Cadenas de Markov

Algunos modelos matriciales, son ejemplos de cadenas de Markov. Esta constituyen modelos de probabilidad dependiendo del tiempo n (sucesion de variables aleatorias) en las que lo que ocurre en el tiempo n solo depende en el tiempo $n-1$. $X_1, X_n : (\Omega, \mathbb{P}) \to S$. Tipicamente

Definición (Conjunto de estados)

Los valores de las variables aleatorias se llaman S.Puede ser finito o numerable.

Ejemplo el tiempo en una ciudad puede ser Soleado, nublado o lluvioso. El cardinal de S es tres $S = S_1, S_2, S_3$ La cadena de Markov es finita, si el número de estados es finito e infinita si es infinito.

Definición (Cadena de Markov)

- $X_n \in S$ Conjunto de estados.
- Propiedad de Markov. Para todo $(x_0, x_1, \ldots, x_n, y)$ $\mathbb{P}(X_{n+1} = y / X_0 = x_0, X_1 = x_1, \ldots X_n = x) = \mathbb{P}(X_{n+1} = y / X_n = x)$ solo depende del estado inicial.
- Estacionaria
  $\mathbb{P}(X_{n+1} = y / X_n = x) = \mathbb{P}(X_1 = y / X_0 = x) = p_{xy}$

Matriz de transición La matriz formada por los p_{xy} se le llama matriz de transición. Por ejemplo: Si hoy es soleado con igual posibilidad mañana sera nuboso o soleado. Si hoy es nuboso con 50% de posibilidades mañana sera soledado con 25% seguira nuboso y con 25% de posibilidades llovera. Si hoy es lluvioso, mañana no sera soleado pero con igual probabilidad sera nuboso o lluvioso.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Propiedades de matriz de transición:

- $p_{ij} \geq 0$.
- para todo $i, \sum_j p_{ij} = 1$. P es una matriz estocástica. La suma de las filas da siempre 1.

Definimos la probabilidad de ir del paso i al paso j en n pasos como $(P_n)_{ij} = P(X_n = j/X_0 = i) = P(X_{m+n} = j/X_m = i)$

Theorem

$$P_n = P^n$$

La prueba usa el llamado análisis del primer paso: Combinar la ley de la probabilidad total con la propiedad de Markov. Por la ley de la probabilidad total:

$(P_n)_{ij} = \mathbb{P}(X_n = j/X_0 = i) = \sum_k \mathbb{P}(X_n = j, X_{n-1} = k/X_0 = i) =$

$\sum_k \mathbb{P}(X_n = j/X_{n-1} = k, X_0 = i)\mathbb{P}(X_{n-1} = k/X_0 = i)$

ahora usamos la propiedades de Markov estacionaria $\mathbb{P}(X_n = j/X_{n-1} = k, X_0 = i) = \mathbb{P}(X_n = j/X_{n-1} = k) = \mathbb{P}(X_1 = j/X_0 = k)$ y por definicion $\mathbb{P}(X_{n-1} = k/X_0 = i) = (P_{n-1})_{ik}$ asi pues

$$(P_n)_{ij} = \sum_k P_{kj}(P_{n-1})_{ik}$$

en otras palabras, usando las reglas del producto de matrices

$$P_n = P_{n-1} \times P$$

aplicando inducción se sigue el enunciado.

Por tanto $P_m = P^m, P_n = P^n, P_{m+k} = P^{m+k} = P^m \times P^k$. De lo que deducimos las ecuaciones de Chapman-Kolmogorov

$$(P_{m+k})_{ij} = \sum_l (P_m)_{il}(P_k)_{lk}$$

Caracterizamos $\pi_1(i) = \mathbb{P}(X_1 = i)$. Del mismo modo si π_0 es una probabilidad inicial $\mathbb{P}(X_0 = i) = (\pi_0)_i$ para calcular π_1 empleamos de nuevo el analisis del primer paso. Usamos la ley de la probabilidad total: $\mathbb{P}(X_1 = j) = \sum \mathbb{P}(X_1 = j/X_0 = i)\mathbb{P}(X_0 = i) = \sum_i p_{ij}(\pi_0)_i = \pi_0 P = P^t \pi_0$

Cadenas de Markov y grafos A una cadena de Markov se le asocia un grafo dirigido en que cada estado es un vertice. El vertice i se enlazan con el vertice j si $p_{ij} > 0$.

