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

```
1: Initialization: C ← 1
Upon event:
2: if random( ) <= 2^{-C} then          ▷ random returns a random number in a range [0, 1)
3:     C ← C + 1
4: 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 $\mathbb{V}\mathrm{ar}\left(2^{C_n}\right) = \frac{1}{2}n(n+1)$. Based on C_n, define an unbiased estimator of n and calculate its variance.

Good luck,
J.L.