THE CLASSIC WORK
NEWLY UPDATED AND REVISED

The Art of Computer Programming

VOLUME 3

Sorting and Searching

Second Edition

DONALD E. KNUTH

that

$$P_k = M^{-N}\big(g(M,N,k) + g(M,N,k+1) + \cdots + g(M,N,N)\big). \qquad (39)$$

Now $C'_N = \sum_{k=0}^{N}(k+1)P_k$; putting this equation together with (36)–(39) and simplifying yields the following result.

Theorem K. *The average number of probes needed by Algorithm L, assuming that all M^N hash sequences (35) are equally likely, is*

$$C_N = \tfrac{1}{2}\big(1 + Q_0(M, N-1)\big) \quad \text{(successful search)}, \qquad (40)$$

$$C'_N = \tfrac{1}{2}\big(1 + Q_1(M, N)\big) \quad \text{(unsuccessful search)}, \qquad (41)$$

where

$$Q_r(M,N) = \binom{r}{0} + \binom{r+1}{1}\frac{N}{M} + \binom{r+2}{2}\frac{N(N-1)}{M^2} + \cdots$$

$$= \sum_{k\ge 0}\binom{r+k}{k}\frac{N}{M}\frac{N-1}{M}\cdots\frac{N-k+1}{M}. \qquad (42)$$

Proof. Details of the calculation are worked out in exercise 27. (For the variance, see exercises 28, 67, and 68.) ▮

The rather strange-looking function $Q_r(M,N)$ that appears in this theorem is really not hard to deal with. We have

$$N^k - \binom{k}{2}N^{k-1} \le N(N-1)\ldots(N-k+1) \le N^k;$$

hence if $N/M = \alpha$,

$$\sum_{k\ge 0}\binom{r+k}{k}\left(N^k - \binom{k}{2}N^{k-1}\right)\Big/M^k \le Q_r(M,N) \le \sum_{k\ge 0}\binom{r+k}{k}N^k/M^k,$$

$$\sum_{k\ge 0}\binom{r+k}{k}\alpha^k - \frac{\alpha}{M}\sum_{k\ge 0}\binom{r+k}{k}\binom{k}{2}\alpha^{k-2} \le Q_r(M,\alpha M) \le \sum_{k\ge 0}\binom{r+k}{k}\alpha^k;$$

that is,

$$\frac{1}{(1-\alpha)^{r+1}} - \frac{1}{M}\binom{r+2}{2}\frac{\alpha}{(1-\alpha)^{r+3}} \le Q_r(M,\alpha M) \le \frac{1}{(1-\alpha)^{r+1}}. \qquad (43)$$

This relation gives us a good estimate of $Q_r(M,N)$ when M is large and α is not too close to 1. (The lower bound is a better approximation than the upper bound.) When α approaches 1, these formulas become useless, but fortunately $Q_0(M, M-1)$ is the function $Q(M)$ whose asymptotic behavior was studied in great detail in Section 1.2.11.3; and $Q_1(M, M-1)$ is simply equal to M (see exercise 50). In terms of the standard notation for hypergeometric functions, Eq. 1.2.6–39, we have $Q_r(M,N) = F(r+1, -N; ; -1/M) = F\left({r+1, -N, 1 \atop 1}\,\middle|\, -\frac{1}{M}\right)$.