

Conexión entre la actualización bayesiana y las cadenas de Markov

En la sección anterior (de estadística bayesiana), vimos cómo la inferencia basada en el **Teorema de Bayes** permite **actualizar** la distribución de nuestra creencia sobre un parámetro conforme llegan nuevos datos. Ahora, mostraremos que el proceso de **multiplicar repetidamente** una distribución inicial por la matriz de transición de una **cadena de Markov** es, en esencia, un **proceso bayesiano de actualización iterativa** cuando el estado es directamente observable.

Nota: Aquí nos referimos a **cadenas de Markov estándar** (no ocultas). En un **Modelo Oculto de Markov (HMM)**, el estado no se observa directamente y el análisis se complica, pero la idea básica de actualización probabilística sigue presente, solo que requiere pasos adicionales (filtrado, etc.).

1. Breve recordatorio de cadenas de Markov y notación

Consideremos una **cadena de Markov** de primer orden con un conjunto finito de estados $\{s_1, s_2, \dots, s_n\}$.

1. Distribución inicial:

- Denotada como $\alpha^{(0)}$.
- Es un **vector de probabilidad** de dimensión $1 \times n$,
 $\alpha^{(0)} = (\alpha^{(0)}(s_1), \alpha^{(0)}(s_2), \dots, \alpha^{(0)}(s_n))$,
donde $\alpha^{(0)}(s_i)$ es la probabilidad de que el proceso comience en el estado s_i .

2. Matriz de transición T :

- Es una **matriz cuadrada** de tamaño $n \times n$.
- Denotamos sus entradas como $T_{i,j}$, donde cada $T_{i,j}$ indica la probabilidad de pasar del estado s_i al estado s_j en un solo paso:
 $T_{i,j} = P(\text{estado siguiente} = s_j \mid \text{estado actual} = s_i)$.
- Filas:** la fila i de T (es decir, $(T_{i,1}, T_{i,2}, \dots, T_{i,n})$) es la **distribución** del siguiente estado si el estado actual es s_i .
- Cada fila i **suma 1**, pues las probabilidades de transición desde s_i a todos los posibles estados se reparten el total de 1.

3. Evolución en el tiempo:

- Sea $\alpha^{(t)}$ la **distribución** sobre los estados en el tiempo t . (Es también un vector fila $1 \times n$.)
- La regla de evolución de la cadena de Markov dice que, en cada paso, aplicamos la matriz de transición:
 $\alpha^{(t+1)} = \alpha^{(t)} T$.
- Explícitamente, para cada estado s_j ,
 $\alpha^{(t+1)}(s_j) = \sum_{i=1}^n \alpha^{(t)}(s_i) \times T_{i,j}$,
lo cual es un **ejemplo directo** de la **ley de la probabilidad total**: se suman las

probabilidades de venir de cada estado s_i ponderado por la probabilidad de transición a s_j .

2. Situación: Conocemos T , pero no sabemos el estado inicial

Supongamos que tenemos certeza absoluta sobre la **matriz de transición** T (esto es, conocemos perfectamente las probabilidades de pasar de un estado a otro), **pero no** sobre el **estado inicial** de la cadena:

- Podríamos decir que "no sabemos si la cadena arranca en s_1 , s_2 , etc."
- Por lo tanto, **asignamos una prior** sobre el estado inicial: por ejemplo, una prior uniforme $\alpha^{(0)}$ o cualquier vector de probabilidad que refleje nuestras creencias iniciales.

Una vez elegida esa prior, **¿cómo evoluciona nuestra creencia sobre el estado** de la cadena con el tiempo? Basta con **multiplicar** sucesivamente $\alpha^{(t)}$ por T .

2.1. Interpretación bayesiana

1. **Prior:** $\alpha^{(t)}$ representa nuestras creencias sobre en qué estado se encuentra la cadena en el paso t .
2. **"Likelihood" de transición:** El hecho de que $X_t = s_i$ pase a $X_{t+1} = s_j$ con probabilidad $T_{i,j}$ se comporta como la "verosimilitud" de obtener s_j desde s_i .
3. **Posterior:** Al combinar (mediante la ley de la probabilidad total) la distribución previa $\alpha^{(t)}$ con la transición dada por T , obtenemos la **nueva distribución** $\alpha^{(t+1)}$.

En otros términos, la multiplicación $\alpha^{(t)} \times T$ es la análoga de:

$$\text{Posterior} = \text{Prior} \times \text{Likelihood},$$

donde no necesitamos un factor de normalización extra porque cada fila de T ya suma 1 (en un problema bayesiano general, ese factor extra aparece explícitamente en el denominador " $\sum_{\theta} \dots$ ").

3. Multiplicación repetida y convergencia a la distribución estacionaria

En una **cadena de Markov ergódica** (irreducible y aperiódica), existe una **distribución estacionaria** π tal que

$$\pi = \pi T,$$

y para cualquier distribución inicial $\alpha^{(0)}$, la iteración $\alpha^{(t+1)} = \alpha^{(t)} T$ **converge** a π conforme $t \rightarrow \infty$.

3.1. Interpretación estadístico-bayesiana de la convergencia

- Si seguimos actualizando indefinidamente (multiplicando por T), la secuencia de distribuciones $\{\alpha^{(t)}\}$ se "mezcla" hasta estabilizarse en π .

