Véletlenített algoritmusok

4. előadás

Tartalomjegyzék:

- elfoglalási probléma, születésnap probléma,
- kupongyűjtő probléma,
- stabil házassági feladat,
- Chernoff korlát (példák),
- forgalomirányítási probléma.

Elfoglalási probléma

n darab megkülönböztethetetlen golyó, m darab különböző láda. A golyókat véletlenszerűen egyenletes eloszlás szerint szétosztjuk a ládákba. Ennél a feladatnál a következő kérdések lehetnek érdekesek számunkra:

- mennyi a golyók maximális száma valamely ládában?
- mennyi a várható értéke adott ládába kerülő golyók számának?

$$x_j \begin{cases} 1, & ha\ j\ az\ i{-}dik\ ládában\ van \\ 0, & ha\ nem \end{cases}$$

$$E\left(\sum_{j=1}^{m} x_j\right) = \sum_{j=1}^{m} E(x_j) = m\frac{1}{m}$$

Nagy valószínűséggel egyetlen ládában sincs k-nál több elem, megfelelően választott k esetén.

Annak a valószínűsége, hogy az i-dik ládában pontosan k golyó van:

$$\binom{m}{k}\left(\frac{1}{m}\right)^k\left(\frac{m-1}{m}\right)^{m-k} \leq \left(\frac{me}{k}\right)^k\left(\frac{1}{m}\right)^k = \left(\frac{e}{k}\right)^k$$

A következő egyenlőtlenség egy felső korlát a binomiális együtthatóra:

$$\Pr(\text{k vagy több golyó}) \leq \left(\frac{e}{k}\right)^k + \left(\frac{e}{k+1}\right)^{k+1} + \cdots + \left(\frac{e}{m}\right)^m \leq$$

Mivel felső becslés kell k+1… helyett k jó, így mértani sorozat lesz.

$$\leq \left(\frac{e}{k}\right)^k + \left(\frac{e}{k}\right)^{k+1} + \cdots + \left(\frac{e}{k}\right)^m < \left(\frac{e}{k}\right)^k \sum_{i=0}^{\infty}\left(\frac{e}{k}\right)^i$$

Legyen $k^* = \lceil\frac{(3ln n)}{(\ln\ln n)}\rceil$, ekkor

$$\Pr(\text{k}^* \text{ vagy több golyó}) \leq \left(\frac{e}{k^*}\right)^{k^*}\frac{1}{1-e/k^*} \leq n^{-2}$$

Az esetek uniójának a valószínűsége nem több mint az összegük valószínűsége:

$$\Pr[\bigcup_{i=1}^{n} \varepsilon_i(k^*)] \ \leq \ \sum_{i=1}^{n} \Pr\,[\varepsilon_i(k^*)] \ \leq \ \frac{1}{n}$$

Köv1.: Ha m golyót m ládába rakunk, akkor

$$\Pr\,(\ \exists \text{ olyan láda, ami } \lceil\frac{(elnn)}{(\ln\ln n)}\rceil \text{ golyónál többet tartalmaz}) < \frac{1}{m}$$

Köv2.: Kevesebb, mint $1 - \frac{1}{n}$ valószínűséggel nincs olyan láda, amiben $k^* = \frac{(e\ln n)}{(\ln\ln n)}$ golyónál több van.

Születésnap probléma

Az előző problémánál *m* golyót véletlenszerűen tettünk *n* ládába. Ennek speciális esete az n=365, amit születésnap problémának nevezünk. Ebben az értelmezésben az év 365 napja megfelel 365 ládának és az *m* ember születésnapját választjuk egymástól függetlenül és egyenletesen a 365 napból. (Ennél a problémánál most eltekintünk a szökőévektől, valamint attól hogy az *m* ember között lehetnek ikrek.)

Milyen nagynak kell lennie az *m*-nek, hogy két ember születésnapja megegyezzen?

Ennek komplementer feladat: mennyi a valószínűsége, hogy nincs két ember, akinek ugyanakkor van a születésnapja?

