TD n°06

15 mars 2018

Exercice 1 - Pour les plus rapides

We are given a biased coin and want to know how biased it is with probability at least $1-\delta$.

1.1 First, we want to determine that the bias is at least ε. How many throws do we need ?

1.2 Now we are given a second biased coin without the bound on the bias. We want to decide how large the bias is with probability at least $1-\delta$. Let p be the probability that the coin turns up head and $\hat{p}_n$ be an estimation on p obtained after n throws.
Find a value t (as a function of n and δ) for which the following relation holds :

$$\mathbf{P}\left\{p \in [\hat{p}_n - t, \hat{p}_n + t]\right\} \geq 1-\delta \ .$$

1.3 The interval considered above is valid for a fixed n. Let us slightly modify the procedure by decomposing it in several successive steps. We perform n_1 throws, then n_2 throws and so on, having a sequence $(n_j)_{j\in\mathbb{N}}$ (how to choose such a sequence is a concern of the next question). Define

$$I_j = \left[\hat{p}_{n_j} - \sqrt{\frac{\ln\left(\frac{2}{\delta/2^j}\right)}{2n_j}}, \hat{p}_{n_j} + \sqrt{\frac{\ln\left(\frac{2}{\delta/2^j}\right)}{2n_j}}\right] \ .$$

Give a lower-bound of $\mathbf{P}\left\{p \in I_j\right\}$ and on $\mathbf{P}\left\{\forall j \in \{1,2,\ldots\}\ p \in I_j\right\}$.

1.4 Explain how you would choose the sequence $(n_j)_{j\in\mathbb{N}}$ and how the procedure terminates.

Exercice 2 - Booster un algorithme randomisé générique

Suppose you are given a randomized polynomial-time algorithm $\mathcal{A}$ for deciding whether $x \in \{0,1\}^*$ is in the language L or not. Suppose it has the following property. If $x \in L$, then $\mathbf{P}\left\{\mathcal{A}(x) = 0\right\} \leq 1/4$ and if $x \notin L$, then $\mathbf{P}\left\{\mathcal{A}(x) = 1\right\} \leq 1/3$. Note that the probability here is taken over the randomness used by the algorithm $\mathcal{A}$ and *not* over the input x. Construct a randomized polynomial-time algorithm $\mathcal{B}$ that is allowed to make independent calls to $\mathcal{A}$ such that for all inputs $x \in \{0,1\}^*$, we have $\mathbf{P}\left\{\mathcal{B}(x) = \mathbf{1}_{x\in L}\right\} \geq 1 - 2^{-|x|}$. Here $\mathbf{1}_{x\in L} = 1$ if $x \in L$ and 0 otherwise, and $|x|$ denotes the length of the bitstring x.

Exercice 3 - Collectionneur de vignettes

Recall the coupon collector problem. Let X be the number of boxes that are bought before having at least one of each coupon. Show that

$$\mathbf{P}\left\{X \geq n \ln n + cn\right\} \leq e^{-c}.$$

In class we proved a similar bound using Chebychev's inequality. Here you are asked to prove this better bound in an elementary way.

Exercice 4 - Tri par seaux

Suppose that we have a set of $n = 2^m$ elements to be sorted and that each element is an integer chosen independently and uniformly at random from the range $[0, 2^k)$, where $k \geq m$ and k is assumed to be known. Bucket sort works in two stages. In the first stage (pre-sorting), we place the elements into n buckets according to some rules. In the second stage, we call a simple sorting algorithm (say, insertion sort with quadratic complexity) within each bucket. Finally, we concatenate the sorted lists from each bucket. For the algorithm to be correct, the pre-sorting must be done in such a way that all the elements of the bucket i-th are inferior to all the elements of the bucket j-th for $i < j$.

4.1 Give a way for pre-sorting (stage 1) that satisfies the stated conditions. We want that for any given element x, the decision which bucket x goes to can be made in constant time (assume, arithmetic operations take constant time).

4.2 Let X_i be a random variable that counts the number of elements in the i-th bin after pre-sorting. Which distribution do the X_i's follow ?

4.3 Show that the expected complexity of Bucket sort is $\mathcal{O}(n)$.

Exercice 5 - Approximation de Poisson

Consider the *Balls and Bins* model once again : we randomly put m balls into n bins. The problem is that random variables X_i representing the number of balls in the i-th bin are not independent. We would like to approximate the *Balls and Bins* model with the Poisson distribution. Here, $Y_1, \dots, Y_n$ are independent random variables each following Poisson distribution with parameter (i.e. expected value) $\mu = m/n$ (Y_i can be viewed as a simplified version of X_i).

5.1 Show that $Y = \sum_{i=1}^{n} Y_i$ follows Poisson distribution and determine its parameter.

5.2 Show that the distribution of $(Y_1, \dots, Y_n)$ conditioned on $Y = m$ is the same as the distribution of $(X_1, \dots X_n)$.

Note : One can, in fact, obtain a slightly more general result. If $(X_1, \dots X_n)$ represents the load (charge) of n bins after throwing k balls at random, and Y_i are n independent random variables that follow Poisson distribution with parameter m/n, then the distribution of $(Y_1, \dots, Y_n)$ conditioned on $Y = k$ is the same as the distribution of $(X_1, \dots X_n)$ independent of value m.

5.3 Let f be a function of n variables that takes values in $\mathbb{R}_+ \cup \{0\}$. Show that

$$\mathbf{E}\left\{f(X_1, \dots, X_n)\right\} \leq e\sqrt{m}\mathbf{E}\left\{f(Y_1, \dots, Y_n)\right\} .$$

You may use the fact that $m! < e\sqrt{m}\left(\frac{m}{e}\right)^m$.

5.4 Call the *Poisson case* the set of events that occur when the number of balls in the bins are taken to be independent Poisson random variables with mean m/n. Call the *Balls and Bins case*, the set of events when m balls are thrown into n bins independently at random. Which function f would you apply to the above result to conclude :

Any event that takes place with probability p in the Poisson case takes place with probability at most $pe\sqrt{m}$ in the Balls and Bins case.

5.5 Re-establish the lower-bound on the maximal load in case $m = n$ using Poisson approximation. More precisely, show that if n balls are thrown independently into n bins, the maximal load will be at least $(\ln n)/(\ln \ln n)$ with probability at least $1 - 1/n$ for sufficiently large n.