About picking colored elements from a set

Daniel Reitzner
August 6, 2009

Let's suppose we have n elements from which k are colored (target). How many elements (in average) we have to search in order to find target one? In following we will prove several theorems concerning this problem. First we will consider a very simple type of search, where, after choosing unmarked element, we put it back to the set we search. We will call this search *naïve search*.

Theorem 1. *Let k be the number of target elements forming a subset of a set of n elements. Then the average number of elements to be searched by a naïve search in order to find marked element is*

$$\bar{n}_{\text{naïve}} = \frac{n}{k}.$$

Proof. The proof is simple, since in every step the probability to pick marked element is $p = k/n$. Hence, the probability, that we will find marked element on the m-th step and no sooner is $P_m = (1-p)^{m-1}p$. Using Lemma 1 from Appendix we find, that

$$\bar{n}_{\text{naïve}} = \sum_{m=1}^{\infty} mP_m = \frac{1}{p} = \frac{n}{k}.$$

□

In the next theorem the question is answered by employing a different kind of search — *memory search*, where every picked (non-target) element is put aside, so the number of elements is decreasing from step to step.

Theorem 2. *Let k be the number of target elements forming a subset of a set of n elements. Then the average number of elements to be searched by a memory search in order to find marked element is*

$$\bar{n}_{\text{memory}} = \frac{n+1}{k+1}. \tag{1}$$

Proof. First, we will determine the probability P_m of not finding a target element in $m-1$ steps but finding one on the m-th step. On the first step the probability of not finding target element is $(n-k)/n$. In the next step the probability is $(n-k-1)/(n-1)$, since the number of elements in both sets — the set of non-target elements and the set of all elements — reduces by one by excluding previously chosen element. In the following step the numbers are decreased again. Finally in the m-th step the probability of finding target element is $k/(n-m+1)$ since in the previous $m-1$ steps the number of all elements was reduced by $m-1$ while the number of target elements k remained the same. So the probability P_m then reads

$$P_m = \frac{k}{n-m+1} \prod_{j=0}^{m-2} \frac{n-k-j}{n-j}.$$

We can rewrite the probability to have form

$$P_m = k\frac{(n-k)!(n-m)!}{n!(n-k-m+1)!}.$$

The average number of elements to be searched is then

$$\bar{n}_{\text{memory}} = \sum_{m=1}^{n-k+1} mP_m = \frac{k(n-k)!}{n!} \sum_{m=0}^{n-k} (m+1) \frac{(n-m-1)!}{(n-k-m)!},$$

where the range in the second sum was shifted by one. The upper boundary of the first sum range comes from the fact, that after $n-k$ unsuccessful searches, there is no unwanted element left and in the following step we are sure to pick a target element. Furthermore, by setting $m = n-k-l$, l parameter, we can rewrite the formula for $\bar{n}_{\text{memory}}$ as

$$\bar{n}_{\text{memory}} = \frac{k(n-k)!}{n!} \sum_{l=0}^{n-k} (n-k-l+1) \frac{(k+l-1)!}{l!}.$$

By using formula from Eq. (8) for the sum enumeration we get

$$\bar{n}_{\text{memory}} = \frac{k(n-k)!}{n!} \frac{(k-1)!}{(n-k)!} \frac{n!}{(k+1)!} (n+1) = \frac{n+1}{k+1},$$

which concludes the proof. □

Corollary 1: If we search for a single target element ($k=1$) then we have

$$\bar{n}_{\text{memory}} = \frac{n+1}{2}, \qquad \bar{n}_{\text{naïve}} = n. \tag{2}$$

So the complexity is of the order $O(n)$.

