Homework 0

Due: N/A

Problem 1

Proceed by induction on the number of steps k. When $k = 1$, then our only item is stored in memory, so the sample is uniformly distributed. Assume our algorithm gives a uniform distribution for the first k steps for an arbitrary k. On the $k+1$st step, we have two cases.

In the first case, by definition of our algorithm, the new item is stored with probability $\frac{1}{k+1}$.

In the second case, the old item is maintained with probability $\frac{k}{k+1}$. By independence of memory storage and our inductive hypothesis, the probability that one of the first k items is stored in memory is equal to

$$\frac{1}{k} * \frac{k}{k+1} = \frac{1}{k+1}$$

Therefore, our sample is uniformly distributed over all $k+1$ items. By induction, our sample is uniformly distributed at all steps.

Problem 2

a. After we see the sth picture, we must increment *count* to $4s$ for the algorithm to terminate. By independence of each reload, the probability that we do not see a new picture in $4s$ steps is $\left(1 - \frac{n-s}{n}\right)^{4s}$. Using the inequality

$$1 - x \leq e^{-x} \qquad \text{(Exercise: prove!)}$$

we get that

$$\left(1 - \frac{n-s}{n}\right)^{4s} \leq (e^{-\frac{n-s}{n}})^{4s} = e^{\frac{-4s(n-s)}{n}}.$$

b. Let E_s be the event: "the algorithm terminate after observing exactly s pictures".

In part (a) we proved that for any $s < n$,

$$P(E_s \mid \cap_{i<s} \bar{E}_i) \leq e^{\frac{-4s(n-s)}{n}}.$$

We now observe that

$$P(E_s) = P(E_s \cap (\cap_{i<s} \bar{E}_i)) = P(E_s \mid \cap_{i<s} \bar{E}_i) \cdot P(\cap_{i<s} \bar{E}_i) \leq P(E_s \mid \cap_{i<s} \bar{E}_i),$$

and the probability that the algorithm terminates before observing n is, by addivity of disjoint events,

$$\begin{aligned} \sum_{s=1}^{n-1} P(E_s) \leq \sum_{s=1}^{n-1} e^{-4s(n-s)/n} &= \sum_{s=1}^{n/2} e^{-4s(n-s)/n} + \sum_{s=n/2+1}^{n-1} e^{-4s(n-s)/n} \\ &\leq 2 \sum_{s=1}^{n/2} e^{-2s} < \frac{2e^{-2}}{1-e^{-2}} < 1/2. \end{aligned}$$