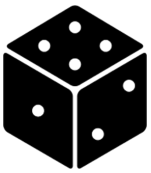
METODI PROBABILISTICI PER L'INFORMATICA

Prof. Massimiliano Goldwurm
6 CFU

Luca Cappelletti
Desilda Toska

Lecture Notes
Year 2017/2018



Magistrale Informatica
Università di Milano
Italy
19 luglio 2018

Indice

1 Probabilità

1.1 Disuguaglianza di Chernoff

$$\mathbb{P}\left(\left|X_{n,p} - np\right| \geqslant n\epsilon\right) < 2e^{-2\epsilon^2 n}$$

Figura 1.1: Disuguaglianza di Chernoff per le binomiali

La **disuguaglianza di Chernoff** è utilizzata quando è necessario calcolare la probabilità di una **funzione** f che:

1. Non è semplice da calcolare, oppure,
2. Non è nota.

1.1.1 Caratteristiche

Possiede 3 distinte caratteristiche:

Variabili di indipendenti

Essa richiede che le **variabili siano indipendenti**, *condizione che la disuguaglianza di Markov non richiede* e nel caso delle disuguaglianza di Chebyshev è necessaria solo l'indipendenza da coppie di variabili casuali.

Non è una disuguaglianza vera

La disuguaglianza di Chernoff **non è una disuguaglianza vera**, ma piuttosto va vista come una tecnica per ottenere limiti esponenziali decrescenti sulle probabilità di coda.

Il valore di n è arbitrario

La disuguaglianza di Chernoff non fornisce un preciso valore di n oltre il quale siamo sicuri che la probabilità della coda sia minore di una δ fissata.

1.1.2 Applicazioni alle binomiali

Nell'analisi degli algoritmi probabilistici occorre spesso valutare la coda di una binomiale, ovvero la probabilità che questa variabili aleatoria disti dalla media per una quantità fissata.

Partendo dalla **disuguaglianza di Chebyshev** per una variabile X con $\mathrm{Var}(X)$ finita:

$$\mathbb{P}\left(\left|X - \mathbb{E}(X)\right| \geqslant a\right) \leqslant \frac{\mathrm{Var}(X)}{a^2} \quad \forall a > 0$$

Figura 1.2: Disuguaglianza di Chebyshev

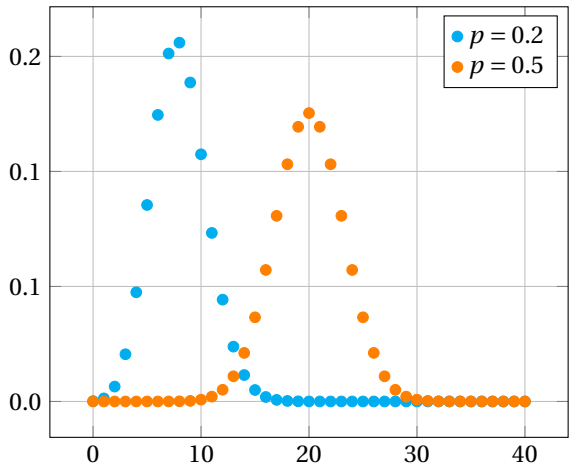


Figura 1.3: Distribuzioni binomiali al variare del parametro p

Nel caso di $X_{n,p}$ binomiale, con media $\mathbb{E}(X) = np$ e varianza $\mathrm{Var}(X) = npq$, $\forall \epsilon > 0$ vale che:

$$\mathbb{P}\left(\left|X_{n,p} - np\right| \geqslant n\epsilon\right) \leqslant \frac{pq}{\epsilon} \cdot \frac{1}{n} = \mathcal{O}\left(\frac{1}{n}\right)$$

Per il **teorema del limite centrale** vale che:

$$\frac{X_{n,p} - np}{\sqrt{npq}} \longrightarrow \mathcal{N}(0,1)$$

Applico il limite ed ottengo:

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{X_{n,p} - np}{\sqrt{npq}} \geqslant \epsilon\right) = \frac{2}{\sqrt{2\pi}} \int_{\epsilon}^{+\infty} e^{-\frac{t^2}{2}}\, dt$$

Vado ad aggiungere il termine $\sqrt{\frac{n}{pq}}$ come coefficiente a ϵ per consentire l'integrazione:

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{X_{n,p} - np}{\sqrt{npq}} \geqslant \sqrt{\frac{n}{pq}} \cdot \epsilon\right) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\frac{n}{pq}}\cdot\epsilon}^{+\infty} e^{-\frac{t^2}{2}}\, dt$$

Maggioro l'integrale:

$$\frac{2}{\sqrt{2\pi}} \int_{\sqrt{\frac{n}{pq}}\cdot\epsilon}^{+\infty} e^{-\frac{t^2}{2}}\, dt \leqslant \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\frac{n}{pq}}\cdot\epsilon}^{+\infty} \frac{t}{\sqrt{\frac{n}{pq}} \cdot \epsilon} e^{-\frac{t^2}{2}}\, dt = \mathcal{O}\left(\frac{1}{\sqrt{n}} e^{-\frac{\epsilon^2}{2pq} n}\right)$$

Il risultato ottenuto, si avvicina allo 0 molto più velocemente del valore $\mathcal{O}\left(\frac{1}{n}\right)$ ottenuto precedentemente tramite la disuguaglianza di Chebyshev.

La disuguaglianza di Chebyshev è più generale, di conseguenza è più debole. Per esempio nel caso della distribuzione gaussiana risulta molto debole, per cui se è necessario avere un'accuratezza maggiore conviene assolutamente usare la disuguaglianza di Chernoff.

1.1.3 Esempio: Riduzione della probabilità di errore di un algoritmo

Supponiamo di voler calcolare una funzione $f_I \to O$ difficile da calcolare e di disporre di un algoritmo probabilistico $\mathscr{A}$ tali che $\forall x \in I, T_{\mathscr{A}} \leqslant p(|x|)$ dove p è un polinomio di grado **piccolo** e che $\mathbb{P}\left(A(x) = f(x)\right) \geqslant \frac{3}{4}$, col valore di destra maggiore di $\frac{1}{2}$ e indipendente da x, altrimenti non potrei determinare il risultato corretto la frequenza con cui appare

Dato un intero $t > 0$, chiamiamo $\mathscr{A}_t$ l'algoritmo probabilistico che ripete $\mathscr{A}$ per un numero t di volte e ne restituisce il valore più frequente, se esso esiste.

L'algoritmo non garantisce che il risultato sia corretto, infatti anche con n iterazioni rimane possibile che l'algoritmo calcoli con frequenza maggiore la soluzione errata.

Algorithm 1: Riduzione della probabilità di errore

input : $x \in I$
output: Valore più frequente o non so

```
1  begin
2      for i ∈ [1,...,t] do
3          Esecuzioni indipendenti di 𝒜 su x;
4          A[i] = 𝒜(x)
5      end
6      if ∃z ∈ (A[1],...,A[t]) : #{j ∈ {1,...,t} : z = A[j]} > t/2 then
7          return z;
8      end
9      else
10         return non so rispondere;
11     end
12 end
```

Quante volte devo iterare l'algoritmo per ridurre la probabilità di errore ad un valore δ?

Utilizzando la disuguaglianza di Chernoff per le binomiali ottengo:

$$\mathbb{P}\left(\mathscr{A}_t \neq f(x)\right) \leqslant \mathbb{P}\left(X_{t,\frac{3}{4}} \leqslant \frac{t}{2}\right) = \mathbb{P}\left(X_{t,\frac{3}{4}} - \frac{3}{4}t \leqslant -\frac{t}{4}\right) \leqslant e^{-2\frac{1}{16}t} = e^{-\frac{1}{8}} \leqslant \delta$$

Calcolo il logaritmo naturale ed ottengo il valore di t per $\delta = 1000^{-1}$.

$$t \geqslant 8\ln\frac{1}{\delta} \Rightarrow t \geqslant 8\ln 1000 = 24\ln 10 \approx 56$$

Se avessimo invece utilizzato la **disequazione di Chebyshev** avremmo invece ottenuto un risultato molto peggiore:

$$\mathbb{P}\left(X_{t,\frac{3}{4}} - \frac{3}{4}t \leqslant -\frac{t}{4}\right) \leqslant \mathbb{P}\left(\left|X_{t,\frac{3}{4}} - \frac{3}{4}t\right| \geqslant \frac{t}{4}\right) \leqslant \frac{3}{t} \Rightarrow t \geqslant 3000$$

1.1.4 Esempio: Lancio di monete

Consideriamo il caso di una variabile aleatoria $X_{n,\frac{1}{2}}$ con media $\mathbb{E}(X) = \frac{n}{2}$:

$$\mathbb{P}\left(\left|X_{n,\frac{1}{2}} - \frac{n}{2}\right| \geqslant \sqrt{n\log n}\right) \leqslant 2e^{-2\log n} = \frac{2}{n^2}$$

Per esempio, posto $n = 100$ si ottiene:

$$\mathbb{P}\left(\left|X_{n,\frac{1}{2}} - 50\right| \geqslant 21\right) \leqslant \frac{1}{5000}$$

Rendendo la maggiorazione più aderente, dividendo il termine destro per due, si ottiene:

$$\mathbb{P}\left(\left|X_{n,\frac{1}{2}} - \frac{n}{2}\right| \geqslant \sqrt{\frac{n\log n}{4}}\right) \leqslant 2e^{\frac{-2\log n}{4}} = \frac{2}{\sqrt{n}}$$

Nuovamente risolvendo per $n = 100$ si ottiene:

$$\mathbb{P}\left(\left|X_{n,\frac{1}{2}} - 50\right| \geqslant 10.5\right) \leqslant \frac{1}{5} \rightarrow \mathbb{P}\left(39 \leqslant X_{n,\frac{1}{2}} \leqslant 61\right) \geqslant \frac{4}{5}$$

1.1.5 Esempio: Intervalli di confidenza

Si vuole valutare la probabilità p di un evento A disponendo di un test f sull'evento A che restituisce un valore:

$$f: \begin{cases} \text{SI} & \text{con probabilità } p \\ \text{NO} & \text{con probabilità } 1-p \end{cases}$$

$\frac{X_{n,p}}{n} = \overline{p}$ è una variabile aleatoria con media $\mathbb{E}\left(X_{n,p}\right) = np$.

$$\mathbb{P}\left(\left|p - \overline{p}\right| \geqslant \delta\right) = \mathbb{P}\left(\left|np - X_{n,p}\right| \geqslant n\delta\right) \leqslant 2e^{-2\delta^2 n} \leqslant t$$

$$-2\delta^2 n \leqslant \log\frac{2}{t} \Rightarrow n \leqslant \frac{1}{2\delta^2}\log\frac{2}{t}$$

Dove t rappresenta la probabilità di errore.

Ponendo $\delta = 0.1$ e $t = 100^{-1}$ si ottiene:

$$n \leqslant 50\log 200 = 50\left(\log 2 + 2\log 10\right) \approx 250 \Rightarrow n \geqslant 250$$

2

Algoritmi probabilistici

Gli algoritmi probabilistici richiedono un numero minore di calcoli degli algoritmi deterministici e quindi determinano una soluzione in un tempo minore. Esistono principalmente due famiglie di algoritmi probabilistici:

1. Per cercare una soluzione esatta.
2. Per cercare una soluzione approssimata.

Per **algoritmi Monte-Carlo** si intendono gli algoritmi della categoria **1-sided error** e **2-sided error**.

Generalmente, per diminuire la probabilità di errore viene reiterato l'algoritmo probabilistico per un numero di volte ottenuto utilizzando la **disequazione di Chernoff** sino a che l'errore è inferiore ad un δ accettabile.

