

# HMMS & ANN

Juan Tornero Lucas

23 de febrero de 2017

## 1. Marco Teórico de las Cadenas de Markov

### Secuencias

Sea un *alfabeto*  $X = \{0, 1, \dots, m-1\}$  de *símbolos*, con  $m$  un natural. Si  $x \in X^{\mathbb{Z}}$ , entonces  $x = \dots x_{-1}x_0x_1\dots$  secuencia infinita. Donde los índices  $i < 0$  denotan el pasado de la secuencia, y los  $i \geq 0$  el futuro, particularmente el índice  $i = 0$  es el primer símbolo desconocido de la secuencia.

En estos términos definimos una palabra  $w \in X^l$  de longitud  $l$ , como una  $l$ -tupla de  $X$ .  $\phi$  denotará la palabra vacía de longitud 0. Una subsecuencia  $s$  es una estructura  $s = (w, (a, b))$ , donde  $w$  es una palabra y  $a, b \in \mathbb{Z}$  tq  $|w| = b - a + 1$ . Así,  $s$  puede también escribirse  $s = s_a \dots s_b$ , y  $a$  representaría el tiempo inicial, y  $b$  el tiempo final. Una secuencia  $x$  contendrá a la subsecuencia  $s$  si  $\forall t \in [a, b], s_t = x_t$ .

El conjunto  $A_s = \{x \in X^{\mathbb{Z}} | x_i = s_i \forall i \in [a, b]\}$  es el conjunto de la secuencias de  $X^{\mathbb{Z}}$  que contienen a  $s$ , si por ejemplo  $s = (\phi, (a, a-1))$ , entonces  $A_s = X^{\mathbb{Z}}$ .

El conjunto  $X^*$  denotará el de todas las palabras.

### Procesos

Un proceso es una medida de probabilidad estacionaria en un espacio de secuencias.

Una medida de probabilidad es una función que asigna probabilidades a conjuntos (el espacio medible de probabilidades), en este caso a conjuntos de secuencias. Sea  $\chi$  la menor colección de subconjuntos de  $X^{\mathbb{Z}}$  tq:

1. Para toda secuencia  $s$ ,  $A_s \in \chi$ .
2.  $\chi$  es cerrado bajo complementos y uniones contables.

El par  $(X^{\mathbb{Z}}, \chi)$  es el espacio medible que asigna probabilidades a los conjuntos  $A_s$  fijados por las subsecuencias  $s$ . Definiremos entonces  $P(s) = P(A_s)$ , tenemos en particular:

$$P(\phi) = P(X^{\mathbb{Z}})$$

En nuestra definición de proceso nos referíamos también a ellos como estacionarios. Si  $D$  es la función desplazamiento  $D : X^{\mathbb{Z}} \rightarrow X^{\mathbb{Z}}$ , que actúa sobre todo  $x \in X^{\mathbb{Z}}$  de manera que  $D(x_t) = x_{t+1}$ , es decir desplaza el tiempo de origen.

Decimos que  $P$  es una medida de probabilidad estacionaria si  $\forall A \in \chi, P(D(A)) = P(A)$ , como  $D$  es de hecho un automorfismo sobre  $X^{\mathbb{Z}}$ , entonces  $P(D^{-1}(A)) = P(D^{-1}(D(A))) = P(A)$

Finalmente podemos definir de forma más formal un proceso  $Q$  Como el espacio de probabilidades estacionario  $(X^{\mathbb{Z}}, \chi, P)$ .

Sea  $w$  una palabra, entonces si  $P$  estacionario  $P(w) = P(s)$ . Además de manera trivial podemos obtener que, si  $W_l$  es el conjunto de las palabras de longitud  $l > 0$ :

$$\sum_{z \in W_l} P(wz) = \sum_{z \in W_l} P(zw) = P(w) \quad (1)$$

Además, para  $w = \phi$ :

$$P(\phi) = 1 \quad (2)$$

El recíproco también es cierto, y cualquier función de  $X^*$  que satisfaga (1) y (2) define un proceso.

**Teorema** Dado  $f : X^* \rightarrow [0, 1]$  que satisface:

$$1. f(\phi) = 1$$

$$2. \forall w \in X^*, f(w) = \sum_{z \in X^*} f(zw) = \sum_{z \in X^*} f(wz)$$

,

existe un único proceso  $Q = (X^{\mathbb{Z}}, \chi, P)$  tq  $\forall w \in X^*, P(w) = f(w)$ .

Demostración ...

