Distributed Algorithms 2022/2023
(practical exercise)

Leader election

1 — Find the expected value for the random variable $X \sim Geo(p)$.

2 — Find the variance of the random variable $X \sim Geo(p)$.

3 — Let $p \in [0,1]$ and $n \geq k \geq 1$, $n, k \in \mathbb{N}$. For what value of the argument a function f takes the maximum value?

a. $f(p) = np(1-p)^{n-1}$,
b. $f(n) = np(1-p)^{n-1}$,
c. $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

4 — Prove that $(1+x)^r \geq 1 + rx$ for $x \geq -1, r \geq 1$.

5 — Prove that $(1+x)^r \leq 1 + rx$ for $x \geq -1, r \in (0,1)$.

6 — Prove that $1 + x \leq e^x$ for $x \in \mathbb{R}$.

7 — Prove that $\frac{x}{e^x} < \frac{1.5}{x^2}$ for $x > 0$.

8 — Let $f_i(n) = n\frac{1}{2^i}\left(1 - \frac{1}{2^i}\right)^{n-1}$. Prove that functions $f_i(2^{i-1})$, $f_{i+1}(2^{i-1})$ are decreasing and function $f_{i-1}(2^i)$ is increasing for $i \geq 2$.

9 — Present the definition of the Lambert W functions family and draw its real branches.

10 — Use Lambert W function to analytically determine real solutions to the equation

$$3^x = x^3 .$$

What is Lambert's W function called in Mathematica?

11 — Have a look at this paper and show that for $x \geq e$

$$\ln x - \ln\ln x < W_0(x) \leq \ln x - \frac{1}{2}\ln\ln x .$$

12 — Completion of the lecture proof. Check that if $K \geq 1, f > 1, u \geq 2$ and

$$3e(K+1)u^{\frac{-1}{2(K+1)}} \geq 1 - \frac{1}{f} \quad \text{then} \quad K \geq \frac{\ln u}{2W_0(\frac{3e}{2}\frac{f}{f-1}\ln u)} - 1 .$$

Data stream analysis: approximate counting

13 — For continuous and independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution given by the density function $f(x)$ and the cumulative distribution function $F(x)$ show that k-th order statistic $X_{k:n}$ has a distribution described by the density function

$$f_k(x) = \frac{F^{k-1}(x)\,[1-F(x)]^{n-k}\,f(x)}{B(k, n-k+1)}\,,$$

where $B(\alpha, \beta)$ denotes beta function. Hint: see this notes.

14 — For n independent random variables $U_1, U_2, \ldots, U_n$ with the uniform distribution: $U_i \sim \mathcal{U}(0,1)$, show that k-th order statistic has distribution $Beta(k, n-k+1)$ and an expected value equal to $k/(n+1)$.

Hint. In this and in the next exercise use different representations of the beta function: given by the definition and by factorial for arguments that are natural numbers.

15 — Let $U_{k:n}$ denote kth order statistic for n independent uniformly distributed random variables with distribution $\mathcal{U}(0,1)$. Show that for the estimator $\hat{n}_k = \frac{k-1}{U_{k:n}}$ and $k \geq 2$ we have $\mathbb{E}\left(\hat{n}_k\right) = n$ and that for for $k \geq 3$ we have

$$\mathbb{V}\mathrm{ar}\left(\hat{n}_k\right) = \frac{n(n-k+1)}{k-2}\,.$$

16 — (Markov's inequality) Let X denote the random variable that takes only non-negative values. Then for all $a > 0$

$$\mathbb{P}\left(X \geq a\right) \leq \frac{\mathbb{E}\left(X\right)}{a}\,.$$

17 — (Chebyshev's inequality) Let X denote a random variable with a finite expected value and a finite, non-zero variance. Show that for any $a > 0$ the following inequality holds:

$$\mathbb{P}\left(|X - \mathbb{E}\left(X\right)| < a\right) > 1 - \frac{\mathbb{V}\mathrm{ar}\left(X\right)}{a^2}\,.$$

Hint: use Markov's inequality.

18 — (Chernoff inequality for sum of Bernoulli trials) Let $X_1, \ldots, X_n$ be independent Bernoulli trials such that $\mathbb{P}\left(X_i = 1\right) = p_i$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}\left(X\right)$. Show that

a) for any $\delta > 0$

$$\mathbb{P}\left(X \geq (1+\delta)\mu\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}\,,$$

b) for any $0 < \rho \leq 1$

$$\mathbb{P}\left(X \leq (1-\rho)\mu\right) \leq \left(\frac{e^{-\rho}}{(1-\rho)^{(1-\rho)}}\right)^{\mu}\,.$$

Hint: see chapter 4.2. in this book.

19 — Using the notations and the inequalities obtained in the previous task, show that for any $0 < \delta < 1$

$$\mathbb{P}\left(|X - \mu| \geq \delta\mu\right) \leq 2e^{-\mu\delta^2/3}\,.$$

20 — Let S_n be the number of heads in n flips of a symmetrical coin. Show that

a) using Chebyshev's inequality we have

$$\mathbb{P}\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{4}{n} ,$$

b) using Chernoff's inequality from the previous task we have

$$\mathbb{P}\left(\left|S_n - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq 2e^{-n/24} .$$

21 — Consider the following algorithm, from which the idea of the HyperLogLog algorithm is derived.

Probabilistic Counter

```
Initialization: C ← 1
Upon event:
if random( ) <= 2^{-C} then        ▷ random returns a random number in a range [0, 1)
    C ← C + 1
end if
```

In other words, when an event occurs, we toss a coin C times, and if each time we get heads, we increment the C counter by one. Otherwise, we do nothing. Let C_n be the value stored in the counter C after observing n events. Show that $\mathbb{E}\left(2^{C_n}\right) = n+2$ and $\mathbb{V}\text{ar}\left(2^{C_n}\right) = \frac{1}{2}n(n+1)$. Based on C_n, define an unbiased estimator of n and calculate its variance.

