UNIVERSIDAD DE GRANADA

MODELOS DE MARKOV OCULTOS
Y
APLICACIONES A LA BIOLOGÍA

XUSHENG ZHENG

Trabajo Fin de Grado

Doble Grado en Ingeniería Informática y Matemáticas

Tutores

Lidia Fernández Rodríguez

FACULTAD DE CIENCIAS

E.T.S. INGENIERÍAS INFORMÁTICA Y DE TELECOMUNICACIÓN

Granada, a 3 de diciembre de 2022

ÍNDICE GENERAL

RESUMEN

Nos basamos en el trabajo desarrollado en [Tur36].

Occaecati expedita cumque est. Aut odit vel nobis praesentium dolorem sed eligendi. Inventore molestiae delectus voluptatibus consequatur. Et cumque quia recusandae fugiat earum repellat porro. Earum et tempora vel voluptas. At sed animi qui hic eaque velit.

Saepe deleniti aut voluptatem libero dolores illum iusto iusto. Explicabo dolor quia id enim molestiae praesentium sit. Odit enim doloribus aut assumenda recusandae. Eligendi officia nihil itaque. Quas fugiat aliquid qui est.

Quis amet sint enim. Voluptatem optio quia voluptatem. Perspiciatis molestiae ut laboriosam repudiandae nihil.

Parte I

PARTE DE MATEMÁTICAS

Esta es una descripción de la parte de matemáticas. Nótese que debe escribirse antes del título

1

PRELIMINARES

2

INTRODUCCIÓN A LAS CADENAS DE MARKOV

En este capítulo vamos a inicializar la teoría de cadenas de Markov, avanzado progresivamente hacia las cadenas de Markov ocultas. En primer lugar, vamos a presentar el concepto de procesos de Markov:

2.1 PROPIEDAD DE MARKOV

Sea $\mathbb{S}$ un conjunto finito de forma $\{s_1, ..., s_n\}$, definimos un proceso estocástico sobre $\mathbb{S}$ como una secuencia de variables aleatorias $\{\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, ...\}$, o $\{\mathcal{X}_t\}_{t=0}^{\infty}$ para acortar, donde cada $\mathcal{X}_t$ es una variable aleatoria que toma valores en $\mathbb{S}$.

A pesar de que el índice t puede representar cualquiera magnitud, lo más común es que represente el tiempo. Si tomamos el índice t como el tiempo, obtenemos la noción de "pasado" y "futuro", esto es, si $t < t'$, entonces $\mathcal{X}_t$ es una variable "pasada" para $\mathcal{X}_{t'}$, mientras que $\mathcal{X}_{t'}$ es una variable "futura" para $\mathcal{X}_t$. Sin embargo, esto no es siempre así, por ejemplo, si el proceso estocástico corresponde al de las secuencias de genomas de un organismo, el conjunto $\mathbb{S}$ estará formado por los cuatro símbolos para las subunidades de nucleótidos $\{A, C, G, T\}$ y las secuenciaciones tienen un significado más espacial que temporal.

Definición 2.1. Un proceso estocástico $\{\mathcal{X}_t\}_{t=0}^{\infty}$ se dice que posee **la propiedad de Markov**, o es un **proceso de Markov**, si para todo $t \geq 1$ y $(u_0, ..., u_{t-1}, u_t) \in \mathbb{S}^{t+1}$ se tiene que:

$$P[\mathcal{X}_t = u_t | \mathcal{X}_0 = u_0, ..., \mathcal{X}_{t-1} = u_{t-1}] = P[\mathcal{X}_t = u_t | \mathcal{X}_{t-1} = u_{t-1}] \tag{2.1}$$

La propiedad de Markov afirma que las distribuciones condicionadas del "estado actual" $\mathcal{X}_t$ depende únicamente del "pasado inmediato" $\mathcal{X}_{t-1}$ y no depende de ninguno de los anteriores estados.

