Uporabne formule

$$H_n = \sum_{k=1}^{n} \frac{1}{k} \ \le\ 1 + O(\log n)$$

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q} \qquad \sum_{n=0}^{b} q^n = \frac{1-q^{b+1}}{1-q}$$

$$\sum_{n=a}^{\infty} q^n = \frac{q^a}{1-q} \qquad \sum_{n=a}^{b} q^n = \frac{q^a - q^{b+1}}{1-q}$$

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + ... + ab^{n-2} + b^{n-1})$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

$$\frac{1}{(1-x)^n} = \sum_{k=0}^{n} \binom{n+k-1}{k} x^k$$

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Izbori

Imamo n oštevilčenih kroglic. Na koliko načinov lahko izberemo k kroglic?

	s pon.	brez pon.
variacije *vrstni red je pomemben*	n^k	$n^{\underline{k}}$
kombinacije *vrstni red ni pomemben*	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Verjetnostni algoritmi za odločitvene probleme

Odgovarjamo na vprašanje $\omega \in \Pi$?
Las Vegas algoritmi vedno vrnejo pravilen odgovor
Monte Carlo algoritmi lahko vrnejo napačen odgovor

- tip 1: $P(\text{yes} \mid \omega \in \Pi) \ge \frac{1}{2}$ $P(\text{yes} \mid \omega \notin \Pi) = 0$
- tip 2: $P(\text{yes} \mid \omega \in \Pi) = 1$ $P(\text{yes} \mid \omega \notin \Pi) \le \frac{1}{2}$
- tip 3: $P(\text{yes} \mid \omega \in \Pi) \ge \frac{3}{4}$ $P(\text{yes} \mid \omega \notin \Pi) \le \frac{1}{4}$

Razredi kompleksnosti odločitvenih problemov

- RP (randomized polynomial time):
  $\exists$ Monte Carlo tipa 1, ki v najslabšem primeru deluje v polinomskem času.
- co-RP:
  $\exists$ Monte Carlo tipa 2, ki v najslabšem primeru deluje v polinomskem času.
- BPP (bounded-error probabilistic polynomial time): $\exists$ Monte Carlo tipa 3, ki v najslabšem primeru deluje v polinomskem času.
- ZPP (zero-error probabilistic polynomial time):
  $\exists$ Las Vegas algoritem, ki deluje v pričakovanem polinomskem času.
  Ali (ekvivalentna definicija): $\exists$ alg, ki v najslabšem primeru deluje v polinomskem času in vedno vrne pravilen odgovor ali "ne vem" in $P(\text{"ne vem"}) < \frac{1}{2}$.

$\text{ZPP} = \text{RP} \cap \text{co-RP}$, $\text{P} \subset \text{ZPP}$, $\text{RP} \cup \text{co-RP} \subset \text{BPP}$

Neenakost Chernoffa

$X_1, \dots, X_n$ neodvisne slučajne spremenljivke, $X_i \in \{0,1\}$, $X = \sum_{i=1}^{n} X_i$, $\mu = E(X)$. Potem za vsak $\delta \in (0,1)$ velja:

$$P(X - \mu \ge \delta\mu) \le e^{-\frac{\delta^2\mu}{2+\delta}} \le e^{-\frac{\delta^2\mu}{3}}$$

$$P(\mu - X \ge \delta\mu) \le e^{-\frac{\delta^2\mu}{2}} \le e^{-\frac{\delta^2\mu}{3}}$$

$$P(|X - \mu| \ge \delta\mu) \le 2e^{-\frac{\delta^2\mu}{3}}$$

Verjetnostni algoritmi za aproksimacijo

Verjetnostni algoritem izračuna (ϵ, δ)-aproksimacijo za V, če vrne X tako, da velja:

$$P(|X - V| \le \epsilon V) \ge 1 - \delta$$

Naj bodo $X_1, \dots X_m$ slučajne spremenljivke, $\mu = E(X_i)$, $Y = \frac{\sum X_i}{m}$. Če je $m \ge \frac{3\ln(2/\delta)}{\epsilon^2\mu}$, potem velja:

$$P(|X - \mu| \ge \epsilon\mu) \le \delta$$

in Y je (ϵ, δ)-aproksimacija za μ.