Data stream analysis: approximate summation

22 — Recall the basics of the exponential distribution.

a) Recall the formula for density and distribution function. Derive the formula for expected value and variance.

b) Suppose you have a generator that returns numbers in the range $[0, 1)$ following a uniform distribution. Present a procedure that will transform the returned values into values that follow the exponential distribution with the parameter λ.
Hint: see pages 28 and 29 in this book.

c) Present and prove the theorem on which the procedure in point b) is based.

23 — Let $S_1, S_2, \ldots, S_n$ be a sequence of independent exponential random variables and $S_i \sim Exp(\lambda_i)$ for $\lambda_i > 0$. Let $\Lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. Show that the first order statistic $S_{\min} = \min\{S_1, S_2, \ldots, S_n\}$ has an exponential distribution with parameter Λ:

$$S_{\min} \sim Exp(\Lambda) .$$

24 — Let X and Y be independent random variables with density functions $f_X(x)$ and $f_Y(y)$, respectively. For $Z = X + Y$ show that

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx .$$

How does this task relate to the next task?
Hint: see convolution of probability distributions.

25 — Assume $\Lambda > 0, m \in \mathbb{N}_{>0}$ and let $S_{\min}^{(1)}, S_{\min}^{(2)}, \ldots, S_{\min}^{(m)}$ will be independent random variables with the same exponential distribution

$$S_{\min}^{(i)} \sim Exp(\Lambda) \,.$$

Show that the variable

$$G_m = \sum_{i=1}^{m} S_{\min}^{(i)}$$

has gamma distribution defined by density function:

$$g_m(x) = \Lambda \frac{(\Lambda x)^{m-1}}{\Gamma(m)} e^{-\Lambda x} \quad for \quad x > 0 \,.$$

Hint 1: use the previous exercise and induction.
Hint 2: $\Gamma(m) = (m-1)!$ for $m \in \mathbb{N}_{>0}$.

26 — Using the notations from the exercise 25 show that for $m \geq 2$ and the estimator defined as

$$\bar{\Lambda}_m = \frac{m-1}{\sum_{i=1}^{m} S_{\min}^{(i)}}$$

we have $\mathbb{E}\left(\bar{\Lambda}_m\right) = \Lambda$.

27 — Using the notations from the exercise 25 and 26 show that for $m \geq 3$ the standard error of the $\bar{\Lambda}_m$ estimator depends only on the m parameter and is expressed by the formula:

$$\mathbb{SE}\left(\bar{\Lambda}_m\right) = \frac{1}{\sqrt{m-2}} \,.$$

28 — Suggest an algorithm that can estimate the average value

$$Av = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_n}{n}$$

for all unique elements of the stream $\mathfrak{M}$. Determine the bias of the proposed estimator.

Hint: note that for $\lambda_1 = \lambda_2 = \ldots \lambda_n = 1$ we have $\mathbb{E}\left(\bar{\Lambda}_m\right) = n$.

Blockchain

29 — Show that for the notations from the task **23** we have:

$$\mathbb{P}\left(S_{min} = S_i\right) = \frac{\lambda_i}{\Lambda}\,.$$

30 — Let the continuous random variable X take values in the range $[0, \infty)$. We say that the distribution of X is memory-less if the following condition holds:

$$\left(\forall x_1, x_2 > 0\right)\left(\mathbb{P}\left(X > x_1 + x_2 | X > x_2\right) = \mathbb{P}\left(X > x_1\right)\right)\,.$$

Show that the exponential distribution

a) satisfies this definition,

b) is the only continuous distribution that satisfies this definition.

31 — We toss a coin until we obtain r tails. Assume that tails and heads appear with probabilities p and q, respectively. Derive the distribution of the random variable X describing the number of heads obtained. What is the name of this distribution? Derive the formula for the expected value and variance.

32 — Recall the definitions and basic properties of the Poisson distribution. Show that the Poisson distribution with parameter μ is the limiting distribution for the binomial distribution $Bin(n, p_n)$ if $\lim_{n\to\infty} np_n = \mu > 0$. You may refer to this book.

33 — (Coupon collector's problem) We have n urns into which we randomly (uniformly) throw balls. Let X be the number of balls thrown until there is at least one ball in each urn. Using the approximation of the number of balls in a given urn by the Poisson distribution, show that for large values of n, we have

$$Pr[X > n \ln n + cn] \approx 1 - e^{-e^{-c}},$$

and then determine the smallest value of c such that for large values of n, the value of X lies in the interval $[n \ln n - cn, n \ln n + cn]$ in 99% of cases. Hint: you may refer to this book.

Self-stabilization

34 — Present the self-stabilizing algorithm for obtaining the maximum independent set developed in laboratory task 11. Prove its correctness and convergence. Derive the tightest possible upper bound on the number of steps until reaching a legal configuration.

35 — (Conservative MM) Propose a modification to the Maximal Matching algorithm presented during the lecture, which will enable it to operate under the additional assumption that each process is either type A or B, and two processes of the same type cannot form a pair. Furthermore, you can assume that the type of a given process is predetermined and can never be changed. Justify the correctness and convergence of the algorithm.

36 — (Liberal MM) Propose a modification to the Maximal Matching algorithm presented during the lecture, which will enable it to operate under the additional assumption that the number of edges incident with a given vertex can be greater than one (so-called b-matching). Justify the correctness of the algorithm. You may rely, for example, on this paper.

Good luck,
J.L.