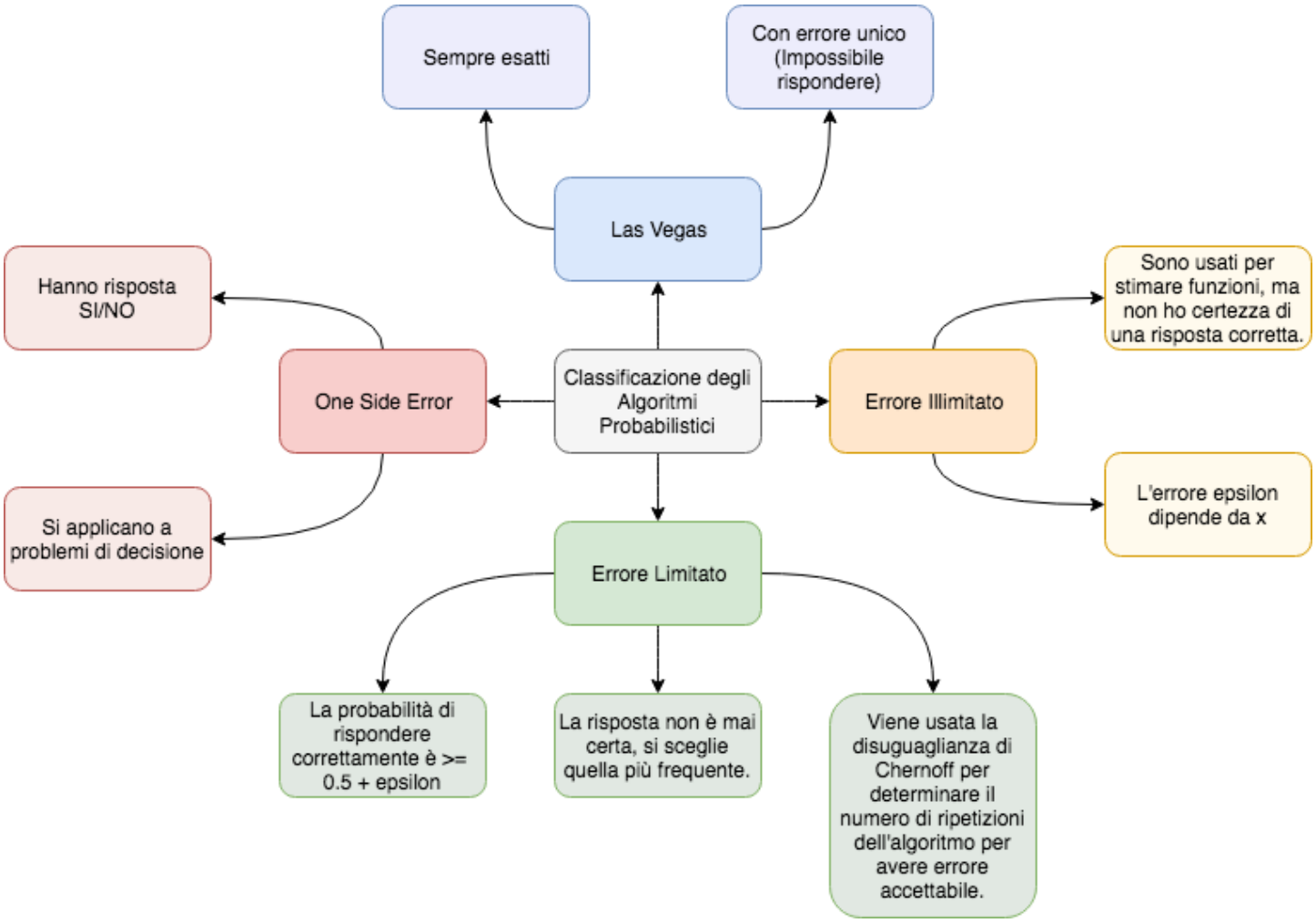


Figura 2.1: Mappa degli algoritmi probabilistici

2.1 1-Sided Error

Hanno risposta SI/NO e vengono applicati a problemi di decisione.

2.1.1 Esempio: Problema di protocollo di comunicazione

Si vuole determinare se due stringhe binarie a e b, lunghe n, salvate su due terminali separati sono uguali senza inviare una stringa completamente da un terminale all'altro.

Un approccio banale risolverebbe in n messaggi il problema.

Un approccio sofisticato utilizza i resti della divisione con numeri primi. Dato un numero primo $p \in \{2,\dots,n^2\}$ con probabilità uniforme, di dimensione $\lceil 2\log_2 n \rceil$ bit.

$$r_a = \text{resto}\left(\frac{a}{p}\right) \in \{0,1,\dots,p-1\} \qquad r_b = \text{resto}\left(\frac{b}{p}\right) \in \{0,1,\dots,p-1\}$$

$$r_a \neq r_b \Rightarrow a \neq b$$

$$r_a = r_b \Rightarrow a \text{ potrebbe essere uguale a } b$$

I valori del resto e di p sono inviati al secondo dei due nodi usando al più $2\lceil 2\log_2 n \rceil$ bit, dove viene ripetuta la divisione ed effettuato il controllo, rispondendo con un bit se diversi o zero se uguali.

Se $\mathscr{A}$ restituisce "diverso" la risposta è certamente corretta, mentre se restituisce "uguale" potrebbe essere sbagliato, con una probabilità pari a:

$$\mathbb{P}\left(\text{errore} \mid \mathscr{A} \text{ dice uguali}\right) = \mathbb{P}\left(p \text{ divide } a-b \;\middle|\; a \neq b\right)$$

Usando il **teorema dei numeri primi** che enuncia:

Teorema 2.1.1 (Teorema dei numeri primi). Se $\text{prim}(m) = \#\{i \in \{2,\dots,m\} : i \text{ è primo}\}$ allora vale che:

$$\text{prim}(m) \approx \frac{m}{\ln m} \qquad \text{ovvero} \qquad \frac{\text{prim}(m)}{m} \ln m \longrightarrow 1$$

Inoltre è noto che:

$$\forall m > 67, \text{prim}(m) > \frac{m}{\ln m}$$

Ogni numero intero di k bit possiede al più $k-1$ divisori primi.

Si determina che la probabilità di scegliere un certo p risulta essere pari a:

$$\mathbb{P}(p) = \frac{1}{\text{prim}\left(n^2\right)}, \qquad \mathbb{P}\,(\text{errore}) = \frac{\#\{i \in \{1,\dots,n^2\}, i \text{ primo e divisore di } a-b\}}{\text{prim}\,(n)^2}$$

$$\mathbb{P}\,(\text{errore}) \leqslant \frac{n-1}{{}^{n^2}/_{2\log_2 n}} \leqslant \frac{2\log_2 n}{n} \longrightarrow 0$$

$$\mathbb{P}\,(\text{errore}) \leqslant 10^{-14}$$

Anche per n piccolo, se $\frac{2\ln n}{n} < \frac{1}{2}$ ripeto l'algoritmo t volte ed ottengo una probabilità di errore pari a $\mathbb{P}\,(\text{errore}) < \frac{1}{2}^t$

2.1.2 Esempio: commutatività di matrici

L'approccio deterministico usualmente termina in $\mathscr{O}\left(k^3\right)$ o $\mathscr{O}\left(k^\alpha\right)$ con $2 < \alpha < 3$.

La procedura probabilistica cerca di determinare se $\boldsymbol{AB} \neq \boldsymbol{BA}$.

$$\mathbb{P}\,(\text{errore}) = \mathbb{P}\left(\underline{\boldsymbol{u}} = \underline{\boldsymbol{b}}, \boldsymbol{AB} = \boldsymbol{BA}\right)$$

Algorithm 2: Commutatività di matrici

input : $\boldsymbol{A} = (A[1], \ldots, A[n]), A \in \mathbb{U}^n$
output: Il k-esimo elemento di A

```
1  begin
2  |  Scegli a ∈ {−1,1}^k a caso in modo uniforme;
3  |  u = (a' · A) B;
4  |  v = (a' · B) A;
5  |  if u ≠ v then
6  |  |  return "Sono diverse";
7  |  end
8  |  else
9  |  |  return "Sono uguali";
10 |  end
11 end
```

$$\boldsymbol{C} = \boldsymbol{AB} - \boldsymbol{BA} \neq 0 \Rightarrow \exists \text{ una colonna di } \boldsymbol{C} : \underline{\boldsymbol{c}}_i \neq \underline{\boldsymbol{0}}$$

Supponiamo che $\underline{\boldsymbol{c}}_i \neq \underline{\boldsymbol{0}}$, sapendo che $\underline{\boldsymbol{a}}\boldsymbol{AB} = \underline{\boldsymbol{a}}\boldsymbol{BA}$. Prendiamo $\underline{\boldsymbol{a}} = \begin{bmatrix} a_1 & a_2 & \ldots & a_k \end{bmatrix}$, $a_i \in \{-1, 1\}$. Il prodotto quando $\underline{\boldsymbol{u}} = \underline{\boldsymbol{v}}$ risulta essere:

$$\underline{\boldsymbol{a}}'\boldsymbol{AB} = \underline{\boldsymbol{a}}'\boldsymbol{BA} \Rightarrow \underline{\boldsymbol{a}}'\underline{\boldsymbol{c}} = 0 \Rightarrow \underline{\boldsymbol{a}}' \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = 0 \Rightarrow \sum_{i=1}^{k} a_i c_i = 0$$

Modifichiamo il vettore:

$$\underline{\boldsymbol{\alpha}} = \begin{bmatrix} -1 & a_2 & \ldots & a_k \end{bmatrix} \qquad \underline{\boldsymbol{\beta}} = \begin{bmatrix} 1 & a_2 & \ldots & a_k \end{bmatrix}$$

Uno di questi due vettori deve coincidere con A.

$$\underline{\boldsymbol{\alpha}}'\boldsymbol{C} \quad \vee \quad \underline{\boldsymbol{\beta}}'\boldsymbol{C} = 0$$

È importante notare che non può capire che entrambe le relazioni siano vere allo stesso tempo poiché implicherebbe, ricordando la premessa $\underline{\boldsymbol{c}}_i \neq 0$, che $\underline{\boldsymbol{c}}_i = 0$.

2.2 Las Vegas

Sono sempre esatti o ritornano 'impossibile determinare soluzione' (spesso rappresentato con un punto di domanda).

2.2.1 Esempio: RSEL

Si tratta di un algoritmo **Las Vegas** utilizzato per estrarre il k-esimo elemento da un vettore definito su dominio $\mathbb{U}$ totalmente ordinato (non esistono termini intercambiabili nell'ordine).

$$T(n,k) \leqslant T_{\text{Quicksort}}(n)$$

Caso pessimo $T(n,k) = \mathcal{O}\left(n^2\right)$

Caso medio $T(n,k) = \mathcal{O}\left(n \log n\right)$

Algorithm 3: RSEL

input : $A, B \in \mathbb{N}^{k \times k}$
output: Il k-esimo elemento di A

```
1  begin
2      if n=2 then
3          Risolvi il problema direttamente.
4      end
5      else
6          Scegli t ∈ {1,2,...,n} a caso in modo uniforme.;
7          Calcola A ⩽ {A[j], A[j] < A[t]}.;
8          Calcola A ⩾ {A[j], j ≠ t, A[j] ⩾ A[t]}.;
9          if #{A_<} ⩽ k−1 then
10             return A[t]
11         else if #{A_<} ⩾ k then
12             return RSEL(A_<, k)
13         else
14             return RSEL(A_>, k − #{A_<} − 1)
15         end
16     end
17 end
```

2.2.2 Esempio: Quicksort

Dato $\mathbb{E}(n)$, numero medio di confronti richiesto da **Quicksort** su un input di n elementi determinato come:

$$\mathbb{E}(n) = (n+1) + \frac{1}{n}\sum_{k=0}^{n-1}\left[\mathbb{E}(k) + \mathbb{E}(n-1-k)\right]$$

Usando il corollario del **teorema delle probabilità totali** che ricordiamo essere:

Definizione 2.2.1 (Formula delle probabilità totali). Sia $(\Omega, \mathscr{F}, P)$ uno spazio di probabilità e $F_1, F_2, \ldots, F_n \in \mathscr{F}$ una partizione finita di ω, $\bigcup_{k=1}^{n} F_k = \Omega$ e $F_h \cap F_k = \emptyset$ se $h \neq k$, tale che $\mathbb{P}(F_k) > 0$ per $k = 1, 2, \ldots, n$. Allora ogni evento $E \in \mathscr{F}$ si ha:

$$\mathbb{P}(E) = \sum_{k=1}^{n} \mathbb{P}(E \mid F_k)\,\mathbb{P}(F_k)$$

Corollario 2.2.1.1 (Corollario sulla media).

$$\mathbb{E}(E) = \sum_{k=1}^{n} \mathbb{E}(E \mid F_k)\,\mathbb{P}(F_k)$$

Ottengo che i **tempo di calcolo che Quicksort** impiegati con un input di dimensione n sono pari a:

$$\mathbb{E}(T_n) = \sum_{i=0}^{n-1} \mathbb{E}(T_n \mid B_i)\,\mathbb{P}(B_i) = n - 1 + \sum_{i=0}^{n-1} \mathbb{E}(T_i) + \mathbb{E}(T_{n-1-i})$$

$$\mathbb{E}(n) = 2(n+1)H_n - 4n$$

Dove H_n è l'n-esimo numero armonico, cioè $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

2.2.3 Numero medio di confronti in RSEL

$$T(n,k) = n - 1 + \frac{1}{n}\sum_{l=k}^{n-1} T(l,k) + \frac{1}{n}\frac{1}{n}\sum_{l=0}^{k-2} T(n-1-l, k-l-1)$$

Con $k \leqslant n, \#\{A_<\} = l$.

