Universitat Politècnica de Catalunya
Facultat d'Informàtica de Barcelona

Degree: Grau en Enginyeria Informàtica **Academic year:** 2025–2026 (Mid-term Exam)
Course: Randomized Algorithms (RA-MIRI) **Date:** October 30, 2025
Time: 2h

1. **(2.5 points)** We have an unbounded number of bins, numbered 1, 2, 3, ... and we have to put n balls into the bins. For the j-th ball, $1 \leq j \leq n$, we decide which bin it has to go to by flipping a fair coin until it shows heads. We assign that ball j to bin i if we needed to flip the coin i times to get the first "heads".

   Questions:

   (a) What is the probability that the j-th ball goes to bin i? Let ρ_j denote the index of the bin into which we have to put the j-th ball. What is the distribution followed by ρ_j? Are $\rho_1, \rho_2, \ldots, \rho_n$ identically distributed? Independent? Justify your answers.

   (b) Let X_i be the number of balls in bin i. What is the distribution followed by X_i? Are $X_1, X_2, \ldots,$ identically distributed? Independent? Justify your answers.

   (c) What is the expected number of balls in the i-th bin? And the variance?

   (d) Let R be the largest index of a bin that contains a ball. Give an expression as simple as possible for
   $$\mathbb{P}[R \leq r]\,.$$
   Show that, for any $\epsilon > 0$, $\mathbb{P}[R \leq (1+\epsilon)\lg n] \to 1$ with high probability.

   (e) Give an upper bound, as strong as possible, for the probability that X_i significantly deviates from $\mathbb{E}[X_i]$.

---

2. **(2.5 points)** Our studies indicate that network attacks occur at a rate of 1 in 1 million network flows. We are deploying a new Intrusion Detection System which

   (a) If a flow is an attack, the software correctly labels it as such with probability 0.995.

   (b) If a flow is not an attack, the software correctly labels it as non-attack with probability 0.999.

   We've just received an alert, the software has labeled a network flow as an attack. What is the probability p that the flow is actually an attack?

(a) Above 99% ($p > 0.99$)

(b) Above 95% and at most 99% ($0.95 < p \leq 0.99$)

(c) Above 50% but at most 95% ($0.5 < p \leq 0.95$)

(d) Above 1% but at most 50% ($0.01 < p \leq 0.5$)

(e) Below 1% ($p < 0.01$)

Justify your answer, note that you need not to compute p exactly.

---

3. **(2.5 points)** We have a randomized algorithm A that runs in linear time and produces the correct answer (yes/no) for an input of size n with probability $\geq 1 - 1/n$. In order to boost the algorithm's probability of being correct, we make t independent executions of A, and return the most frequent answer; we call this "amplified" algorithm A'.

   (a) What is the probability that we get a correct answer with A', as a function of t?

   (b) What is the cost of A' if we want it to return the correct answer with probability at least $1 - 1/n^2$?

   (c) Our colleague has designed a deterministic algorithm B with cost $\Theta(n^2)$. He claims that A' is not competitive against B if you require low probability of error. According to our colleague, A' has cost $\Omega(n^2)$ if we want the probability of being correct to be $\geq 1 - 1/n^c$, for large values of c, say for $c > c_0$. Which is the value c_0 such that achieving a probability of error $\leq n^{-c}$ and running time $\Theta(n \cdot t) = o(n^2)$ if feasible?

---

4. **(2.5 points)** In the model of balls & bin, assume that the n balls are put into the m bins by choosing the bins u.a.r. for each ball.

   (a) What is the probability that exactly k bins contain $\geq r$ balls (and thus all other bins contain $< r$ balls)? Give the exact answer.

   (b) What is the expected number of bins that contain $\geq r$ balls? Give the exact answer, then an approximation for the regime in which $\lambda = n/m = \Theta(1)$.
