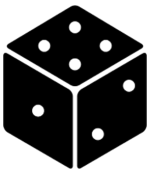
METODI PROBABILISTICI PER L'INFORMATICA

Prof. Massimiliano Goldwurm
6 CFU

Work In Progress

Lecture Notes
Year 2017/2018



Magistrale Informatica
Università di Milano
Italy
10 giugno 2018

Indice

1 Probabilità

1.1 Disuguaglianza di Chernoff

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

Figura 1.1: Disuguaglianza di Chernoff per le binomiali

La **disuguaglianza di Chernoff** utilizza 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

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:

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

Figura 1.2: Disuguaglianza di Chebyshev

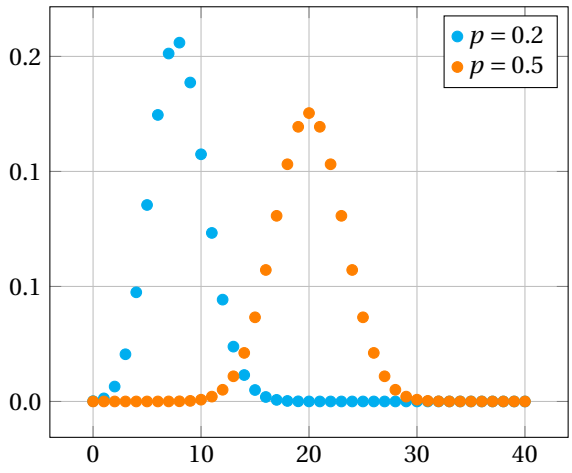


Figura 1.3: Distribuzioni binomiali al variare del parametro p

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

2.1 Classificazione degli algoritmi probabilistici

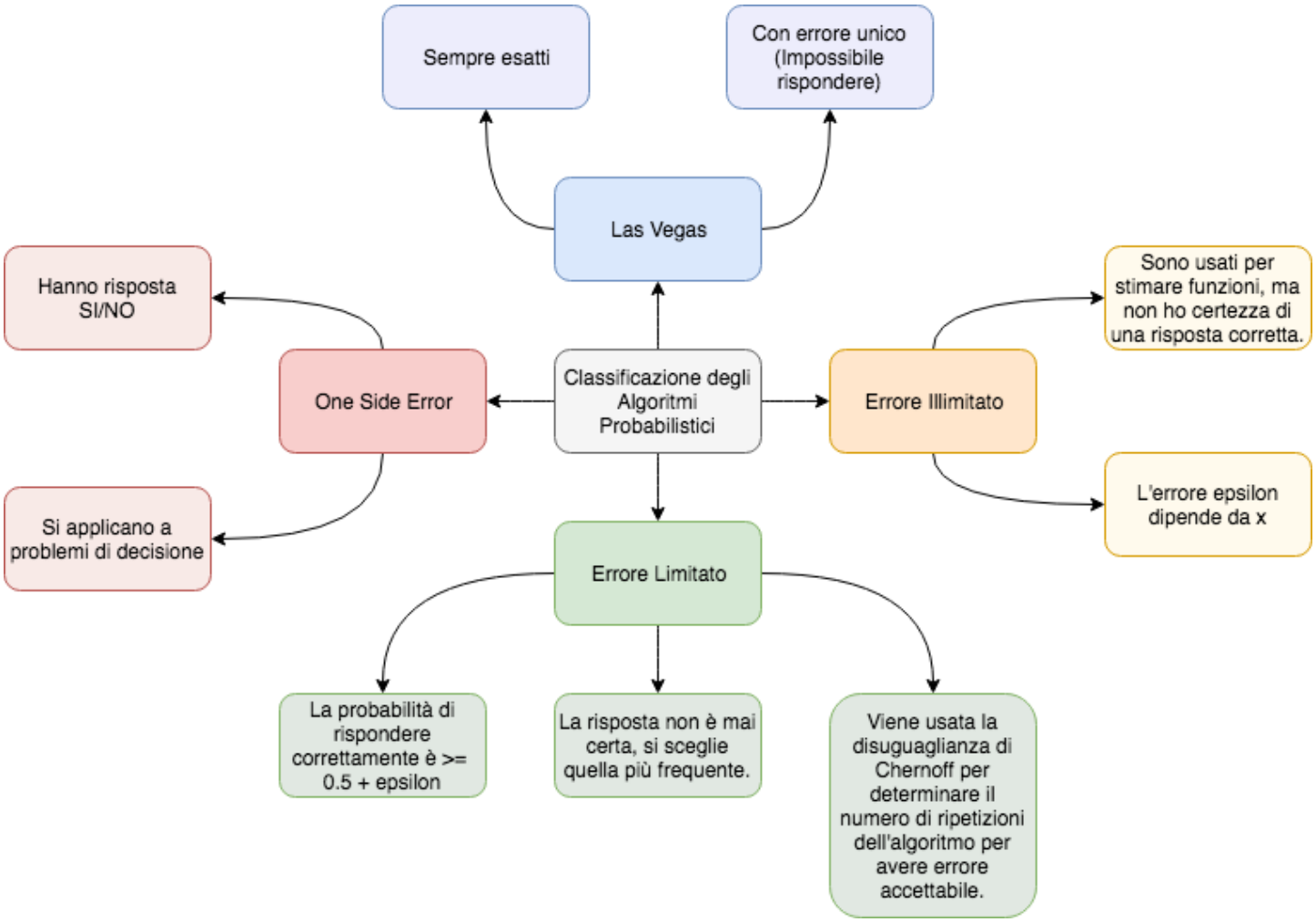


Figura 2.1: Mappa degli algoritmi probabilistici

2.1.1 Las Vegas

Esempio

2.1.2 1-Sided Error

Esempio

2.1.3 2-Sided Error

Esempio

2.1.4 Errore Illimitato

Esempio

2.2 Metodi per l'eliminazione dell'errore

2.3 Metodi per l'eliminazione dell'errore

3

Catene di Markov

3.1 Proprietà di Grafi, matrici e vettori stocastici

3.1.1 Cos'è una matrice stocastica?

3.1.2 Cos'è una matrice primitiva?

3.1.3 Proprietà di autovalori ed autovettori di una matrice stocastica