Objetos de estudio en C.M

- Quien es π_k.
- Tipos de estados.
- Frecuencias de visita f_{ik}, tiempo entre visitas τ_k.
- Distribución a largo plazo.

Existen tres comportamientos típicos

1. Ciclos.
2. $S = T \cup A$ Transitorios y absorbentes. Ejemplo Ruina del jugador.
3. Existe una distribución límite. Ejemplo Urna de Ehrenfest.

Aviso de notacion. Se usa ademas de $(P_n)_{ij} = p_{ij}(n)$.

Estados accesibles, clases

- Un estado j es accesible desde i si y solo si $p_{ij}(n) > 0$ para algún n.
- Dos estado se comunican si i es accesible desde j y j es accesible desde i. i se comunica con j ($i \sim j$ es una clase de equivalencia).
- Si todos los estados en S se comunican decimos que la C.M es irreducible

- Un subconjunto $C \subset S$ es cerrado si $\forall i \in C j \notin C$ $p_{ij} = 0$
- Un subconjunto cerrado C es irreducible si no contiene subconjuntos cerrados. Todos los estados en C se comunican.
- $C.M$ es irreducible si solo si S es el único conjunto cerrado.

Ejemplos de cadena reducible.

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}, P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

En el primer caso tenemos dos clases equivalencia $[1], [2]$.
En el segundo caso tenemos dos clases equivalencia $[1, 2], [3, 4]$.
En el tercero tenemos tres clases de equivalencia $[1, 2], [3], [4]$.
Nota: Todos los estados dentro de la misma clase de equivalencia comparten las mismas propiedades.

Periodicidad

- un estado i es periodo de periodo $d \in \mathbb{N} \setminus 1$ si $p_{ii}(n) \neq 0$ si y solo si $n = md$, m entero.
- un estado no periódico es aperiodico

$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, prototipo de cadena no aperiódica $P^2 = I$.

Gráficamente i es periódico si los unicos caminos en el grafo que empiezan en i y vuelven a i tienen md pasos.

- Si C.M es irreducible todos los estados tienen el mismo periodo (o son aperiódicos).
- Si todos los estados son aperiódicos decimos que C.M es aperiódica.

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, P^3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

Tiene periodo 3. Dibujar el grafo.

Estados comunicados tienen el mismo periodo

Si i, j se comunican, existen n_1, n_2 tales que $p_{ij}(n_1) \neq 0 \neq p_{ji} \neq 0$.
Supongamos que sean periódicos y $d(i), d(j)$ los periodo de i y de j respectivamente. En toda la discusion $m_h \in \mathbb{Z}$
Por definición

$$p_{ii}(n_1 + n_2) \neq 0 \neq p_{jj}(n_1 + n_2)$$

asi que $n_1 + n_2 = m_1 d(i) = m_2 d(j)$
$p_{ii}(n_1 + n_2 + d(j)) \neq 0 \neq p_{jj}(n_2 + n_1 + d(i))$ asi que
$n_1 + n_2 + d(j) = m_3 d(i), n_1 + n_2 + d(i) = m_4 d(j)$
Por tanto $d(i) = m_5 d(j) = m_5 m_6 d(i)$ y $d(i) = d(j)$. Ejercicio:
Finalmente nos aseguramos de que si i es periódico j es periódico tambien: Si $p_j j(k) \neq 0, p_{ii}(n_1 + n_2 + k) \neq 0$ asi que
$n_1 + n_2 + k = m_1 d(i)$. por lo que $k = m_1 d(i)$ asi que $d(j) \geq d(i)$.

Tipos de Estado:Retorno

Definimos la probabilidad de retorno del estado i a si mismo como: $f_i = f_{ii} = \mathbb{P}[(X_n = i \text{para algun } n > 0/X_0 = i)]$.

- Recurrentes $f_i = 1$ $(\sum_{n=1}^{\infty}(P^n)_{ii} = \infty)$
- Transitorios $f_i < 1$ $(\sum_{n=1}^{\infty}(P^n)_{ii} < \infty)$
- Absorbentes $p_{ii} = 1$. Los estados absorbentes son recurrentes.
- Ser recurrente o Transitorio es una posibilidad de la clase.

