Exercise session 8 – Prob. & Stat. 2 — Nov 21, 2023

Balls&bins

1. Czech (and Slovak) Birth number BN (rodné číslo) consists of six digits that encode the date of birth (& sex), plus four digits, let us assume these are assigned at random (discuss!). Occasionally, these four digits are used as a password.

Find/approximate minimum k such that in a group of k people there is more than 50 % chance that some two people have the same last four digits of their BN.

2. Suppose that n balls are thrown independently and uniformly at random into n bins.

(a) Find the conditional probability that bin 1 has one ball given that exactly one ball fell into the first three bins.

(b) Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.

(c) Write an expression for the probability that bin 1 receives more balls than bin 2. Hint: start with probability that bin 1 and bin 2 got the same number of balls.

(A formula with a sum is good enough, but you may also try to approximate it – using $1 - x \doteq e^{-x}$.)

3. Our analysis of Bucket sort in the last lecture assumed that $n = 2^k$ elements were chosen independently and uniformly at random from the range $[0, 2^\ell)$. Suppose instead that n elements are chosen independently from the range $[0, 2^\ell)$ according to a distribution with the property that any integer $x \in [0, 2^\ell)$ is chosen with probability at most $a/2^\ell$ for some fixed constant $a > 0$. Show that, under these conditions, Bucket sort still runs in a linear expected time.

4. The following problem models a simple distributed system wherein agents contend for resources but "back off" in the face of contention. Balls represent agents, and bins represent resources. The system evolves over rounds. Every round, balls are thrown independently and uniformly at random into n bins. Any ball that lands in a bin by itself is served and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we finish when every ball is served.

(a) If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?

(b) Suppose that every round the number of balls served was exactly the expected number of balls to be served. Show that all n balls would be served in $O(\log\log n)$ rounds. (Hint: If x_j is the expected number of balls left after j rounds, show and use that $x_{j+1} \leq x_j^2/n$. Estimate x_1 directly.)

(Hint2: $(1-t)^n \geq 1 - tn$ for $t \leq 1$.)

5. Consider the following modification of the balls-and-bins process. Let $n = 2^k$ (so $k = \log_2 n$). We k times choose a uniformly random bin, then put one ball into this bin and the next $\frac{n}{k} - 1$ consecutive bins (wrapping back to the beginning, if we go over the last bin). Assume k divides n, for simplicity.

So, again, we distribute n balls, but we do it in $\log_2 n$ rounds, each involving just one random choice.

Argue that the maximum load in this case is only $O(\log\log n/\log\log\log n)$ with probability that approaches 1 as $n \to \infty$.

6. Consider the probability that every bin receives exactly one ball when n balls are thrown randomly into n bins.

(a) Give an upper bound on this probability using the Poisson approximation. (Warning: we did not cover this in class yet!)

(b) Determine the exact probability of this event.

(c) Show that these two probabilities differ by a multiplicative factor that equals the probability that a Poisson random variable with parameter n takes on the value n. Explain why this is implied by a theorem from the lecture.