*Ejemplo de Proceso.* Para ilustrar el concepto de proceso, pondremos un pequeño ejemplo basado en un dado 'ideal' de seis caras,  $X = \{1, 2, 3, 4, 5, 6\}$  y cualquier palabra de longitud  $l$  tendrá asociada una probabilidad de  $P(w) = \frac{1}{6}^l$ . En particular, si la palabra es de longitud  $l = 0$ , entonces  $P(w) = \frac{1}{6}^0 = 1$ . Si además tenemos

otra palabra  $z \in X$  entonces, a partir de (1) vemos que  $f(wz) = f(zw) = \frac{1}{6}^l \left( \sum_{i=1}^6 \frac{1}{6} \right) = \frac{1}{6} = f(w)$ . Por lo tanto cumple tanto (1) como (2) y es un proceso. *cqd*

## Estados de un Proceso

Supongamos que tenemos un proceso  $Q = (X^{\mathbb{Z}}, \chi, P)$  del que conocemos algunos símbolos recientes. Si  $s = (w, (a, b))$  es una subsecuencia conocida, con  $b = -1$ ,  $s$  induce una distribución condicional en un espacio futuro (DCF)  $(\chi, F)$  del tipo  $P(\cdot|s) = P(\cdot|A_s)$ . Esta CDF condensa todo nuestro conocimiento sobre las posibilidades futuras del proceso  $Q$ , por lo que al conocer esta distribución, en cierto sentido, podemos olvidar todo lo acontecido en el proceso anteriormente. Esta es la base con la que decimos que  $P(\cdot|s)$  es un estado.

Ahora ampliaremos esta noción. Primero definimos  $\chi^+ = \{x^+ = x_0, x_1, \dots | \forall i \geq 0, x_i \in \chi\}$ , y respectivamente  $\chi^- = \{x^- = \dots, x_{-2}, x_{-1} | \forall i \leq 0, x_i \in \chi\}$ . Las CDF están condicionadas por secuencias  $s = (w, (a, b))$  con  $w$  palabra y  $w \in X^+$ , por lo que estas subsecuencias pueden o no acabar en la última observación conocida (también conocidas como *historias*), pero nos centraremos únicamente en este último caso, a este tipo de subsecuencias las llamaremos.

Si dos historias,  $z$  e  $y$  acaban determinando la misma DCF  $P(\cdot|z) = P(\cdot|y)$ , entonces decimos que  $y$  y  $z$  son equivalentes, es decir podemos definir la clase de equivalencia  $y \sim z = \{y \in X^+ | y \sim z\}$ .

Sea  $\pi$  la función proyección sobre las clases de equivalencia de  $\chi^+$ , es decir  $\pi(z) = \bar{z}$ . Para toda palabra futura  $w$  se cumplirá entonces  $P(s|z) = P(s|\bar{z})$ .

Utilizando las clases de equivalencia podemos olvidarnos de la historia concreta que ha generado la DCF. Así, imaginemos la historia  $w$ , cuya clase de equivalencia  $\bar{w}$  induce una DCF sobre  $x_0$ . Así, si escogemos un símbolo  $k$  adecuado para  $x_0$ , tenemos la nueva palabra  $z = wk$ , y tenemos entonces de nuevo una clase de equivalencia  $\bar{z}$  que inducirá una DCF sobre  $x_1$ , a las diferentes clases de equivalencia  $\bar{z}$  que se obtienen a partir de las DCF inducidas por  $\bar{w}$  las llamamos *estados del proceso*.

Usando el ejemplo del dado, en este caso todas las historias son equivalentes, luego hay un solo estado del proceso, el cual es visitado recurrentemente. Daremos a continuación una definición más formal de los estados de un proceso.

Sea  $s_l = x_{-l} \dots x_{-1}$ , con  $l > 0$  y  $R_l = \{s_l | P(s_l) > 0\}$ . Dada la historia  $s_l$ , para cualquier evento futuro  $A$ :

$$P(A|s_l) = \frac{P(A \cap A_{s_l})}{P(A_{s_l})},$$

En concreto, si  $s_l = x^-$ :

$$P(A|x^-) = \lim_{l \rightarrow -\infty} P(A|s_l).$$

Si  $x^-$  tiene una historia  $s_l$  con probabilidad nula, entonces  $P(A|x^-) = 0$ . Denominaremos al conjunto de historias con probabilidad nula  $N = \{x^- \in X^{\mathbb{Z}} | P(x^-) = 0\}$ .

Un estado de proceso es un DCF que se da al condicionar en una historia de probabilidad no nula. El conjunto de todos los estados del proceso para un proceso dado es:

$$PS = \{P(\cdot|w) | w \in R\} \cup \{P(\cdot|x^-) \in \chi^- \setminus N\}$$