TD n°05

8 mars 2018

Exercice 1 - Borne de Chernoff pour des variables aléatoires continues

Let X be an arbitrary random variable with $0 \leq X \leq 1$ and $\mathbf{E}\{X\} = p$. Consider the random variable $Y \in \{0, 1\}$ with $\mathbf{P}\{Y = 1\} = p$.

1.1 Show that for any $\lambda > 0$, $\mathbf{E}\left\{e^{\lambda X}\right\} \leq \mathbf{E}\left\{e^{\lambda Y}\right\}$.

1.2 Using this fact, show that the Chernoff bound we saw in class still holds if we replace the condition $X_i \in \{0, 1\}$ by $X_i \in [0, 1]$.

Exercice 2 - Sampling in a Rectangle

Let $P \subset \mathbb{Z}^2$ of size n. Our objective is to be able to quickly answer queries of the form "what is the fraction of points in P that are in the rectangle $r = [a_1, b_1] \times [a_2, b_2]$?" We write $r[P] = \frac{|P \cap r|}{n}$ for this fraction. We consider a simple data structure to approximate $r[P]$ efficiently for any query r. The data structure is just a random subset $S \subset P$ of size m. On query r, the estimate for $r[P]$ we output is $\frac{|S \cap r|}{m}$. The structure S defines an ε-approximation if for all queries r, we have $|r[P] - \frac{|S \cap r|}{m}| \leq \varepsilon$. What m should we take to obtain an ε-approximation with probability $1 - \delta$?

Exercice 3 - Pour les plus rapides

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

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

3.2 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 :

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

3.3 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 $\mathbf{P}\{p \in I_j\}$ and on $\mathbf{P}\{\forall j \in \{1, 2, \ldots\} \ p \in I_j\}$.

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

Exercice 4 - 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 $\mathbf{P}\{\mathcal{A}(x) = 0\} \leq 1/4$ and if $x \notin L$, then $\mathbf{P}\{\mathcal{A}(x) = 1\} \leq 1/3$. 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 $\mathbf{P}\{\mathcal{B}(x) = \mathbf{1}_{x \in L}\} \geq 1 - 2^{-|x|}$. Here $\mathbf{1}_{x \in L} = 1$ if $x \in L$ and 0 otherwise, and $|x|$ denotes the length of the bitstring x.

Exercice 5 - 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

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

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