

Teoria probabilităților - Curs 7

Olariu E. Florentin

Aprilie, 2016

Table of contents

- 1 Metoda probabilistică
 - Problema satisfiabilității
 - Aplicații în teoria grafurilor
- 2 Variabile aleatoare continue
 - Variabile aleatoare continue
 - Repartiții continue remarcabile
- 3 Bibliography

Metoda probabilistică

- Metoda probabilistică este o tehnică folosită pentru demonstrarea existenței unor obiecte matematice (combinatorii) cu anumite proprietăți.
- Ideea de bază a acestei metode constă în a demonstra că probabilitatea existenței unui obiect cu proprietățile cerute este strict pozitivă, ceea ce înseamnă că un asemenea obiect există.
- Pentru aceasta, construim un spațiu de probabilitate peste mulțimea obiectelor implicate și arătăm ca probabilitatea obiectului respectiv este nenulă.

Metoda probabilistică

Theorem 0.1

(Principiul mediei) Dacă X este o variabilă aleatoare discretă cu $M[X] \geq \alpha$, atunci $P\{X \geq \alpha\} > 0$.

Primele exemple ale aplicării metodei probabiliste sunt legate de chestiunea satisfiabilității.

Proposition 0.1

Fie $\mathcal{F} = \{C_1, C_2, \dots, C_m\}$ o familie de m clauze. Există o asignare a valorilor de adevăr a variabilelor booleene implicate, astfel ca numărul de clauze satisfăcute să fie cel puțin

$$\sum_{i=1}^m \left(1 - 2^{-|C_i|}\right) \geq m \left(1 - 2^{-h}\right),$$

unde $h = \min_i |C_i|$.

Problema satisfiabilității

proof: Imaginăm următorul experiment aleator abstract: fiecărei variabile booleene x îi asignăm independent valoarea 1 (*adevărat*) sau 0 (*fals*) cu aceeași probabilitate 0.5. Definim variabilele aleatoare

$$X_i = \begin{cases} 1, & \text{dacă } C_i \text{ este adevărată} \\ 0, & \text{dacă } C_i \text{ este falsă} \end{cases}$$

Probabilitatea ca să fie adevărată clauza C_i este egală cu probabilitatea ca măcar unul dintre cei l_i literali ai ei să fie adevărat, adică $(1 - 2^{-|C_i|})$.

Numărul de clauze satisfăcute este egal cu $X = \sum_{i=1}^n X_i$ și, atunci

$$M[X] = \sum_{i=1}^n M[X_i] = \sum_{i=1}^n (1 - 2^{-|C_i|}) \geq m (1 - 2^{-h}).$$

Concluzia urmează conform Teoremei 0.1. ■

Problema satisfiabilității

Corollary 1.1

Orice instanță a problemei k -SAT cu un număr de clauze mai mic strict decât 2^k este satisfiabilă. (O instanță a problemei k -SAT are în fiecare clauză exact k literali.)

proof: Reluând argumentul de mai sus, cu $|C_i| = k, \forall i = \overline{1, m}$:

$$M[X] = \sum_{i=1}^m M[X_i] = \sum_{i=1}^m \left(1 - 2^{-|C_i|}\right) = m \left(1 - 2^{-k}\right) > m - 1,$$

obținem de aici că $P\{X \geq m\} = P\{X > m - 1\} > 0$ - deci există o asignare a valorilor de adevăr care să satisfacă toate cele m clauze. ■

Problema satisfacibilității

- Marginea indicată în acest rezultat este cea mai bună posibilă deoarece putem defini o instanță a problemei k -SAT cu 2^k clauze care să fie nesatisfiabilă: de exemplu familia tuturor clauzelor având k literali definiți peste o mulțime de k variabile booleene.
- În mod similar se poate demonstra

Proposition 1.1

O familie de clauze $\mathcal{F} = \{C_1, \dots, C_m\}$ este satisfiabilă dacă

$$\sum_{i=1}^m 2^{-|C_i|} < 1.$$

Aplicații în teoria grafurilor

- Următoarele două rezultate sunt aplicații ale metodei probabiliste în teoria grafurilor. Introducem aceste rezultate cu două definiții mai degrabă informale.
- Fie $G = (V, E)$ un graf, o mulțime de noduri $S \subseteq V$ se numește *stabilă* dacă oricare două noduri din S nu sunt adiacente: $uv \notin E$, $\forall u, v \in S$. Dacă (A, B) este o partiție a nodurilor lui G , *tăietura* generată de această partiție este mulțimea de muchii în cross între A și B :
$$E(A, B) = \{uv \in E : u \in A, v \in B\}.$$
- Este de menționat că atât problema determinării unei mulțimi stabile de cardinal maxim cât și cea a determinării unei tăieturi cu număr maxim de muchii sunt probleme NP-hard.

