Exercise 4

Claim:
For the Reservoir Sampling (Vitter, 1985) algorithm holds the following: after l elements, S has $min(s, l)$ elements and each picked with probability $\frac{s}{l}$.
Proof:
Let $\sigma =< a_1, \ldots, a_m >$ be a data stream over $[n]$.
Case $l \leq s$: In that case the only first l elements of the data stream are added to S and after l steps: $|S| = l$. Probability $\frac{s}{l}$ equals 1 for $s > l$. And every element in S is picked with probability 1 since every element of the first l is picked.
Case $l > s$: In that case s many elements are added to S and hence after l steps: $|S| = s$. Together with the first case we have: after l steps $|S| = min(s, l)$. After s processed elements S consists of the first s elements of the stream. We know that there is at least a $s + 1$-th step. In that step with probability $\frac{s}{s+1}$ an element of S is picked uniformly at random and replaced with the currently processed element. Hence an element of S gets replaced with probability $\frac{s}{s+1}$ and that means after that processing step the probability for an element to be picked and added (with replacement) to S is $\frac{s}{s+1}$. In general we can say that after l processing steps the probability for an element of the stream to be in S is $\frac{s}{l}$.

$\square$