- 数组长度为 N，抽样个数为 k，$i \in [k+1, N]$，**注意 i 不是数组下标，第 i 个数字是 $A[i-1]$**
- 每次生成的随机数 $j \in [0, i)$，这个区间长度为 i，按照古典概型，$A[i-1]$ 被放进长度为 k 的水塘里的概率是 $P\{j < k\} = \frac{k}{i}$
- 为什么 i 从 $k+1$ 开始，而不能是 k? 因为生成的随机数 j 落在 $[0, k-1]$ 内时会把 $A[i-1]$ 放入水塘。如果 $i = k$，水塘内就可能出现两个 $A[k-1]$，这是不对的。因此必须保证 $A[i-1]$ 位于水塘之外，即 $i > k$，才能避免这种错误。
- $A[i-1]$ 被放进水塘后有可能被后面的数字替换掉，那么 $A[i-1]$ 最终能留在水塘中的概率是多少？对于 $A[i-1]$ 后面的数字 $A[i'-1] (i' > i)$，先生成一个随机数 $j' \in [0, i')$，如果 $A[i-1]$ 能被 $A[i'-1]$ 替换，就说明 $j' = j$。即，$A[i-1]$ 被 $A[i'-1]$ 替换掉的概率为 $P\{j' = j\} = \frac{1}{i'}$。如果 $A[i-1]$ 最终留在了水塘中，就说明它没有被后面的数字替换掉，概率为：

  $$\frac{k}{i}(1 - \frac{1}{i+1})(1 - \frac{1}{i+2}) \cdots (1 - \frac{1}{N}) = \frac{k}{N}, i \in [k+1, N)$$

  其中，

  $$\begin{aligned}
  &\frac{k}{i} = A[i-1]\text{被选进水塘的概率} \\
  &1 - \frac{1}{i+1} = A[i-1]\text{不被}A[i]\text{替换掉的概率} \\
  &1 - \frac{1}{i+2} = A[i-1]\text{不被}A[i+1]\text{替换掉的概率} \\
  &\cdots \\
  &1 - \frac{1}{N} = A[i-1]\text{不被}A[N-1]\text{替换掉的概率}
  \end{aligned}$$

  当 $i = N$ 时，也就是要把 $A[N-1]$ 放进水塘，概率是 $\frac{k}{N}$。由于后面已经没有数字了，所以不用考虑被替换的情况。

  当 $i \leqslant k$ 时，$A[i-1]$ 位于水塘内，相当于被选进水塘的概率 $= 1$，并且只能被水塘外面的数字，即 $A[k]$ 及其之后的数字替换掉，被替换概率 $= (1 - \frac{1}{k+1})(1 - \frac{1}{k+2}) \cdots (1 - \frac{1}{N}) = \frac{k}{N}$，所以留在水塘中的概率仍是 $\frac{k}{N}$。