Polinomi

Naj bo $\mathbb{F}$ polje. Stopnja polinoma $p \in \mathbb{F}[x_1, \dots, x_n]$ je $\deg(p(x_1, \dots, x_n)) = \deg(p(x, \dots, x))$

Predstavitev s polinomi

terka	$T = (a_0, \dots, a_n) \ \mapsto\ p_T(x) = \sum_{i=0}^{n} a_i x^i$
terka alternativa	$T = (a_0, \dots, a_n) \ \mapsto\ p_T(x_0, \dots, x_n) = \sum_{i=0}^{n} a_i x_i$
množica	$M = \{a_0, \dots, a_n\} \ \mapsto\ p_M(x) = \prod_{i=0}^{n} (x - a_i)$
množica terk	$\{T_0, \dots, T_m\} \ \mapsto\ p(x,y) = \prod_{i=0}^{m} (y - p_{T_i}(x))$
množica množic	$\{M_0, \dots, M_m\} \ \mapsto\ p(x,y) = \prod_{i=0}^{m} (y - p_{M_i}(x))$

Želimo ugotoviti ali je $A = B$. Skonstruiramo polinoma p_A in p_B.

```
za i = 0, ..., k
    r ← nakljucna vrednost iz S^n
    ce p_A(r) ≠ p_B(r):
        vrni NE
vrni DA
```

$$P(\text{DA}|A \ne B) \le \left(\frac{d}{|S|}\right)^k$$

Schwartz-Zippelov izrek

Naj bo $p \in \mathbb{F}[x_1, \dots, x_n]$ in $\deg(p) = d \ge 0$. Naj bo $S \subseteq \mathbb{F}$ poljubna končna podmnožica. Za naključno izbiro (enakomerno) $r \in S^n$ velja:

$$P(p(r) = 0) \le \frac{d}{|S|}$$

Verjetnost

Verjetnost na $(\Omega, \mathcal{F})$ je preslikava $P : \mathcal{F} \to \mathbb{R}$ z lastnostmi:

- $P(A) \ge 0$ za $\forall A \in \mathcal{F}$
- $P(\Omega) = 1$
- Za paroma nezdružljive (disjunktne) dogodke $\{A_i\}_{i=1}^{\infty}$ velja *števna aditivnost*
  $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$
- $P(\emptyset) = 0$
- P je končno aditivna.
- P je *monotona*: $A \subseteq B \implies P(A) \le P(B)$
- $P(A^\complement) = 1 - P(A)$
- P je *zvezna*:
  $$A_1 \subseteq A_2 \subseteq \cdots \implies P(\bigcup_{i=1}^{\infty}) = \lim_{i \to \infty} P(A_i)$$
  $$B_1 \supseteq B_2 \supseteq \cdots \implies P(\bigcap_{i=1}^{\infty}) = \lim_{i \to \infty} P(B_i)$$

Matematično upanje

Za slučajno spremenljivko $X : \Omega \to \mathbb{Z}$

$$E(X) = \sum_{c \in \mathbb{Z}} c P(X = c)$$

Lastnosti

$$E(f(X)) = \sum_{c \in \mathbb{Z}} f(c) P(X = c)$$

Linearnost: za poljubne sl. sprem $X_1, \dots, X_n$ velja:

$$E(a_1X_1 + \dots a_nX_n) = a_1E(X_1) + \dots + a_nE(X_n)$$

Če ima $|X|$ mat. up., ga ima tudi X in velja

$$|E(X)| \le E(|X|)$$

Če obstaja $E(X^2)$ in $E(Y^2)$, obstaja tudi $E(XY)$ in velja:

$$|E(XY)| \le E(|XY|) \le \sqrt{E(X^2)E(Y^2)}$$

Disperzija (varianca)

$$D(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

Lastnosti:

- $D(X) \ge 0$
- $D(X) = 0 \iff P(X = E(X)) = 1$
- $D(aX) = a^2 D(X)$

Standardna diviacija/odklon:

$$\sigma(X) = \sqrt{D(X)}$$

zanjo velja $\sigma(aX) = |a|\sigma(X)$.