Legyen továbbra is *n* a napok száma az évben, *i* az emberek száma.

$$\text{Pr(nincs két ember, akinek u.akkor van a szül. napja)} = \frac{n(n-1)(n-2)\ldots(n-i-1)}{n^i} =$$

$$= 1(1-\tfrac{1}{n})(1-\tfrac{2}{n})\ldots\,(1-\tfrac{i-1}{n})$$

Ha x kicsi, akkor $1 - x \leq e^{-x}$.

$$\text{Pr(nincs két ember, akinek u.akkor van a szül. napja)} = \prod_{i=2}^{m}\left(1-\frac{i-1}{n}\right) \leq \prod_{i=2}^{m} e^{-(i-1)/n} =$$

$$= e^{-m(m-1)/2n}$$

Kupongyűjtő probléma

Adott *n* féle kupon (érme). Minden esetben azonos valószínűséggel kapunk belőlük. Hány érmét kell összegyűjteni, hogy minden fajtából kapjunk?

Legyen X véletlen változó, az összes fajta érme összegyűjtéséhez szükséges legkevesebb kísérlet száma. Határozzuk meg X várható értékét: E(X)=?

Az *i*-dik kísérlet sikeres, ha olyan érmét kapunk, amit az előző (*i-1*)-dik kísérletben még nem. Az első (még nincs érménk) és az utolsó (nem végzünk több kísérletet) kísérlet mindig sikeres. Osszuk fázisokra a kísérlet sorozatot. Az első fázis akkor kezdődik, amikor megkapjuk az első érmét, az *i*-dik fázis az *i*-dik sikeres kísérlettel kezdődik. Az *i*-dik fázis vége az *(i+1)*-dik sikeres kísérlet.

Példa: 4 érme esetén, ha az érméket az 1,2,3,4 számokkal jelöljük, kaphatjuk a következő kísérlet sorozatot: $\underline{1}$ 1 $\underline{2}$ $\underline{3}$ 3 1 2 $\underline{4}$

Ahol az aláhúzott számok a sikeres kísérletek, az azonos színű számok azonos fázisba tartoznak.

Definiáljuk az X_i véletlen változót, ahol $0 \leq i \leq n-1$ az i-dik fázisban lévő kísérletek száma (másképp: egy fázis hossza). Így X-et a következő képpen kapjuk:

$$X = \sum_{i=0}^{n-1} X_i$$

Így X várható értéke:

$$E(X) = \sum_{i=1}^{n} E(X_i)$$

(i+1)-re van *i* darab különböző érménk, és *(n-i)* amilyet még nem kaptunk.

Tehát

$$1 - p = \frac{n-i}{n}$$

valószínűséggel megyünk tovább.

Ekkor a következőt kapjuk

$$E(X_i) = p + 2(1-p)p + 3(1-p)^2 p + \cdots$$

Felhasználva, hogy $1 - p = q$, amiből $p = 1 - q$

$$E(X_i) = (1-q)[1 + 2q + 3q^2 + \cdots + (n+1)q^n]$$

X_i véletlen változó geometriai eloszlású, p paraméterrel.

X_i várható értéke $\frac{1}{p}$ és szórásnégyzete $\frac{(1-p)}{p^2}$.

$$E(X) = \sum_{i=0}^{n-1} E(X_i) = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{i=1}^{n} \frac{1}{i} = nH_n$$

Ahol H_n az n-dik Harmonikus szám, ami aszimptotikusan egyenlő $\ln n + \Theta(1)$, így kapjuk hogy

$$E(X) \approx n \ln n$$

Létezik ennél egy élesebb korlát is. Legyen

$$m = n \ln n + cn$$

Ha E>0 , akor kapjuk, hogy

$$\Pr(X > m) \leq 1 - e^{-e^{-c}}$$

Ez nagyon gyorsan tart $\ln n$-hez.