Por conveniencia, introducimos la notación $\mathcal{X}_j^k$ para denotar los estados $\mathcal{X}_i$ con $j \leq i \leq k$. Con esta notación, podemos reescribir la Definición 1.1 como sigue: Un proceso estocástico $\{\mathcal{X}_t\}$ es un **proceso de Markov** si, para todo $(u_0, ..., u_{t-1}, u_t) \in \mathbb{S}^{t+1}$ es cierto que:

$$P[\mathcal{X}_t = u_t | \mathcal{X}_0^{t-1} = u_0...u_{t-1}] = P[\mathcal{X}_t = u_t | \mathcal{X}_{t-1} = u_{t-1}] \tag{2.2}$$

Para cualquier proceso estocástico $\{\mathcal{X}_t\}$ y cualquiera secuencia $(u_0, ..., u_{t-1}, u_t) \in \mathbb{S}^{t+1}$, tenemos que por definición de probabilidad condicionada:

$$P[\mathcal{X}_0^t = u_0...u_t] = P[\mathcal{X}_0 = u_0] \cdot \prod_{i=0}^{t-1} P[\mathcal{X}_{i+1} = u_{i+1} | \mathcal{X}_0^i = u_0...u_i]$$

Sin embargo, si consideramos un proceso de Markov, entonces la fórmula anterior se reduce a:

$$P[\mathcal{X}_0^t = u_0...u_t] = P[\mathcal{X}_0 = u_0] \cdot \prod_{i=0}^{t-1} P[\mathcal{X}_{i+1} = u_{i+1} | \mathcal{X}_i = u_i] \tag{2.3}$$

En probabilidad, es usual referirse con el nombre **cadena de Markov** a un proceso de Markov $\mathcal{X}_t$ donde el parámetro t toma únicamente valores discretos. En este trabajo, pondremos nuestra atención en los casos donde t toma valores en $\mathbb{Z}_+$.

En (2.3) vemos la importancia del valor:

$$P[\mathcal{X}_{t+1} = u | \mathcal{X}_t = v]$$

al que podemos identificar como una función de tres variables: el estado "actual" $v \in \mathbb{S}$, el estado "siguiente" $u \in \mathbb{S}$ y el "tiempo actual" $t \in \mathbb{Z}_+$. Así, teniendo en cuenta que $\mathbb{S} = \{s_1, ..., s_n\}$ definimos para todo tiempo $t \in \mathbb{Z}_+$:

$$a_{ij}(t) := P[\mathcal{X}_{t+1} = s_i | \mathcal{X}_t = s_j] \tag{2.4}$$

Por tanto, $a_{ij}(t)$ es la probabilidad de realizar una transición desde el estado actual s_j al estado siguiente s_i en el instante t.

Definición 2.2. Sea $\mathcal{X}_t$ una cadena de Markov, la matriz cuadrada de dimensión n, $A(t) = [a_{ij}(t)]$, es la **matriz de transición** de $\mathcal{X}_t$ en el instante t. Una cadena de Markov es **homogénea** si $A(t)$ es constante para todo $t \in \mathbb{Z}_+$, en otro caso, es **no homogénea**.

Definición 2.3. Sea $\mathcal{X}_t$ una cadena de Markov tomando valores en un conjunto finito $\mathbb{S} = \{s_1, ..., s_n\}$, y sea $A(t)$ su matriz de transición en el instante t. Entonces, $A(t)$ es una **matriz estocástica** (por columna) para todo t, esto es:

$$a_{ij}(t) \in [0,1], \forall i, j \in \{1, ..., n\}, t \in \mathbb{Z}_+$$
$$\sum_{i=1}^{n} a_{ij}(t) = 1, \forall j \in \{1, ..., n\}, t \in \mathbb{Z}_+$$

Lema 2.4. Sea A una matriz positiva de dimensión $n \times n$:

1. A es estocástica si y solo si 1 es un valor propio de A^t con vector propio $\mathbf{1} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^t$.
2. si A es estocástica, entonces para todo valor propio λ, se cumple que $|\lambda| \leq 1$.

