Probability and Computation: Problem sheet 1 Solutions

Submit by 2pm Friday 31st Jan via moodle or at student reception

Question 1. *Suppose that there are n coupons to be collected from the cereal box. Every morning you open your cereal box and get only one coupon. Each coupon appears with the same probability in the box, i.e. $1/n$. Let T the number of cereal boxes you need to open to collect all the coupons.*

(i) Compute $\mathbf{E}[\,T\,]$

(ii) Compute $\mathbf{Var}\,[\,T\,]$

Solution: *Let T_i be the waiting time between the $(i-1)$-th coupon and the i-coupon. If we have $i-1$ coupons then the probability we obtain a new coupon is $\frac{n-i+1}{n}$. We conclude that T_i has geometric distribution with mean $n/(n-i+1)$. Notice that T_i are independent random variables. Now compute*

$$\mathbf{E}[\,T\,] = \mathbf{E}[\,T_1\,] + \ldots + \mathbf{E}[\,T_n\,]\,,$$

and

$$\mathbf{Var}\,[\,T\,] = \mathbf{Var}\,[\,T_1\,] + \ldots + \mathbf{Var}\,[\,T_n\,]\,.$$

Question 2. *Consider the Balls-into-Bins setting where we have n balls and n bins. We assume that n is large enough. We are going to prove that the maximum load is whp. at least $c\log n/\log\log n$ for some $c > 0$.*

(i) Let $Y_j(k)$ be the random variable that indicates that bin j receives at least k balls. Prove that for any $k \le n$, it holds that

$$\mathbf{P}[\,Y_j(k) = 1\,] \ge \frac{e^{-2}}{k^k}.$$

(ii) Show it exist $c > 0$ such that for $k^* = \left\lfloor \frac{c\log n}{\log\log n} \right\rfloor$, we have

$$\mathbf{P}[\,Y_j(k^*) = 1\,] \ge n^{-1/3},$$

(iii) Argue that for any $k \le n$, and any bins i, j we have

$$\mathbf{P}[\,Y_j(k)Y_i(k) = 1\,] \le \mathbf{P}[\,Y_i(k) = 1\,]\,\mathbf{P}[\,Y_j(k) = 1\,]$$

(iv) Let $Y = \sum_{j=1}^{n} Y_j(k^*)$. Check that $\mathbf{E}[\,Y\,] \ge n^{2/3}$ and that $\mathbf{Var}\,[\,Y\,] \le n$

(v) Conclude that $\mathbf{P}[\,Y = 0\,] \le n^{-1/3}$

Solution:

(i) The probability bin j receives exactly k balls is

$$\binom{n}{k}\frac{1}{n^k}\left(1-\frac{1}{n}\right)^{n-k} \ge \left(\frac{n}{k}\right)^k \frac{1}{n^k} e^{-2(n-k)/n} = \frac{1}{k^k} e^{-2(n-k)/n} \ge \frac{e^{-2}}{k^k} \tag{1}$$

(ii) Note that

$$(k^*)^{k^*} = \exp(k^* \log k^*) \leq \exp\left(\frac{c\log n}{\log\log n} \log\left\{\frac{c\log n}{\log n \log n}\right\}\right)$$

the term inside the exponential is

$$\frac{c\log n}{\log\log n} \times (\log c + \log\log n - \log\log\log n)$$

Here we use that n is large enough. In such a case, $\log\log n > \frac{1}{2}\log\log\log n$, and $\frac{3}{2}\log\log n > \log c$. Then

$$\frac{c\log n}{\log\log n} \times (\log c + \log\log n - \log\log\log n) \leq \frac{c\log n}{\log\log n} \times (2\log\log n) = 2c\log n.$$

Then

$$\mathbf{P}[\, Y_j(k^*) = 1\,] \geq \frac{e^{-2}}{(k^*)^{k^*}} \geq e^{-2}\exp(-2c\log n) = \frac{e^{-2}}{n^{2c}}.$$

Now, choose c small, e.g. $c = 1/10$, to get the result.

(iii) If $Y_j(k)$ it means that at least k balls are in bin j, meaning that there are less balls for the other bins.

(iv) The variable Y counts the number of bins with more than k^* balls.

$$\mathbf{E}[\, Y\,] = n\mathbf{E}[\, Y_1(k^*)\,] \geq n \times n^{-1/3} = n^{2/3}.$$

Also

$$\begin{aligned}
\mathbf{Var}[\, Y\,] = \mathbf{E}[\, Y^2\,] - \mathbf{E}[\, Y\,]^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{E}[\, Y_j(k^*)Y_i(k^*)\,] - \mathbf{E}[\, Y_j(k^*)\,]\,\mathbf{E}[\, Y_i(k^*)\,] \\
&\leq \mathbf{E}[\, Y\,] + \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{1}_{\{i=j\}}\left(\mathbf{E}[\, Y_j(k^*)Y_i(k^*)\,] - \mathbf{E}[\, Y_j(k^*)\,]\,\mathbf{E}[\, Y_i(k^*)\,]\right) \\
&\overset{(iii)}{\leq} \mathbf{E}[\, Y\,] \leq n
\end{aligned}$$

(v)

$$\mathbf{P}[\, Y = 0\,] = \mathbf{P}[\, |Y - \mathbf{E}[\, Y\,]| = \mathbf{E}[\, Y\,]\,] \leq \mathbf{P}[\, |Y - \mathbf{E}[\, Y\,]| \geq \mathbf{E}[\, Y\,]\,] \leq \frac{\mathbf{Var}[\, Y\,]}{\mathbf{E}[\, Y\,]^2} \leq \frac{n}{n^{4/3}} = n^{-1/3}.$$

Question 3. *Consider a $0,1$ matrix A of n rows and columns. A row shift consists in selecting a row and moving all its elements one step to the right, except the last one that moves to the first column.*

Example: Consider the following matrix A

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

then a shift in the second row gives

$$A' = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

We want to apply a sequence of row shifts that minimises the maximum column sum.

For example, the columns sum in A are $(4,2,1,3)$, and in A' are $(3,2,2,3)$. So A has maximum column sum equals to 4, and A' has maximum column sum equals to 3. We can verify that 3 is the optimal solution.

At this point you can decide to stop reading and to think in a randomised algorithm for this problem, and we actually encourage you to do so...

If you just want to enjoy the show, we propose the following algorithm: For each row i, choose independently $X_i \in \{0, 1, 2, \ldots, n-1\}$ uniformly at random and apply X_i shifts to row i.

(i) Let Q be the total number of 1's in the matrix. What is the best possible solution you can get for this problem?

(ii) After the application of our randomised algorithm, what is the expected sum of the first column?

(iii) Study the upper tail of the sum of the first column. Consider two cases: $Q \geq n \log n$ and $Q < n \log n$.

(iv) What can you say about the behaviour of this randomised algorithm? Should we use this algorithm or should we try to find an optimal deterministic algorithm?

Solution:

Solution given in problem class.

Hint (Collected hints for the exercises)**.**

Q1: Find a way of writing T as a sum of independent random variables

Q2(i): Hint 1: Compute the probability of getting exactly k balls in the bin.
Hint 2: Use that $1 - x \geq e^{-2x}$ for $x \in [0, 1/2]$ and that $\binom{n}{k} \geq (n/k)^k$.

Q3(iii): Use a Chernoff Bound to get the upper tail probablity of one column and then use the Union Bound.