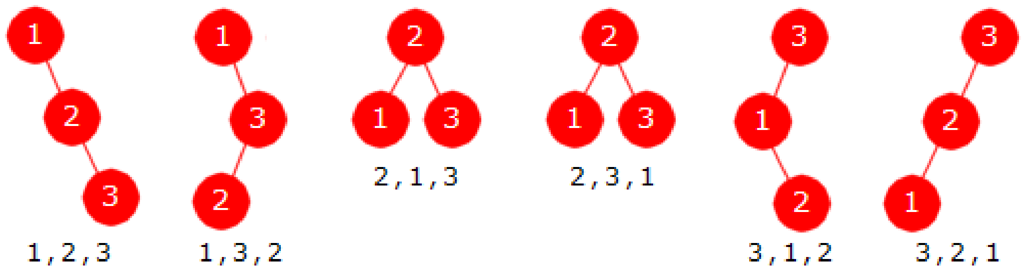
Algoritmer og datastrukturer - Avsnitt 5.2.15 - Algoritmeanalyse

Punkt 5.2.15.6 - Gjennomsnittlig antall noder i en nodes venstre subtre i binære søketrær med *n* forskjellige verdier

En node i et binærtre kan ha et tomt venstre subtre, et venstre subtre med én node, et venstre subtre med to noder, osv. Vi tenker oss nå at hver node har en variabel med navn *vAntall*. Den skal inneholde antallet noder i nodens venstre subtre. La S_n være den gjennomsnittlige *vAntall*–summen der summen tas over alle nodene i treet og der gjennomsnittet tas over de *n*! binære søketrærne vi får ved å lage ett tre for hver permutasjon av tallene fra 1 til *n*. Hvis *n* = 0 har vi et tomt tre. Per definisjon sier vi at $S_0 = 0$. I flg. eksempel er *n* = 3. Figuren viser de seks trærne som de seks permutasjonene av tallene fra 1 til 3 gir:



Det første treet over gir *vAntall*–sum lik 0, i det neste treet blir den 1 (*vAntall* i 3-noden er 1), så blir det 1, 1, 2 og til slutt 3 (= 2 + 1) i det siste treet. Gjennomsnittlig sum blir (0 + 1 + 1 + 1 + 2 + 3) / 6 = 8/6.

La *n* > 0. Hver av verdiene fra 1 til *n* vil ha samme sannsynlighet for å ligge i rotnoden. La $S_n(k)$ være gjennomsnittlig *vAntall*–sum for de binære søketrærne som har *k*, $1 \leq k \leq n$, som rotverdi. I et slikt tre vil de *k* – 1 verdiene som er mindre enn *k* ligge i rotnodens venstre subtre, *k* i rotnoden og resten (*n* – *k* verdier) i høyre subtre. Gjennomsnittlig *vAntall*–sum for de to subtrærne er da S_{k-1} og S_{n-k}. Men rotnoden får *k* – 1 noder i sitt venstre subtre. Dermed:

$$(1) \quad S_n(k) \quad = \quad S_{k-1} + k - 1 + S_{n-k} \ .$$

Siden alle verdiene har samme sannsynlighet for å ligge i roten, får vi

$$(2) \quad S_n \quad = \quad \frac{1}{n}(S_n(1) + S_n(2) + S_n(3) + \ . \ . \ . \ . + S_n(n)) \ .$$

Ved å bruke (1) i (2) får vi så :

$$(3) \quad S_n \quad = \quad \frac{2}{n}(S_0 + S_1 + \ . \ . \ . \ . + S_{n-1}) + \frac{1}{n}(0+1+2+ \ . \ . \ . \ n-1) \ .$$

Lengst til høyre i (3) har vi summen av tallene fra 1 til n – 1 som er lik $\frac{n(n-1)}{2}$. Dermed:

(4) $S_n \;=\; \frac{2}{n}(S_0 + S_1 + \;.\;.\;.\;.\; + S_{n-1}) + \frac{n-1}{2}$.

Dette er samme differensligning som vi hadde i Avsnitt 5.2.15.1 (om gjennomsnittlig nodedybde) bortsett fra at leddet lengst til høyre her er halvparten av det leddet vi hadde i 5.2.15.1. Dermed kan vi bruke samme utledning og svaret blir:

(5) $S_n \;=\; (n+1)H_n - 2n$

For tilfellene *n* = 1, 2, 3 og 4 får vi flg. verdier:

$$S_1 = (1+1)H_1 - 2\cdot 1 = 2\cdot 1 - 2 = 0$$
$$S_2 = (2+1)H_2 - 2\cdot 2 = 3\cdot\frac{3}{2} - 4 = \frac{9-8}{2} = \frac{1}{2}$$
$$S_3 = (3+1)H_3 - 2\cdot 3 = 4\cdot\frac{11}{6} - 6 = \frac{44-36}{6} = \frac{8}{6}\ .$$
$$S_4 = (4+1)H_4 - 2\cdot 4 = 5\cdot\frac{25}{12} - 8 = \frac{125-96}{12} = \frac{29}{12}$$

Vi ser at S_3 stemmer med det vi fant tidligere.