Demostración.

1. Es suficiente con observar que la condición de estocaticidad para una matriz positiva A es equivalente a que $A^t \cdot \mathbf{1} = \mathbf{1}$ y tener en cuenta que A y A^t tienen los mismos valores propios.
2. Sea v un vector propio asociado a λ, por ser A positiva y estocástica, tenemos que:
$$|\lambda| \sum_{i=1}^{n} |v_i| = \sum_{i=1}^{n} |\lambda v_i| = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} v_j \right| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} |v_j| =$$
$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij} \right) |v_j| = \sum_{j=1}^{n} |v_j|$$
Puesto que $\sum_{r=1}^{n} |v_r| \neq 0$, tenemos que $|\lambda| \leq 1$.

Para continuar con los estudios de las cadenas de Markov presentamos el siguiente conjunto:

Definición 2.5. El **n-símplex estándar** es el subconjunto de $\mathbb{R}^{n+1}$ dado por:

$$\Delta^n = \{(t_1, ..., t_{n+1})^t \in \mathbb{R}^{n+1} \mid \sum_{i=1}^{n+1} t_i = 1 \text{ y } t_i \geq 0 \text{ para todo } i\}$$

Lema 2.6. Sea $\{\mathcal{X}_t\}$ una cadena de Markov con valores en $\mathbb{S} = \{s_1, ..., s_n\}$, y sea $A(t)$ su matriz de transición en el instante t. Supongamos que el estado inicial $\mathcal{X}_0$ se distribuye de acuerdo con $c^0 \in \Delta^{n-1}$, esto es:

$$P[\mathcal{X}_0 = s_i] = c_i^0, \forall i \in \{1, ..., n\}$$

Entonces, para todo $t \geq 0$, el estado $\mathcal{X}_t$ se distribuye de acuerdo con:

$$c^t = A(t-1) \cdots A(1) A(0) c^0 \tag{2.5}$$

Demostración. Sea $s_i \in \mathbb{S}$, por 2.3 tenemos que:

$$P[\mathcal{X}_t = s_i] =$$
$$\sum_{u_0 u_1 \dots u_{t-1} \in \mathbb{S}^t} P[\mathcal{X}_0 = u_0] \cdot \prod_{i=0}^{t-2} P[\mathcal{X}_{i+1} = u_{i+1} | \mathcal{X}_i = u_i] \cdot P[\mathcal{X}_t = s_i | \mathcal{X}_{t-1} = u_{t-1}]$$

$$= \sum_{u_0 u_1 \ldots u_{t-1} \in \mathbb{S}^t} c_{u_0} \cdot a_{u_1,u_0}(0) \cdots a_{u_{t-1},u_{t-2}}(t-2) \cdot a_{s_i,u_{t-1}}(t-1)$$
$$= \sum_{u_0 u_1 \ldots u_{t-1} \in \mathbb{S}^t} a_{s_i,u_{t-1}}(t-1) \cdot a_{u_{t-1},u_{t-2}}(t-2) \cdots a_{u_1,u_0}(0) \cdot c_{u_0}$$

Notemos que esta última expresión es justamente la componente i-ésima de $c^t = A(t-1)\cdots A(1)A(0)c^0$ escrita en forma extensa. □

Ejemplo 1. En este ejemplo presentamos una variación del juego de cartas "blackjack". En este caso, tenemos un dado de cuatro caras con valores 0, 1, 2 y 3, y con probabilidad uniforme en cada lanzamiento. Un jugador lanza el dado de forma repetida y $\mathcal{X}_t$ representa el valor acumulado tras t lanzamientos. Si el total es igual a nueve, el jugador gana; en otro caso se considera que pierde. Podemos asumir que el resultado de cada lanzamiento es independiente de los lanzamientos anteriores.

