Probabilistiska metoder

Oberoende mängder och klickar

Låt G vara en graf med hörnen V och kanterna E. En delmängd W av hörnen säges vara *oberoende* om den inte har några kanter, och den bildar en *klick* om det finns kanter mellan samtliga hörn i W.

Exempel. Antag att ett antal personer antingen känner varandra eller är främlingar. I en grupp om 6 personer finns det då *antingen* tre personer som alla känner varandra inbördes *eller* tre personer som alla är främlingar till varandra.

Bevis. Låt gruppen bestå av A, B, C, D, E och F. A har då antingen minst tre bekanta eller tre främlingar. Antag att A har tre bekanta, säg B, C och D. Antingen finns det då två av dem, säg B och C, som känner varandra, och då bildar A, B och C en klick. I annat fall är B, C och D en oberoende mängd. Motsvarande händer, om A har minst tre främlingar.

Vi undrar hur många hörn en grupp måste ha för att det skall finnas en stor oberoende mängd eller en stor klick.

Sats. I en grupp med n hörn finns det antingen en oberoende mängd av storlek s eller en klick av storlek t under förutsättning att

$$n \geq 2^{s+t} - 1$$

Bevis. Beviset är induktion över $s + t$. Om det inte finns en oberoende mängd av ordningen 2, så finns det kanter mellan samtliga hörn i grafen, och om det inte finns en klick av ordningen 2, så finns det inga kanter i grafen. Satsen är alltså sann för s =1, 2 och alla t, samt för t = 1, 2 och alla s. Dessutom är den enligt ovan sann för $s = t = 3$, eftersom $2^6 - 1 > 6$.

Antag nu att satsen är sann för alla s och t med $s + t = k - 1$. Tag nu en mängd med $2^k - 1$ element och välj ett godtyckligt s; $1 < s < k - 1$ och sätt $t = k - s$. Fixera ett element A i denna mängd. A har $2^k - 2$ grannar. Antingen har A en kant till minst hälften av dessa grannar, eller också har A ingen kant till minst hälften av dessa grannar. Enligt induktionsantagandet har vi bland grannarna i det första fallet antingen en oberoende mängd av ordningen $s - 1$ eller en klick av ordningen t. Tillsammans med A har vi därför antingen en oberoende mängd av ordningen s eller en klick av ordningen t. I det andra fallet vet vi att grannarna innehåller antingen en oberoende mängd av ordningen s eller en klick av ordningen $t - 1$. Även i detta fall innehåller alltså hela mängden antingen en oberoende mängd av ordningen s eller en klick av ordningen t. Satsen gäller alltså för alla k. ♦

Vi har alltså visat att för en graf med n hörn kan satsen gälla för

$$s = t \approx \frac{\log n}{2}$$

***t* = 2 log *n* + 3**

Härnäst skall vi visa att $s = t \approx 2 \cdot \log n$ inte är möjligt.

Det är svårt att explicit konstruera en mängd som motsäger satsen. I stället arbetar vi med slumpvisa mängder och visar att det existerar en graf utan någon klick av storlek *t* och utan någon oberoende mängd av storlek *t*.

Vi skapar en graf där varje par av hörn har en kant med sannolikheten ½. Definiera för varje delmängd α av storlek *t* den stokastiska variabeln

$X_\alpha = 1$ om α är en klick av storlek t
$X_\alpha = 0$ annars.

och låt $X = \sum X_\alpha$ vara antalet klickar av storlek *t*. Det finns $\binom{n}{t}$ delmängder α och $\binom{t}{2}$ kanter i varje delmängd. Sannolikheten för att α är en klick är $2^{-\binom{t}{2}}$ så

$$E[X] = E\left[\sum X_\alpha\right] = \sum E[X_\alpha] = \sum_{\varepsilon} 2^{-\binom{t}{2}} = \binom{n}{t} \cdot 2^{-\binom{t}{2}}$$

Vi sätter $t = 2\log n + 3$ och uppskattar:

$$E[X] < n^t \cdot 2^{-\binom{t}{2}} = \left(\frac{n}{2^{(t-1)/2}}\right)^t = \frac{1}{2}$$

På samma sätt är

E[antal oberoende mängder av storlek *t*] < ½.

