Radomized Algorithms Problemes 3 Fall 2019. To hand Problem 15 on the 15th of October. To hand the remaining ones (or mail it before) October 29

15.- (*) (**Recall exercise 5**) In cryptographic applications we often need to generate random integer $r \in \{1, \ldots, n\}$ together with the factorization of r. Note the obvious method of just generating r u.a.r. and then factoring r, is not efficient for large r as factoring is not known to be in P. Consider the following algorithm:

generate an integer sequence $n \geq s_1 \geq s_2 \geq \cdots \geq s_t = 1$
  chose u.a.r. $s_1 \in \{1, \ldots, n\}$
  until reaching 1 **do**
    chose s_i u.a.r. from $\{1, \ldots, s_{i-1}\}$
select the set $S = \{s_i | s_i \text{ is prime}\}$
do $r = \prod_{s_i \in S} s_i$
if $r \leq n$ **output** r with prob $= r/n$
  else FAIL

You can generate u.a.r. $r \in \{1, \ldots, n\}$ by using biased coins as in Problem 4. To test for primality we can either use the deterministic algorithm AKS (Agrawal–Kayal–Saxena) ($O(\lg n)^6$)) or the faster probabilistic algorithms from Miller-Rabin a Monte-Carlo, one side error that runs in ($O(\ln n)^4$)).

(a) Suppose we represent the sequence $\{s_i\}$ obtained by the algorithm as a vector $(m_1, \ldots, m_n)$, where m_i is the number of times the number j occurs in the sequence (for ex. if $n = 10$ and the sequence is $s_1 = 8, s_2 = 5, s_3 = 5, s_4 = 1$the vector would be $(1, 0, 0, 0, 2, 0, 0, 1, 0, 0)$.) Show that the probability of generating the sequence is given by $\prod_{j=2}^{n} (\frac{1}{j})^{m_j} (1 - \frac{1}{j})$. (*Hint: the entire generating process can be thought of as a tossing of a sequence of coins*)

(b) Show (from the previous item) that the algorithm outputs each $r \in \{1, \ldots, n\}$ with equal probability α_n/n where $\alpha_n = \prod_p (1 - 1/p)$, and the product is over all primes $p \leq n$.

(c) A standard theorem from number theory says that $\alpha_n^{-1} \sim 1.8 \ln n$. Suppose we repeat the algorithm until it outputs a r. What is the expected number of repetitions needed?

(d) The running time of each trial of the algorithm is dominated by the primality test. Show that the expected number of primality test is bounded by $H_n = \Theta(\ln n)$.

(e) Show that the expected number of primality tests performed before an output r is produced is $O(\lg^2 n)$. Justify why it is not enough to multiply the expectations obtained in items (c) and (d).

16.- Let X a r.v. such that $X = Rand(1, n)$. Find $\mathbf{Var}[X]$.

17.- Sigui $a_1, a_2, \ldots, a_n$ una llista de n enters diferents. Direm que a_i i a_j estan invertits si $i < j$ però $a_i > a_j$. L'algorisme de la bombolla intercanvia parells d'elements invertits fins que tots els elements són ordenats. Suposeu que l'entrada a l'algorisme és una permutació amb n enters diferents (agafada d'una distribució uniforme).

(a) Digueu quin és el nombre esperat μ d'inversions que realitzarà l'algorisme de la bombolla.

(b) Calculeu la variança del nombre d'inversions. Si X representa la variable aleatòria del nombre d'inversions, X està concentrada al voltant de μ?

18.- (MU 3.15) Let the random variable X be representable as a sum of random variables $X = \sum_{i=1}^{n} X_i$. Show that, if $\mathbf{E}[X_i X_j] = \mathbf{E}[X_i]\,\mathbf{E}[X_j]$ for every pair of i and j with $1 \leq i < j \leq n$, then $\mathbf{Var}[X] = \sum_{i=1}^{n} \mathbf{Var}[X_i]$.

19.- (MU 3.15) (*) Suppose that we flip a fair coin n times to obtain n random bits. Consider all $m = \binom{n}{2}$ pairs of these bits in some order. Let Y_i be the exclusive or $\oplus$ of the i-th pair of bits, and let $Y = \sum_{i=1}^{m} Y_i$ the number of Y_i that equal 1.

(a) Show that each $Y_i = 0$ with prob $= 1/2$ (therefore, $Y_i = 1$ with probability also $1/2$)

(b) Show that Y_i are not mutually independent.

(c) Show that $\mathbf{E}[Y_i Y_j] = \mathbf{E}[Y_i]\,\mathbf{E}[Y_j]$.

(d) Find $\mathbf{Var}[Y]$.

(e) Use Chebyshev to bound $\mathbf{Pr}[|Y - \mathbf{E}[Y]| \geq n]$.

20.- (*) (MU 4.9) Suppose that we can obtain independent samples $X_1, X_2, \ldots X_n$ of a random variable X and that we want to use these samples to estimate $\mathbf{E}[X]$. Using t samples, we use $(\sum_{i=1}^{n} X_i/t)$ for our estimate of $\mathbf{E}[X]$. We want the estimate to be within $\epsilon\mathbf{E}[X]$ from the true value of $\mathbf{E}[X]$ with probability $\geq (1 - \delta)$. We develop an alternative approach that requires only having a bound on the $\mathbf{Var}[X]$. Let $r = \sqrt{\mathbf{Var}[X]}/\mathbf{E}[X]$.

(a) Show using Chebyshev that $O(r^2/\epsilon^2\delta)$ samples are sufficient to solve the problem.

(b) Suppose that we need only a weak estimate that is within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$, with probability at least $3/4$. Argue that. $O(r^2/\epsilon^2)$ samples are enough for this weak estimate.

(c) Show that, by taking the median of $O(\lg(1/\delta))$ weak estimates, we can obtain an estimate within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$ with probability at least $(1 - \delta)$. Conclude that we need only $(r^2 \lg(1/\delta)/\epsilon^2)$ samples.