2.2.4 Esempio: Protocollo 10

Si tratta di un algoritmo **Las Vegas** per la comunicazione, che vuole stabilire se esiste un termine condiviso tra le due liste: $\exists i : x_i = y_i$.

$$\begin{aligned} A &: x_i, x_2, \ldots, x_{10}, \qquad x_i \in \{0,1\}^n \qquad \forall i \in 1, \ldots, 10 \\ B &: y_i, y_2, \ldots, y_{10}, \qquad y_i \in \{0,1\}^n \qquad \forall i \in 1, \ldots, 10 \end{aligned}$$

1. A genera 10 numeri primi $p_1, p_2, \ldots, p_10$, anche uguali, poiché estratti a caso in modo uniforme in $[0, n^2]$.

   A calcola $r_i = \text{resto}\left(\frac{x_i}{p_i}\right) \quad \forall i, \ldots, 10$

   A invia le 10 coppie di resti e numeri primi.

2. B calcola $s_i = \text{resto}\left(\frac{y_i}{p_i}\right) \quad i = 1, \ldots, 10.$

   Se sono tutti diversi allora B risponde "NO".

   Altrimenti sia $j = \min\{i : r_i = s_i\}$ e B invia $\left(y_j, J\right)$ ad A.

3. In quest'ultimo caso (2) A verifica se $x_j = y_j$ e risponde "SI", altrimenti risponde "non so".

L'algoritmo termina in $10n$ messaggi.

L'algoritmo nel caso di risposta "NO" richiede $10 \cdot n \cdot \left(2\log_2 n\right) + 1 = 40\log_2 n + 1$ e nel caso "SI" oppure "?" richiede $40\log n + 4 + n + 1$.

Le risposte "SI" e "NO" sono corrette.

2.2.5 Probabilità di errore del protocollo 10

Primo Caso $\mathbb{P}\left(? \mid \forall i\, x_i \neq y_i\right) = \mathbb{P}\left(\exists l \mid r_l = s_l : \forall i\, x_i \neq y_i\right) \leqslant \sum_{l=1}^{1} 0\mathbb{P}\left(r_l = s_l : \forall i\, x_i \neq y_i\right) = 10 \cdot \frac{2\ln n}{n} = \frac{10\ln n}{n} \Rightarrow n \geqslant 40\ln n$

Secondo Caso

$$\begin{aligned} \mathbb{P}\left(? \mid \exists l : x_l = y_l\right) &\leqslant \mathbb{P}\left(\exists j < l : r_j = s_j \;\middle|\; l = \#\left\{i : x_i = y_i\right\}\right) \\ &\leqslant \mathbb{P}\left(\exists j < 10 : r_j = s_j \;\middle|\; 10 = \min\left\{i : x_i = y_i\right\}\right) \\ &\leqslant g \cdot \mathbb{P}\left(r_1 = s_1 \mid x_1 \neq y_1\right) \\ &\leqslant g \cdot \frac{2\ln n}{n} = 18 \cdot \frac{\ln n}{n} \leqslant \frac{1}{2} \\ &\Rightarrow n \geqslant 40\ln n \end{aligned}$$

2.3 Errore limitato (2-Sided Error)

$$f : I \to O, \qquad \exists \epsilon > 0 \quad \mathbb{P}\left(A(x)) f(x)\right) \geqslant \frac{1}{2} + \epsilon$$

La probabilità di rispondere correttamente è $\mathbb{P}$ (corretto) $\geqslant \frac{1}{2} + \epsilon$. Non si è mai assolutamente certi della risposta ottenuta ma si sceglie quella che appare più frequentemente. Generalmente si usa la **disuguaglianza di Chernoff** per determinare il numero di ripetizioni dell'algoritmo per avere un errore accettabile.

L'errore ϵ non dipende da x.

2.4 Errore Illimitato

$$f : I \to O, \qquad \exists \epsilon > 0 \quad \mathbb{P}\left(A(x)) f(x)\right) \geqslant \frac{1}{2} + \epsilon(x)$$

Sono una tipologia di algoritmi probabilistici usati per stimare funzioni ma **non ho mai** la certezza di una risposta corretta.

L'errore ϵ dipende dalla x.

3

Matrici e grafi

3.1 Periodo di nodi connessi bi-direzionalmente

Teorema 3.1.1 (Periodo di nodi connessi bi-direzionalmente). Dato un grafo orientato G consideriamo due nodi distinti i, j tali che $i \leftrightarrow j$. Allora $d(i) = d(j)$.

Periodo di nodi connessi bi-direzionalmente. Consideriamo un ciclo C qualsiasi di lunghezza s passante per j, un cammino C_1 da j a i lungo u ed un cammino C_2 da i a j di lunghezza v.

Dimostriamo che il periodo di i è divisore di C I cammini C_1 e C_2 formano un ciclo passante per i e quindi $d(i)$ ne è, per definizione, divisore. Possiamo costruire un ciclo di dimensione maggiore, combinando C_1, C_2 e C: nuovamente, per definizione, $d(i)$ ne è divisore.

Essendo $d(i)$ divisore sia di $u+v$ che $u+v+s$ è certamente divisore anche di s.

Dimostriamo che $d(i) = d(j)$ Il massimo dei divisore di C è $d(j)$ *per definizione*, per cui:

$$d(i) \leqslant d(j)$$

Ripetendo il ragionamento analogamente su j e i si ottiene anche la seconda disequazione:

$$d(j) \leqslant d(i)$$

Da cui la tesi:

$$d(i) = d(j)$$

□

3.2 Proprietà dei coefficienti di matrici irriducibili

Teorema 3.2.1 (Proprietà dei coefficienti di matrici irriducibili). Data una matrice irriducibile di periodo d, scelto un indice i arbitrario, per ogni indice j esistono due interi $r_j, q_j : 0 \leqslant r_j < d \quad q_j > 0$ tali che:

1. Se $a_{ij}^{(s)} > 0$ allora $s - \lfloor \frac{s}{d} \rfloor = r_j$
2. $a_{ij}^{(r_j+nd)} > 0 \quad \forall n \geqslant q_j$

3.3 Relazione tra matrici primitive e irriducibili

Teorema 3.3.1 (Relazione tra matrici primitive e irriducibili). Una matrice non negativa $\boldsymbol{A}$ è primitiva se e solo se $\boldsymbol{A}$ è irriducibile e aperiodica.

Relazione tra matrici primitive e irriducibili. **Se matrice è primitiva allora è irriducibile e aperiodica**

Per definizione, tutte le matrici primitive sono irriducibili e aperiodiche.

Se una matrice è irriducibile e aperiodica allora è primitiva

Supponiamo che $\boldsymbol{A}$ sia irriducibile e aperiodica. Utilizzando le **proprietà dei coefficienti di matrici irriducibili**, notiamo che tutti i coefficienti r_j in questo caso sono **nulli**. Per la seconda proprietà, essendo $r_j = 0$ e $d = 1$ deve valere che:

$$a_{ij}^{(n)} > 0 \quad \forall n \geqslant q_j$$

È quindi possibile scegliere un n tale per cui:

$$\boldsymbol{A}^n > 0$$

Per cui $\boldsymbol{A}$ è **primitiva**. □

3.4 Teorema di Perron-Frobenius per le matrici primitive

Teorema 3.4.1 (Teorema di Perron-Frobenius per le matrici primitive). Sia $\boldsymbol{A} \geqslant 0$ una matrice primitiva. Allora esiste un autovalore di $\boldsymbol{A}$ λ tale che:

1. λ è reale e strettamente positivo.
2. Tutti gli autovalori diversi da λ sono minori in modulo.
3. λ ammette autovettori destri e sinistri strettamente positivi.
4. λ è radice semplice dell'equazione caratteristica (ha molteplicità algebrica unitaria).

3.5 Teorema di Perron-Frobenius per le matrici irriducibili

Teorema 3.5.1 (Teorema di Perron-Frobenius per le matrici irriducibili). Sia $\boldsymbol{A} \geqslant 0$ una matrice irriducibile. Allora esiste un autovalore di $\boldsymbol{A}$ λ che rispetta **le proprietà del teorema di Perron-Frobenius per le matrici primitive** con l'eccezione che tutti gli autovalori diversi da λ sono minori **o uguali** in modulo.

3.6 Proprietà di autovalore unitario delle matrici stocastiche

Teorema 3.6.1 (Proprietà di autovalore unitario delle matrici stocastiche). Sia $\boldsymbol{P}$ una matrice stocastica di dimensione r. Allora 1 è autovalore di $\boldsymbol{P}$ e ammette come corrispondente autovettore destro il vettore $\underline{\boldsymbol{e}}$, di dimensione r tale che $\underline{\boldsymbol{e}}^T = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$.

Proprietà di autovalore unitario delle matrici stocastiche. Sia $\boldsymbol{P} = \begin{bmatrix} p_{ij} \end{bmatrix}$. Le righe di una **matrice stocastica** hanno somma unitaria, per cui vale che:

$$\boldsymbol{P}\underline{\boldsymbol{e}} = \underline{\boldsymbol{e}}$$

Da cui segue che 1 è un autovalore di $\boldsymbol{P}$ e che $\underline{\boldsymbol{e}}$ è un autovettore destro di $\boldsymbol{P}$ corrispondente a 1. □

3.7 Proprietà degli autovalori delle matrici stocastiche

Teorema 3.7.1 (Proprietà degli autovalori delle matrici stocastiche). Se λ è un autovalore di una matrice stocastica allora $|\lambda| \leqslant 1$.

Proprietà degli autovalori delle matrici stocastiche. Sia $\boldsymbol{P} = \begin{bmatrix} p_{ij} \end{bmatrix}$ una matrice stocastica di dimensione r e sia λ un suo autovalore. Consideriamo un autovettore sinistro $\underline{\boldsymbol{v}}$ di $\boldsymbol{P}$ corrispondente a λ. Allora $\underline{\boldsymbol{v}}^T\boldsymbol{P} = \lambda\underline{\boldsymbol{v}}^T$ e siccome la somma delle righe di una matrice stocastica è unitaria, segue che $|\lambda|\sum_{j=1}^{r}|v_j| \leqslant \sum_{j=1}^{r}|v_j| \Rightarrow |\lambda| \leqslant 1$. □

3.8 Teorema di Perron-Frobenius per le matrici stocastiche

Teorema 3.8.1 (Teorema di Perron-Frobenius per le matrici stocastiche). Se $\boldsymbol{P}$ è una matrice stocastica primitiva, allora 1 è il suo autovettore di **Perron-Frobenius** ed inoltre, per qualche $0 \leqslant \epsilon < 1$, abbiamo che:

$$\boldsymbol{P}^n = \underline{\boldsymbol{e}}\underline{\boldsymbol{v}}^T + \mathcal{O}\left(\epsilon^n\right)$$

dove $\underline{\boldsymbol{v}}$ è l'autovettore sinistro di $\boldsymbol{P}$ corrispondente a 1 mentre $\underline{\boldsymbol{e}}$ ne è l'autovettore destro.

4

Catene di Markov

4.1 Definizione di catena di Markov

4.1.1 Stati

Definizione 4.1.1 (Stati). Un insieme finito $S = \{1, 2, \dots, k\}$ di elementi detto insieme di **stati**.