- Esta π es el **estado estacionario** que describe la probabilidad a largo plazo de estar en cada estado.
- Desde un punto de vista bayesiano, equivale a decir que, partiendo de una prior muy incierta, tras muchas transiciones (y asumiendo que observamos o conocemos la regla de evolución en cada paso), nuestra creencia llega a estabilizarse en π .

Atención: Si la cadena **no** es ergódica (por ejemplo, si la matriz de transición no conecta todos los estados o hay periodicidad), es posible que no haya convergencia a una única distribución o que existan múltiples distribuciones estacionarias. De manera análoga, en un escenario bayesiano con información incompleta o no identificable, puede que la actualización no conduzca a un único "consenso".

4. "Truco" de la multiplicación de matrices como actualización bayesiana

En estadística bayesiana, la fórmula de actualización (en su forma más general) es:

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}.$$

Si enfocamos la cadena de Markov como un proceso donde θ es "el estado actual" y " D " consiste en "avanzar un paso al siguiente estado", la matriz T guarda las probabilidades condicionales de la transición. Entonces:

$$\alpha^{(t+1)} = \underbrace{\alpha^{(t)}}_{\text{prior sobre } X_t} \times \underbrace{T}_{\text{likelihood de } X_{t+1}}, (\text{implícitamente normalizado, pues cada fila de } T \text{ suma } 1).$$

4.1. Comparación conceptual

- **Bayes clásico:**
 $\text{Posterior} \propto \text{Likelihood} \times \text{Prior}.$
- **Cadena de Markov:**
 $\alpha^{(t+1)} = \alpha^{(t)} \times T.$

PROF

La multiplicación vector-fila $\alpha^{(t)}$ por la matriz T (cuyas filas suman 1) encapsula la misma **lógica**: estamos combinando la distribución previa con las probabilidades condicionales de transición para obtener la distribución posterior.

5. Ejemplo ilustrativo (muy breve)

Imaginemos una cadena de Markov con 3 estados $\{s_1, s_2, s_3\}$ y la matriz de transición:

$$T = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}.$$

5.1. Prior desconocida sobre el estado inicial

No sabemos en qué estado arrancó la cadena, así que fijamos la prior inicial $\alpha^{(0)} = (0.3, 0.5, 0.2)$. Es decir:

- 30% de probabilidad de iniciar en s_1 ,
- 50% en s_2 ,
- 20% en s_3 .

5.2. Un paso de actualización

Para ir al **siguiente** instante (tiempo $t=1$):

$$\alpha^{(1)} = \alpha^{(0)} T.$$

Calculándolo explícitamente (multiplicación fila \times matriz):

$$\begin{aligned} \alpha^{(1)}(s_1) &= 0.3 \cdot 0.7 + 0.5 \cdot 0.3 + 0.2 \cdot 0.2 \\ &= 0.21 + 0.15 + 0.04 \\ &= 0.40, \\ \alpha^{(1)}(s_2) &= 0.3 \cdot 0.2 + 0.5 \cdot 0.4 + 0.2 \cdot 0.3 \\ &= 0.06 + 0.20 + 0.06 \\ &= 0.32, \\ \alpha^{(1)}(s_3) &= 0.3 \cdot 0.1 + 0.5 \cdot 0.3 + 0.2 \cdot 0.5 \\ &= 0.03 + 0.15 + 0.10 \\ &= 0.28. \end{aligned}$$

Así, $\alpha^{(1)} = (0.40, 0.32, 0.28)$. Eso es, en esencia, nuestra **posterior** sobre el estado tras un paso de transición.

PROF

5.3. Convergencia a la distribución estacionaria

Si repetimos $\alpha^{(t+1)} = \alpha^{(t)} T$ muchas veces (en esta cadena, que es ergódica), obtendremos una distribución π que satisface $\pi = \pi T$. Esta π es la **distribución estacionaria**, y a ella convergen todas las distribuciones iniciales $\alpha^{(0)}$. De modo análogo, en un escenario bayesiano, si seguimos recibiendo evidencia coherente, nuestra **posterior** puede estabilizarse en una región de alta verosimilitud.

6. Conclusión

1. **Multiplicar** un vector de probabilidades $\alpha^{(t)}$ por la **matriz de transición** T en una cadena de Markov **equivale** a un **proceso de actualización** de creencias probabilísticas:
 $\alpha^{(t+1)} = \alpha^{(t)} T.$

2. En la **interpretación bayesiana**, $\alpha^{(t)}$ es la **distribución posterior** sobre el estado en el tiempo t , y el producto con T refleja la **regla de Bayes** (ley de probabilidad total) para el paso de t a $t+1$.
3. Si la cadena es **ergódica**, este proceso **converge** a la distribución estacionaria π , es decir, a la creencia "a largo plazo" sobre en qué estado se encuentra el sistema.
4. Si **no** hay ergodicidad, la convergencia puede no existir o depender de la prior, al igual que en un escenario bayesiano donde los datos no discriminan suficientemente entre hipótesis.

En síntesis, el "**truco**" de seguir multiplicando la distribución sobre estados por la matriz de transición es un **caso particular** (y fundamental) de la **actualización bayesiana**: cada "paso" re-calcula la probabilidad de los estados futuros a partir de la probabilidad de los estados presentes y las probabilidades condicionales de transición (que ejercen el rol de "likelihood"). Cuando la cadena es ergódica, este procedimiento converge a un punto fijo (la distribución estacionaria), que resulta análogo a cómo, en inferencia bayesiana, una gran cantidad de datos puede llevar a una posterior muy concentrada (o estable) en una región específica.