Vi får den gjennomsnittlige verdien på *vAntall* hvis vi deler med *n* :

I gjennomsnitt i et binært søketre med n forskjellige verdier vil en nodes venstre subtre ha $(1+\frac{1}{n})H_n - 2$ *noder.*

Vi kan også analyser dette slik:

I et binært søketre kan antall noder i en nodes venstre subtre variere fra 0 (et tomt subtre) og oppover. La A_n^m være det gjennomsnittlige antallet noder (i et binært søketre med *n* forskjellige verdier) som har et venstre subtre med *m* noder der 0 ≤ *m* ≤ *n* – 1.

La $A_n^m(k)$ være det gjennomsnittlige antallet for de trærne som har *k* som rotverdi der 1 ≤ *k* ≤ *n*. Hvis *k* er rotverdi, vil det være *k* – 1 noder i rotnodens venstre subtre og *n* – *k* noder i høyre subtre. Dermed:

(6) $A_n^m(k) \;=\; A_{k-1}^m + 1 + A_{n-k}^m$ hvis *k* = *m* + 1 og

(7) $A_n^m(k) \;=\; A_{k-1}^m + A_{n-k}^m$ hvis *k* ≠ *m* + 1.

Vi får spesielt at $A_k^m = 0$ for $k \le m$ siden et tre må ha minst m + 1 noder for at det skal ha en venstre subtre med m noder. Hvis $k = m + 1$, vil rotnodens venstre subtre inneholde m noder. Dermed 1-tallet i (6).

Siden alle verdiene har samme sannsynlighet for å ligge i roten, får vi

$$(8) \quad A_n^m = \frac{1}{n}(A_n^m(1) + A_n^m(2) + A_n^m(3) + \ldots + A_n^m(n)) .$$

Ved å bruke (6) i (7) får vi så :

$$(9) \quad A_n^m = \frac{2}{n}(A_0^m + A_1^m + \ldots + A_{n-1}^m + 1) \text{ eller}$$

$$(10) \quad n A_n^m = 2(A_0^m + A_1^m + \ldots + A_{n-1}^m + 1).$$

Erstatter vi n med $n - 1$ i (10) får vi:

$$(11) \quad (n-1) A_{n-1}^m = 2(A_0^m + A_1^m + \ldots + A_{n-2}^m + 1).$$

Differensen mellom (10) og (11) gir:

$$(12) \quad n A_n^m = (n+1) A_{n-1}^m \quad \text{eller}$$

$$(13) \quad \frac{A_n^m}{n+1} = \frac{A_{n-1}^m}{n} = \frac{A_{n-2}^m}{n-1} = \cdots = \frac{A_{m+1}^m}{m+2}$$

A_{m+1}^m er det gjennomsnittlige antallet noder som har et venstre subtre med m noder for trær med m + 1 noder. Det er (m + 1)! trær med m + 1 noder og blant dem m! stykker som har med m + 1 som rotnodeverdi og dermed m noder i venstre subtre. Dermed blir

$$(14) \quad A_{m+1}^m = \frac{m!}{(m+1)!} = \frac{1}{m+1}$$

Ved hjelp av (13) får vi da

$$(15) \quad A_n^m = (n+1)\frac{A_{m+1}^m}{m+2} = \frac{n+1}{(m+1)(m+2)}$$

For tilfellene m = 0,1, 2 og 3 får vi flg. verdier:

$$A_n^0 = \frac{n+1}{2}, \quad A_n^1 = \frac{n+1}{6}, \quad A_n^2 = \frac{n+1}{12}, \quad A_n^3 = \frac{n+1}{20}$$

Det gjennomsnittlige antallet noder i venstre subtre er dermed:

$$(16) \quad (\frac{n+1}{n})\sum_{m=0}^{n-1}\frac{m}{(m+1)(m+2)} \;=\; (1+\frac{1}{n})H_n - 2$$

Det er egentlig rett frem å vise at summen på venstre side gir høyre side som svar. Her nøyer vi oss med å sjekke at det stemmer f.eks. for n = 4:

$$\frac{5}{4}(\frac{0}{2}+\frac{1}{6}+\frac{2}{12}+\frac{3}{20}) = \frac{5}{4}(\frac{10+10+9}{60}) = \frac{12}{48}$$

$$(1+\frac{1}{4})H_4 - 2 = \frac{5}{4}(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}) - 2 = \frac{5}{4}\frac{25}{12} - 2 = \frac{125-96}{48} = \frac{29}{48}$$