4.1.2 Distribuzione iniziale

Definizione 4.1.2 (Distribuzione iniziale). Una distribuzione di probabilità $\mu : S \to \mathbb{R}$ definita su un insieme di stati S, tale che:

$$\mu(i) \geqslant 0 \quad \forall i \in S \qquad \sum_{i \in S} \mu(i) = 1$$

4.1.3 Matrice di transizione

Definizione 4.1.3 (Matrice di transizione). Una matrice **stocastica** $\boldsymbol{P}$ con indici in un insieme di stati S, detta **matrice di transizione** , ovvero una famiglia di coefficienti tale che:

$$p_{ij} \geqslant 0 \qquad \forall i, j \in S \qquad \sum_{j \in S} p_{ij} = 1 \quad \forall i \in S$$

Definizione 4.1.4 (Catena di Markov). Una **Catena di Markov finita e omogenea** con spazio degli stati S, distribuzione iniziale μ e matrice di transizione $\boldsymbol{P}$ è una sequenza di variabili aleatorie $\{X_n\}_{n \in \mathbb{N}}$ a valori in S tale che:

1. $\mathbb{P}(X_0 = i) = \mu(i) \quad \forall i \in S$
2. La probabilità di trovarsi in uno stato è data unicamente dallo stato precedente, proprietà nota anche come **proprietà di Markov**.
3. Un elemento p_{ij} della matrice rappresenta la probabilità di spostarsi nello stato j dallo stato i.

4.2 Probabilità di transizione

Teorema 4.2.1 (Probabilità di transizione). Supponendo che la probabilità di trovarsi nello stato i $\mathbb{P}_\mu(X_k = i)$ sia non nulla, la probabilità di trovarsi nello stato j dopo n passi è pari a:

$$\mathbb{P}_\mu\left(X_{k+n} = j \mid X_k = i\right) = p_{ij}^{(n)}$$

Probabilità di transizione. **Caso $n = 0$:**

Per $n = 0$, $p_{ij}^{(0)}$ coincide con il **coefficiente di Kronecker** σ_{ij}: senza muoversi si rimane certamente nello stato k.

Caso $n = 1$:

Per $n = 1$ il coefficiente $p_{ij}^{(1)}$ coincide con il coefficiente della **matrice di transizione**, per cui per definizione è vero.

Caso $n \geqslant 1$:

Ragionando per induzione, ricordando che **per ogni tripla di eventi** A, B, C vale che:

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(B \mid C) \cdot \mathbb{P}(A \mid B \cap C)$$

Si ottiene così la catena di uguaglianze:

$$\begin{aligned}
\mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_k = i\right) &= \mathbb{P}_\mu\left(X_{k+n+1} = j \cap \exists l \in S : X_{k+1} = l \mid X_k = i\right) \\
&= \sum_{l \in S} \mathbb{P}_\mu\left(X_{k+n+1} = j \cap X_{k+1} = l \mid X_k = i\right) \\
&= \sum_{l \in S} \mathbb{P}_\mu(X_{k+1} = l \mid X_k = i) \cdot \mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_{k+1} = l \cap X_k = i\right)
\end{aligned}$$

Siccome $\mathbb{P}_\mu(X_{k+1} = l \mid X_k = i) > 0 \Rightarrow \mathbb{P}_\mu(X_{k+1} = l, X_k = i) > 0$, per la **proprietà di Markov** si ottiene:

$$\begin{aligned}
\mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_k = i\right) &= \sum_{l \in S} \mathbb{P}_\mu(X_{k+1} = l \mid X_k = i) \cdot \mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_{k+1} = l\right) \\
&= \sum_{l \in S} p_{il} \cdot \mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_{k+1} = l\right)
\end{aligned}$$

Che **per ipotesi di induzione** diventa:

$$\mathbb{P}_\mu\left(X_{k+n+1} = j \mid X_k = i\right) = \sum_{l \in S} p_{il} \cdot p_{lj}^{(n)} = p_{ij}^{(n+1)}$$

□

4.3 Funzione di Green

Definizione 4.3.1 (Funzione di Green). La **funzione di Green** associata a una matrice stocastica $\boldsymbol{P}$ è una matrice di funzioni $\boldsymbol{G}(z)$ data da:

$$\boldsymbol{G}(z) = (\boldsymbol{I} - \boldsymbol{P}z)^{-1} = \sum_{n=0}^{+\infty} \boldsymbol{P}^n z^n$$

definita per ogni $z \in \mathbb{C}$ tale che $|z| \leqslant 1$.

4.4 Tempi di prima entrata

Definizione 4.4.1 (Tempi di prima entrata). Prendendo in considerazione una catena di Markov finita ed omogenea $\{X_n\}$, con spazio degli stati S e matrice di transizione $\boldsymbol{P}$. Per ogni $j \in S$ denotiamo con τ_j la variabile aleatoria che rappresenta il minimo numero di passi $n \geqslant 1$ tale che $X_n = j$, cioè il tempo di attesa necessario per entrare in j per la prima volta dopo l'istante 0:

$$\tau_j = \min\left\{ n \geqslant 1 | X_n = j \right\}, \quad \tau_j \in \mathbb{N}_+ \cup \{+\infty\}$$

4.5 Probabilità di ingresso

Definizione 4.5.1 (Probabilità di ingresso). Il coefficiente f_{ij} denota la probabilità di ingresso in uno stato j partendo da uno stato i. Essi possiedono alcune caratteristiche:

$$\begin{aligned} f^{(0)}(i,j) &= 0 \\ f^{(n)}(i,j) &= \mathbb{P}_i\left(\tau_j = n\right) \quad \forall n \in \mathbb{N}_+ \\ f(i,j) &= \sum_{n \geqslant 1} f^{(n)}(i,j) = \mathbb{P}_i\left(\tau_j < \infty\right) = 1 - \mathbb{P}_i\left(\tau_j = \infty\right) \end{aligned}$$

I coefficienti sulla diagonale principale f_{ii} sono detti **coefficienti di probabilità di rientro**.

4.6 Relazione tra probabilità di transizione e di ingresso

Teorema 4.6.1 (Relazione tra probabilità di transizione e di ingresso). Per transitare da i a j in n passi è necessario entrare in j per la prima volta in $k \leqslant n$ passi e poi rientrarvi in $n-k$ passi.

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{ij}^{(n-k)}$$

Relazione tra probabilità di transizione e di ingresso. Considerando **l'evento condizionato** su $\tau_j = k$:

$$p_{ij}^{(n)} = \mathbb{P}_i\left(X_n = j\right) = \sum_{k=1}^{n} \mathbb{P}_i\left(X_n = j \;\middle|\; \tau_j = k\right) \mathbb{P}_i\left(\tau_j = k\right)$$

Procediamo ad ottenere il primo coefficiente della sommatoria: se $\mathbb{P}_i\left(\tau_j = k\right) \neq 0$ per la **proprietà di Markov** vale che:

$$\mathbb{P}_i\left(X_n = j \;\middle|\; \tau_j = k\right) = \mathbb{P}_i\left(X_n = j \;\middle|\; X_k = j\right) = p_{jj}^{(n-k)}$$

Il secondo coefficiente già lo conosciamo: si tratta della definizione di coefficiente di probabilità di ingresso, $f_{ij}^{(k)} = \mathbb{P}_i\left(\tau_j = k\right)$.

Sostituendo nella sommatoria ottengo la relazione che volevamo raggiungere:

$$p_{ij}^{(n)} = \sum_{k=1}^{n} p_{jj}^{(n-k)} f_{ij}^{(k)}$$

□

4.7 Funzione generatrice di probabilità di ingresso

Definizione 4.7.1 (Funzione generatrice di probabilità di ingresso). La **relazione tra probabilità di transizione e di ingresso** stabilisce una convoluzione tra sequenze e può quindi essere trasformata in un prodotto tra **funzioni generatrici**.

$$F_{ij}(z) = \sum_{n=0}^{+\infty} f_{ij}^{(n)} z^n$$

La serie converge per $z = 1$ ed il suo raggio di convergenza è maggiore o uguale a 1.

4.8 Relazione tra f. di Green e f. gen. di prob. di ingresso

Teorema 4.8.1 (Relazione tra funzione di Green e funzione generatrice delle probabilità di ingresso). Per ogni $i, j \in S$, se $i \neq j$:

$$G_{ij}(z) = F_{ij}(z) G_{jj}(z)$$

Nel caso $i = j$ abbiamo invece:

$$G_{jj} = \frac{1}{1 - F_{jj}(z)}$$

4.9 Stati ricorrenti

Teorema 4.9.1 (Stati ricorrenti). Per ogni stato $i \in S$ le seguenti proprietà sono equivalenti:

1. i è ricorrente
2. $f_{ii} = 1$
3. $\sum_{n \geqslant 0} p_{ii}^{(n)} = +\infty$

4.10 La somma delle probabilità di transienza è limitata

Osservazione 4.10.1 (La somma delle probabilità di transienza è limitata). Per ogni $i, j \in S$, se j è transiente allora vale che:

$$\sum_{n \geqslant 0} p_{ij}^{(n)} \leqslant +\infty$$

4.11 La ricorrenza è proprietà delle classi

Teorema 4.11.1 (La ricorrenza è proprietà delle classi). La ricorrenza è una proprietà delle classi. Ovvero, se $i \in S$ è uno stato ricorrente e C è la sua classe allora ogni stato $j \in C$ è ricorrente.

La ricorrenza è proprietà delle classi. Data una classe C, siano $i, j \in C$ due stati distinti. Allora esistono due valori $a, b > 0$ tali che $a = p_{ij}^{(p)}$ e $b = p_{ji}^{(s)}$ per qualche $r, s \in \mathbb{N}$ tali che:

$$p_{jj}^{(n+r+s)} \geqslant b a p_{ii}^{(n)}$$

Se i è ricorrente, osservando che **la somma delle probabilità di transienza è limitata**, se i è **ricorrente** la serie è **divergente**. Per la maggiorazione appena svolta anche $p_{jj}^{(n+r+s)}$ diverge, per cui j è ricorrente.

Per **simmetria**, se i fosse ricorrente anche j lo sarebbe. □

4.12 Ereditarietà della ricorrenza

Teorema 4.12.1 (Ereditarietà della ricorrenza). Se uno stato i è ricorrente e $i \to j$, allora anche j è ricorrente e inoltre $f_{ij} = f_{ji} = 1$

4.13 Relazioni tra stati ricorrenti e stati essenziali

Teorema 4.13.1 (Relazione tra stati ricorrenti e stati essenziali). Ogni stato ricorrente è essenziale.

Teorema 4.13.2 (Relazione tra stati ricorrenti e stati essenziali in una catena di Markov). In una catena di Markov finita e omogenea ogni stato essenziale è ricorrente.

4.14 Probabilità di rientro per stati transienti

Teorema 4.14.1 (Probabilità di rientro per stati transienti). Per ogni $i, j \in S$, se j è transiente allora:

$$P_{ij}^{(n)} = \mathcal{O}\left(\epsilon^n\right)$$

Per qualche $0 < \epsilon < 1$.

4.15 Tempi medi di rientro

Definizione 4.15.1 (Tempi medi di rientro). Supponendo $X_0 = i$ ricorrente, definiamo il tempo medio di rientro nello stato i:

$$\mathbb{E}_i(\tau_i) = \sum_{n\geqslant 1} n\mathbb{P}_i(\tau_i = n) = \sum_{n\geqslant 1} n f_{ii}^{(n)}$$

Teorema 4.15.2 (Il tempo medio di nodi ricorrenti è finito). Per ogni nodo $i \in S$ ricorrente vale che:

$$\mathbb{E}_i(\tau_i) < +\infty$$

Il tempo medio di nodi ricorrenti è finito. Per il **teorema della probabilità di rientro per stati transienti**, possiamo limitarci a provare che $f_{ii}^{(n)} = \mathcal{O}(\epsilon^n)$.

Essendo inoltre i essenziale, per la **relazione tra stati essenziali e ricorrenti**, possiamo restringere l'insieme degli stati S alla classe C di i.

