

Reservoir Sampling

Algorithm Steps:

1. For first K elements, put them into the sample set S .
2. For the i th element, put it into S , with probability $x = \frac{K}{i}$, where i is N in this case. Then randomly replace an element in S with probability $\frac{1}{K}$

Proof:

When $N \leq K$, each element has probability of 1.

When $N = K+1$, there are 2 cases:

When current element is selected with probability $\frac{K}{N}$, then each old element remains in the sample set S , with probability $\frac{K}{N} \left(1 - \frac{1}{K}\right)$.

When current element isn't selected with probability $\left(1 - \frac{K}{N}\right) = \frac{N-K}{N}$, then each old element remains in the sample set S , with probability $\frac{N-K}{N} * 1$.

Therefore, each old element remains in the the sample set S , with probability

$$\frac{K}{N} \left(1 - \frac{1}{K}\right) + \frac{N-K}{N} = \frac{N-1}{N} = \frac{K}{N}$$