Homework #9
Algorithms I
600.463
Spring 2017

Due on: Thursday, May 4th, 11:59pm
Late submissions: will NOT be accepted
Format: Please start each problem on a new page.
Where to submit: On Gradescope, under HW9
Please type your answers; handwritten assignments will not be accepted.
To get full credit, your answers must be explained clearly,
with enough details and rigorous proofs.

May 15, 2017

Problem 1 (20 points)

You are given a biased coin but you dont know what the bias is. Can you simulate a fair coin? Please prove the correctness of your solution.
Answer:
First of all, let's denote the probability that the given biased tosses "head" is p and thus the probability of "tail" is $1 - p$. Assume $0 < p < 1$ here for a valid solution. Our solution is like the following:

1. Toss the coin twice in a row.
2. If both tosses face the same, redo the step 1 until two tosses face differently .
3. If two tosses are "head" and "tail", return "head"; Elseif two tosses are "tail" and "head", return "tail".

Proof of Correctness:
Let's consider the four cases:

- $Pr[(\text{"}head\text{"}, \text{"}head\text{"})] = p^2$

- $Pr[(\text{"}tail\text{"}, \text{"}tail\text{"})] = (1-p)^2$
- $Pr[(\text{"}head\text{"}, \text{"}tail\text{"})] = p(1-p)$
- $Pr[(\text{"}tail\text{"}, \text{"}head\text{"})] = (1-p)p$

From the method described above, let's consider $Pr[\text{"}head\text{"}]$

$$\begin{aligned} Pr[\text{"}head\text{"}] &= Pr[firsttossis\text{"}head\text{"}|\text{two tosses are different}] \\ &= \frac{Pr[(\text{"}head\text{"}, \text{"}tail\text{"})]}{Pr[(\text{"}head\text{"}, \text{"}tail\text{"})] + Pr[(\text{"}tail\text{"}, \text{"}head\text{"})]} \\ &= \frac{p(1-p)}{p(1-p) + (1-p)p} = 1/2 \end{aligned}$$

Similarly $Pr[\text{"}tail\text{"}]$ is also 1/2 and thus $Pr[\text{"}head\text{"}] = Pr[\text{"}tail\text{"}]$.

Problem 2 (20 points)

The following approach is often called *reservoir sampling*. Suppose we have a sequence of items passing by one at a time. We want to maintain a sample of one item with the property that it is uniformly distributed over all the items that we have seen at each step. Moreover, we want to accomplish this without knowing the total number of items in advance or storing all of the items that we see. Consider the following algorithm, which stores just one item in memory at all times. When the first item appears, it is stored in the memory. When the k-th item appears, it replaces the item in memory with probability $1/k$. Explain why this algorithm solves the problem. (Exercise 2.18 from the Book "**Probability and Computing:** randomized algorithms and probabilistic analysis", Mitzenmacher and Upfal 2009. `https://catalyst.library.jhu.edu/catalog/bib_5738970`)

Answer:

We need to show for a total of n elements, by reservoir sampling, each element can be sampled with equal probability $1/n$. Let's prove the correctness by induction.

- Let's consider some base case: when $n = 1$, the first element is sampled with probability 1; when $n = 2$, the second element is sampled with probability $1/2$, and thus the probability that the first element is finally sampled is $1 - 1/2$; when $n = 3$, the third element is sampled with probability $1/3$ to replace the previous stored element. The second element is sampled with probability $1/2 * (1 - 1/3) = 1/3$. The first element is sampled with probability $1 * (1 - 1/2) * (1 - 1/3) = 1/3$.

- Assume that when the total number of element is $n = j$, each element is sampled with probability $1/j$.

- Let's consider the came when $n = j + 1$. So for the first j elements, each element can be sampled with probability $1/j$. When the $(j + 1)$-th arrives, with probability $1/(j + 1)$ to replace the previous one. So for the j-th element, there was with $1/j$ probability to be sampled and have $1 - 1/(j + 1)$ to be maintained when $j + 1$-th element arrives. So j-th element can be sampled with probability of $\frac{1}{j} * (1 - \frac{1}{j+1}) = \frac{1}{j+1}$. Similarly, the same argument applies to all other element from 1st to $(j - 1)$-th. Thus, each element is sampled with $\frac{1}{j+1}$ probability.