Tenemos entonces que $\{\mathcal{X}_t\}$ toma valores en el conjunto $\mathbb{S} := \{0, 1, ..., 8, W, L\}$ de cardinalidad 11. Sea $\mathcal{Y}_t$ el resultado del lanzamiento en el instante t:

$$P[\mathcal{Y}_t = 0] = P[\mathcal{Y}_t = 1] = P[\mathcal{Y}_t = 2] = P[\mathcal{Y}_t = 3] = 1/4$$

Examinemos ahora la distribución de $\mathcal{X}_t$, por definición sabemos que $\mathcal{X}_t = \mathcal{X}_{t-1} + \mathcal{Y}_t$, excepto el caso de que $\mathcal{X}_{t-1} + \mathcal{Y}_t = 9$, consideraremos $\mathcal{X}_t = W$ (ganar); y si $\mathcal{X}_{t-1} + \mathcal{Y}_t > 9$, consideraremos $\mathcal{X}_t = L$ (perder). Si $\mathcal{X}_{t-1} = W$ o L, consideraremos que el juego está acabado y $\mathcal{X}_t = \mathcal{X}_{t-1}$. Estas observaciones se pueden resumir en las siguientes reglas:

- Si $\mathcal{X}_{t-1} \leq 5$:
$$P[\mathcal{X}_t = \mathcal{X}_{t-1}] = P[\mathcal{X}_t = \mathcal{X}_{t-1} + 1] =$$
$$= P[\mathcal{X}_t = \mathcal{X}_{t-1} + 2] = P[\mathcal{X}_t = \mathcal{X}_{t-1} + 3] = 1/4$$
- Si $\mathcal{X}_{t-1} = 6$:
$$P[\mathcal{X}_t = 6] = P[\mathcal{X}_t = 7] = P[\mathcal{X}_t = 8] = P[\mathcal{X}_t = W] = 1/4$$
- Si $\mathcal{X}_{t-1} = 7$:
$$P[\mathcal{X}_t = 7] = P[\mathcal{X}_t = 8] = P[\mathcal{X}_t = W] = P[\mathcal{X}_t = L] = 1/4$$
- Si $\mathcal{X}_{t-1} = 8$:
$$P[\mathcal{X}_t = 8] = P[\mathcal{X}_t = W] = 1/4$$
$$P[\mathcal{X}_t = L] = 1/2$$
- Si $\mathcal{X}_{t-1} = W$ o L:
$$P[\mathcal{X}_t = \mathcal{X}_{t-1}] = 1$$

$\{\mathcal{X}_t\}$ es una cadena de Markov pues la distribución de $\mathcal{X}_t$ depende únicamente del valor de $\mathcal{X}_{t-1}$ y no de cómo se ha alcanzado dicho valor. Notemos que las probabilidades anteriores no dependen de t, con lo cual la matriz de transición de $\mathcal{X}_t$ es una matriz fija y $\mathcal{X}_t$ es homogénea.

La matriz de transición de $\mathcal{X}_t$ es entonces una matriz 11×11 dado por:

$$A = \begin{pmatrix} 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 1/4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 1 \end{pmatrix}$$

Es natural que el juego comience con el valor inicial igual a cero. Por lo tanto, la distribución de $\mathcal{X}_0$ está representada por $c_0 \in \mathbb{R}^{11}$ con un 1 en la primera componente y ceros en el resto. Aplicando repetidamente la fórmula 2.5 obtendremos las distribuciones de $\mathcal{X}_1$, $\mathcal{X}_2$, etc. Así, sea c_t la distribución de $\mathcal{X}_t$, tenemos:

$$c_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad c_1 = Ac_0 = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \\ \vdots \\ 0 \end{pmatrix} \qquad c_2 = Ac_1 = \begin{pmatrix} 1/16 \\ 1/8 \\ 3/16 \\ 1/4 \\ 3/16 \\ 1/8 \\ 1/16 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Cabe destacar que si examinamos la distribución c_t, entonces tenemos que $P[\mathcal{X}_t \in \{0..,8\}]$ tiende a cero conforme $t \to \infty$. Esto es natural pues el juego terminará eventualmente en victoria (W) o en pérdida (L) y todos los otros estados son transitorias. Esta idea, nos introduce al siguiente apartado.