Definiamo una nuova catena di Markov sugli stati C, rendendo i assorbente. Definiamo la nuova matrice di transizione $\tilde{\boldsymbol{P}}$ della catena come:

$$\tilde{p}_{kj} = p_{kj} \quad k \neq i$$
$$\tilde{p}_{ij} = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

Tutti gli stati $j \neq i$ sono **transienti** e quindi:

$$\sum_{j\neq i} \tilde{p}_{kj}^{(n)} = \mathcal{O}\left(\epsilon^n\right)$$

Ne segue allora che, per ogni intero $n \geqslant 2$:

$$\begin{aligned} f_{ii}^{(n)} &= \sum_{k,j\in Cmk,j\neq i} p_{ik}\tilde{p}_{kj}^{(n-2)}p_{ji} \\ &= \mathcal{O}\left(\epsilon^n\right)\cdot \sum_{k,j\in Cmk,j\neq i} p_{ik}p_{ji} = \mathcal{O}\left(\epsilon^n\right) \end{aligned}$$

□

Teorema 4.15.3 (Tempi medi di rientro di matrici irriducibili). Se $\boldsymbol{P}$ è **irriducibile**, per ogni nodo $i, j \in S$ vale che:

$$\mathbb{E}_i(\tau_i) = \sum_{n\geqslant 1} n f_{ij}^{(n)} < +\infty$$

5

Catene di Markov ergodiche

5.1 Catena ergodica

Teorema 5.1.1 (Catena ergodica). Una catena di Markov finita $\{X_n\}$ su un insieme di stati S si dice **ergodica** se esiste un vettore stocastico $\underline{\boldsymbol{\pi}}^* = \left(\pi_i^*\right)_{i\in S}$ tale che, per ogni $i, j \in S$:

$$\lim_{n\to+\infty} \mathbb{P}_i\left(X_n = j\right) = \lim_{n\to+\infty} p_{ij}^{(n)} = \pi_j^*$$

Equivale a richiedere che per ogni distribuzione $\underline{\boldsymbol{\mu}}$ definita su S e ogni $j \in S$:

$$\lim_{n\to+\infty} \left(\underline{\boldsymbol{\mu}}^T \boldsymbol{P}^n\right)_j = \pi_j^*$$

Una catena di Markov $\{X_n\}$ è **ergodica** se al crescere di n la distribuzione limite della variabile X_n esiste ed è indipendente dalla distribuzione iniziale $\underline{\boldsymbol{\mu}}$.

5.2 Distribuzioni stazionarie

Definizione 5.2.1 (Distribuzioni stazionarie). Data una catena di Markov finita $\{X_n\}$ sull'insieme di stati S, con matrice di transizione $\boldsymbol{P}$, chiamiamo **distribuzione stazionaria** un vettore $\pi \in \mathbb{R}^k$ che soddisfa le seguenti proprietà:

1. $\underline{\boldsymbol{\pi}}$ è un vettore **stocastico**, ovvero $\pi_i \geqslant 0 \quad \forall i \in S$ e la somma è pari a 1.
2. $\underline{\boldsymbol{\pi}}$ è autovettore sinistro di $\boldsymbol{P}$ corrispondente a 1, ovvero $\underline{\boldsymbol{\pi}}^T \boldsymbol{P} = \underline{\boldsymbol{\pi}}^T$

Un vettore con tali proprietà verifica l'equazione:

$$\underline{\boldsymbol{\pi}}^T \boldsymbol{P}^n = \underline{\boldsymbol{\pi}}^T \quad \forall n \in \mathbb{N}$$

Per cui vale che:

$$\mathbb{P}_{\underline{\boldsymbol{\pi}}}\left(X_n = i\right) = \pi_i \quad \forall i \in S$$

Se $\underline{\boldsymbol{\pi}}$ coincide con la distribuzione iniziale della catena, la probabilità di trovarsi nei vari stati rimane sempre la stessa durante l'evoluzione del processo.

Definizione 5.2.2 (Coefficienti di una distribuzione stazionaria). Poiché ogni catena finita ammette sempre una classe **ricorrente**, senza perdita di generalità supponiamo che lo stato 1 sia ricorrente. Allora, $\forall i \in S$ definiamo:

$$\rho_i = \sum_{n=0}^{+\infty} \mathbb{P}_1\left(X_n = i, \tau_1 > n\right)$$

Definizione 5.2.3 (Proprietà dei coefficienti di una distribuzione stazionaria). I coefficienti di una distribuzione stazionaria godono di 3 proprietà:

1. Ogni ρ_i è finito.
2. $\rho_1 = 1$
3. Ogni ρ_i rappresenta il numero medio di entrate nello stato i nell'intervallo di tempo $\{0, 1, \ldots, \tau_1 - 1\}$, supponendo $X_0 = 1$.

Teorema 5.2.4 (Esistenza di distribuzione stazionaria). Sia $\{X_n\}$ una catena di Markov sugli stati $\{1,2,\dots,k\}$ con matrice di transizione $\boldsymbol{P}$ e supponiamo che lo stato 1 sia ricorrente. Allora il vettore $\underline{\boldsymbol{\pi}}$ è una distribuzione stazionaria per $\{X_n\}$:

$$\underline{\boldsymbol{\pi}} = \left[\frac{\rho_1}{\mathbb{E}_1(\tau_1)} \quad \frac{\rho_2}{\mathbb{E}_1(\tau_1)} \quad \cdots \quad \frac{\rho_k}{\mathbb{E}_1(\tau_1)} \right]$$

Esistenza di distribuzione stazionaria. È sufficiente provare che il vettore $\underline{\boldsymbol{\rho}}$ è autovettore sinistro di $\boldsymbol{P}$ in corrispondenza dell'autovalore 1.

Consideriamo prima il caso $j \neq 1$ e proviamo che $\rho_j = \left(\underline{\boldsymbol{\rho}}^T \boldsymbol{P}\right)_j$.

Procediamo condizionando la definizione di ρ_j condizionando sullo stato raggiunto al passo n-esimo e successivamente applichiamo la condizione di Markov:

$$\begin{aligned}
\rho_j &= \sum_{n=1}^{+\infty} \mathbb{P}_1\left(X_n = j, \tau_1 > n\right) \\
&= \sum_{n=1}^{+\infty} \mathbb{P}_1\left(X_n = j, \tau_1 > n-1\right) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \mathbb{P}_1\left(X_n = j, X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \mathbb{P}_1\left(\, X_n = j \;\middle|\; X_{n-1} = i, \tau_1 > n-1\right) \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right)
\end{aligned}$$

Se $i \neq 1$ e $n > 1$ il primo termine, per la proprietà di Markov, coincide con $p_i j$ e negli altri casi il secondo termine è nullo. Ne segue che:

$$\begin{aligned}
\rho_j &= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} p_{ij} \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{i=1}^{k} p_{ij} \sum_{n=1}^{+\infty} \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{i=1}^{k} p_{ij} \rho_i = \left(\underline{\boldsymbol{\rho}}^T \boldsymbol{P}\right)_j
\end{aligned}$$

Consideriamo ora il caso in cui $j = 1$. Poiché lo stato è ricorrente:

$$\begin{aligned}
\rho_1 &= \sum_{n=1}^{+\infty} \mathbb{P}_1\left(X_n = 1, \tau_1 = n\right) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \mathbb{P}_1\left(X_n = 1, X_{n-1} = i, \tau_1 = n\right) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} \mathbb{P}_1\left(\, X_n = 1 \mid X_{n-1} = i, \tau_1 > n-1\right) \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{n=1}^{+\infty} \sum_{i=1}^{k} p_{i1} \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{i=1}^{k} p_{i1} \sum_{n=1}^{+\infty} \mathbb{P}_1\left(X_{n-1} = i, \tau_1 > n-1\right) \\
&= \sum_{i=1}^{k} p_{i1} \rho_i = \left(\underline{\boldsymbol{\rho}}^T \boldsymbol{P}\right)_1
\end{aligned}$$

□

5.3 Catena primitiva

Definizione 5.3.1 (Insieme dei vettori stocastici). Sia $\mathbb{R}_+$ l'insieme dei reali non negativi. Per ogni $k \in \mathbb{N}$ denotiamo con M_k l'insieme dei **vettori stocastici** a k **componenti**:

$$M_k = \left\{ \underline{\boldsymbol{v}} \in \mathbb{R}_+^k \middle| \sum_{i=1}^{k} v_i = 1 \right\}$$

Definizione 5.3.2 (Distanza di variazione totale). Per ogni $\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}} \in M_k$ chiamiamo **distanza di variazione totale** tra i due vettori:

$$d_{TV}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}}) = \frac{1}{2} \sum_{i=1}^{k} |u_i - v_i|$$

Definizione 5.3.3 (Coefficienti di matrice stocastica). Data una matrice stocastica $\boldsymbol{P}$ di dimensione $k \times k$ e un qualsiasi $j \in \{1, 2, \ldots, k\}$, denotiamo con $\alpha(j)$ e $\beta(\boldsymbol{P})$ i valori:

$$\alpha(j) = \min p_{ij} \qquad \beta(\boldsymbol{P}) = 1 - \sum_{j=1}^{k} \alpha j$$

Dove $\beta(\boldsymbol{P})$ rispetta le seguenti proprietà:

1. $0 \leqslant \beta(\boldsymbol{P}) \leqslant 1$ per ogni matrice stocastica $\boldsymbol{P}$.
2. $\beta(\boldsymbol{P}) < 1$ se e solo se $\boldsymbol{P}$ possiede una colonna a valori tutti positivi.
3. $\beta(\boldsymbol{P}) < 1$ se $\boldsymbol{P} > 0$

Lemma 5.3.4 (Relazione tra matrici e vettori stocastici). Per ogni matrice stocastica $\boldsymbol{P} \in \mathbb{R}_+^{k\times k}$ e ogni coppia di vettori stocastici $\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}} \in M_k$ abbiamo:

$$\left\|\underline{\boldsymbol{u}}^T\boldsymbol{P} - \underline{\boldsymbol{v}}^T\boldsymbol{P}\right\|_1 \leqslant \beta(\boldsymbol{P})\left\|\underline{\boldsymbol{u}} - \underline{\boldsymbol{v}}\right\|_1$$

Relazione tra matrici e vettori stocastici. Per ogni $j = 1, 2, \ldots, k$ è possibile scrivere:

$$\begin{aligned}
\left(\underline{\boldsymbol{u}}^T\boldsymbol{P}\right)_j - \left(\underline{\boldsymbol{v}}^T\boldsymbol{P}\right)_j &= \sum_{i=1}^{k}(u_i - v_i)\,p_{ij} \\
&= \sum_{i=1}^{k}|u_i - v_i|\,p_{ij} - \left[\sum_{i=1}^{k}(|u_i - v_i| - (u_i - v_i))\,p_{ij}\right] \\
&\leqslant \sum_{i=1}^{k}|u_i - v_i|\,p_{ij} - \alpha(j)\left[\sum_{i=1}^{k}(|u_i - v_i| - (u_i - v_i))\right] \\
&= \sum_{i=1}^{k}|u_i - v_i|\left(p_{ij} - \alpha(j)\right)
\end{aligned}$$

In modo analogo si prova che $\left(\underline{\boldsymbol{v}}^T\boldsymbol{P}\right)_j - \left(\underline{\boldsymbol{u}}^T\boldsymbol{P}\right)_j \leqslant \sum_{i=1}^{k}|u_i - v_i|\left(p_{ij} - \alpha(j)\right)$ e quindi si ottiene:

$$\left|\left(\underline{\boldsymbol{u}}^T\boldsymbol{P}\right)_j - \left(\underline{\boldsymbol{v}}^T\boldsymbol{P}\right)_j\right| \leqslant \sum_{i=1}^{k}|u_i - v_i|\left(p_{ij} - \alpha(j)\right)$$

Questo implica:

$$\begin{aligned}
\left\|\underline{\boldsymbol{u}}^T\boldsymbol{P} - \underline{\boldsymbol{v}}^T\boldsymbol{P}\right\|_1 &= \sum_{j=1}^{k}\left|\left(\underline{\boldsymbol{u}}^T\boldsymbol{P}\right)_j - \left(\underline{\boldsymbol{v}}^T\boldsymbol{P}\right)_j\right| \\
&\leqslant \sum_{j=1}^{k}\sum_{i=1}^{k}|u_i - v_i|\left(p_{ij} - \alpha(j)\right) \\
&\leqslant \sum_{i=1}^{k}|u_i - v_i|\sum_{j=1}^{k}\left(p_{ij} - \alpha(j)\right) \\
&\leqslant \sum_{i=1}^{k}|u_i - v_i|\,\beta(\boldsymbol{P}) \\
&\leqslant \beta(P)\left\|\underline{\boldsymbol{v}} - \underline{\boldsymbol{u}}\right\|_1
\end{aligned}$$

□

Teorema 5.3.5 (Relazione tra potenze di matrici e vettori stocastici). Per ogni matrice stocastica $\boldsymbol{P} \in \mathbb{R}_+^{k\times k}$ primitiva esistono una costante $C > 0$ e un valore $0 < \epsilon < 1$ tali che per ogni coppia di vettori stocastici $\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}} \in M_k$:

$$\left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^n - \underline{\boldsymbol{v}}^T \boldsymbol{P}^n \right\|_1 \leqslant C\epsilon^n$$

Relazione tra potenze di matrici e vettori stocastici. Poiché $\boldsymbol{P}$ è primitiva esiste $t \in \mathbb{N}$ tale che $\boldsymbol{P}^t > 0$ e di conseguenza il suo coefficiente $0 \leqslant \beta(\boldsymbol{P}^n) < 1$.

