This is a reference version of an exam that was offered online via Gradescope. Out of necessity, some questions are formatted differently here than on the actual exam.

1. A *fugue* (pronounced “fyoog”) is a highly structured style of musical composition that was popular in the 17th and 18th century. A fugue begins with an initial melody, called the *subject*, that is repeated several times throughout the piece.

   Suppose we want to design an algorithm to detect the subject of a fugue. We will assume a *very* simple representation as an array $F[1..n]$ of integers, each representing a note in the fugue as the number of half-steps above or below middle C. (We are deliberately ignoring all other musical aspects of real-life fugues, like multiple voices, timing, rests, volume, and timbre.)

   (a) Describe an algorithm to find the length of the longest prefix of F that reappears later as a substring of F. The prefix and its later repetition must not overlap.

   (b) In many fugues, later occurrences of the subject are ***transposed***, meaning they are all shifted up or down by a common value. For example, the subject $(3,1,4,1,5,9,2)$ might be transposed transposed down two half-steps to $(1,-1,2,-1,3,7,0)$.

   Describe an algorithm to find the length of the longest prefix of F that reappears later, *possibly transposed*, as a substring of F. Again, the prefix and its later repetition must not overlap.

   For example, if the input array is

   $$\overbrace{3,1,4,1},5,9,2,6,5,\overbrace{3,1,4,1},-1,2,-1,3,7,0,1,4,2$$

   then your first algorithm should return 4, and your second algorithm should return 7.

2. A network of secret agents, numbered from 1 to n, have established the following randomized communication protocol. At a precise prearranged signal, each agent sends a message to exactly one of the *other* $n-1$ agents, chosen independently and uniformly at random.

   As a running example, if $n = 3$, then agents 1 and 2 might both send their messages to agent 3, while agent 3 sends their message to agent 1.

   (a) An agent is *bored* if no other agent sends them a message. (In the example scenario, agent 2 is bored.) Let $B(n)$ denote the expected number of bored agents; thus, $B(n)/n$ is the expected *fraction* of agents that are bored. ***Prove*** that $\lim_{n\to\infty} B(n)/n = 1/e$.

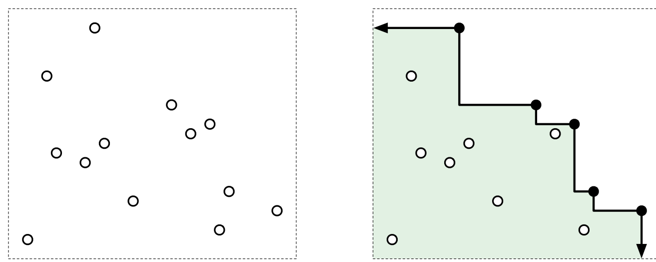
   (b) An agent is *swamped* if more than one other agent sends them a message. (In the example scenario, agent 3 is swamped.) Let $S(n)$ denote the expected number of swamped agents. What is $\lim_{n\to\infty} S(n)/n$?

   (c) Suppose each agent can *accept* at most one message. Thus, each swamped agent accepts one of the messages sent to them, chosen arbitrarily, and rejects the rest. (In the example scenario, exactly one message is rejected.) Let $R(n)$ denote the expected number of rejected messages. What is $\lim_{n\to\infty} R(n)/n$?

   [Hint: The World’s Most Useful Limit is useful.]

3. Suppose we are given a directed flow network $G = (V, E)$ where every edge has capacity 1, together with an integer k. Describe and analyze an algorithm to identify k edges in G such that after deleting those k edges, the value of the maximum (s, t)-flow in the remaining subgraph is as small as possible. *[Hint: First consider the case $k = 1$.]*

4. Let S be an arbitrary set of n points in the plane with distinct x- and y-coordinates. A point p in S is ***Pareto-optimal*** if no other point in S is both above and to the right of p. The ***staircase*** of S is the set of all points in the plane (not just in S) that have at least one point in S both above and to the right. All Pareto-optimal points lie on the boundary of the staircase.