Stabil házasság probléma

A stabil házasság problémában adott *n* férfi ($Y_1,\ldots,Y_n$), *n* nő ($X_1,\ldots,X_n$) és mindegyikük valahogyan rangsorolja az ellentétes nem tagjait; ez az illető személy *preferencia listája*. Így például akkor mondjuk, hogy az Y_1 (férfi) jobban kedveli vagy preferálja X_1-t X_2-höz képest, ha Y_1 preferencia listáján X_1 előrébb van, mint X_2. A házasság vagy párosítás (jelöljük *M*-el), egy egy-az-egyhez megfeleltetés egy férfi és egy nő között.
Egy *M* párosítás instabil ha vannak olyan Y_1, Y_2 férfiak és X_1, X_2 nők, hogy $(Y_1,X_1) \in M$, $(Y_2,X_2) \in M$, de Y_2 preferálja X_1-et X_2-höz képest, és X_1 preferálja Y_2-t Y_1-hez képest. Egy *M* párosítás stabil, ha nem instabil.
Formálisan:

$$\exists\ (X_i, Y_j)\ pár\ és\ (X_k, Y_l) pár$$
$$instabil, ha\ X_i\ preferencia\ listáján\ Y_l\ >\ Y_j\ és\ Y_l\ listájában\ X_i\ >\ X_k$$

AJÁNLAT-ALGORITMUS:

1. Veszünk egy férfit, abból a csoportból akiknek aktuálisan nincs párja.
2. Ajánlatot tesz a preferencia listája alapján a legkedveltebb nőnek, aki még nem utasította vissza.
3. Ha a nő jobb ajánlatot kapott, mint az aktuális, akkor csere, az addigi választottja pedig visszakerül az aktuálisan nincs párja csoportba.
4. Ha még nincs minden személynek párja, akkor 1., egyébként VÉGE.

Ez az algoritmus mindig terminál és stabil házasság lesz az eredmény.
Bizonyítás:

1. Legyen (X,Y) és (A,B) párosítások. Tegyük fel, hogy X-B felbontja!
2. Ehhez X preferencia listájában B-nek előrébb kell lennie, mint Y, és B preferencia listájában X-nek előrébb kell lennie, mint A.
3. Ez azt jelenti, hogy X a B-nek tett ajánlatot valamikor Y előtt, de B valamikor az eljárás során talált jobb párt. Ez azt jelenti, hogy A jobb X-nél, ami ellentmondás.

Futási idő: Hány ajánlatot tesz ez az algoritmus? $O(n^2)$

Átlagos eset vizsgálat: Átlagosan hány ajánlat?

A késleltetett döntések elvét itt úgy módosítjuk, hogy a férfi mindig véletlenszerűen választ a nők közül, akik még nem utasították vissza. Így figyelni kell, hogy ki utasította vissza, tehát függés van.

Egy amnéziás algoritmus esetén a férfi mindig véletlenszerűen választ egy nőt és annak tesz ajánlatot. Az algoritmus akkor ér véget, amikor minden nő kapott ajánlatot.

Látható, hogyha a férfiakat behelyettesíthetjük a golyókkal, a nőket pedig a ládákkal, akkor az eddigiekhez hasonló problémát kapunk és így alakul a futási idő is.

Várható futási idő: $n \ln n$.

$$P(futás > n \ln n + Cn) < 1 - e^{-e^{-c}}$$

Chernoff-korlát

Bernoulli eloszlású, korlátos valószínűségi változók összegének közelítésére.

Legyen $X_1,\dots,X_n$ Bernoulli eloszlású független kísérlet.
Legyen továbbá $1 \leq i \leq n$.

$$\Pr(X_i = 1) = p_i$$
$$\Pr(X_i = 0) = 1 - p_i$$

Legyen

$$X = \sum_{i=1}^{n} X_i$$

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p_i = \mu$$

Mennyire térhetünk el ettől a μ-től? Mennyi a valószínűsége, hogy az X meghaladja az $(1+\delta)\mu - t$?

Tétel: A fenti összegekre és bármely δ>0-ra:

$$\Pr(X > (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)\mu^{(1+\delta)\mu}}\right)^{\mu}$$

Jelölés: $F^{+}(\mu, \delta)$.

Köv.: Ha $\delta > 2e - 1$, akkor $F^{+}(\mu, \delta) \leq 2^{-(1+\delta)\mu}$.