Per ogni $n \in \mathbb{N}\ \exists q \in \mathbb{N}\ r \in \{0, 1, \ldots, t-1\} : n = qt + r$.

Applicando quindi il **lemma di relazione tra matrici e vettori stocastici**, si ottiene:

$$\begin{aligned}
\left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^n - \underline{\boldsymbol{v}}^T \boldsymbol{P}^n \right\|_1 &= \left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^r \left(\boldsymbol{P}^t\right)^q - \underline{\boldsymbol{v}}^T \boldsymbol{P}^r \left(\boldsymbol{P}^t\right)^q \right\|_1 \\
&\leqslant \left(\beta(\boldsymbol{P}^t)\right)^q \left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^r - \underline{\boldsymbol{v}}^T \boldsymbol{P}^r \right\|_1 \\
&\leqslant 2\left(\beta(\boldsymbol{P}^t)\right)^{\frac{n-r}{t}} \\
&\leqslant 2\left(\beta(\boldsymbol{P}^t)\right)^{-1} \left[\left(\beta(\boldsymbol{P}^t)\right)^{\frac{1}{t}}\right]^n \\
&\leqslant C\epsilon^n
\end{aligned}$$

Dove $C = 2\left(\beta(\boldsymbol{P}^t)\right)^{-1}$ e $\epsilon = \left(\beta(\boldsymbol{P}^t)\right)^{\frac{1}{t}} < 1$. □

5.4 Proprietà di catene di Markov primitive

Teorema 5.4.1 (Proprietà di catene di Markov primitive). Sia $\{X_n\}$ una catena di Markov con matrice di transizione primitiva sull'insieme di stati S. Allora valgono le seguenti proprietà:

1. $\{X_n\}$ possiede una sola distribuzione stazionaria $\underline{\boldsymbol{\pi}}^*$.
2. $\pi_i^* = \frac{1}{\mathbb{E}_i(\tau_i)} \quad \forall i \in S$
3. $\{X_n\}$ è ergodica e $\lim_{n\to+\infty} \mathbb{P}_i\left(X_n = j\right) = \pi_j^*$

Proprietà di catene di Markov primitive. Sia $\boldsymbol{P}$ la matrice di transizione della catena. Per la **relazione tra potenze di matrici e vettori stocastici** sappiamo che per ogni coppia di distribuzioni $\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}} \in S$ vale che:

$$\left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^n - \underline{\boldsymbol{v}}^T \boldsymbol{P}^n \right\|_1 \to 0 \quad \text{per } n \to +\infty$$

Se $\underline{\boldsymbol{u}}$ e $\underline{\boldsymbol{v}}$ sono distribuzioni stazionarie per la catena allora $\underline{\boldsymbol{u}}^T \boldsymbol{P}^n = \underline{\boldsymbol{u}}^T$ e $\underline{\boldsymbol{v}}^T \boldsymbol{P}^n = \underline{\boldsymbol{v}}^T$ e sostituendo nella relazione otteniamo:

$$\left\| \underline{\boldsymbol{u}}^T - \underline{\boldsymbol{v}}^T \right\|_1 = 0$$

Per cui $\underline{\boldsymbol{u}} = \underline{\boldsymbol{v}}$. La catena pertanto ammette un'unica distribuzione stazionaria π^*.

Tutti gli stati sono **ricorrenti**, per cui è possibile costruire la distribuzione π^* a partire da qualsiasi stato $i \in S$ invece che solamente da 1.

Ne segue che $\pi_i^* = \frac{1}{\mathbb{E}_i(\tau_i)} \quad \forall i \in S$.

Assumendo infine $\underline{\boldsymbol{v}} = \underline{\boldsymbol{\pi}}^*$, per ogni distribuzione $\underline{\boldsymbol{u}} \in S$ vale che:

$$\left\| \underline{\boldsymbol{u}}^T \boldsymbol{P}^n - \underline{\boldsymbol{\pi}}^{*T} \right\|_1 \to 0 \quad \Rightarrow \quad \left(\underline{\boldsymbol{u}}^T \boldsymbol{P}^n\right)_j \to \pi_j^* \quad \forall j \in \{1, 2, \ldots, k\}$$

Che è la definizione di catena ergodica. □

5.5 Esempio: catena irriducibile non primitiva

Prendendo in considerazione una catena formata da soli due stati con una matrice di transizione:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Chiaramente $\boldsymbol{P}$ è irriducibile di periodo 2 e si verifica subito che $p_{12}^{(n)} = 1$ per n dispari e $p_{12}^{(n)} = 0$ per n pari.

Al crescere di n quindi la sequenza non ammette limite., per quanto è possibile dimostrare una forma di convergenza più debole, data da una media dei valori di $p_{ij}^{(n)}$.

5.6 Catene riducibili

Nel caso di catene con un grado di connessione che contiene più componenti fortemente connesse non è possibile dare una risposta univoca ed il comportamento dipende dalla relazione di connessione tra le varie componenti e dalla periodicità dei loro elementi. È possibile provare che se esiste una sola classe essenziale e quest'ultima è aperiodica allora la catena è ergodica, viceversa allora la catena non può essere ergodica.

Teorema 5.6.1 (Catene riducibili ergodiche). Se la catena di Markov $\{M_k\}$ possiede una sola classe essenziale C e quest'ultima è **aperiodica** allora, per ogni $j \in C$ e ogni $i \in S$:

$$\lim_{n \to +\infty} p_{ij}^{(n)} = \frac{1}{\sum_{n \geq 1} n f_{ij}^{(n)}}$$

La catena risulta pertanto **ergodica**.

Teorema 5.6.2 (Relazione tra classi essenziali e catene ergodiche). Una catena di Markov finita è **ergodica** se e solo se possiede un'unica classe essenziale e quest'ultima è **aperiodica**.

6 Catene reversibili

6.1 Catena di Markov reversibile

Definizione 6.1.1 (Catena di Markov reversibile). Una catena di Markov $\{X_n\}_n$ definita sugli spazi S con matrice di transizione $\boldsymbol{P}$ si dice **reversibile** se esiste un **vettore stocastico** $\underline{\boldsymbol{\pi}}$ definito in S tale che:

$$\pi_i p_{ij} = \pi_j p_{ji}$$

O equivalentemente per qualsiasi sequenza di finita di stati:

$$\mathbb{P}_{\underline{\boldsymbol{\pi}}}\left(X_0 = i, X_1 = j\right) = \mathbb{P}_{\underline{\boldsymbol{\pi}}}\left(X_0 = j, X_1 = i\right)$$

Il vettore $\underline{\boldsymbol{\pi}}$ è detto **distribuzione reversibile** per la catena $\{X_n\}_n$.

Teorema 6.1.2 (Distribuzioni reversibili stazionarie). Se $\underline{\boldsymbol{\pi}}$ è una distribuzione reversibile per una catena $\{X_n\}_n$ allora $\underline{\boldsymbol{\pi}}$ è anche una distribuzione stazionaria per $\{X_n\}_n$.

Distribuzioni reversibili stazionarie. Siano S e $\boldsymbol{P}$ rispettivamente, l'insieme degli stati e la matrice di transizione della catena. Poiché $\underline{\boldsymbol{\pi}}$ è una **distribuzione reversibile**, per ogni $j \in S$ abbiamo che:

$$\left\{\underline{\boldsymbol{\pi}}^T \boldsymbol{P}\right\}_j = \sum_{i=1}^{m} \pi_i p_{ij} = \sum_{i=1}^{m} \pi_i p_{ji} = \pi_j$$

Di conseguenza $\underline{\boldsymbol{\pi}}^T \boldsymbol{P} = \underline{\boldsymbol{\pi}}^T$ e quindi $\underline{\boldsymbol{\pi}}$ risulta stazionaria. □

Teorema 6.1.3 (Matrice simmetrica irriducibile bistocastica). Sia $\{X_n\}_n$ una catena di Markov con matrice di transizione $\boldsymbol{P}$ irriducibile e bistocastica. Se $\{X_n\}_n$ è reversibile allora $\boldsymbol{P}$ è simmetrica.

Matrice simmetrica irriducibile bistocastica. Infatti, essendo $\boldsymbol{P}$ irriducibile e bistocastica, la sua unica distribuzione stazionaria è quella uniforme $\underline{\boldsymbol{\pi}} = \begin{bmatrix} \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{bmatrix}$ dove k è la dimensione della matrice $\boldsymbol{P}$.

Inoltre per la reversibilità della catena $\underline{\boldsymbol{\pi}}$ è anche l'unica distribuzione reversibile.

Di conseguenza, $\pi_i \boldsymbol{P}_{ij} = \pi_j \boldsymbol{P}_{ji}$ per ogni coppia di indici i, j il che implica $\boldsymbol{P}_{ij} = \boldsymbol{P}_{ji}$. □

6.2 Passeggiate a caso su grafi

Teorema 6.2.1 (Catena delle passeggiate a caso su grafo non orientato). Se G è un grafo non orientato connesso allora la catena delle passeggiate a caso su G gode delle seguenti proprietà:

1. La catena è **irriducibile** e **reversibile**.
2. La sua **periodicità** è al più 2.
3. La sua distribuzione stazionaria è data dal vettore $\pi = \begin{bmatrix} \frac{d_1}{2m} & \frac{d_2}{2m} & \cdots & \frac{d_k}{2m} \end{bmatrix}$
4. Per ogni nodo i di G il tempo medio di rientro in i è dato da $\mathbb{E}_i(\tau_i) = 2\frac{m}{d_i}$ dove m è il numero di lati di G e d_i è il grado di i.

6.3 Passeggiate in un cammino semplice

Teorema 6.3.1 (Lunghezza di una passeggiata). Compiendo una passeggiata a caso in un grafo formato da un cammino semplice di k nodi il numero medio di un passi necessari per raggiungere una estremità del grafo a partire da un vertice qualunque è minore o uguale a k^2.

6.4 Problema 2-CNF SODD

Il problema consiste nella determinazione della soddisfacibilità di formule booleane in forma normale seconda congiunta.

La procedura è un algoritmo **one-sided error** e si procede ripetendo un certo numero di volte un ciclo principale di istruzioni nel quale partendo da un assegnamento casuale A di valori alle variabili in V si verifica se A soddisfa la formula booleana data Φ. In caso affermativo è stato identificato un assegnamento che rende vera a formula, altrimenti si procede a modificare il valore di una variabile che compare in una clausola non correntemente soddisfatta di A e si ripete con il nuovo assegnamento.

Il ciclo interno viene ripetuto $2k^2$ volte per portare la probabilità di errore a $\frac{1}{2}$, dato che per il teorema della **lunghezza di una passeggiata a caso** il tempo medio per raggiungere un'estremità del grafo da un vertice è k^2 e per la disuguaglianza di Markov:

$$\mathbb{P}\left(u_i > 2k^2\right) \leqslant \frac{\mathbb{E}\left(u_i\right)}{2k^2} \leqslant \frac{1}{2}$$

7

Monte Carlo Markov chain

7.1 Cosa sono i metodi MCMC

Tipicamente nella **generazione casuale** si utilizzano catene di Markov definite su un insieme di elementi S, da cui si vuole estrarre un elemento a caso. Talvolta però, l'insieme S non è ben definito oppure è difficile calcolare la probabilità dei suoi elementi.

