

Výběr z existujících dat

dají se použít některé metody výběrových sítí

Máme náhodný soubor U různých jednotek, rozsah souboru je N .

Výběrový soubor je libovolná podmnožina $S \subset U$, je výsledkem realizace výběrového experimentu.

Pravděpodobnostní výběr (experiment) je výběr dle předepsaných pravděpodobností $P(S)$, $S \subset U$.

Soubor pravděpodobností $\{P(S); S \subset U\}$, takový, že $\sum_{S \subset U} P(S) = 1$, se nazývá výběrový plán.

Každé jednotce v náhodném souboru můžeme přiřadit jedinečné nějaké číslo (index) - soubor těchto čísel se nazývá prostor výběru.

Výběrové metody pak můžeme aplikovat buď přímo na jednotky nebo na prostor výběru.

dále pojmy:

indikator zahrnutí jednotky i do výběru S :

$$I_i(S) = \begin{cases} 1 & S \ni i \\ 0 & S \not\ni i \end{cases}, \quad 1 \leq i \leq N.$$

rozsah výběru: $K(S) = \sum_{i=1}^N I_i(S)$ - obecně je to náhodná veličina.

chceme-li výběr o pevném rozsahu k , předepíšeme výběrový plán

$$P(S) = 0 \text{ po n. } S \subset U, \text{ kde } K(S) \neq k$$

pravděpodobnost zahrnutí jednotky i do výběru:

$$\pi_i = \sum_{\{S \subset U; S \ni i\}} P(S)$$

rovnoměrný výběr: π_i je konstantní po n. i

pozor: neplatí $\sum_{i=1}^N \pi_i = 1$, ale $\sum_{i=1}^N \pi_i = EK$, kde k je rozsah výběru!

dh:
$$\sum_{i=1}^N \pi_i = \sum_{i=1}^N \sum_{S \ni i} P(S) = \sum_{i=1}^N \sum_{S \subset U} I_i(S) P(S) = \sum_{S \subset U} \underbrace{\sum_{i=1}^N I_i(S)}_{K(S)} P(S) = EK(S)$$

po výběr o pevném rozsahu k platí $\sum_{i=1}^N \pi_i = k$

pravděpodobnosti vyřazení jednotky i - značí se α_i

mají význam a přírady, že výběr se realizuje jako posloupnost

losů jednotek ze souboru - tedy platí $\sum_{i=1}^N \alpha_i = 1$

nyní chceme zav. prstý maticový výběr o pevném rozsahu k :

je definován pravděpodobnostmi $P(S) = \frac{1}{\binom{N}{k}}$, je-li $K(S) = k$
 $= 0$ jinak

tj. všechny k -lice z daných N prvků mají stejnou pravděpodobnost

příjmě $\pi_i = \frac{k}{N}$ pro vs. $1 \leq i \leq N$

dh.
$$\pi_i = \sum_{S \ni i} P(S) = \binom{N-1}{k-1} \frac{1}{\binom{N}{k}} = \frac{\frac{(N-1)!}{(k-1)!(N-k)!}}{\frac{N!}{k!(N-k)!}} = \frac{k}{N}$$

$$\alpha_i = \frac{1}{N}$$

Jak se da' představit? pohárováním

sekvencím výběrem

nějakou obecnou metodou, kde se položí

$$\alpha_i = \frac{1}{N}$$

Výběr postrokováním:

necht' jednotlivé soubory (nebo grupa) jsou obsaženy ve složkách
vektoru A

pro $i = N, N-1, \dots, N-k+1$

vygeneruji j rovnoměrně na $\{1, \dots, i\}$

pohodím obsah $A(i)$ s obsahem $A(j)$

následek je ve složkách $A(N-k+1), \dots, A(N)$

poznámka: je-li $k = N$, je výsledkem náhodná permutace
složek vektoru A

dh: $\pi_i = \frac{1}{N}$ (i se dostala do S postrokováním s posledním prvkem)

$+ \frac{N-1}{N} \cdot \frac{1}{N-1}$ (————— předposledním)

$+ \frac{N-1}{N} \cdot \frac{N-2}{N-1} \cdot \frac{1}{N-2}$ (————— předim od konce)