Sätt nu

p_i = Pr(exakt *i* stycken klickar)
q_i = Pr(exakt *i* stycken oberoende mängder)

Då är

$$\sum_{i>0} p_i \le \sum_{i>0} i \cdot p_i = E[klickar] < \frac{1}{2}$$

$$\sum_{i>0} q_i \le \sum_{i>0} i \cdot q_i = E[ober] < \frac{1}{2}$$

och

$$\Pr(klick \vee ober) \le \Pr(klick) + \Pr(ober) = \sum_{i>0} p_i + \sum_{i>0} q_i < 1$$

så sannolikheten för ingen klick eller ingen oberoende mängd av storlek t är positiv. Men detta innebär precis att det finns en graf utan klick och utan oberoende mängd av ordningen t, då $t = 2 \log n + 3$. ♦

Anmärkning. Uppskattningen av $E[X]$ var mycket grov. Vi tappade en faktor $t!$, så sannolikheten för klick eller oberoende mängd var i själva verket

$$< O(1/t!) = O(n^{c - \log \log n})$$

Det är alltså ovanligt med så stora klickar och oberoende mängder.

t* = log *n

Vad händer för $t = \log n$? Vårt mål är att visa att $E[X]$ och $V[X]$ är stora men att p_0 är litet.

$$E[X] = \binom{n}{t} \cdot 2^{-\binom{t}{2}} = \frac{n^t}{t!} \cdot 2^{-\binom{t}{2}} \cdot \prod_{i=1}^{t-1} \left(1 - \frac{i}{n}\right)$$

Produkten på slutet uppskattas genom Taylorutveckling av logaritm- och exponentialfunktionerna så att vi får:

$$E[X] = \frac{n^t}{t!} \cdot 2^{-\binom{t}{2}} \cdot \left(1 - \frac{t \cdot (t-1)}{2n} + O\left(t^4 / n^2\right)\right)$$

Med Stirlings formel kan vi skriva om resten till

$$E[X] \approx \frac{n^t}{t!} \cdot 2^{-\binom{t}{2}} = \frac{1}{\sqrt{2\pi t}} \cdot \left(\frac{ne}{t \cdot 2^{(t-1/2)}}\right)^t$$

Sätter vi in $t = \log n$ får vi

$$E[X] \approx \sqrt{\frac{n}{2\pi \log n}} \cdot \left(\frac{ne}{\log n \cdot \sqrt{n}}\right)^{\log n} = \sqrt{\frac{n}{2\pi \log n}} \cdot \left(\frac{e\sqrt{n}}{\log n}\right)^{\log n}$$

och detta växer uppenbarligen mycket fort.

Vi skall beräkna variansen $V[X] = \sigma^2 = E[X^2] - E[X]^2$ och utnyttja Tjebyshovs olikhet

$$\Pr[|X - m| \geq k\sigma] \leq \frac{1}{k^2}$$

för att få en uppskattning av p_0.

$$E[X^2] = E\left[\left(\sum_{\alpha} X_{\alpha}\right)^2\right] = E\left[\sum_{\alpha_1,\alpha_2} X_{\alpha_1} \cdot X_{\alpha_2}\right] = \sum_{\alpha_1,\alpha_2} E[X_{\alpha_1} \cdot X_{\alpha_2}]$$

där summan tas över alla par av delmängder av storlek t.