In questi casi si utilizzano di metodi MCMC, che consistono nel definire una catena **irriducibile** e **aperiodica** sull'insieme di dati S che abbia $\underline{\boldsymbol{\pi}}$ come distribuzione stazionaria, in modo tale che eseguendo un numero di stati n abbastanza grande, qualunque sia lo stato iniziale X_0, la probabilità $\mathbb{P}(X_n = i)$ approssima $\pi_i \; \forall i \in S$.

7.2 Generazione di insiemi indipendenti

Definiamo una catena di Markov per generare a caso in modo uniforme un insieme indipendente in un grafo non orientato $G = (V, E)$. Sia S la famiglia di tutti gli insiemi indipendenti in G e sia Z_G la cardinalità di S e consideriamo $\underline{\boldsymbol{\pi}}$ una distribuzione uniforme.

Si sceglie un nodo $v \in V$ a caso in maniera uniforme e se $v \in A$ allora si toglie v da A. Altrimenti se A non possiede nemmeno nodi adiacenti a v, esso viene aggiunto ad A. In caso contrario, infine, non si modifica A. Lo stato A modificato viene chiamato B.

Teorema 7.2.1 (La catena di Markov per generazione di indipendent set è reversibile). La matrice di transizione $\boldsymbol{P}$ definita come:

1. Se i due set A e B hanno più di un elemento di differenza non è possibile transitare in un'iterazione da uno all'altro.
2. Se gli insiemi A e B hanno esattamente un elemento di differenza, la probabilità di transitare da uno all'altro è uniforme $\frac{1}{k}$.
3. Se gli insiemi coincidono, la probabilità di non cambiare insieme è pari a 1 meno la somma delle probabilità di cambiare insieme, cioè l'uniforme moltiplicata per il numero di insiemi a distanza 1 da A.

Con notazione formale dei coefficienti, usando la notazione della sottrazione insiemistica $\#\left\{\frac{A \cup B}{A \cap B}\right\}$:

$$p_{AA'} = \begin{cases} 0 & \#\left\{\frac{A \cup B}{A \cap B}\right\} > 1 \\ \frac{1}{k} & \#\left\{\frac{A \cup B}{A \cap B}\right\} = 1 \\ 1 - \frac{\#\left\{C \in S \mid \#\left\{\frac{A \cup C}{A \cap C}\right\} = 1\right\}}{k} & A = B \end{cases}$$

è **irriducibile** e **aperiodica**. Inoltre, la distribuzione $\underline{\boldsymbol{\pi}}$ è **reversibile** per la catena $\{X_n\}$ associata.

In particolare, gli elementi dell'insieme sono:

La catena di Markov per generazione di indipendent set è reversibile. Possiamo costruire un cammino dallo stato A allo stato B rimuovendo prima i nodi dell'insieme $A \setminus B$, ottenendo lo stato $A \cup B$, e poi aggiungendo i nodi del set $B \setminus A$: per la definizione della catena ognuno di questi passi ha probabilità non nulla e quindi la probabilità dell'intero cammino è positiva.

Per un opportuno $n \in \mathbb{N}$ quindi $\boldsymbol{P}^n$ è positiva ed è pertanto **irriducibile**.

La matrice è inoltre **aperiodica** Poiché esiste sempre una probabilità uniforme $\frac{1}{k}$ di spostarsi da un determinato nodo $\{a\}$ ad un nodo $\{b\}$, compreso $\{a\} \to \{a\}$: il periodo pertanto è unitario.

Per la **reversibilità** della distribuzione $\underline{\boldsymbol{\pi}}$ è sufficiente osservare che $\boldsymbol{P}$ è simmetrica e $\underline{\boldsymbol{\pi}}$ è uniforme:

$$\pi_A p_{AB} = \pi_B p_{BA}$$

□

La catena pertanto definisce un **algoritmo MCMC corretto**.

7.3 Campionatori di Gibbs

Dati due insiemi finiti V, R di cardinalità $k, q > 1$, denotiamo con R^V il set delle funzioni a valori da V a R, di cardinalità q^k: un campionatore di Gibbs estrae secondo una distribuzione fissata $\underline{\boldsymbol{\pi}}$ un elemento da tale insieme.

Viene definito tramite una catena di Markov sull'insieme di stati:

$$S = \left\{A \in R^V : \pi(A) > 0\right\}$$

Il nuovo stato della catena viene ottenuto estraendo a caso in modo uniforme un elemento $v \in V$ e scegliendo un nuovo valore $c \in R$ con probabilità $\underline{\boldsymbol{\pi}}$ condizionata a conservare uguali ad A i valori degli altri elementi di V.

Il nuovo stato B differisce da A al più per il solo valore attribuito all'elemento v estratto, ed esiste sempre una probabilità non nulla che B coincida con A.

Teorema 7.3.1 (La distribuzione di un campionatore di Gibbs è reversibile). La distribuzione $\underline{\boldsymbol{\pi}}$ è **reversibile** per la catena di Markov con matrice di transizione $\boldsymbol{P}$ definita per un campionatore di Gibbs.

Chiamando π_v la probabilità $\pi(C \in R^V : C(u) = A(u)\ \forall u \neq v)$

$$p_{AB} = \begin{cases} 0 & \exists u, v \in V : A(u) \neq B(U) \wedge A(v) \neq B(v) \\ \frac{\pi_B}{k\pi_v} & !\exists v \in V : A(v) \neq B(v) \\ \frac{\pi_A}{k\sum_{v \in V} \pi_v} & A = B \end{cases}$$

7.4 Generazione di colorazioni su grafi

Si tratta di un classico campionatore di Gibbs per la generazione casuale di colorazioni di un grafo: il problema è definito da un grafo non orientato $G = (V, E)$ di k nodi e da un insieme di colori $Q = \{1, 2, \ldots, q\}$.

Definizione 7.4.1 (q-colorazione). Una q-colorazione di un grafo G è una funzione $f : V \to Q$ tale che:

$$f(u) \neq f(v) \quad \forall \{u, v\} \in E$$

Cioè è una funzione che assegna valori diversi ai nodi, per ogni tupla di nodi connessi.

Denotiamo con $Z_{G,q}$ il numero di q-colorazioni di G e con $\underline{\boldsymbol{\pi}}$ rappresentiamo la distribuzione uniforme sulle q-colorazioni, ovvero la funzione $\pi : Q^V \to [0, 1]$ tale che, $\forall f \in Q^V$:

$$\pi(f) = \begin{cases} \frac{1}{Z_{G,q}} & \text{Se } f \text{ è una } q\text{-colorazione di } G \\ 0 & \text{altrimenti} \end{cases}$$

L'obbiettivo è quello di generare una funzione $f \in Q^V$ secondo la distribuzione di probabilità $\underline{\boldsymbol{\pi}}$.

Si definisce quindi una catena di Markov sull'insieme di stati $S = \left\{ f \in Q^V \,\middle|\, f \text{ è } q\text{-colorazione di } G \right\}$ e si procede a simulare la catena per n passi. È necessaria determinare una q-colorazione iniziale che funge da stato iniziale.

Si sceglie uniformemente a caso $v \in V$ secondo la distribuzione uniforme, si determina l'insieme:

$$U_f(v) = \left\{c \in Q : f(w) \neq c \text{ per ogni } w \text{ vicino di } v\right\}$$

e si sceglie a caso un $f(v) = c$ appartenente ad esso secondo la distribuzione uniforme.

La matrice di transizione della catena è definita come:

$$P_{f,g} = \begin{cases} 0 & \text{se } f \text{ e } g \text{ si differenziano per il valore attribuito ad almeno 2 nodi} \\ \frac{1}{k}\frac{1}{\#\{U_f(v)\}} & \text{se } f \text{ e } g \text{ si differenziano per il valore attribuito ad un solo nodo } v \end{cases}$$

7.5 Algoritmo di Metropolis

si tratta di una procedimento per generare a caso un elemento secondo una distribuzione fissata.

Sia S un insieme finito e $\underline{\boldsymbol{\pi}}$ una distribuzione definita su S tale che $\pi(i) > 0 \, \forall \, i \in S$. Definiamo inoltre un grafo non orientato $G = (S, E)$ che gode delle seguenti proprietà:

1. S sia l'insieme dei nodi di G.
2. G sia connesso.
3. Il grado di G sia limitato da una opportuna costante.

Possiamo procedere a definire la matrice di transizione $\boldsymbol{P}$ come:

$$p_{ij} = \begin{cases} 0 & i \neq j \wedge \{i,j\} \notin E \\ \frac{1}{d_i} \min\left\{\frac{\pi_j d_i}{\pi_d i d_j}, 1\right\} & \{i,j\} \in E \\ 1 - \frac{1}{d_i} \sum_{l:\{i,l\}\in E} \min\left\{\frac{\pi_l d_i}{\pi_d i d_l}, 1\right\} & i == j \end{cases}$$

Dove:

1. Se i nodi non hanno un arco non vi è possibilità di transitarvi.
2. Se i nodi hanno un arco tra loro, la probabilità è pari all'uniforme di entrare in i, $\frac{1}{d_i}$ moltiplicata per il rapporto dell'estrazione dell'elemento j e dell'elemento i limitato a 1, cioè $\min\left\{\frac{\pi_j d_i}{\pi_i d_j}, 1\right\}$.
3. Se i nodi coincidono, la probabilità di rimanere nello stesso è 1 meno la somma delle probabilità di andare in uno qualsiasi dei nodi adiacenti.

dove d_i è il grado del generico nodo i.

La catena è **irriducibile** Poiché il grafo associato G è connesso, ed inoltre la distribuzione $\underline{\boldsymbol{\pi}}$ è reversibile per la catena, infatti:

$$\pi_i p_{ij} = \pi_j p_{ji}$$

7.5.1 La procedura

Si inizia da un nodo i e si estrae a caso in modo uniforme un nodo $l \in S$ tale che $\{i,l\} \in E$, quindi si confrontano le probabilità $\pi_l d_i \geqslant \pi_i d_l$: se la disuguaglianza è vera, il prossimo nodo $j = l$, altrimenti si sceglie a caso un nuovo nodo $j \in l, i$ con le seguenti probabilità:

$$\mathbb{P}(l) = \frac{\pi_l d_i}{\pi_i d_l} \qquad \mathbb{P}(i) = 1 - \mathbb{P}(l)$$

8

Analisi della velocità di convergenza

8.1 La problematica

Non tutte le distribuzioni di probabilità sono uguali ed è quindi lecito chiedersi dopo quanti n passi la catena approssima effettivamente la desiderata distribuzione in un modo che la loro differenza sia minore di una quantità fissata.

8.2 Applicazioni della variazione totale

Teorema 8.2.1 (Variazione totale tra distribuzioni). Se $\underline{\boldsymbol{\mu}}$ e $\underline{\boldsymbol{\nu}}$ sono due distribuzioni di probabilità definite sullo stesso insieme finito S allora la massima differenza tra le due distribuzioni valutate su un qualunque sottoinsieme di S è la **variazione totale**:

$$d_{TV}(\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\nu}}) \max_{A \subseteq S} \left| \underline{\boldsymbol{\mu}}(A) - \underline{\boldsymbol{\nu}}(A) \right|$$

Teorema 8.2.2 (Relazione tra variazione totale e probabilità). Se X e Y sono due variabili aleatorie definite sullo stesso spazio di probabilità a valori in un insieme finito S e hanno distribuzione $\underline{\boldsymbol{\mu}}$ e $\underline{\boldsymbol{\nu}}$ rispettivamente, allora:

$$d_{TV}(\underline{\boldsymbol{\mu}}, \underline{\boldsymbol{\nu}}) \leqslant \mathbb{P}(X \neq Y)$$

8.3 Convergenza di matrice stocastica primitiva

Teorema 8.3.1 (Convergenza di matrice stocastica primitiva). Sia $\boldsymbol{P}$ una matrice stocastica primitiva di dimensione $k \times k$, con distribuzione stazionaria $\underline{\boldsymbol{\pi}}$ e sia $t \in \mathbb{N}$ il minimo intero tale che $\boldsymbol{P}^t > 0$. Allora, $\forall \epsilon > 0$ e ogni vettore stocastico $\underline{\boldsymbol{\mu}}$ di dimensione k, si verifica che $d_{TV}(\underline{\boldsymbol{\mu}}\boldsymbol{P}^n, \underline{\boldsymbol{\pi}}) \leqslant \epsilon$ e per tutti gli $n \in \mathbb{N}$ tali che:

$$n \geqslant t\left(1 + \frac{\log\epsilon}{\log\beta(\boldsymbol{P}^t)}\right)$$