$+ \dots = \frac{k}{N}$

Seleční výběr:

po případy, kdy chceme ve výběru zachovat stejné uspořádání,
jako bylo v náhodném souboru

algoritmus: for $i = 1$ to N do

vygeneruji $U \in R(0, 1)$

if $U \leq \frac{k}{N-i+1}$ then výběru $A(i)$
 $k = k-1$ fi

od

nebo: $i = 0$

repeat $i = i + 1$

vygeneruj $U \sim \mathcal{R}(0, 1)$

if $U \leq \frac{k}{N-i+1}$ then vyber $A(i)$

$k = k - 1$ fi

until $k = 0$

$$\text{dk: } \pi_1 = P\left(U \leq \frac{k}{N-1+1}\right) = \frac{k}{N}$$

$$\pi_2 = \left(1 - \frac{k}{N}\right) \frac{k}{N-1} + \frac{k}{N} \cdot \frac{k-1}{N-1} = \frac{(N-k)k + k(k-1)}{N(N-1)} = \frac{k(N-1)}{N(N-1)} = \frac{k}{N}$$

$$\begin{aligned} \pi_3 &= \left(1 - \frac{k}{N}\right) \left(1 - \frac{k}{N-1}\right) \frac{k}{N-2} + \left(1 - \frac{k}{N}\right) \frac{k}{N-1} \frac{k-1}{N-2} + \\ &+ \frac{k}{N} \left(1 - \frac{k-1}{N-1}\right) \frac{k-1}{N-2} + \frac{k}{N} \cdot \frac{k-1}{N-1} \cdot \frac{k-2}{N-2} = \frac{k}{N} \quad \text{ald.} \end{aligned}$$

Výběry při nestýpných pravděpodobnostech káknuli:

pro předepsané pravděpodobnosti káknuli π_i .

α nich se vyčísloje pravděpodobnosti vyčíslení α_i .

obecně je spóčilak múně byt des ténke, doporučuje se aproximace

$$\alpha_i = n \pi_i \left(1 + \frac{\pi_i}{2}\right)$$

nebo

$$\alpha_i = n \pi_i \left(1 - \frac{\pi_i}{2}\right)^{-1}$$

hde n se určí tak, aby platilo $\sum_{i=1}^N \alpha_i = 1$

např. po posly' náhodný' výběr $\bar{\pi}_i = \frac{k}{N}$

$$d_i = \pi \frac{k}{N} \left(1 + \frac{k}{2N}\right) = \pi \frac{k(2N+k)}{2N^2}$$

$$\sum_{i=1}^N d_i = \pi \frac{k(2N+k)}{2N} = 1 \Rightarrow \pi = \frac{2N}{k(2N+k)}$$

$$d_i = \frac{2N}{k(2N+k)} \cdot \frac{k(2N+k)}{2N^2} = \frac{1}{N}$$

Jak se realizují taky s pravděpodobnostmi d_i ?

metoda kumulativních součtů:

napočítáme součty $S_0^* = 0, S_1^* = d_1, \dots, S_m^* = \sum_{i=1}^m d_i, \dots, S_N^* = 1$

generujeme $U \sim \mathcal{R}(0,1)$

určíme j tak, aby $S_{j-1}^* < U \leq S_j^*$

vybereme jednotku j , položíme $d_j = P(S_{j-1}^* < U \leq S_j^*) = S_j^* - S_{j-1}^*$

Postupný' výběr:

konáme posloupnost nerovnost'ých taků s pravděpodobnostmi d_i

pokud vyřádneme jednotku vybranou už dříve, kamilneme ji

a vybereme místo ní jinou, dokud nemáme k různým

Teoreticky je lepší kamilaci' výběr:

vyřádneme-li jednotku vybranou už dříve, kamilneme celý

výběr a začneme znovu

dokud nemáme všechny jednotky různé

Systematický výběr:

$$\text{vezmeme } e = \left\lceil \frac{N}{h} \right\rceil$$

Vzvolíme náhodně jednu jednotku mezi prvními e v řadě souboru

přičiníme k této jednotce bereme do výběrového souboru každou

e -tou, dále pokračujeme do konce řady souboru

takto můžeme dostat k nebo $k+1$ jednotek

Náhodný systematický výběr:

Před provedením systematického výběru se náhodně soubor nejprve permuluje

Oblastní výběr:

Náhodně soubor se rozdělí na oblasti, v každé oblasti

se realizují nějaký výběrový plán nezávisle na ostatních

můžeme se rozhodnout a realizovat i výběrový plán po oblasti

Vicestupňový výběr:

po každém stupni se používá vlastní výběrový plán,

realizuje se postupně

můžeme se kombinovat s oblastním