- Si todos los estados son recurrentes la C.M se dice recurrente.
- Si tiene estados absorbentes y de los transitorios es accesible algún absorbente la C.M absorbente.
- Aperiódica y recurrente se dice ergódica.

Caracterización de estados transitorios

LLamemos $\mathcal{I}(X_n)$ a la función indicatriz del estado i. Es decir $\mathcal{I}(X_n) = 1$ si $X_n = i$, 0 sino. Visitas a $i \sim V_i = \sum_{n=0}^{\infty} \mathcal{I}(X_n)$.

i recurrente implica $E[V_i/X_0 = i] = \infty$.

$$\infty = E[V_i/X_0 = i] = \sum E[\mathcal{I}(X_n)/X_0 = i]$$

$$= \sum 1\mathbb{P}(X_n = i/X_0 = i) = \sum_{n=0}^{\infty}(P_n)_{ii} = \sum_{n=0}^{\infty}(P^n)_{ii} = \sum p_{ii}(n)$$

Si es transitorio $\mathbb{P}(V_i = k/X_0 = i) = f_i^{k-1}(1 - f_i)$ (Hay $k - 1$ retornos) Entonces $E[V_i/X_0 = i] =$

$$= \sum_{k=1}^{\infty} kP(V_i = k/X_0 = i) = \sum_{k=1}^{\infty} kf_i^{k-1}(1 - f_i) = (1 - f_i)\sum_{k=1}^{\infty} kf_i^{k-1}$$

$$= (1 - f_i)(\sum f_i^k)' = (1 - f_i)(\frac{1}{(1 - f_i)})' = \frac{1 - f_i}{(1 - f_i)^2} = \frac{1}{1 - f_i}$$

Si dos estados se comunican ambos son transitorios

Supongamos que i es transitorio y se comunica con j y como antes $p_{ij}(n_1) > 0$ $p_{ji}(n_2) > 0$ y r cualquier otro entero. Sea $\alpha = p_{ij}(n_1)p_{ji}(n_2)$. Ahora por Chapman-Kolmogorov o directamente por las reglas de la probabilidad. $p_{ii}(n_1 + n_2 + r) \geq p_{ij}(n_1)p_{jj}(r)p_{ji}(n_2) = \alpha p_{jj}(r)$ Ahora como i es transitorio por el teorema de la página anterior

$$\infty > \sum_r p_{ii}(n_1 + n_2 + r) \geq \alpha \sum_r p_{ii}(r)$$

Por lo que $\sum_r p_i i(r) < \infty$ i es transitorio.

Clasificación de C.M

Teorema de Clasificacion de C.M. Para toda C.M $S = T \cup C_1 \cup C_2 \cup C_r$ donde las clases T_1, T_2 son estados transitorios y las clases C_1, C_2, C_r son cerradas e irreducibles. Si S es finito no puede estar siempre en los estados transitorios asi que acabara en una de las clases C_R. Estudiaremos el proceso de ser absorbido por las clases irreducibles y que ocurre una vez que estas en una clase irreducible.

- Existen estados transitorios y absorbentes (Reducibles).
- Existe una sola clase (la cadena es irreducible).

La urna de Ehrenfest

Tenemos N bolas repartidas en dos urnas. En cada tic del reloj, al azar movemos una bola de una urna a otra.
Opcion 1: S numero de bolas en la urna blanca.
$p_{i(i-1)} = \frac{i}{N}, p_{i(i+1)} = 1 - \frac{i}{N}$
Estados: S Partes de $\mathbb{N}$ listas de bolas. Con esta formulación tenemos 2^N estados pero la matriz es doblemente estocastica. Cual sera su distribución estacionaria?

Cadenas absorbentes

Supongamos que $S = T \cup A$, T transitorios, A absorbentes. En el caso de que haya estados absorbentes, la distribucion limite depende de las distribuciones iniciales. Pero podemos preguntarnos otras cosas:

- Probabilidad de absorción por el estado k (absorbente) si empezamos con el estado i (transitorio)
- Tiempo de llegada. Cuanto tardamos para ser absorbidos por algun estado si empezamos en i. LLamamos a esta variable aleatoria $T(X_n) = n$ si $X_n \in A$ $X_k \in S$ para todo $k < n$.