Aplicații în teoria grafurilor

Proposition 2.1

Fie $G = (V, E)$ un graf cu n noduri și m muchii. Există o bipartiție (A, B) a lui G astfel încât

$$|E(A, B)| \geq \frac{m}{2}.$$

dem.: Considerăm următorul algoritm aleator care construiește o bipartiție (A, B) a lui V :

*$A \leftarrow \emptyset$;
for $(i = 1, n)$
 cu probabilitate $\frac{1}{2}$ adaugă v la A ;
 $B \leftarrow V \setminus A$;*

Aplicații în teoria grafurilor

Pentru fiecare muchie $uv \in E$ definim o variabilă Bernoulli

$$X_{uv} = \begin{cases} 1, & \text{dacă } uv \in E(A, B), \\ 0, & \text{dacă } uv \notin E(A, B) \end{cases}$$

Evident că

$$\begin{aligned} P\{X_{uv} = 1\} &= P\{(u \in A \text{ și } v \in B) \text{ sau } (u \in B \text{ și } v \in A)\} = \\ &= P\{u \in A\} \cdot P\{v \in B\} + P\{u \in B\} \cdot P\{v \in A\} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}. \end{aligned}$$

În plus, $X = \sum_{uv \in E} X_{uv}$ este o variabilă aleatoare care numără câte muchii conține tăietura generată de bipartiția (A, B) și, atunci

$$M[X] = \sum_{uv \in E} M[X_{uv}] = \frac{m}{2} \Rightarrow P\left\{X \geq \frac{m}{2}\right\} > 0.$$

Astfel, va trebuie să existe o partiție cu proprietățile din enunț. ♣

Aplicații în teoria grafurilor

Proposition 2.2

Fie $G = (V, E)$ un graf cu n noduri și $m \geq n/2$ muchii. Atunci G are o mulțime stabilă (sau independentă) de noduri de cardinal cel puțin n^2/m .

dem.: Fie $d = \frac{2m}{n}$ gradul mediu în G ($\sum_{v \in V} d_G(v) = 2m$); deoarece

$m \geq n/2$, avem $(1 - d^{-1}) \geq 0$. Construim o mulțime stabilă folosind următorul algoritm aleator ($V = \{v_1, v_2, \dots, v_n\}$):

- for ($i = \overline{1, n}$)
 cu probabilitate $(1 - d^{-1})$ șterge nodul v_i ;
 // se elimină și toate muchiile incidente cu v .
- for ($e = uv \in E$)
 șterge muchia e și nodul u sau v ;

Aplicații în teoria grafurilor

După execuția acestui algoritm, nodurile rămase formează o mulțime independentă. După pasul 1 un nod rămâne în graf cu probabilitate d^{-1} ; fie X_i o variabilă egală cu 1 dacă nodul v_i a rămas în graf și 0 altfel. Notăm $X = \sum_{i=1} X_i$ numărul de noduri rămase după pasul 1.

Deoarece fiecare X_i este o variabilă Bernoulli cu media d^{-1} , obținem $M[X] = nd^{-1}$.

Fie acum Y_{uv} , $uv \in E$, un alt set de variabile Bernoulli: $Y_{uv} = 1$ dacă muchia uv rămâne în graf după pasul 1 și 0 altfel. $Y = \sum_{uv \in E} Y_{uv}$ este numărul de muchii rămase în graf după primul pas.

Evident că $P\{Y_{uv} = 1\} = d^{-2}$ (probabilitatea ca nici u , nici v să nu fie șterse), deci $M[Y_{uv}] = d^{-2}$ și $M[Y] = md^{-2} = \frac{n}{2d}$.