Convergenza di matrice stocastica primitiva. Sappiamo che, per una catena definita su un insieme finito di stati S, con matrice di transizione $\boldsymbol{P}$ **primitiva** e distribuzione **stazionaria** $\underline{\boldsymbol{\pi}}$ vale:

$$\lim_{n\to+\infty} \left\| \underline{\boldsymbol{\mu}}\boldsymbol{P}^n - \underline{\boldsymbol{\pi}} \right\|_1 = 0$$

Poiché $\boldsymbol{P}$ è primitiva esiste un intero $t \in \mathbb{N}$, minore del numero di stati, tale che $\boldsymbol{P}^t > 0$ e pertanto il **coefficiente ergodico** $\beta(\boldsymbol{P}^t) < 1$. Per il **lemma di relazione tra potenze di matrici e vettori stocastici**, inoltre, vale che:

$$\left\| \underline{\boldsymbol{\mu}}\boldsymbol{P}^t - \underline{\boldsymbol{\pi}} \right\|_1 \leqslant \beta\left(\boldsymbol{P}^t\right) \left\| \underline{\boldsymbol{\mu}} - \underline{\boldsymbol{\nu}} \right\|_1$$

Quindi, $\forall n \in N : n > t$, ponendo $n = mt + r$ e con $r \in \{0, 1, \ldots, t-1\}$ si ottiene:

$$\begin{aligned} d_{TV}\left(\underline{\boldsymbol{\mu}}\boldsymbol{P}^n, \underline{\boldsymbol{\pi}}\right) &= \frac{1}{2}\left\| \underline{\boldsymbol{\mu}}\boldsymbol{P}^t - \underline{\boldsymbol{\pi}} \right\|_1 \\ &= \left\| \underline{\boldsymbol{\mu}}\boldsymbol{P}^r \cdot \boldsymbol{P}^{tm} - \underline{\boldsymbol{\pi}} \right\|_1 \\ &\leqslant \frac{1}{2}\beta\left(\boldsymbol{P}^t\right)^m \left\| \underline{\boldsymbol{\mu}}\boldsymbol{P}^r - \underline{\boldsymbol{\pi}} \right\|_1 \\ &\leqslant \beta\left(\boldsymbol{P}^t\right)^{\frac{n}{t}-1} \end{aligned}$$

Ovvero, per ogni $n \in N$ tale che:

$$n \geqslant t\left(1 + \frac{\log\epsilon}{\log\beta(\boldsymbol{P}^t)}\right)$$

la distanza totale è minore di un dato ϵ. □

8.4 Mixing time

Definizione 8.4.1 (Mixing time). Per ogni $i \in S$ ed ogni $n \in \mathbb{N}$ sia $\boldsymbol{P}_i^n$ la i-esima riga di $\boldsymbol{P}^n$. Definiamo allora:

$$\Delta_i(n) = d_{TV}\left(\boldsymbol{P}_i^n, \pi\right) \qquad \Delta(n) = \max\{\Delta_i(n) : i \in S\}$$

Inoltre, $\forall \epsilon > 0$ definiamo la funzione **mixing time** della catena come:

$$\tau(\epsilon) = \min\{n \in \mathbb{N} : \Delta(n) \leqslant \epsilon\}$$

Essa rappresenta la miglior funzione che garantisce un'approssimazione alla distribuzione stazionaria con errore al più ϵ.

8.5 Accoppiamento

Definizione 8.5.1 (Accoppiamento). Dato un insieme finito di stati S e una matrice **stocastica** $\boldsymbol{P}$, un **accoppiamento** è una catena di Markov $\{Z_n\}$ sull'insieme degli stati $S \times S$, tale che $Z_n = (X_n, Y_n)$ per ogni $n \in \mathbb{N}$ e $\forall i, j, l \in S$:

$$\begin{aligned} \mathbb{P}\left(X_{n+1} = j \,\middle|\, X_n = i, Y_n = l\right) &= p_{ij} \\ \mathbb{P}\left(Y_{n+1} = j \,\middle|\, X_n = l, Y_n = i\right) &= p_{ij} \end{aligned}$$

Le due variabili possono essere viste come due catene $\{X_n\}, \{Y_n\}$ che si muovono in parallelo ad ogni passo.

8.6 Generatore di indipendent set di dimensione fissata

Consideriamo un grafo non orientato $G = (V, E)$ di n nodi, con grado massimo Δ e sia $k \in \mathbb{N}$ un intero tale che:

$$k \leqslant \frac{n}{3(\Delta + 1)}$$

Denotiamo l'insieme degli indipendent set di G di dimensione k con:

$$S_k = \{ A \subseteq V \mid \#A = k, \forall u, v \in A, \{u, v\} \notin E \}$$

Sia inoltre $Z_k(G)$ la cardinalità di S_k. Vogliamo generare un elemento $A \in S_k$ con probabilità uniforme:

$$\pi(A) = \frac{1}{Z_k G}$$

È sempre possibile costruire un elemento $A \in S_k$ in $\mathcal{O}\left(n^2\right)$ passi.

8.6.1 La procedura

Si sceglie a caso e uniformemente un elemento $v \in A$ e un elemento$w \in V$m quindi se $w \notin A \wedge (A \setminus \{v\}) \cup \{w\} \in S_k$ allora lo stato successivo è definito come $B = (A \setminus \{v\}) \cup \{w\}$.

8.6.2 La catena

Quindi, per ogni $A, B \in S_k$ la probabilità di transizione è data da:

$$P_{AB} = \begin{cases} 0 & \#\left\{\frac{A \cup B}{A \cap B}\right\} > 2 \\ \frac{1}{kn} & \#\left\{\frac{A \cup B}{A \cap B}\right\} = 2 \\ 1 - \frac{\#\left\{C \in S_k \mid \#\left\{\frac{A \cup C}{A \cap C}\right\} = 1\right\}}{k} & A = B \end{cases}$$

La matrice così definita è **primitiva** e la distribuzione uniforme $\underline{\boldsymbol{\pi}}$ coincide con la **distribuzione stazionaria** della catena.

8.7 Velocità di convergenza della colorazione di grafi

Consideriamo una versione ciclica del campionatore di Gibbs per le q-colorazioni, cioè nella quale tutti i nodi vengono ricolorati dopo un numero fissato di passi.

Dato un grafo non orientato $G = (V, E)$ di k nodi $V = \{v_1, v_2, \ldots, v_k\}$ e un insieme di colori $Q = \{1, 2, \ldots, q\}$ e un insieme di colorazioni $Q = \{1, 2, \ldots, q\}$. Sia S l'insieme delle q-colorazioni di G e sia $Z_{G,q}$ il loro numero.

Sia infine d il grado di G, cioè il massimo tra i gradi dei suoi nodi. Supponiamo che il numero dei colori sia abbastanza grande, per esempio $q > d$: in questo modo siamo sicuro che sia facile determinare una q-colorazione di G.

8.7.1 La procedura

La procedura definisce una catena di Markov **non omogenea** $\{X_n\}$. Al primo passo si sceglie un elemento $\{X_0\}$ qualsiasi, quindi si eseguono n passi nella catena. All'i-esimo passo si ricolora il nodo v_j dove $j = 1 + [i-1]_k$, scegliendo il nuovo colore in modo uniforme sull'insieme dei colori compatibili con quelli dei nodi adiacenti a v_j.

8.7.2 Velocità di convergenza

Teorema 8.7.1 (Velocità di convergenza). Dato un insieme di colori $Q = \{1, 2, \ldots, q\}$, sia $G = (V, E)$ un grafo non orientato di k nodi, sia d il grado di G e sia $\{X_n\}$ la catena di Markov descritta sopra sulle q-colorazioni di G. Denotiamo inoltre con $\mu^{(n)}$ la distribuzione di probabilità di X_n e con $\underline{\boldsymbol{\pi}}$ la distribuzione uniforme sull'insieme delle q-colorazioni di G.

Se $q > 2d^2$, allora esiste una costante $C > 0$ dipendente solo da d e q tale che per ogni $\epsilon > 0$ vale:

$$d_{TV}(\mu^{(n)}, \pi) \leqslant \epsilon$$

per tutti gli $n \in \mathbb{N}$ che soddisfano:

$$n > Ck\left(\log k + \log \frac{1}{\epsilon}\right)$$

9 Conteggio

9.1 Conteggio approssimato di colorazioni

9.1.1 Cosa è un RPTAS

Definizione 9.1.1 (RPTAS). Si definisce **RPTAS** uno schema di approssimazione probabilistico pienamente polinomiale per una funzione $F : I \to \mathbb{N}$ è un algoritmo probabilistico che su input (x, ϵ) che soddisfa le seguenti condizioni:

1. Per ogni $\epsilon > 0$ esiste un polinomio $p_\epsilon(y)$ tale che l'algoritmo su input (x, ϵ) richiede un tempo minore o uguale a $p_\epsilon(n)$ dove $n = |x|$.
2. Il valore $R(x, \epsilon)$ restituito dall'algoritmo su input (x, ϵ) verifica la relazione:

$$\mathbb{P}\left((1-\epsilon)\,F(x) \leqslant R(x,\epsilon) \leqslant (1+\epsilon)\,F(x)\right) \geqslant \frac{2}{3}$$

9.1.2 Applicazione

Si considera un $G = (V, E)$ e da un numero q di colori. Denotiamo con k il numero dei nodi del grafo, d il suo grado ed m il numero dei suoi lati. La coppia (k, m) rappresenta la dimensione dell'istanza del problema.

Consideriamo l'insieme dei lati $E = \{e_1, e_2, \ldots, e_m\}$ del grafo e per ogni $j \in \{0, 1, \ldots, m\}$ definiamo:

$$G_j = \left(V, \{e_1, \ldots, e_j\}\right) \quad \ldots \quad G_0 = (V, \emptyset)$$

Denotiamo inoltre Z_j come il numero di q-colorazioni di G_j, in particolare $Z_0 = q^k$, poiché nel grafo G_0 ogni assegnamento di colori ai nodi è una q-colorazione.

Il valore di Z_m che vogliamo calcolare può essere ottenuto mediante:

$$Z_m = Z_0 \cdot \frac{Z_1}{Z_0} \cdot \frac{Z_2}{Z_1} \cdots \frac{Z_m}{Z_{m-1}}$$

Denotiamo ora con x_j e y_j i due nodi del lato e_j, per ciascun $j = \{1, 2, \ldots, m\}$. Allora:

$$Z_j = \#\left\{ f \in Q^V \,\middle|\, f \text{ è } q - \text{colorazione di } G_{j-1} \text{ tale che } f(x_j) \neq f(y_j) \right\}$$

Quindi $\frac{Z_j}{Z_{j-1}}$ è la frequenza di colorazioni di G_{j-1} che sono valide anche per G_j. Ne segue che:

$$\frac{Z_j}{Z_{j-1}} = \mathbb{P}\left(f(x_j) \neq f(y_j)\right)$$

9.1.3 La sequenza

Per ogni grafo G_{j-1} si generano diverse q-colorazioni casuali usando il metodo dei campionatore di Gibbs e si verifica quante di queste attribuiscono colori diversi ai nodi x_j e y_j. Si ottiene così una stima Y_j della probabilità $\frac{Z_j}{Z_{j-1}}$.

Moltiplicando tutti questi valori tra di loro insieme alla quantità $Z_0 = q^k$ si ottiene R, un'approssimazione di Z_m:

$$R = q^k \prod_{j=1}^{m} Y_j = \prod_{j=1}^{m} \frac{H_j}{l}$$

Dove H_j è il numero di q-colorazioni di G_{j-1} che attribuiscono colori diversi ai nodi x_j e y_j

Il procedimento descritto dipende da due parametri, l e n:

l numero di q-colorazioni casuali di G_{j-1}.

n numero di passi principali eseguiti da ciascun campionatore di Gibbs per determinare la q-colorazione casuale corrispondente.

Teorema 9.1.2 (RPTAS per q-colorazioni). Assumendo la notazione illustrata sopra, supponiamo che $d \geqslant 2$ e $q > 2d^2$. Allora esiste un **RPTAS** per il problema di determinare il numero minimo di colorazioni di G.