Teoría General Facill pero Computacionalmente duro. Hay que calcular inversas de matrices etc,
Casos particulares: La misma filosofia produce cálculos muy sencillos. Discutimos primero el caso de la ruina del jugador y luego lo axiomatizamos.

Ruina del jugador 1

Sea ρ_j Probabilidad de ganar Toni si empieza con j euros y Maria con $n-j$. Es decir $\rho_k = P(X_T = n : X_0 = j)$. Realizamos el Análisis del primer paso sobre $\rho_j =$

$$
\begin{aligned}
&= p\mathbb{P}(X_T = n : X_1 = j+1, X_0 = j) \\
&+ (1-p)\mathbb{P}(X_T = n : X_1 = j+1, X_0 = j-1) \\
&= p\mathbb{P}(X_T =: X_1 = j+1) + (1-p)\mathbb{P}(X_T = n : X_1 = j-1) \\
&= p\rho_{j+1} + (1-p)\rho_{j-1}
\end{aligned}
$$

Las condiciones iniciales son $\rho_0 = 0, \rho_n = 1$. Acopladas a:

$$\rho_j = p\rho_{j+1} + (1-p)\rho_{j-1}$$

que es una sucesion en diferencias como por ejemplo la de Fibonacci. Se puede resolver de muchas maneras.

Sucesión de diferencias

Se puede resolver usando ecuaciones en diferencias.
Alternativamente usamos un truco habitual en probabilidad: estudiar la distribución de las diferencias $d_j = X_{j+1} - X_j$.
Observamos que como $p + 1 - p = 1$, sale que

$$\rho_j = p\rho_j + (1-p)\rho_j, \rho_j = p\rho_{j+1} + (1-p)\rho_{j-1}$$

Buscamos las ecuaciones para $d_j = \rho_j - \rho_{j-1}$. Sumando y restando

$$0 = p(\rho_{j+1} - \rho_j) + (1-p)(\rho_{j-1} - \rho_j) = pd_j - (1-p)d_{j-1}$$

Es decir la sucesión de diferencias satisface una recurrencia de primer orden,

$$d_j = \frac{1-p}{p} d_{j-1}, d_1 = \rho_1,$$

Ruina del jugador 2

Ahora observamos que

$$d_1 = \rho_1, \rho_k = \sum d_j = \sum_{k=0}^{j-1} (\frac{1-p}{p})^k \rho_1, 1 = \rho_n = \sum_{k=0}^{n} (\frac{1-p}{p})^k \rho_1.$$

De la última despejamos $\rho_1 = \dfrac{1}{\sum_{k=1}^{n} (\frac{1-p}{p})^k}$ y obtenemos la conclusión final.

$$\boxed{\rho_j = \frac{\sum_{k=0}^{j} (\frac{1-p}{p})^k}{\sum_{k=0}^{n} (\frac{1-p}{p})^k} = \frac{1-(\frac{1-p}{p})^j}{1-(\frac{1-p}{p})^n}}$$

Nota: Estamos asumiendo implicitamente que $p \neq \frac{1}{2}$. Donde falla la prueba? Que pasa con $p = \frac{1}{2}$

Tiempo de duración. Argumentamos de manera similar
$\tau_j = E[T/X_0 = j]$

$$\begin{aligned}\tau_j &= pE[T/X_1 = j+1, X_0 = j] + (1-p)E[T/X_1 = j-1, X_0 = j] \\ &= pE[T/X_1 = j+1] + (1-p)E[T/X_1 = j-1]\end{aligned}$$

Ahora somos cuidadosos porque para pasar de X_1 a X_0 gastamos un unidad de tiempo.

$$1+pE[T/X_0 = j+1]+(1-p)E[T/X_0 = j-1] = 1+p\tau_{j+1}+(1-p)\tau_{j-1}$$

Ahora las condiciones de contorno son simétricas $\tau_0 = 0, \tau_n = 0$. Resolvemos la ecuación en diferencias de nuevo usando d_j para el caso $p = \frac{1}{2}$. En este caso resuelve

$$d_k = d_1 - 2(k-1)$$

y volviendo a los tiempos de retorno: $\boxed{\tau_j = j(N-j)}$

Estructura de Matrices absorbentes

Suponemos que los estados de $0, r$ son transitorios (no absorbentes) y de $r+1, a, k$ son absorbentes. De este modo $P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$

