This is the first graded homework exercise set.

Regulations:

- There will be a total of three special exercise sets during this semester.
- You are expected to solve them carefully and then write a nice and complete exposition of your solutions using \LaTeX. Please submit your solutions as a pdf or ps file to heinarss@inf.ethz.ch (filename: [nethz user name]_ghw1.pdf/ps).
- You are welcome to discuss the exercises with your colleagues, but we expect each of you to hand in your own, individual writeup.
- Your solutions will be graded. Each solution will account for 10% of your final grade for the course (so 30% of the grade in total).

Due date: Monday, October 13th, 2014 at 23:59

Exercise 1

Consider the coupon collector problem presented in the lecture (see Section 2.2 of the lecture notes for reference). Like there let $X$ be the number of coupons you need to buy until you have a complete set of $n$ coupons. In the lecture you saw that Markov’s inequality yields

$$\Pr[X \geq 2nH_n] \leq \frac{1}{2}$$

and using Chebyshev you can prove

$$\Pr[X \geq 2nH_n] \leq \frac{\pi^2}{6(\log n)^2}.$$ 

Now, for $\varepsilon > 0$, prove the following statement

$$\Pr[X \geq (1 + \varepsilon)nH_n] \leq \frac{1}{n^\varepsilon} + o\left(\frac{1}{n^\varepsilon}\right).$$

**Hint:** For Markov and Chebyshev you only needed the expectation and variance of $X$. For this bound you need to study the process with slightly more detail. If your proof is longer than one page you should consider finding a shorter proof.
Exercise 2

You're going on a long vacation and you have a quite large backpack of size $n$. Because you're packing in the last minute you just greedily pack your items into the bag one by one in the order in which they appear on the floor of your room. Since you are not a stylistic person, and tend to buy things randomly, your items have the nice property that their sizes are i.i.d. random variables $X_i$, $i \in \mathbb{N}^1$. The expected size of your items is $\mu > 0$ and their variance is $\sigma^2$.

Now let $I$ denote the total number of items you managed to pack into your backpack. Since you pack greedily you stop when you reach an item which does not fit into the backpack. Since you also happen to be a mathematician/computer scientist you would of course like to know if $I$ is concentrated around its mean as the size of your backpack tends to infinity. For a function $f := f(n)$, such that $\omega(1) \ll f \ll \sqrt{n}$ show that

$$\Pr\left[I = (1 \pm f(n)^{-1})\frac{n}{\mu}\right] = 1 - O\left(\frac{f(n)^2}{n}\right).$$

Exercise 3

You find an old computer tape which stores a large vector of numbers $(x_1, \ldots, x_m)$ where $0 \leq x_i \leq n - 1$. Naturally, you want to find out how many of the $x_i$ are unique (formally you want to compute the size of $D := \{j \mid \exists i: x_i = j\}$). The only computer which can still read the tape is the old Cray X-MP/28 supercomputer in the cellar of the CAB. But this machine does not have enough memory to store a list of all elements and you want to avoid scanning the tape multiple times (rewinding takes a lot of time and might damage the old tape).

The problem seems unsolvable and you are about to give up. Luckily you are taking a course on randomized algorithms and on one of the exercise sheets you find the following randomized algorithm:

Algorithm 1.1 Approximating the number of unique entries in a vector

| Input: | a vector $x = (x_1, \ldots, x_m)$ of elements from $\mathbb{Z}_n$ |
| Output: | an estimate for the number of unique entries in $x$ |
| $p :=$ | a prime (not much) larger than $10n$ |
| $a :=$ | $\text{u.a.r.} \{0, \ldots, p - 1\}$ |
| $b :=$ | $\text{u.a.r.} \{0, \ldots, p - 1\}$ |
| $c := \infty$ | |
| for $i = 1, \ldots, m$ do | |
| | if $c > a \cdot x_i + b \mod p$ then |
| | $c := a \cdot x_i + b \mod p$ |
| return | $p/(c + 1)$ |

\footnote{Yes, you own $\aleph_0$ many items.}
You are a bit sceptical. So before basing important decisions on the result of the algorithm you decide to analyse it: For integers \( r \) and \( j \) satisfying \( 0 \leq r \leq p \) and \( 1 \leq j \leq p \) let \( X_{j,r} \) denote the indicator random variable for the event that \((a \cdot j + b \mod p) < r\). Finally let \( Y_r = \sum_{j \in D} X_{j,r} \).

(a) Prove that \( X_{i,r} \) and \( X_{j,r} \) are independent whenever \( i \neq j \).

(b) Prove that \( \Pr [Y_r > 0] \leq \frac{|D|r}{p} \) and \( \Pr [Y_r = 0] \leq \frac{p}{|D|r} \).

(c) Let \( A \) denote the output of Algorithm 1.1. Prove that

\[
\Pr [A \geq 3|D|] \leq \frac{1}{3} \quad \text{and} \quad \Pr \left[A \leq \frac{|D|}{3}\right] \leq \frac{1}{2.9}.
\]

(d) You are not satisfied by these bounds. Design an improved algorithm which uses at most \( k \) times the memory of Algorithm 1.1 and whose output \( A' \) satisfies

\[
\Pr \left[\frac{|D|}{3} \leq A' \leq 3|D|\right] \geq 1 - \frac{C}{k},
\]

for some absolute constant \( C > 0 \).

*Hint:* Run \( k \) independent instances of the algorithm in parallel and consider the median result.

**Exercise 4** *(10 points)*

A balanced cut \( C = (X,Y) \) in a graph \( G = (V,E) \) is a partition of the vertex set into two parts \( X \) and \( Y \) such that \( |X| = |Y| = |V|/2 \) and \( X \cup Y = V \) (we assume that \( |V| \) is even). The size of \( C \) is the number of edges of \( G \) which have one endpoint in \( X \) and the other one in \( Y \). We say that a balanced cut is maximum if it has maximal size among all balanced cuts of \( G \). We denote the size of a maximum cut of \( G \) by \( c(G) \).

Let \( G_{n,1/2} \) be the random graph generated by including each edge with probability \( 1/2 \) independently. Show that for \( n \) large enough

\[
\Pr \left[c(G_{n,1/2}) \geq \frac{n^2}{8} + 6n^{3/2}\right] \leq e^{-n}.
\]