2.2 DINÁMICAS DE CADENAS DE MARKOV ESTACIONARIAS

En esta sección, vamos a centrarnos en el estudio de dinámicas de cadenas de Markov cuyas matrices de transición son constantes. Comenzamos con la siguiente definición:

Definición 2.7. Sea A una matriz estocástica de dimensión $n \times n$, un vector $\pi \in \Delta^{n-1}$ es una distribución estacionaria de A si $A\pi = \pi$.

Supongamos que $\{\mathcal{X}_t\}$ es una cadena de Markov con matriz de transición A. Hemos visto que, dependiendo de la distribución inicial, el proceso resultante $\{\mathcal{X}_t\}$ puede tener distintas distribuciones a lo largo del tiempo. Sin embargo, si A tiene una distribución estacionaria π, entonces $A^t\pi = \pi$ para todo t. Por ello, si $\mathcal{X}_0$ tiene la distribución π, por 2.5, $\mathcal{X}_t$ tiene también la distribución π para todo t. Podemos observar así la importancia de las distribuciones estacionarias.

Parte II

PARTE DE INFORMÁTICA

3

SECCIÓN TERCERA

El siguiente código es un ejemplo de coloreado de sintaxis e inclusión directa de código fuente en el texto usando `minted`.

```haskell
-- From the GHC.Base library.
class  Functor f  where
    fmap        :: (a -> b) -> f a -> f b

    -- | Replace all locations in the input with the same value.
    -- The default definition is @'fmap' . 'const'@, but this may be
    -- overridden with a more efficient version.
    (<$)        :: a -> f b -> f a
    (<$)        =  fmap . const

-- | A variant of '<*>' with the arguments reversed.
(<**>) :: Applicative f => f a -> f (a -> b) -> f b
(<**>) = liftA2 (\a f -> f a)
-- Don't use \$ here, see the note at the top of the page

-- | Lift a function to actions.
-- This function may be used as a value for `fmap` in a `Functor` instance.
liftA :: Applicative f => (a -> b) -> f a -> f b
liftA f a = pure f <*> a
-- Caution: since this may be used for `fmap`, we can't use the obvious
-- definition of liftA = fmap.

-- | Lift a ternary function to actions.
liftA3 :: Applicative f => (a -> b -> c -> d) -> f a -> f b -> f c -> f d
liftA3 f a b c = liftA2 f a b <*> c


{-# INLINABLE liftA #-}
{-# SPECIALISE liftA :: (a1->r) -> IO a1 -> IO r #-}
{-# SPECIALISE liftA :: (a1->r) -> Maybe a1 -> Maybe r #-}
{-# INLINABLE liftA3 #-}
```

```
{-# SPECIALISE liftA3 :: (a1->a2->a3->r) -> IO a1 -> IO a2 -> IO a3 -> IO r #-}
{-# SPECIALISE liftA3 :: (a1->a2->a3->r) ->
                                Maybe a1 -> Maybe a2 -> Maybe a3 -> Maybe r #-}

-- | The 'join' function is the conventional monad join operator. It
-- is used to remove one level of monadic structure, projecting its
-- bound argument into the outer level.
join              :: (Monad m) => m (m a) -> m a
join x            =  x >>= id
```

Vivamus fringilla egestas nulla ac lobortis. Etiam viverra est risus, in fermentum nibh euismod quis. Vivamus suscipit arcu sed quam dictum suscipit. Maecenas pulvinar massa pulvinar fermentum pellentesque. Morbi eleifend nec velit ut suscipit. Nam vitae vestibulum dui, vel mollis dolor. Integer quis nibh sapien.

BIBLIOGRAFÍA

[Tur36] Alan M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42):230–265, 1936.