Vi grupperar summorna så att vi först summerar över par med gemensam skärningsmängd β, sedan över alla skärningsmängder med storlek u och slutligen över alla u.

$$E[X^2] = \sum_{u} \sum_{\beta;|\beta|=u} \left(\sum_{a_1,a_2;a_1 \cap a_2 = \beta} E[X_{\alpha_1} \cdot X_{\alpha_2}] \right)$$

Det finns $\binom{n}{u}$ sätt att välja en skärningsmängd med u element. När detta är gjort, finns det $\binom{n-u}{t-u}$ sätt att utvidga den till α_1 och därefter $\binom{n-t}{t-u}$ sätt att välja α_2. Summan har därför

$$\binom{n}{u} \cdot \binom{n-u}{t-u} \cdot \binom{n-t}{t-u}$$

stycken termer.

α_1 och α_2 har vardera $\binom{t}{2}$ par av hörn. Gemensamt har de $\binom{u}{2}$ hörnpar. Sannolikheten att varje sådant par har en kant är ½. Variablerna X_{α} uppfyller

$$X_{\alpha_1} \cdot X_{\alpha_2} = 1 \qquad \text{om både } \alpha_1 \text{ och } \alpha_2 \text{ är klickar}$$
$$X_{\alpha_1} \cdot X_{\alpha_2} = 0 \qquad \text{annars.}$$

$$\Pr(X_{\alpha_1} \cdot X_{\alpha_2} = 1) = E[X_{\alpha_1} \cdot X_{\alpha_2}] = 2^{-2\binom{t}{2}+\binom{u}{2}}$$

Alltså blir

$$E\left[X^2\right] = \sum_u \binom{n}{u} \cdot \binom{n-u}{t-u} \cdot \binom{n-t}{t-u} \cdot 2^{-2\binom{t}{2}+\binom{u}{2}} =$$
$$= \sum_{u=0}^{t} \frac{n!}{u! \cdot (t-u)!^2 \cdot (n+u-2t)!} \cdot 2^{-2\binom{t}{2}+\binom{u}{2}}$$

Kvoten mellan termerna $u + 1$ och u är för $u > 1$

$$\frac{(t-u)^2}{(u+1)\cdot(n+u+1-2t)} \cdot 2^u,$$

så första termen är störst och termerna avtar mycket snabbt i början. Om vi bara tar med de två första så får vi

$$E\left[X^2\right] \approx \frac{n!}{(t!)^2 \cdot (n-2t)!} \cdot 2^{-2\binom{t}{2}} \cdot \left(1 + \frac{t^2}{(n+1-2t)}\right) =$$
$$= \frac{n^{2t}}{(t!)^2} \cdot 2^{-2\binom{t}{2}} \cdot \prod_{i=1}^{2t-1}\left(1 - \frac{i}{n}\right) \cdot \left(1 + \frac{t^2}{(n+1-2t)}\right)$$

Produkten av de två sista faktorerna kan som ovan uppskattas till

$$1 - \frac{t\cdot(t-1)}{n} + O\left(\frac{t^4}{n^2}\right)$$

och detta gäller även om vi tar med alla termerna. Uppskattningen blir alltså:

$$E\left[X^2\right] = \frac{n^{2t}}{(t!)^2} \cdot 2^{-2\binom{t}{2}} \cdot \left(1 - \frac{t\cdot(t-1)}{n} + O\left(t^4/n^2\right)\right)$$

Tjebyshovs olikhet ger nu

$$\Pr[X=0] \le \Pr\left[\left|X - E[X]\right| \ge E[X]\right] \le \frac{\sigma^2}{E[X]^2} = \frac{E\left[X^2\right] - E[X]^2}{E[X]^2}$$

Eftersom

$$E[X]^2 = \frac{n^{2t}}{(t!)^2} \cdot 2^{-2\binom{t}{2}} \cdot \left(1 - \frac{t\cdot(t-1)}{n} + O\left(t^4/n^2\right)\right)$$

kan vi förkorta bort de stora faktorerna och få

$$\Pr[X=0] \le O\left(t^4/n^2\right)$$

och för $t = \log n$

$$\Pr[X = 0] \leq O\left(\log^4 n / n^2\right)$$