Y observamos que

$$P^2 = \begin{pmatrix} Q^2 & QR + R \\ 0 & I \end{pmatrix}, P^n = \begin{pmatrix} Q^n & (I + Q + \dots Q^{n-1})R \\ 0 & I \end{pmatrix}$$

La matriz $W = \lim_{n\to\infty}((I + Q + \dots Q^{n-1}) = (I - Q)^{-1}$ se llama matriz fundamental codifica casi toda la información que necesitamos

Nota: Necesitamos que el maximo de los autovalores de Q es estrictamente menor que uno. Es cierto?

Directamente por definición

- $W_{ij} = \sum_{n=1}^{T} (P_n)_{ij} = E[V_j / X_0 = i]$ el numero medio de visitas al estado transitorio j si empezamos en i
- Sea $U = WR$, $U_{ik} = \mathbb{P}(X_T = k; X_0 = i)$
- $\tau_i = \sum_{j=1}^{r} W_{ij}$ donde τ_i es el tiempo de absorción si estamos en el estado i.

Las formulas matriciales son cruciales para el análisis de Cadenas de Markov con muchas entradas pero a menudo se pueden resolver con análisis del primer paso.

Tiempos Esperados

Como veiamos en la clasificación de estados transitorios y recurrentes. $\mathcal{J}(X_n)$ es la funcion indicatriz de el evento $X_n = j$.

$$W_{ij} = E[\sum_{n=0}^{T-1} \mathcal{J}(X_n)/X_0 = i]$$

Observación:Si $n < T$ X_n siempre es igual a un estado transitorio por lo que si $n < T$ $\sum_{j=1}^{r} \mathcal{J}(X_n) = 1$ es un evento de probabilidad 1.

Y cambiando el orden de la sumas, con probabilidad 1,

$$\sum_{j=1}^{r} \sum_{n=0}^{T-1} \mathcal{J}(X_n) = \sum_{n=0}^{T-1} \sum_{j=1}^{r} \mathcal{J}(X_n) = \sum_{n=0}^{T-1} 1 = T.$$

Introduciendo este evento en la esperanza:

$$\sum^{r} W_{ij} = E[\sum^{r} \sum^{T-1} \mathcal{J}(X_n)/X_0 = i] = E[T/X_0 = i] = \tau_i$$

Distribución estacionaria

Decimos que π es una distribución estacionaria si

$$P^t\pi = \pi, \pi P = \pi$$

Theorem

Toda cadena de Markov tiene una distribución estacionaria

Por definición las distribuciones estacionarias son autovectores con autovalor 1 de P^t.
Como P es estocastica si $\xi = (1, 1, , 1, 1, 1)$ $(P\xi)_k = \sum_j p_{ij}1 = 1$

asi pues $P\xi = \xi$ es decir $\lambda = 1$ es autovalor de P. Es decir

$$0 = \det(P - 1I) = \det(P^t - I)$$

Asi pues 1 es autovalor P^t asi pues al menos hay una distribución estacionaria.

Decimos que P es regular (primitiva) si existe un n tal que $(P^n)_{ij} > 0$. π es un una distribución de equilibrio (distribución límite)[1] si existe π_0 tal que

$$\pi = \lim_{n\to\infty} \pi_0 P^n$$

Theorem (Teorema de Perron)

Si P es regular tiene una única distribución de equilibrio π Ademas para cualquier distribución inicial $\lim \pi_0 P^n = \pi$

Sea $Q = (P^t)^n$, y $S \subset \mathbb{R}^n$ el simplice $S = \{(\xi_1, \dots, \xi_n) : \xi_i \geq 0, \sum_{i=1}^{n} (\xi)_i = 1\}$. Idea de la prueba: Demostrar que Q es una contracción de conjuntos del espacio $S - \pi$ en $S - \pi$.