Neodvisnost

Diskretno porazdeljeni sl. sprem. X in Y sta noedvisni, če velja:

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

za vse i, j.

Nekoreliranost

Sl. sprem. X in Y sta nekorelirani, če velja:

$$E(XY) = E(X)E(Y)$$

$$X, Y \text{ neodvisni} \implies X, Y \text{ nekorelirani}$$

Če imata X in Y, je nekoreliranost ekvivalentna zvezi:

$$D(X + Y) = D(X) + D(Y)$$

Neenakost Markova

Če je X ne negativna sl. sprem. z mat. up., potem je

$$P(|X| \ge a) \le \frac{E(|X|)}{a} \quad \forall a > 0$$

Neenakost Čebiševa

Če ima X disperzijo, je

$$P(|X - E(X)| \ge a\sigma(X)) \le \frac{1}{a^2} \quad \forall a > 0$$

oziroma za $\varepsilon := a\sigma(X)$

$$P(|X - E(X)| \ge \varepsilon) \le \frac{D(X)}{\varepsilon^2}$$

Znani problemi

Perfect matching

Naj bo G graf. Popolno ujemanje je podmnožica povezav $M \subseteq E(G)$, tako da vejla

$$\forall e, f \in M \ :\ e \cap f = \emptyset \quad \text{in} \quad \bigcup_{e \in M} e = V(G)$$

G imam popolno ujemanje $\iff \det(A_G) \ne 0$

$$A_G = [a_{ij}]_{i,j=1}^{n} \qquad a_{ij} = \begin{cases} x_ij & \text{če } ij \in E(G), i < j \\ -x_ij & \text{če } ij \in E(G), i > j \\ 0 & \text{sicer} \end{cases}$$

Min/max prerez (Min/max cut)

Naj bo G graf. Prerez je particija $V(G)$ na U in $V(G) \setminus U$ tako da se minimizira/maksimizira število povezav med U in $V(G) \setminus U$.

```
alg rand_min_cut
vhod: graf G
G_0 ← G
i ← 0
dokler V(G_i) > 2:
    e ← nakljucna povezava v G_i
    G_{i+1} ← G_i \ e // skrcitev povezave e
    i ← i + 1
u, v ← V(G_{n-2}) // n = |V(G)|
U = {w ∈ V(G) | w je bil skrcen v u}
vrni (U, V(G) \ U)
```

Algoritem `rand_min_cut` vrne min. prerez z verjetnostjo $\frac{2}{n(n-1)}$.

Nakjučni grafi

Erdős-Rényi model

$G \in G(n,p)$ je nakjučni Erdős-Rényi graf z n vozlišči, za katere velja, da je vsak par povezan z verjetnostjo p.

$$E(\text{št. povezav v } G(n,p)) = \binom{n}{2} p$$

$$E(\deg v) = (n-1)p$$

Pravimo, da ima naključni graf neko lastnost *skoraj gotovo* (*almost surely*), če velja

$$\lim_{n \to \infty} P(G \in G(n,p) \text{ ima lastnost}) = 1$$

Naj bo p odvisen od n in $G \in G(n,p)$. Velja:

- $np < 1 \implies G$ je skoraj gotovo nepovezan z komponentami velikosti $O(\log n)$
- $np = 1 \implies G$ ima skoraj gotovo komponento velikosti $n^{\frac{2}{3}}$
- $np = c > 1 \implies G$ ima skoraj gotovo veliko komponento velikosti dn, kjer je $d \in (0,1)$
- $np < (1 - \varepsilon)\ln n \implies G$ je skoraj gotovo nepovezan z izoliranimi vozlišči.
- $np > (1 + \varepsilon)\ln n \implies G$ je skoraj gotovo povezan.

