Teoria probabilităţilor

Olariu E. Florentin

Aprilie, 2014

Table of contents

- Metoda probabilistică
  - Problema satisfiabilităţii
  - Aplicaţii în teoria grafurilor
- Variabile aleatoare continue
  - Variabile aleatoare continue
  - Repartiţii continue remarcabile
- Bibliography

Metoda probabilistică

- Metoda probabilistica este o tehnica folosita pentru demonstrarea existenţei unor obiecte matematice (combinatorii) cu anumite proprietăţi.
- Ideea de bază a acestei metode constă în a demonstra că probabilitatea unui obiect cu proprietăţile cerute este strict pozitivă ceea ce înseamnă că un asemenea obiect există.
- Pentru aceasta construim un spaţiu de probabilitate peste mulţimea obiectelor implicate şi arătăm ca probabilitatea obiectului respectiv este nenula.

Metoda probabilistică

Teorema 1

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

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

Propoziţia 1

Fie $\mathcal{F} = \{C_1, C_2, \ldots, C_m\}$ 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 puţin

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

unde $l = \min\limits_{1 \leqslant i \leqslant m} |C_i|$.

Problema satisfiabilităţii

dem.: Imaginăm următorul experiment aleator abstract: fiecărei variabile booleene x îi asignăm independent valoarea 1 (*adevărat*) sau 0 (*fals*) cu aceeaşi 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 clauza C_i să fie adevărată este egală cu probabilitatea ca cel puţin unul dintre cei l_i literali ai ei să fie adevărat, adică $\left(1-2^{-|C_i|}\right)$. Numărul de clauze satisfăcute este egal cu $X=\sum_{i=1}^{n} X_i$ şi

$$M[X] = \sum_{i=1} M[X_i] = \sum_{i=1} \left(1-2^{-|C_i|}\right) \geqslant m\left(1-2^{-l}\right).$$

Concluzia urmează conform Teoremei 1. ∎

Problema satisfiabilităţii

Corolar 1

Orice instanţă a problemei k-SAT cu mai puţin de k clauze este satisfiabilă.

dem.: Reluând argumentul de mai sus, cu $|C_i| = k$, $\forall i$:

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

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

■

Problema satisfiabilităţii

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

Propoziţia 2

O familie de clauze $\mathcal{F}$ este satisfiabilă dacă

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

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

Propoziția 3

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

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

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

$A \leftarrow \varnothing$;

for $(i = \overline{1, n})$

  cu probabilitate $\frac{1}{2}$ **adaugă** v **la** A;

$B \leftarrow V \setminus A$;

Aplicaţii 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

$$P\{X_{uv} = 1\} = P\{(u \in A \text{ şi } v \in B) \text{ sau } (u \in B \text{ şi } v \in A)\} =$$

$$= P\{u \in A\} \cdot P\{v \in B\} + P\{u \in B\} \cdot P\{v \in A\} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}.$$

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

$$M[X] = \sum_{uv \in E} M\left[X_{uv}\right] = \frac{m}{2} \Rightarrow P\left\{X \geqslant \frac{1}{2}\right\} > 0.$$

Astfel, va trebuie să existe o partiţie cu proprietăţile din enunţ.

Aplicaţii în teoria grafurilor

Propoziţia 4

Fie $G = (V, E)$ un graf cu n noduri şi $m \geqslant \frac{n}{2}$ muchii. Atunci G are o mulţime stabilă (sau independentă) de noduri de cardinal cel puţin $\frac{n^2}{m}$.

dem.: Fie $d = \frac{2m}{n}$ gradul mediu în G ($\sum_{v \in V} d_G(v) = 2m$); deoarece $m \geqslant \frac{n}{2}$ avem $(1 - d^{-1}) \geqslant 0$. Construim o mulţime stabilă folosind următorul algoritm aleator ($V = \{v_1, v_2, \ldots, v_n\}$):

1. **for** $(i = \overline{1, n})$

   cu probabilitate $\left(1 - d^{-1}\right)$ **şterge nodul** v_i ; // se elimină şi toate muchiile incidente cu v.

2. **for** $(e = uv \in E)$

   şterge muchia e **şi nodul** u **sau** v;

Aplicaţii în teoria grafurilor

După execuţia acestui algoritm, nodurile rămase formează o mulţime 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 şi 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}, obţinem $M[X] = nd^{-1}$.
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 şi 0 altfel.
$Y = \sum_{uv \in E} Y_{uv}$ este numărul de muchii rămse în graf după primul pasul. Evident că $P\{Y_{uv} = 1\} = d^{-2}$ (probabilitatea ca nici u, nici v să nu fie şterse), deci $M[Y_{uv}] = d^{-2}$ şi $M[Y] = md^{-2} = \dfrac{n}{2d}$.