[1] En la literatura a veces se usa distribución de equilibrio para la del Teorema de Perron

Prueba del teorema de Perron para matrices estocasticas

Existe un ϵ tal que $Q(S) \subset \pi + (1-\epsilon)(S-\pi)$

- $\eta = Q(\xi)$. $Q(S) \subset S$. $\sum_j \eta_j = \sum_j \sum_i \sum p_{ij}\xi_i = 1$
- Para todo $\eta \in Q(S)$ para todo i $\eta_i > 0$.

  $$\eta_i = \sum p_{ij}\xi_i \geq \max_i p_{ij}\xi_i > 0$$

  porque $p_{ij} > 0$, $\xi_i > 0$.
- Sea η $\min \eta_i > 0$. Entonces existe $\xi \in S, \epsilon > 0$ tal que $\eta_i = \pi_i + (1-\epsilon)(\xi_i - \pi_i)$ Si fuera cierto despejamos $\xi_i = \frac{1}{1-\epsilon}(\eta_i - \epsilon\pi_i) > 0$ if and only if $\epsilon < \frac{\eta_i}{\pi_i}$ Por otro lado $\sum \xi_i = \frac{1}{1-\epsilon}(\sum \eta_i - \epsilon \sum \pi_i) = \frac{1-\epsilon}{1-\epsilon} = 1.$
- $Q(S-\pi) \subset (1-\epsilon)(S-\pi)$ Se sigue del anterior usando por primera vez que $Q\pi = \pi$ y eligiendo ϵ por compacidad.
- $Q^n(S) \subset \pi + (1-\epsilon)^n(S-\pi)$

- La condicion de ser primitiva tambien es suficiente. Una manera de verlo es que si existe el limite *nico* para n grande la matriz πP^n tiene aproximadamente π_∞ en sus filas, que es positiva.
- Existen muchos criterios que prueban que una matriz es regular. Por ejemplo es facil probar que si $p_{ii} \neq 0$ para todo i y P es irreducible P es regular. (Ejercicio).
- Por otro lado la definición nos dice que m puede ser muy grande. Wielant probó que es suficiente comprobarlo con $m \leq n^2 - 2n + 2$. Si $n = 2, m = 4$, Si $n = 3, m = 5$
- Cualquier distribución límite es estacionaria y viceversa. Pero puede ser que la distribución estacionaria π solo sea el límite de $\pi_0 P$ para $\pi_0 = \pi$.
- Aperiodica e Irreducible implica regular.

Theorem

Si P tiene una única distribución limite $\pi = \pi_1, \pi_2, \pi_n$

$$P_\infty = \lim P^n = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_d \\ \pi_1 & \pi_2 & \dots & \pi_d \\ \pi_1 & \pi_2 & \dots & \pi_d \end{pmatrix}$$

Theorem

Si P es irreducible existe una única distribución estacionaria.

La demostración se sigue del hecho de que si P es irreducible $\tilde{P} = \frac{1}{2}(I + P)$ es regular: En efecto

$$\tilde{P}^n = \frac{1}{2^n} \sum_{k=1}^{n} \binom{n}{k} P^k$$

pero por ser irreducible para todo i, k existe un n (no necesariamente el mismo) tal que $(P_n)_{ik} > 0$. Ahora bien $\tilde{P}$ y P tienen las mismas distribuciones estacionarias.

Suma de Cesaro(Medias en Tiempo)

Theorem

Sea P irreducible y π su distribución estacionaria. Definimos la sucesión de matrices

$$\tilde{P}_n = \frac{1}{n}\sum_{i=0}^{n} P^k$$

Entonces para toda distribucion π_0 $\pi_0 \tilde{P}_n \to \pi$

Sea $L = \lim_{m\to\infty} \tilde{P}_m$, $\tilde{P}_m$ cualquier subsucesión. Observamos que

$$\pi_0 \tilde{P}_n P = \pi_0(\tilde{P}_n + \frac{1}{n}(P^{n+1} - I)$$

$\lim_{n\to\infty} \frac{1}{n}(P^{n+1} - I) = 0$ asi que $\pi_0 LP = \pi_0 L$, i.e $\pi_0 L = \pi$, la única distribución estacionaria de P.

Teorema ergódico para cadenas de Markov

Interpretemos esta suma de Césaro $\pi(u)$ probabilisticamente como la frecuencia media a la larga de visitas al estado u.

$$\mathbb{E}(\text{Numero de visitas a } u \text{ hasta } n) = \sum_{k=0}^{n} 1 p_{uu}(k)$$

nos da el numero de visitas esperado desde 0 hasta N que sabemos dividido por N converge a $\pi(u)$.
Por otro lado sea $\tau_u = E(T : X_T = u / X_0 = u)$ la esperanza del tiempo de primer retorno.