Aplicații în teoria grafurilor

Fie Z numărul de noduri rămase în graf după pasul 2. Deoarece $Z \geq X - Y$ (mai sunt șterse încă cel mult Y noduri):

$$M[Z] \geq M[X] - M[Y] = \frac{n}{d} - \frac{n}{2d} = \frac{n}{d} = \frac{n^2}{2m}. \clubsuit$$

- Următorul rezultat folosește noțiunea de *turneu* (un digraf $D = (V, A)$ cu proprietatea că între orice două noduri există exact un singur arc: $\forall u, v \in V, |\{\vec{uv}, \vec{vu}\} \cap A| = 1$).
- Denumirea vine din aceea că nodurile pot fi asimilate unor jucători și fiecare pereche de jucători se confruntă o singură dată: $\vec{uv} \in A$ numai dacă u îl bate pe v .
- Un turneu D are proprietatea P_k dacă pentru orice mulțime de k jucători există un alt jucător care îi bate pe toți ($k < |V|$).
- Pentru un k dat, există un turneu cu proprietatea S_k ? Metoda oferă și o idee despre câte noduri trebuie să aibă turneul.

Aplicații în teoria grafurilor

Proposition 2.3

Fie $k \in \mathbb{N}^*$, dacă $\binom{n}{k} < (1 - 2^{-k})^{k-n} < 1$, atunci există un turneu cu n noduri care să aibă proprietatea P_k .

dem.: Construim mai întâi un turneu aleator astfel:

$A \leftarrow \emptyset$;

for($\{u, v\} \subseteq V$)

cu probabilitate $\frac{1}{2}$ adaugă uv sau $\bar{u}\bar{v}$ la A ;

Fie $M \subseteq V$, $|M| = k$; probabilitatea ca un nod $v \notin M$ să domine toate nodurile din M este 2^{-k} , deci probabilitatea ca să nu le domine pe toate cele din M este $(1 - 2^{-k})$. Evenimentele că două noduri distincte u , respectiv $v \neq u$ nu domină nodurile din M sunt independente, astfel probabilitatea ca mulțimea M să nu fie dominată de nici un nod din afara ei este $(1 - 2^{-k})^{n-k}$.

Aplicații în teoria grafurilor

Acum putem estima probabilitatea ca nici o mulțime S , de cardinal k , să nu fie dominată:

$$\leq \binom{n}{k} (1 - 2^{-k})^{n-k} < 1,$$

deci probabilitatea ca măcar una dintre mulțimile de cardinal k să fie dominate este nenulă.

Fie acum $f(k)$ numărul minim de noduri ale unui turneu care are proprietatea P_k . Se poate arăta ([Alon08]) că $f(k) = \mathcal{O}(k^2 2^k)$. ♣

Evenimente aleatoare

- În cazul în care $|\Omega| \geq |\mathbb{R}|$ (i.e., cardinalul lui Ω este cel puțin continuu), evenimentele aleatoare se definesc diferit față de cazul discret.
- Diferența constă în aceea că nu orice submulțime $A \subseteq \Omega$ este în mod necesar eveniment aleator.
- Familia evenimentelor aleatoare este o σ -algebră $\mathcal{A} \subseteq \mathbb{P}(\Omega)$:
 - $\emptyset, \Omega \in \mathcal{A}$;
 - dacă $A_1, A_2 \in \mathcal{A}$, atunci $A_1 \cap A_2 \in \mathcal{A}$;
 - dacă $(A_n)_{n \geq 1} \subseteq \mathcal{A}$, atunci $\bigcup_{n \geq 1} A_n \in \mathcal{A}$.
- Iar funcția de probabilitate este definită numai pe submulțimile din \mathcal{A} (cu axiomele cunoscute):

$$P : \mathcal{A} \rightarrow [0, 1].$$

Variabile aleatoare continue

- O funcție $X : \Omega \rightarrow \mathbb{R}$ este numită **variabilă aleatoare** dacă pentru orice J interval din \mathbb{R} , $X^{-1}(J) \in \mathcal{A}$.
- O variabilă aleatoare $X : \Omega \rightarrow \mathbb{R}$ se numește **continuă** dacă are funcția de repartiție continuă.
- Câteodată, această definiție se referă la toate cazurile când $X(\Omega)$ este de cardinal continuu.
- Distribuția (repartiția) unei astfel de variabile poate fi dată prin **funcția de repartiție**:

$$F : \mathbb{R} \rightarrow [0, 1], F(a) = P(X \leq a),$$
- sau prin **funcția de densitate** (de masă), $f : \mathbb{R} \rightarrow [0, +\infty)$, astfel încât funcția de repartiție F poate fi descrisă astfel:

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(t) dt.$$

Variabile aleatoare continue

- Cu ajutorul funcției de densitate putem calcula (dacă integralele corespunzătoare există)