Tétel: A fenti összegekre és bármely δ>0-ra:

$$\Pr(X > (1-\delta)\mu) \leq e^{\frac{-\mu\delta^2}{2}}.$$

Jelölés: $F^{-}(\mu, \delta)$.

Példa: Egy focicsapat minden mérkőzésén $\frac{1}{3}$ valószínűséggel veszít, és *n* fordulós a bajnokság.
Mi a valószínűsége, hogy elérik az 50%-os teljesítményt?
(Vagy tesztírás, kérdésenként 3 válasz lehetőséggel, *n* kérdés esetén.)

$$\mu = \frac{n}{3} \qquad \delta = \frac{1}{2} \qquad \frac{n}{2} = (1 + \frac{1}{2})\frac{n}{3}$$

$$F^{+}\left(\frac{n}{3}, \frac{1}{2}\right) = \Pr(X > (1+\delta)\mu) \leq 2^{-\frac{3}{2}\frac{n}{3}} = 2^{-\frac{n}{2}}$$

$$\mu = \frac{2}{3}n \qquad \delta = \frac{1}{4} \qquad 1-\delta = \frac{3}{4} \qquad \frac{n}{2} = (1-\frac{1}{4})\frac{2}{3}n$$

$$F^{-}\left(\frac{2}{3}n, \frac{1}{4}\right) = \Pr(X < (1-\delta)\mu) \leq 2^{-\frac{3}{4}\frac{2}{3}n} = 2^{-\frac{1}{2}n}$$

Def.: Bármely pozitív μ-re és ϵ-ra $\Delta^{+}(\mu, \epsilon)$ az a δ, amire $F^{+}(\mu, \Delta^{+}(\mu, \epsilon)) = \epsilon$.
Hasonlóképpen $\Delta^{-}(\mu, \epsilon)$ az a δ, amire $F^{-}(\mu, \Delta^{-}(\mu, \epsilon)) = \epsilon$.

Mennyi az az eltérés, amit még megengedhetünk?

$$2^{-(1+\delta)\mu} = \epsilon \qquad \delta = \ln \epsilon$$

$$\log_2 \epsilon = -(1+\delta)\mu$$

$$\frac{1}{\log_2 \epsilon} = (1+\delta)\mu$$

$$\delta = \frac{1}{\log_2 \epsilon\ \mu} - 1$$

Forgalomirányítási probléma (permutációs routing)

Adott egy gráf topológiájú számítógép hálózat. A hálózaton csomagküldés van a kezdőállomásról a célállomásra. A csúcsok egyértelműen azonosítva vannak egy számmal 1-től N-ig. Jelöljük v_i-vel a csomagok kiinduló pontját, és d(i)-vel az i-dik csomag célpontját. $1 \leq i \leq N$, minden csomópont csak egy csomag célpontja lehet, tehát a célpontok egy permutációt alkotnak. Az csomagok útvonala, a csomópontok egy sorozata a kiinduló pontjuktól a célpontjukhoz. A forgalomirányítási probléma egy algoritmusa minden csomaghoz hozzárendel egy útvonalat.

Az útvonal követése során a csomagnak alkalmanként várnia kell egy köztes csomópontnál, mert a következő csúcs az útvonalán foglalt, mert egy másik csomagot továbbít. Tehát minden csomópontnak tartalmaznia kell egy sort és rendelkeznie kell egy sorba állítási stratégiával, hogy a csomagok közötti konfliktusokat feloldja és egyértelmű legyen, hogy melyik csomag következik, ha szabaddá válik a szükséges csomópont.

Hanyag algoritmus esetén az utat v_i-ből csak d(i) alapján határozza meg.

Tétel: Ha N csúcs van és minden fokszám d, akkor minden determinisztikus algoritmushoz létezik olyan input, hogy a késedelem $\geq \sqrt{\frac{N}{d}}$.