Theorem (Teorema ergódico para cadenas de Markov Irreducibles)

$\pi(u)\tau(u) = 1$

Sean k_1, k_2, k_i tal que $X_{k_i} = u$. El cardinal de estos k_i menores que N es el numero de visitas a u es decir $N\pi(u)$ y $k_{i+1} - k_i \approx \tau_u$. Por tanto tomando N cercano a los k_i $N = \sum_{k_i \leq N} (k_{i+1} - k_i) \approx \tau_u N \pi(u)$.

Page Rank:El algoritmo de Brin y Page

Objetivo: Ordenar páginas de web por relevancía. PageRank es una marca registrada y patentada por Google el 9 de enero de 1999 que ampara una familia de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos (o páginas web) indexados por un motor de búsqueda"

- Es un orden estático. Se calcula un valor PR que se asigna a cada página.
- Es un orden democrático. La estructura de enlaces determina PR.
- Sin embargo los enlaces se ponderan dependiendo de la relevancia de la página a la que conecta. (No es lo mismo que te enlace elpais que el blog de Faraco).
- Orden Recursivo
- 1998 habia 150 millones de páginas y 1.5 billones de enlaces.

Cuantificando

Idea genial 1

Calculo recursivo de $PR(W_1)$. Por página W_2 que enlaza con W_1 se añade un PR de $\frac{PR(W_2)}{n(W_2)}$ donde $n(W_2)$ son el número de enlaces que salen de W_2. Si W_j enlaza con W_i decimos que $W_j \sim W_i$. Observamos que j puede ser i.

$$PR(W_j) = \sum_{W_i \sim W_j} \frac{PR(W_i)}{n(W_i)}$$

A partir de ahora identificamos $W_i \equiv i$. Si ahora definimos la matriz $P = p_{ij}$ como $p_{ij} = \frac{1}{n(i)}$ Si $i \sim j$ y 0 en otro caso.
El vector $\vec{p} = (PR(i))_{i=1}^{n}$ satisface

$$p = pP = P^t p$$

Ejemplo:Yahoo, Amazon, Yahoo

- De Yahoo enlanzas a Yahoo y a Amazon 2 enlaces su peso es $\frac{1}{2}$
- De Amazon a Yahoo y a Microsoft 2 enlaces su peso es $\frac{1}{2}$
- De Microsoft uno a Amazon 1 enlaces su peso es 1

Por tanto $P(X_1 = Y/X_0 = Y) = \frac{1}{2}, P(X_1 = Y/X_0 = A) = \frac{1}{2}, P(X_1 = A/X_0 = Y) = \frac{1}{2}, P(X_1 = A/X_0 = M) = 1, P(X_1 = M/X_0 = A) = \frac{1}{2})$

$$Y = 1, A = 2, M = 3 : P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$$

En este caso n es grandisimo por lo que la manera mas facil de llegar a p es considerarlo como un sistema dinámico.
Patente de Google Page Rank. The rank of a page can be interpreted as the probability that a surfer will be at the page after following a large number of forward links Es decir estamos buscando la distribución de equilibrio de una cadena de Markov".

Enemigos

- Paginas sin hiperenlaces de salida (Dangling) Producen ceros en las columnas.
- Spider traps. Ciclos.
- Dead ends. Estados absorbentes. Un solo hipervínculo asi mismo.

La matriz P no es regular (puede ser reducible y/o periódica).

Idea genial 2 En cada paso el surfero que esta en la pagina W_i tiene dos opciones.

- Con probabilidad d sigue el hipervínculo con las propiedades indicadas por la matriz P.
- Con probabilidad $1-d$ se va al buscador y escoge cualquier otra página web aleatoriamente con la misma probabilidad $\frac{1}{n}$.

d dumpling factor La nueva matriz de transición es

$$\tilde{P} = dP + (1-d)\frac{1}{n}E$$

donde $E = \sum_{i,j} e_i \otimes e_j$

$\tilde{P}$ es regular !!!!(es positiva) Asi que existe $p = \lim_{n\to\infty} p_0 \tilde{P}^n$. El valor clásico de $d = 0.85$

