(M. Dietzfelbinger, June 1, 2018)

Offen geblieben: Summation der Terme der mittleren Summe

$$S = \sum_{1\le i<k<j\le n} \frac{1}{j-i+1}.$$

Wir stellen sie in der nachfolgenden $(k-1)\times(n-k)$-Matrix dar (für $k \le n/2$):

$$\begin{pmatrix}
\frac{1}{k+1} & \frac{1}{k+2} & \cdots & \frac{1}{n-k} & \frac{1}{n-k+1} & \frac{1}{n-k+2} & \cdots & \cdots & \frac{1}{n-1} & \frac{1}{n} \\
\frac{1}{k} & \frac{1}{k+1} & \frac{1}{k+2} & \cdots & \frac{1}{n-k} & \frac{1}{n-k+1} & \frac{1}{n-k+2} & \cdots & \frac{1}{n-2} & \frac{1}{n-1} \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\
\frac{1}{4} & \frac{1}{5} & \cdots & \frac{1}{k+1} & \frac{1}{k+2} & \cdots & \frac{1}{n-k} & \frac{1}{n-k+1} & \frac{1}{n-k+2} & \cdots \\
\frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{k} & \frac{1}{k+1} & \frac{1}{k+2} & \cdots & \frac{1}{n-k} & \frac{1}{n-k+1} & \frac{1}{n-k+2}
\end{pmatrix}$$

Einteilung in drei Teile:

Teil I, Summe B_1: Einträge strikt unterhalb der Diagonalen mit den $\frac{1}{k+1}$-Einträgen,

Teil II, Summe B_2: Einträge ab der Diagonalen mit den $\frac{1}{k+1}$-Einträgen (startet in linker oberer Ecke) bis zur Diagonalen mit den $\frac{1}{n-k}$-Einträgen (endet zwei Positionen links von rechter unterer Ecke),

Teil III, Summe B_3: Einträge strikt oberhalb der Diagonalen mit den $\frac{1}{n-k}$-Einträgen.

Setze $\alpha := k/n$. Dann gilt $0 < \alpha \le \frac{1}{2}$. Offenbar gilt:

$$B_1 = \frac{k-2}{k} + \frac{k-3}{k-1} + \frac{k-4}{k-2} + \cdots + \frac{1}{3} < k-2.$$

Weiter gilt (beachte $\frac{1}{u+1} + \frac{1}{u+2} + \cdots + \frac{1}{v} \le \int_u^v \frac{dt}{t} = \ln(v) - \ln(u) = \ln(\frac{u}{v})$)

$$\begin{aligned}
B_3 &= \frac{1}{n} + \frac{2}{n-1} + \frac{3}{n-2} + \cdots + \frac{k-1}{n-k} + \frac{k-1}{n-k+1} \\
&\le \left(\frac{n+1}{n} - 1\right) + \left(\frac{n+1}{n-1} - 1\right) + \cdots + \left(\frac{n+1}{n-k+1} - 1\right) \\
&= (n+1)\left(\frac{1}{n} + \frac{1}{n-1} + \ldots \frac{1}{n-k+1}\right) - k \\
&\le (n+1)\ln\left(\frac{n}{n-k}\right) - k
\end{aligned}$$

Wir erhalten, weil $\ln(1/(1-\alpha)) \le \ln 2 < 1$:

$$B_1 + B_3 < (n+1)\ln\left(\frac{n}{n-k}\right) - 2 = (n+1)\ln\left(\frac{1}{1-\alpha}\right) - 2 < -n\ln(1-\alpha).$$

(Achtung: Der Logarithmus ist negativ!) Schließlich:

$$\begin{aligned}
B_2 &= (k-1)\left(\frac{1}{k+1}+\frac{1}{k+2}+\frac{1}{k+3}+\cdots+\frac{1}{n-k}\right)\\
&\le (k-1)\ln\left(\frac{n-k}{k}\right) < \alpha n \ln\left(\frac{n-k}{k}\right)\\
&= \alpha n \ln\left(\frac{1-\alpha}{\alpha}\right) = n\cdot(\alpha\ln(1-\alpha)-\alpha\ln(\alpha)).
\end{aligned}$$

Also:

$$\begin{aligned}
S &= B_1 + B_2 + B_3\\
&\le n\cdot(-\ln(1-\alpha)+\alpha\ln(1-\alpha)-\alpha\ln(\alpha))\\
&= n\cdot(-(1-\alpha)\ln(1-\alpha)-\alpha\ln(\alpha))\\
&= n\cdot H_e(\alpha),
\end{aligned}$$

wobei $H_e(\alpha)$ die Entropie der Verteilung $(\alpha, 1-\alpha)$ zur Basis e ist. Bezüglich der in der Informatik üblichen binären Entropie $H_2(\alpha) = -(1-\alpha)\log_2(1-\alpha) - \alpha\log_2(\alpha)$ ergibt sich mit Hilfe der Beziehung $H_e(\alpha) = (\ln 2)H_2(\alpha)$:

$$S \le n\cdot(\ln 2)H_2(\alpha).$$

Insgesamt, weil $\alpha \le \frac{1}{2}$:

$$\mathbf{E}(C_k) \le n\cdot(2+(2\ln 2)H_2(\alpha)).$$

Der schlechteste Fall tritt für $\alpha = \frac{1}{2}$, also $k = n/2$, auf. Dies ist der Fall des Medians. Dann ist $H_2(\frac{1}{2}) = 1$, und $\mathbf{E}(C_{n/2}) \le (2+(2\ln 2))n \approx 3.3863n$.

Theorem Die erwartete Vergleichsanzahl von Quickselect mit $k = \alpha n$ ist $\le (2+(2\ln 2)H_2(\alpha))\cdot n$.

Der nächste Abschnitt 4.4 zeigt, dass mit einem anderen Algorithmus die Vergleichsanzahl auf etwa $\frac{3}{2}n$ verbessert werden kann.