A hiperkocka egy négydimenziós kocka, aminek $N=2^n$ csúcsa van. Jelölje a csúcsokat egy n hosszú bitsorozat, tehát egy csúcs $\overline{X} \in \{0,1\}^n$. Két csúcs akkor van összekötve, ha csak 1 bitben térnek el egymástól (Hamming-távolságuk 1).

Bitrögzítő stratégia: ismert a kiinduló pont (X) és a célpont (Y), mint n hosszú bitsorozat:

$$(X_1, \dots, X_n) \to (Y_1, \dots, Y_n)$$

Az útvonalat úgy határozzuk meg, hogy mindig a legbalrább eső különböző bitet állítjuk át.

Például: 1.) (1011) → (0000) esetén az útvonal: (1011) →(0011) →(0001) →(0000)
2.) (101010) →(001111) esetén: (101010) →(001010) →(001110) →(001111).

Ha véletlenített algoritmust használunk akkor a várható lépésszám, kisebb, mint $\sqrt{\frac{N}{n}}$. Ez az algoritmus, használ egy egyszerű kétfázisú sémát. Ez a séma választ minden v_i-re egy közbeeső σ(i) csomópontot:

I. Elküldi a ebbe a σ(i) csomópontba,
II. Ezután σ(i)-ből d(i)-be küldi.

Minden fázis és minden csomag a bitrögzítő stratégiát használja az útvonala meghatározásához.

Az I. fázis elemzése (a II. fázis az első fázis inverze):

Ha két csomagnak egy ideig közös útvonala van, aztán az egyik eltér, akkor később már nem fognak találkozni.

Lemma: Legyen a v_i útvonala az (e_1, e_2, …, e_k) élek sorozata. Legyen S azon csomagok halmaza (v_i kivételével), akiknek az útja az (e_1, e_2, …, e_k)-ba belemetsz. Ekkor v_i késedelme ≤ |S|.

Bizonyítás: minden késést valakihez hozzá tudunk rendelni.

Késedelem: t-j, ahol a csomag a t időpontban az e_j élen megy át.

Amikor v_i késedelme x-ről x+1-re nő, akkor van legalább egy olyan S-beli csomag, amely keresztezné ugyanazt az élet mint v_i, ugyanabban a lépésben, mert ha nem így lenne, akkor v_i kapna engedélyt az él keresztezésére és nem növekedne a késedelme.

Végeredményképpen, minden késedelmet külön ponthoz rendelünk.

Legyen H_{ij} véletlen változó a következőképpen:

$$H_{ij} \begin{cases} 1, ha\ j\ útja\ belemetsz\ i\ útjába \\ 0, különben \end{cases}$$

A teljes késés az i csúcsra:

Késedelem(i) = $\sum_{j=1}^{n} H_{ij}$

A v_i csomag késésre, a Chernoff-korláttal egy felső korlátot kaphatunk $\sum_{j=1}^{n} H_{ij}$ mellett.

Legyen T_k azon utak száma, amik e_k-n átmennek.

$$E\left(\sum_{j=1}^{N} H_{ij}\right) \leq E\left(\sum_{k=1}^{g} T_k\right)$$

Várható úthossz? $\frac{n}{2}$, mert minden bit $\frac{1}{2}$ valószínűséggel változik.

Várható élek száma? $\frac{Nn}{2}$, N csúcs, n bit

$$E(T_k) = \frac{\frac{Nn}{2}}{Nn} = \frac{1}{2}$$

$$E\left(\sum_{j=1}^{N} H_{ij}\right) \leq \frac{n}{2}$$

$$Pr\left(\sum_{j=1}^{N} H_{ij} > 6n\right) \leq 2^{-12\frac{n}{2}} = 2^{-6n}$$

$$Pr(bármely\ csomag\ késése > 6n) \leq 2^{-5n}$$

<u>Tétel:</u> Legalább $1 - 2^{-5n}$ valószínűséggel, minden csomag eléri a közbeeső csomópontot az I. fázisban 7n vagy kevesebb lépésben.

<u>Tétel:</u> Legalább $1 - \left(\frac{1}{N}\right)$ valószínűséggel, minden csomag célba ér 14n vagy kevesebb lépésben.

Készítette: Bobák György