3.1.4 Teoremi sulla periodicità

3.2 Definizione di una catena di Markov

3.3 Proprietà fondamentali sulle catene

3.4 Stati ricorrenti

3.4.1 Prima definizione

3.4.2 Seconda definizione

3.4.3 Terza definizione

3.4.4 Stati essenziali

3.4.5 Gli stati ricorrenti sono stati essenziali

3.5 Tempi medi di rientro

3.5.1 Caso degli stati ricorrenti (*)

3.6 Definizione di catena ergodica

3.7 Esistenza di distribuzioni stazionarie (*)

3.8 Ergodicità delle catene con matrici di transizione primitive (*)

4

Applicazioni algoritmiche

4.1 Catene reversibili

4.2 Passeggiate a cso su grafi

4.2.1 Algoritmo probabilistico per 2-SODD

4.3 Metodo MCMC (Monte Carlo Markov Chain): rappresentazione di algoritmo e proprietà

4.3.1 Generazione di insiemi indipendenti in grafi (anche di dimensione fissata)

4.4 Algoritmo di Metropolis

4.5 Campionatore di Gibbs

5

Velocità di convergenza degli algoritmi MCMC

5.1 La problematica

5.2 Come stimare la velocità di convergenza

5.3 Stimare il numero di passi

5.4 Metodo generale basato sul coefficiente ergodico

6 Coupling

6.1 Metodo di accoppiamento (Coupling method)

6.1.1 Caso del campionatore di set indipendenti di dimensione fissa

6.1.2 Campionatore di colorazioni

7 Colorazioni di grafi

7.1 Stima del numero di colorazioni di un grafo (algoritmo ed enunciati)