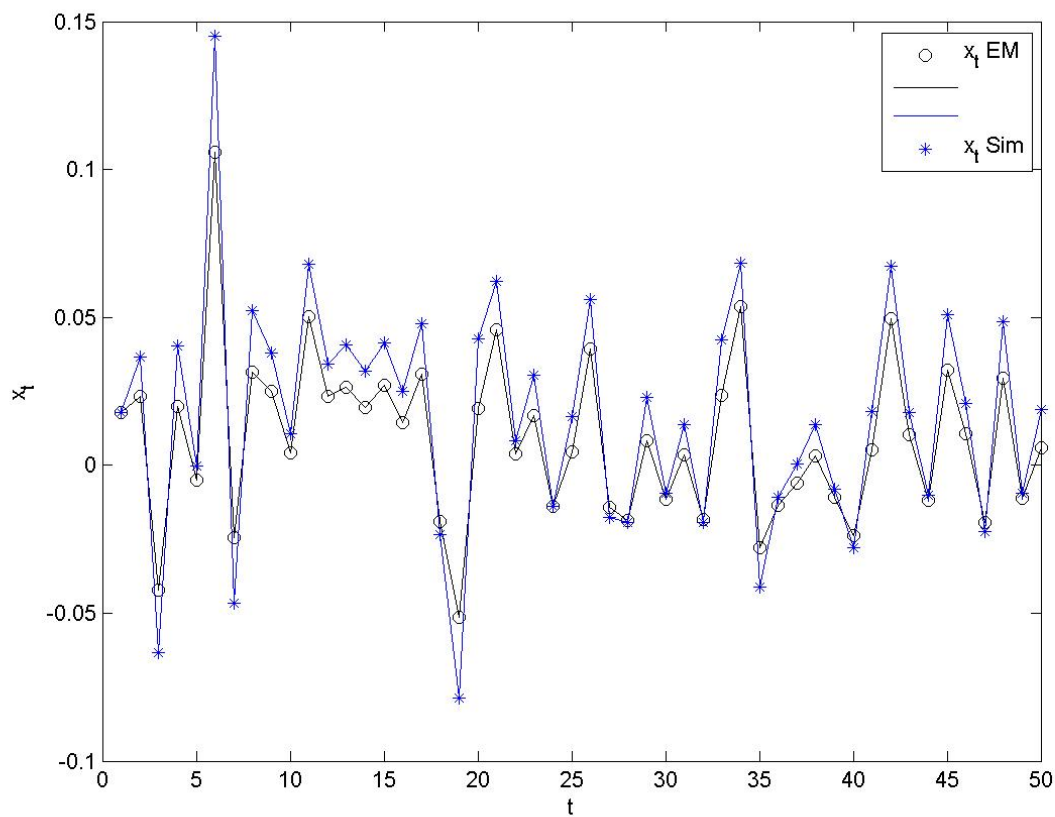


Figura 5.1: Trayectoria variable de estados.

En la figura (5.1) se tiene la trayectoria de los estados para el modelo dado en ([6]) simulado para el modelo real, para esta estimación, utilizamos los siguientes valores: $\beta = 0.95$, $\theta = -1.5$, $\rho = -0.139$, $\sigma = 1$, $\eta = 0.0348$, $\bar{x} = 0.0179$ y $x_1 = 0.0179$ y la trayectoria de los estados dada por el algoritmo EM para el modelo estimado.

Conclusiones

En este trabajo se desarrolla el metodo propuesto por [3] que combina los algoritmos EM, Viterbi y Filtro de Partículas para obtener la identificación de modelos donde la dinámica del proceso es desconocida. En la aplicación de este algoritmo usamos un problema económico conocido para obtener una aproximación de dicho modelo.

En el estudio, se tiene un conjunto de datos que tomamos como vector de observaciones para empezar a construir nuestro modelo, gracias al filtro de Partículas estimamos el vector de estados, luego, obtenemos los mejores x en el vector de estados usando el Algoritmo de Viterbi, estos serán nuestros centros en el modelo y finalmente aplicamos el algoritmo EM para la obtención de los parámetros θ del modelo.

Desde el punto de vista de los resultados dados por el algoritmo podemos observar que la estimación del modelo se aproxima al del modelo real como se puede observar en la imagen del capitulo 5, sin embargo, se debe destacar que esta solución depende mucho del valor inicial de los parámetros, es necesaria una buena inicialización para llevar a cabo el desarrollo de la aplicación obteniendo una buena aproximación, se ha hecho uso del entorno de programación MATLAB.

REFERENCIAS BIBLIOGRÁFICAS

- [1] N.D. CARUSOA, M.C. SANZIEL, M. PORTAPILA Y H. POWER. *Utilizacion de RBFS para la Interpolacion de Datos Batimetricos.*,Argentina, (Noviembre 2009).
- [2] A. P. DEMPSTER, N. M. LAIRD AND D. B. RUBIN. *Maximum Likelihood from Incomplete Data via the EM Algorithm.* Journal of the Royal Statistical Society. (1977).
- [3] R.BHUSHAN GOPALUNI. *Identification of Nonlinear Processes with known Model Structure Under Missing Observations.* The International Federation of Automatic Control.(2008).
- [4] R.BHUSHAN GOPALUNI, ADRIAN G. WILLS AND THOMAS B. SCHON. *Particle Filter Approach to Nonlinear System Identification under Missing Observations with a Real Application.*
- [5] WILLIAN MENDENHALL. *Estadística Matemática con aplicaciones.*
- [6] S. SCHMITT-GROHEA, M. URIBE *Solving dynamic general equilibrium models using a second-order approximation to the policy function,* Journal of Economic Dynamics and Control. (20
- [7] N. SESHADRI AND C.W. SUNDBERG. *List Viterbi Decoding Algorithms with Applications.*, IEEE Transactions on Communications. Febrero (1994).