Corollary 2: If we search between all elements ($k=n$) then we, not surprisingly, have

$$\bar{n}_{\text{memory}} = \bar{n}_{\text{naïve}} = 1. \tag{3}$$

Corollary 3: If we search between half elements being target ($k = n/2$ with n even) then we have

$$\bar{n}_{\text{memory}} = 2\frac{n+1}{n+2}, \qquad \bar{n}_{\text{naïve}} = 2. \tag{4}$$

In the limit of large n we get also $\bar{n}_{\text{memory}} \to 2$. This means that the complexity is of the order $O(1)$ even though in the worst case we have to search through $n/2+1$ elements.

Remark. Conclusion of corollary 3 about limit for large n can be easily obtained also in the following way. The probability of finding marked element is $p = 1/2$, which is the same as probability of finding non-marked element. So finding target element after m steps has the probability

$$P_m = \left(\frac{1}{2}\right)^{m-1} \frac{1}{2} = \left(\frac{1}{2}\right)^{m}. \tag{5}$$

Then the average number of steps to find target element is

$$\bar{n} = \sum_{m=1}^{\infty} \frac{m}{2^m}. \tag{6}$$

The sum can be evaluated in the following manner:

$$\bar{n} = 2\bar{n} - \bar{n} = \sum_{m=1}^{\infty} \frac{m}{2^{m-1}} - \sum_{m=1}^{\infty} \frac{m}{2^m} = 1 + \sum_{m=1}^{\infty} \frac{m+1}{2^m} - \sum_{m=1}^{\infty} \frac{m}{2^m} = 1 + \sum_{m=1}^{\infty} \frac{1}{2^m} = 2. \tag{7}$$

Corollary 4: By setting $k = sn$ we see, that for $n \to \infty$ we have $\bar{n}_{\text{naïve}} = \bar{n}_{\text{memory}} = 1/s$.

Appendix

Lemma 1. *Let* $P_m = (1-p)^{m-1}p$ *for* $0 < p \leq 1$, *then*

$$\bar{n} = \sum_{m=1}^{\infty} m P_m = \frac{1}{p}.$$

Proof. First note, that the equation is trivial for $p = 1$ and so we can now restrict ourselves to $0 < p < 1$. Using the following (for $0 < q < 1$)

$$\sum_{m=1}^{\infty} m q^{m-1} = \left[\sum_{m=1}^{\infty} q^m\right]'_q = \left[q\sum_{m=0}^{\infty} q^m\right]'_q = \left[\frac{q}{1-q}\right]'_q = \frac{1}{(1-q)^2}$$

we have that

$$\bar{n} = p\sum_{m=1}^{\infty} m(1-p)^{m-1} = \frac{p}{[1-(1-p)]^2} = \frac{1}{p}.$$

□

Lemma 2. *Let*

$$\sigma_m = \sum_{l=0}^{m}(n-k-l+1)\frac{(k+l-1)!}{l!},$$

then

$$\sigma_m = \frac{(k-1)!}{m!}\frac{(k+m)!}{(k+1)!}[(n+1)+k(n-k-m)]. \tag{8}$$

Proof. We will use mathematical induction to prove the Lemma. For $m = 0$ we get

$$\sigma_0 = (n-k+1)\frac{(k-1)!}{0!}\frac{(k+1)!}{(k+1)!} = \frac{(k-1)!}{0!}\frac{(k+0)!}{(k+1)!}[(n+1)+k(n-k-0)]$$

and so for $m = 0$, Eq. (8) holds.

Let us suppose, that Eq. (8) holds for general m, then for σ_{m+1} we get

$$\begin{aligned}\sigma_{m+1} &= (n-k-m)\frac{(k+m)!}{(m+1)!} + \sigma_m \\ &= \frac{(k-1)!}{(m+1)!}\frac{(k+m)!}{(k+1)!}[(n-k-m)(k+1)k + (m+1)(n+1) + (m+1)k(n-k-m)]\end{aligned}$$

After short evaluation we get

$$\sigma_{m+1} = \frac{(k-1)!}{(m+1)!}\frac{(k+m)!}{(k+1)!}[k+(m+1)]\{(n+1)+k[n-k-(m+1)]\},$$

which is exactly Eq. (8) for $m+1$.

□