Aplicații în teoria grafurilor

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

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

- 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.

Aplicaţii în teoria grafurilor

Propoziţia 5

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** $\vec{uv}$ **sau** $\vec{uv}$ **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 mulţimea M să nu fie dominată de nici un nod din afara ei este $(1-2^{-k})^{n-k}$

Aplicaţii în teoria grafurilor

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

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

deci probabilitatea ca măcar una dintre mulţimile 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 $|\Omega| \geqslant |\mathbb{R}|$ (i.e., cardinalul lui Ω este cel puţin continuu), evenimentele aleatoare se definesc diferit faţă de cazul discret.
- Diferenţa constă în aceea că nu orice submulţime $A \subseteq \Omega$ este în mod necesar eveniment aleator.
- Familia evenimentelor aleatoare este o σ-algebră $\mathcal{A} \subseteq \mathcal{P}(\Omega)$:
  - $\varnothing, \Omega \in \mathcal{A}$;
  - dacă $A_1, A_2 \in \mathcal{A}$, atunci $A_1 \cap A_2 \in \mathcal{A}$;
  - dacă $(A_n)_{n \geqslant 1} \subseteq \mathcal{A}$, atunci $\bigcup_{n \geqslant 1} A_n \in \mathcal{A}$.
- Iar funcţia de probabilitate este definită numai pe submulţimile din $\mathcal{A}$ (cu axiomele cunoscute):

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

Variabile aleatoare continue

- O funcţie $X : \Omega \to \mathbb{R}$ este numită **variabilă aleatoare** dacă

$$\text{pentru orice } J \text{ interval din } \overline{\mathbb{R}}, X^{-1}(J) \in \mathcal{A}.$$

- O variabilă aleatoare $X : \Omega \to \mathbb{R}$ se numeşte **continuă** dacă are funcţia de repartiţie continuă.
- Câteodată această definiţie se referă la toate cazurile când $X(\Omega)$ este de cardinal continuu.
- Distribuţia (repartiţia) unei astfel de variabile poate fi dată prin **funcţia de repartiţie**:

$$F : \mathbb{R} \to [0,1], F(a) = P(X \leqslant a),$$

- sau prin **funcţia de densitate (de masă)**, $f : \mathbb{R} \to [0, +\infty)$, astfel încât funcţia de repartiţie F poate fi descrisă astfel:

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

Variabile aleatoare continue

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

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

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

- Probabilităţile asociate unei variabile aleatoare continue se calculează astfel:

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

Variabile aleatoare continue

- **Observaţie:** dacă F este continuă, $P(X = a) = F(a) = 0$ şi $P(a \leqslant X < b) = P(a < X \leqslant b) = P(a < X < b)$.
- Pentru o variabilă aleatoare dată $X : \Omega \to \mathbb{R}$ operaţia de **standardizare** constă în următoarea transformare a variabilei X:

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

- Noua variabilă este "standard" pentru ca are

$$M[Y] = 0 \text{ şi } D^2[Y] = 1.$$

Distribuţia (legea) normală.

- Este o distribuţie notată $N(\mu, \sigma^2)$ cu funcţia 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

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

Legea normală standard.

- **Distribuţia** $N(0,1)$ se numeşte **normală standard**. Valorile unei variabile distribuite normal au următoarea împrăştiere:

  %68 se găsesc la cel mult o deviaţie standard faţă de medie;
  %95 se găsesc la cel mult două deviaţii standard faţă de medie;
  %99.7 se găsesc la cel mult trei deviaţii standard faţă de medie;

Distribuţia Student (sau t).

- Este o distribuţie notată $t(r)$ cu funcţia de densitate

$$f(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}},$$

unde $\Gamma(y) = \int\limits_0^{+\infty} x^{y-1} e^{-x}\, dx$.

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

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

Distribuţia Fisher-Snedecor (sau F).

- Este o distribuţie notată $F(r_1, r_2)$ cu funcţia de densitate

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

unde $B(x, y) = \int\limits_0^1 u^{x-1}(1-u)^{y-1}\, du.$

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

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

Distribuția χ^2.

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

$$f(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-n/2},$$

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

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

$$M[X] = n \text{ și } D^2[X] = 2n$$

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*, Scahaum's Outline Series, McGraw-Hill, 1965.
- Motwani R., P. Raghavan *Randomized Algorithms*, Cambrifge 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.