A set of points in the plane and its staircase (shaded), with Pareto-optimal points in black.

   (a) Describe and analyze an algorithm that identifies the Pareto-optimal points in S in $O(n \log n)$ time.

   (b) Suppose each point in S is chosen independently and uniformly at random from the unit square $[0, 1] \times [0, 1]$. What is the ***exact*** expected number of Pareto-optimal points in S? *[Hint: What is the probability that the leftmost point in S is Pareto-optimal?]*

Some Useful Inequalities

Suppose X is the sum of random indicator variables $X_1, X_2, \ldots, X_n$.
For each index i, let $p_i = \Pr[X_i = 1] = \mathrm{E}[X_i]$, and let $\mu = \sum_i p_i = \mathrm{E}[X]$.

- **Markov's Inequality:**

$$\Pr[X \geq x] \leq \frac{\mu}{x} \qquad \text{for all } x > 0 \text{, and therefore...}$$

$$\Pr[X \geq (1+\delta)\mu] \leq \frac{1}{1+\delta} \qquad \text{for all } \delta > 0$$

- **Chebyshev's Inequality:** If the variables X_i are pairwise independent, then...

$$\Pr[(X-\mu)^2 \geq z] < \frac{\mu}{z} \qquad \text{for all } z > 0 \text{, and therefore...}$$

$$\Pr[X \geq (1+\delta)\mu] < \frac{1}{\delta^2 \mu} \qquad \text{for all } \delta > 0$$

$$\Pr[X \leq (1-\delta)\mu] < \frac{1}{\delta^2 \mu} \qquad \text{for all } \delta > 0$$

- **Higher Moment Inequalities:** If the variables X_i are $2k$-wise independent, then...

$$\Pr[(X-\mu)^{2k} \geq z] = O\left(\frac{\mu^k}{z}\right) \qquad \text{for all } z > 0 \text{, and therefore...}$$

$$\Pr[X \geq (1+\delta)\mu] = O\left(\frac{1}{\delta^{2k}\mu^k}\right) \qquad \text{for all } \delta > 0$$

$$\Pr[X \leq (1-\delta)\mu] = O\left(\frac{1}{\delta^{2k}\mu^k}\right) \qquad \text{for all } \delta > 0$$

- **Chernoff's Inequality:** If the variables X_i are fully independent, then...

$$\Pr[X \geq x] \leq e^{x-\mu}\left(\frac{\mu}{x}\right)^x \qquad \text{for all } x \geq \mu \text{, and therefore...}$$

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu/3} \qquad \text{for all } 0 < \delta < 1$$

$$\Pr[X \leq (1-\delta)\mu] \leq e^{-\delta^2\mu/2} \qquad \text{for all } 0 < \delta < 1$$

- **The World's Most Useful Inequality:** $1 + x \leq e^x$ for all x
- **The World's Most Useful Limit:** $\lim_{n\to\infty} \left(1 + \frac{\alpha}{n}\right)^n = e^\alpha$ for any real number α.

Hashing Properties

$\mathcal{H}$ is a set of functions from some universe $\mathcal{U}$ to $[m] = \{0, 1, 2, \ldots, m-1\}$.

- **Universal:** $\Pr_{h\in\mathcal{H}}[h(x) = h(y)] \leq \frac{1}{m}$ for all distinct items $x \neq y$
- **Near-universal:** $\Pr_{h\in\mathcal{H}}[h(x) = h(y)] \leq O\left(\frac{1}{m}\right)$ for all distinct items $x \neq y$
- **Strongly universal:** $\Pr_{h\in\mathcal{H}}[h(x) = i \text{ and } h(y) = j] = \frac{1}{m^2}$ for all distinct $x \neq y$ and all i and j
- **2-uniform:** Same as strongly universal.
- **Ideal Random:** Fiction.