Scale-free property

$$\text{delež vozlišč stopnje k} \approx k^{-\gamma}$$

Barabási-Albert model

Začnemo z m vozlišči in dodajamo vozlišča. Vsakič, ko dodamo vozlišče v dodamo še povezave do ostalih voozlišč z verjetnostjo:

$$P(v \sim u) = \frac{\deg u}{\sum_w \deg w}$$

B.A. model ima *scale-free property* in sicer:

$$\text{delež vozlišč stopnje k} = \frac{2m(m+1)}{k(k+1)(k+2)} \approx k^{-3}$$

Markovske verige

$$\Omega \ \dots \ \text{mnočica stanj}$$
$$X_t \ \dots \ \text{stanje v času } t$$

Markovska veriga je zaporedje slučajnih spremenljivk $X = X_0, X_1, \dots$ za katere velja, da je verjetnost prehoda odvisna le od trenutnega stanja:

$$P(X_{i+1} = x | X_0 = x_0, X_1 = x_1, \dots, X_i = x_i) = \\ = P(X_{i+1} = x | X_i = x_i)$$

Lahko zahtevamo še, da je verjetnost prehoda neodvisna od časa (*time homogeneous*):

$$P(X_i + 1 = x | X_i = y) = P(X_1 = x | X_0 = y)$$

Od zdaj naprej se bomo osredotočili na končno množico stanj $\Omega = \{x_1, \dots, x_n\}$.

$$\mathbf{P} = [p_{ij}]_{i,j=1}^{n} \ \dots \ \text{prehodna matrika}$$
$$p_{ij} \ \dots \ \text{verjetnost prehoda iz } x_i \text{ v } x_j$$

Porazdelitev stanj v času t:

$$q(t) = \begin{bmatrix} q_1(t) & \dots & q_n(t) \end{bmatrix} \qquad P(X_t = x_i) = q_i(t)$$

velja

$$q(t) = q(t-1)\mathbf{P} = q(0)\mathbf{P}^t$$

Stacionarna porazdelitev

$\pi = \left[\pi_1 \dots \pi_n\right]$ je stacionarna porazdelitev, če velja

$$\pi \mathbf{P} = \pi \qquad \sum_i \pi_i = 1$$

Oznake:

- $f_{i,j}$ verjetnost da $X_t = x_j$ za nek (dovoj velik) t, če je $X_0 = x_i$
- $h_{i,j}$ pričakovano število korakov, da pridemo iz stanja x_i v stanje x_j. (*hitting time*)
- $N(i,t,q(0))$ pričakovano število obiskov stanja x_i po t korakih, če je začetna porazdelitev $q(0)$.

Markovska veriga je *irreducible* $\iff \forall i,j \; f_{ij} > 0$. Za take verige velja:

- $\exists$ enolično določena stacionarna porazdelitev
- $f_{i,i} = 1$ in $h_{i,i} = \frac{1}{\pi_i}$
- $\lim_{t\to\infty} \frac{N(i,t,q(0))}{t} = \pi_i$

Markovska veriga je *aperiodična*, če $\nexists c \in \{2,3,\dots\}$, ki deli vse dolžine ciklov (v grafu prehodov med stanji). Za take verige velja še:

- $\lim_{t\to\infty} q(0)P^t = \pi$

Markovske verige v grafu

G povezan graf. Obravnavamo naključne sprehode kot Markovske verige.

$$\mathbf{P} = \left[p_{uv}\right]_{u,v\in V(G)} \qquad p_{uv} = \begin{cases} \frac{1}{\deg(u)} & \text{če } uv \in E(G) \\ 0 & \text{sicer} \end{cases}$$

Velja:

$$\pi = \left[\pi_v\right]_{v\in V(G)} \qquad \pi_v = \frac{\deg(v)}{2|E(G)|} \qquad h_{v,v} = \frac{1}{\pi_v}$$
$$\forall uv \in E(G) \;:\; h_{u,v} < 2|E(G)|$$

Pričakovano število korakov, da obiščemo vsa vozlišča je $4(|V(G)| - 1)|E(G)|$.