- Se corre el vector de rangos primero sin los datos descolgados y luego se añaden.
- Se lanza solo cada varios meses. (mínimo tres)
- Solo se deja correr varias iteraciones Ejemplo $n = 332$ millones. Iteraciones 52.
- Problema de tamaño. Si el numero de paginas Web es $n = 10^9$ a priori $\tilde{P} \sim 10^{18}$ demasiado para calcular $\tilde{P}^{45}$ con la capacidad computacional actual. En la práctica se usa que P tenia ceros y se define

$$\tilde{P}_{k+1} = dM\tilde{P}_k + (1-d)\frac{1}{n}E$$

  aqui M es la matriz original (sin el factor E).
- El algoritmo se intenta sabotear de mil maneras. Por ejemplo mediante link farms (granjas de enlaces).

Cadenas de Markov con infinitos estados

En el caso de que el conjunto de estados S es numerable pero finito aparece un nuevo fenḿeno llamada la recurrencia nula. Comentamos primero el camino aleatorio 1 dimensional que es una cadena del jugador sin limite de capital ni de deuda. Gráfica: avanzamos a derecha o izquierda con probablidad $p, 1-p$. $S = \mathbb{Z}$. La cadena es irreducible.

- Si $p \neq \frac{1}{2}$ todos los estados son transitorios $(\sum_{n=0}^{\infty} p_{ii}(n) = \infty)$
- Si $p = \frac{1}{2}$ todos los estados son recurrentes pero τ_i el tiempo esperado de retorno es ∞. (el estado es recurrente nulo).

Observamos que podemos decir que $X_n = k - m$ donde k es el número de caras y m el número de cruces, $n = k + m$. Por tanto si $X_0 = 0, X_n = 0$ si solo si $k = m$. Asi para impares $2n - 1 = 2k$ no tiene solución $p_{00}(2n-1) = 0$ y $p_{00}(2n) = \binom{2n}{n}(p(1-p))^n$.

Recurrencia en el camino aleatorio

Ahora usamos la formula de Stirling $n! \approx n^n e^{-n}\sqrt{2\pi n}$ y obtenemos

$$\binom{2n}{n} = \frac{2n!}{n!n!} \approx \frac{4^n n^{2n} e^{-2n}\sqrt{4\pi n}}{n^{2n} e^{-2n} 2\pi n} = \frac{4^n}{\sqrt{\pi n}}$$

Por tanto

$$\sum p_{00}(n) = \sum \frac{(4p(1-p))^n}{\sqrt{\pi n}} < \infty$$

si y solo si $4p(1-p) < 1$ i.e $p \neq \frac{1}{2}$ Por tanto para $p = \frac{1}{2}$ la cadena es recurrente. Sin embargo $\lim_{n\to\infty} p_{00}(n) = \lim_{n\to\infty} \frac{1}{\sqrt{\pi n}} = 0$ asi que el tiempo de retorno es ∞[2]. Nota: En dimensión dos el camino simétrico es recurrrente pero para $n \geq 3$ es transitivo. Asi pues un borracho humano volverá de casa al bar pero un ave borracha no llegará al nido. Por eso los pajaros no beben alcohol.

[2] El teorema ergódico tambien se cumple para infinitos estados

Pasando al continuo (En la escala parabólica)

Hemos definido el camino aleatorio lo hacemos con pasos de longitud uno cada instante de longitud uno. La transparencia sugiere que tendemos a alejarnos $\sqrt{n}$ en n pasos. Cambiamos la escala cada intervalo de tiempo $2t = h^2$ es decir para un intervalo de tiempo t avanzamos $h = \sqrt{2t}$

Observamos que

$\pi_x(t+t_0) = \frac{1}{2}(\pi_{x+h}(t_0) + \pi_{x-h}(t_0))$ Restando $\pi_x(t_0)$ a ambos lados y dividiendo por t

$$\frac{\pi_x(t+t_0) - \pi_x(t_0)}{t} = \frac{1}{\underbrace{2t}_{=h^2}}(\pi_{x+h}(t_0) + \pi_{x-h}(t_0) - 2\pi_x(t_0))$$

Pasando al limite cuando t y h tienden a cero llegamos a que la función $u(x,t) = \pi_x t$ satisface

$u_t = u_{xx}$ la ecuación del calor que se ve en E.D.Ps.