Kvikksortering

Sammenligninger Vår kvikksortering – *Programkode* 1.3.9 *h*) – er basert på metodene sParter0 (*Programkode* 1.3.9 *f*) og parter0 (*Programkode* 1.3.9 *a*). Anta at vi starter med en tilfeldig permutasjon av tallene fra 1 til n. Først velges midterste verdi. Alle de n verdiene kan ligge der med samme sannsynlighet, dvs. med sannsynlighet $1/n$. Deretter byttes den med den bakerste verdien og sParter0 utføres på de $n - 1$ første verdiene med den bakerste som *skilleverdi*. Siden sParter0 arbeider med $n - 1$ forskjellige og tilfeldige verdier, vil det i gjennomsnitt bli $n - ½$ sammenligninger. Så flyttes skilleverdien til sin rett sorterte plass og kvikksorteringen fortsetter på hver av de to sidene.

La s_n være antall sammenligninger som i gjennomsnitt blir utført for å sortere en tabell med n forskjellige tall. Vi tar kun med sammenligninger der tabellelementer inngår. Anta at tallet i er skilleverdi. For at det skal finne sin rette plass utføres i gjennomsnitt $n - ½$ sammenligninger. Etterpå er det $i - 1$ verdier til venstre for i og de sorteres vha. s_{i-1} sammenligninger. Tilsvarende $n - i$ verdier til høyre for i og s_{n-i} sammenligninger. Siden alle tall kan være skilleverdi får vi gjennomsnittet ved å summere fra 1 til n og så dele med n:

(1) $$s_n = \left(n - \frac{1}{2}\right) + \frac{1}{n}\left(\sum_{i=1}^{n} s_{i-1} + \sum_{i=1}^{n} s_{n-i}\right),\ n \geq 2$$

De to summene er like siden det er de samme leddene som summeres hver sin vei:

(2) $$s_n = \left(n - \frac{1}{2}\right) + \frac{2}{n}\sum_{i=1}^{n} s_{i-1}\,,\ n\ \geq 2$$

Vi ganger først med n på begge sider av likhetstegnet i (2):

(3) $$ns_n = n\left(n - \frac{1}{2}\right) + 2\sum_{i=1}^{n} s_{i-1}\,,\ n\ \geq 2$$

Så bytter vi ut n med $n - 1$ på alle steder i (3):

(4) $$(n-1)s_{n-1} = (n-1)\left(n - \frac{3}{2}\right) + 2\sum_{i=1}^{n-1} s_{i-1}\,,\ n\ \geq 3$$

Neste skritt er å ta differansen mellom (3) og (4). På venstre side blir det:

(5) $$ns_n - (n-1)s_{n-1}$$

og på høyre side:

(6) $$n\left(n - \frac{1}{2}\right) - (n-1)(n - \frac{3}{2}) + 2\left(\sum_{i=1}^{n} s_{i-1} - \sum_{i=1}^{n-1} s_{i-1}\right),\ n\ \geq 3$$

Legg merke til at siste ledd i første sum i (6) er s_{n-1} og siste i den andre er s_{n-2}. Differansen mellom de to summene blir derfor s_{n-1}. Dermed blir (6) det samme som:

(7) $$2n - \frac{3}{2} + 2s_{n-1}$$

Dermed gir (5) og (7) oss:

(8) $$ns_n - (n-1)s_{n-1} = \ 2n - \frac{3}{2} + 2s_{n-1}\,,\ n\ \geq 3$$

Dette kan gjøres om til flg. differensligning:

(9) $$ns_n = (n+1)s_{n-1} + 2n - \frac{3}{2}\,,\ n\ \geq 3$$

Neste skritt er å dele med $n(n + 1)$ på begge sider av likhetstegnet i (9):

$$(10)\qquad \frac{s_n}{n+1} = \frac{s_{n-1}}{n} + \frac{2}{n+1} - \frac{3}{2}\frac{1}{n(n+1)}\ ,\ n\ \geq 3$$

Ligningen i (10) der $\frac{s_n}{n+1}$ er uttrykt vha. $\frac{s_{n-1}}{n}$, gjelder for $n \geq 3$. På samme måte kan vi uttrykke $\frac{s_{n-1}}{n}$ vha. $\frac{s_{n-2}}{n-1}$, osv. til $\frac{s_3}{4}$ vha. $\frac{s_2}{3}$. Det vil gi oss flg. ligning:

$$(11)\qquad \frac{s_n}{n+1} = \frac{s_2}{3} + 2\left(\frac{1}{n+1} + \frac{1}{n} + \cdots + \frac{1}{4}\right) - \frac{3}{2}\left(\frac{1}{n(n+1)} + \frac{1}{(n-1)n} + \cdots \frac{1}{3\cdot 4}\right)$$

Den første summen $\frac{1}{n+1} + \frac{1}{n} + \cdots + \frac{1}{4}$ kan skrives som $\frac{1}{n+1} + H_n - \frac{11}{6}$ der H_n er det n-te harmoniske tallet, dvs. summen av det inverse av heltallene fra 1 til n. I den andre summen inngår generelt leddet $\frac{1}{k(k+1)}$ og det kan skrives som $\frac{1}{k} - \frac{1}{k+1}$. Dermed blir det en såkalt forkortningsrekke og hele summen reduseres til $\frac{1}{3} - \frac{1}{n+1}$. I (11) inngår også s_2, dvs. antallet som trengs for å sortere to tall. Det holder normalt med én sammenligning, men slik vår algoritme er satt opp vil det gå med i gjennomsnitt $\frac{3}{2}$ stykker, dvs. $s_2 = \frac{3}{2}$. Hvis vi gjør disse forenklingene i (11) og så ganger med $n + 1$ på begge sider, får vi det ønskede resultatet:

$$s_n = 2(n+1)H_n - \frac{11}{3}n - \frac{1}{6}\ ,\ n\ \geq 2$$

Ombyttinger I metoden parter0 (*Programkode* 1.3.9 *a*) kalles metoden *bytt*(). Men hvor mange ganger kalles den? Vi har der generelt et tabellintervall med lengde $n - 1$ som inneholder tallene fra 1 til n bortsett fra det tallet blant dem som er skilleverdi. Hvis f.eks. $n = 10$, kan det være de 9 tallene 8, 2, 5, 9, 4, 1, 6, 10, 3 med $s = 7$ som skilleverdi. I et slikt tilfelle vil antall ombyttinger være det samme som antall tall blant de $s - 1$ første som er større enn s. Her vil det bli 2 ombyttinger siden det er 2 tall blant de 6 første som er større enn 7. Hvis en tar gjennomsnittet for alle valg av skilleverdier, vil antallet bli $\frac{n-2}{6}$. I tillegg kommer to ombyttinger i metoden sParter0. Dermed $\frac{n-2}{6} + 2$ i gjennomsnitt. La o_n være gjennomsnittlig antall ombyttinger i kvikksortering. Da får vi på samme måte som for sammenligninger, at o_n blir lik:

$$o_n = \left(\frac{n-2}{6} + 2\right) + \frac{1}{n}\left(\sum_{i=1}^{n} o_{i-1} + \sum_{i=1}^{n} o_{n-i}\right),\ n \geq 2$$

Det kan løses på samme måte som over og vi får flg. resultat:

$$o_n = \frac{1}{3}(n+1)H_n + \frac{5}{9}n - \frac{11}{18}\ ,\ n\ \geq 2$$

Vi ser at det i gjennomsnitt blir seks ganger så mange sammenligninger som ombyttinger.