Pregled najpogostejših porazdelitev

Porazdelitev	Oznaka	Opis	$E(X)$	$D(X)$	Izvor
Bernoullijeva	$\mathrm{Ber}(p)$	$P(X=0)=1-p$ $P(X=1)=p$	p	pq	Indikator dogodka
Binomska	$\mathrm{Bin}(n,p)$	$P(X=k)=\binom{n}{k}p^k q^{n-k}$	np	npq	Število uspešnih izidov v n neodvisnih poskusih; vsota n neodv. Bernoullijevih sl. spr.
Geometrijska	$\mathrm{Geo}(p)$	$P(X=k)=pq^{k-1}$ $k=1,2,\ldots$	$\frac{1}{p}$	$\frac{q}{p^2}$	Število poskusov do prvega uspešnega izida
Negativna binomska	$\mathrm{NegBin}(n,p)$	$P(X=k)=\binom{k-1}{n-1}p^n q^{k-n}$ $k=n,\,n+1,\,\ldots$	$\frac{n}{p}$	$\frac{nq}{p^2}$	Število poskusov do n-tega uspešnega izida; vsota n neodv. geom. sl. spr.
Poissonova	$\mathrm{Poi}(\lambda)$	$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$ $k=0,1,\ldots$	λ	λ	Število telefonskih klicev, nesreč ipd. v določenem času
Hipergeometrijska	$\mathrm{Hip}(s;r,n)$ $\mathrm{Hip}(r;s,n)$	$P(X=k)=\frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}}$	$\frac{rs}{n}$	$\frac{rs(n-r)(n-s)}{n^2(n-1)}$	Število rdečih kroglic v vzorcu velikosti s, če je v škatli skupaj n kroglic, od tega r rdečih
Diskretna enakomerna na množici $M=\{x_1,\ldots x_n\}$	$\mathrm{Enak_d}(M)$	$P(X=x_k)=\frac{1}{n}$ $P(X\in A)=\frac{\lvert A\cap M\rvert}{\lvert M\rvert}$	$\bar{x}:=\frac{\sum_{k=1}^{n}x_k}{n}$	$\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2=$ $=\frac{\sum_{k=1}^{n}x_k^2-n\bar{x}^2}{n}$	Slepi izbor
Enakomerna na intervalu	$\mathrm{Enak_c}[a,b]$	$p_X(x)=\frac{1}{b-a},\ a\le x\le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	Slepi izbor
Normalna	$\mathrm{N}(\mu,\sigma)$	$p_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	Če je X vsota veliko (vsaj 30) neodvisnih sl. spr., je približno $X\sim\mathrm{N}(\mu,\sigma)$, kjer je $\mu=E(X)$ in $\sigma=\sqrt{D(X)}$.
Standardizirana normalna	$\mathrm{N}(0,1)$	$p_X(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ $P(a<X<b)=\Phi(b)-\Phi(a)$	0	1	$X\sim\mathrm{N}(\mu,\sigma)\Rightarrow\frac{X-\mu}{\sigma}\sim\mathrm{N}(0,1)$
Eksponentna	$\mathrm{Exp}(\lambda)$	$p_X(x)=\lambda e^{-\lambda x},\ x>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Čas čakanja na dogodek
Gama	$\mathrm{Gama}(n,\lambda)$	$p_X(x)=\frac{\lambda^n x^{n-1}e^{-\lambda x}}{\Gamma(n)}$ $x>0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	Za $n\in\mathbb{N}$: čas n-te pojavitve dogodka
Hi kvadrat	$\chi^2(n)=$ $\mathrm{Gama}(\frac{n}{2},\frac{1}{2})$	$p_X(x)=\frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}$ $x>0$	n	$2n$	Vsota kvadratov n neodvisnih stand. normalnih slučajnih spremenljivk

Opomba: $q=1-p$.