Media:
$$M(X) = \int_{-\infty}^{+\infty} tf(t) dt.$$

Dispersia:
$$D^2(X) = \int_{-\infty}^{+\infty} [t - M(X)]^2 f(t) dt.$$

- Probabilitățile asociate unei variabile aleatoare continue se calculează astfel:

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(t) dt$$

Variabile aleatoare continue

- **Observație:** dacă F este continuă, $P(X = a) = F(a) - F(a) = 0$ și $P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.
- Pentru o variabilă aleatoare $X: \Omega \rightarrow \mathbb{R}$, dată, operația de standardizare constă în următoarea transformare a variabilei X :

$$Y = \frac{X - M[X]}{D[X]}.$$

- Noua variabilă este "standard" pentru că are

$$M[Y] = 0 \text{ și } D^2[Y] = 1.$$

Distribuția (legea) normală.

- Este o distribuție notată $N(\mu, \sigma^2)$ cu funcția de densitate

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

- Dacă $X : N(\mu, \sigma^2)$, atunci

$$M[X] = \mu \text{ și } D^2[X] = \sigma^2.$$

Legea normală standard.

- **Distribuția $N(0, 1)$ se numește normală standard. Valorile unei variabile distribuite normal au următoarea împrăștiere:**

%68 se găsesc la cel mult o deviație standard față de medie;

%95 se găsesc la cel mult două deviații standard față de medie;

%99.7 se găsesc la cel mult trei deviații standard față de medie;

Distribuția Student (sau t).

- Este o distribuție notată $t(r)$ cu funcția de densitate

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

unde $\Gamma(y) = \int_0^{\infty} x^{y-1} e^{-x} dx$.

- Pentru o variabilă aleatoare X distribuită Student cu $r > 2$ grade de libertate avem

$$M[X] = 0 \text{ și } D^2[X] = \frac{r}{r-2}.$$

Distribuția Fisher-Snedecor (sau F).

- Este o distribuție notată $F(r_1, r_2)$ cu funcția de densitate

$$f(x) = \frac{\sqrt{\frac{(r_1 x)^{r_1} r_2^{r_2}}{(r_1 x + r_2)^{r_1 + r_2}}}}{xB\left(\frac{r_1}{2}, \frac{r_2}{2}\right)},$$

unde $B(x, y) = \int_0^1 u^{x-1}(1-u)^{y-1} du.$

- Pentru o variabilă aleatoare X distribuită $F(r_1, r_2)$ avem

$$M[X] = \frac{r_2}{r_2 - 2}, r_2 > 2 \text{ și } D^2[X] = \frac{2r_2^2(r_1 + r_2 - 2)}{r_1(r_2 - 2)^2(r_2 - 4)}, r_2 > 4.$$

Distribuția χ^2 .

- Este o distribuție notată $\chi^2(n)$ ($n \in \mathbb{N}^* \setminus \{1\}$) cu funcția de densitate

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right) \sigma^n} x^{n/2-1} e^{-x/2\sigma^2}, & x > 0 \\ 0, & x \leq 0 \end{cases},$$

unde $\Gamma(y) = \int_0^{\infty} x^{y-1} e^{-x} dx$.

- Pentru o variabilă aleatoare X distribuită $\chi^2(n)$ avem

$$M[X] = n\sigma^2 \text{ și } D^2[X] = 2n\sigma^4$$

Despre examenul parțial din săptămâna a 8-a.

- 1 Examenul parțial durează aproximativ o oră și constă din exerciții de tipul celor rezolvate la seminar și din rezultatele teoretice demonstrate la curs (cele scrise cu negru).
- 2 La examen puteți folosi cursurile disponibile pe pagina <http://profs.info.uaic.ro/~olariu/>, în format electronic sau imprimat. Acestea sunt singurele materiale care sunt permise în timpul examenului.
- 3 În lipsa unui motiv foarte întemeiat, fiecare student este obligat să vină la examen cu semianul său.

Bibliography



Alon, N., J. H. Spencer, *The probabilistic method*, Wiley, 2008.



Bertsekas, D. P., J. N. Tsitsiklis, *Introduction to Probability*, Athena Scietific, 2002.



Lipschutz, S., *Theory and Problems of Probability*, Schaum's Outline Series, McGraw-Hill, 1965.



Motwani R., P. Raghavan *Randomized Algorithms*, Cambridge University Press, 1995.



Ross, S. M., *A First Course in Probability*, Prentice Hall, 5th edition, 1998.



Spencer, J. H., *Ten lectures on the probabilistic method*, SIAM, 1994.



Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, 1996.