Exercise 1. Prove the other side of the Chernoff Bound. That is

$$Pr(X < (1-\delta)\mu) \leq \exp[-\delta^2\mu/2] .$$

solution : We use Markov inequality on the strictly decreasing function $h(x) = e^{tx}$ where $t < 0$. We have

$$\Pr(X < (1-\delta)\mu) = \Pr\left(h(X) > h((1-\delta)\mu)\right) \leq E[h(X)]/h\left((1-\delta)\mu\right)$$

The expectation of $h(X_i)$ is $(1-p_i) + p_i e^t = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)}$ so we get

$$\Pr\left(X < (1-\delta)\mu\right) \leq e^{\mu\left[(e^t-1)-t(1-\delta)\right]} .$$

The minimum exponent is obtained for $e^t - (1-\delta) = 0$, that is, $t = \log(1-\delta)$. With this t we obtain

$$\Pr\left(X < (1-\delta)\mu\right) \leq e^{-\mu[\delta+(1-\delta)\log(1-\delta)]} .$$

To conclude, we need to show that

$$(1-\delta)\log(1-\delta) \geq -\delta + \delta^2/2 .$$

One way to prove it is to use the fact that for $\delta \in (-1, 1)$

$$\log(1-\delta) = \sum_{i>0} -\delta^i/i ,$$

so

$$(1-\delta)\log(1-\delta) = -\delta + \sum_{i>1} \frac{\delta^i}{i(i-1)} .$$

Exercise 2. Consider a biased coin with the probability of getting heads being an unknown parameter p, which is known to be at least a, for some $a > 0$. A natural procedure for estimating the coin bias is to flip the coin n times, and estimate p as the fraction of times it lands on head. Denote this estimate by p' and suppose that for a given parameter we want to have

$$\Pr[|p-p'| > \epsilon p] < \delta .$$

How many flips do we need in function of a, δ and ϵ ? Using a calculator/computer, compute this number for $a = 0.1, \epsilon = 0.1$ and $\delta = 0.01$.

solution : Let X be the sum of n Bernoulli variable of parameter p. X is a Binomial law of parameter n and p. Moreover

$$\Pr[|p-p'| > \epsilon p] = \Pr[|X - np| > \epsilon np] .$$

Using Chernoff bound we obtain

$$\Pr[|X-\mu| > \epsilon\mu] \leq 2e^{\frac{-\epsilon^2\mu}{3}}$$

So we must choose n such that

$$\frac{-\epsilon^2 np}{3} \leq \log(\delta/2) ,$$

that is,

$$n \geq \frac{-3\log(\delta/2)}{a\epsilon^2} \geq \frac{-3\log(\delta/2)}{p\epsilon^2} .$$

With the given parameter we need to choose

$$n \geq 15995 .$$

Exercise 3. Suppose we are given a sequence of n distinct numbers $a_1, \ldots, a_n$ and we want to compute the median of the sequence. There is a deterministic linear time algorithm for this problem, but it uses $\Omega(n)$ memory which can be problematic for huge n.

The goal of this exercise is to analyze an algorithm that estimates the median using random sampling and whose required storage only depends on the quality of the estimate and not on n. The algorithm works as follows, select uniformly and independently at random k numbers of the sequence then output the median of those k numbers.

Define the rank of an element as its position in the sorted sequence. For instance the rank of the minimum, maximum and median element will be $1, n$ and $\lfloor n/2 \rfloor$ respectively.

Suppose that we want the sampling algorithm to return an element x whose rank is approximately $n/2$ with high probability. More precisely, we want

$$\Pr\left[\frac{n}{2}(1-\epsilon) \leq \text{rank}(x) \leq \frac{n}{2}(1+\epsilon)\right] \geq 1 - \delta\ .$$

What k do we need to get the desired confidence?

Hint: Treat each inequality independently and try to express the probability in such a way that we can apply the Chernoff bound.

solution : The problem being almost symmetrical we just need to compute k such that

$$\Pr\left[\text{rank}(x) > (1+\epsilon)\frac{n}{2}\right] \leq \delta/2\ .$$

Let introduce a function f such that $f(x)$ is equal to 1 if $\text{rank}(x) \leq (1+\epsilon)\frac{n}{2}$ and is equal to 0 otherwise. If we call x_i the $i-$th randomly selected number, the median of the x_i for $i \in \{1, \ldots, k\}$ will be greater than $(1+\epsilon)\frac{n}{2}$ if and only if

$$\sum_i f(x_i) < k/2\ .$$

But this sum is given by a Binomial random variable X with parameters k and $p = (1+\epsilon)/2$. Using Chernoff bound, we get

$$\Pr(X < k/2) = \Pr(X < (1-a)(1+\epsilon)k/2) \leq e^{-a^2\mu/2}\ ,$$

with $\mu = (1+\epsilon)k/2$ and $1 - a = 1/(1+\epsilon)$. So $a = \epsilon/(1+\epsilon)$ and

$$\Pr(X < k/2) \leq e^{-\epsilon^2(1+\epsilon)k/4(1+\epsilon)^2} \leq e^{-\frac{\epsilon^2 k}{4(1+\epsilon)}}\ .$$

To get the desired confidence, we need to choose k such that

$$e^{-\epsilon^2 k/4(1+\epsilon)} \leq \delta/2$$

that is,

$$k \geq \frac{-\log(\delta/2)4(1+\epsilon)}{\epsilon^2}\ .$$