ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Foundations of Data Science — Assignment date: Wednesday, November 17th, 2021, 10:15
Fall 2021 — Due date: Wednesday, November 17th, 2021, 12:00

Midterm Exam – CM 1 2

This exam is closed book, closed notes. You are allowed to bring one A4 sheet (both sides of it) of hand-written and not photocopied notes ("cheat sheet"). No electronic devices of any kind are allowed. There are four problems. Choose the ones you find easiest and collect as many points as possible. We do not necessarily expect you to finish all of them. Good luck!

Name: ______________________________

Problem 1	/ 10
Problem 2	/ 10
Problem 3	/ 10
Problem 4	/ 10
Total	/40

Problem 1 (Even Moments of Subgaussian RV). [10pts]

(i) [5pts] Let Z be a non-negative random variable. Show that

$$\mathbb{E}[Z] = \int_0^\infty \mathrm{Prob}(Z > z)dz. \tag{1}$$

(ii) [5pts] Let X be a σ^2-subgaussian random variable. Show that for even integers $k = 2m$,

$$\mathbb{E}[X^k] \le C(k)\sigma^k, \tag{2}$$

and find the expression for $C(k)$. (HINT: Use the formulation of the mean above. HINT: $\int_0^\infty x^{2m-1}e^{-x^2/2}dx = 2^{m-1}(m-1)!$)

(space for problem 1)

(space for problem 1)

(space for problem 1)

Problem 2 (Generating fair coin flips from rolling the dice)**.** [10pts] Suppose $X_1, X_2, \ldots$ are the outcomes of rolling a possibly loaded die multiple times. The outcomes are assumed to be iid. Let $\mathbb{P}(X_i = m) = p_m$, for $m = 1, 2, \ldots, 6$, with p_m unknown (but non-negative and summing to one, clearly). By processing this sequence we would like to obtain a sequence $Z_1, Z_2, \ldots$ of *fair* coin flips.

Consider the following method: We process the X sequence in successive pairs, (X_1X_2), (X_3X_4), (X_5X_6), mapping $(3, 4)$ to 0, $(4, 3)$ to 1, and all the other outcomes to the empty string λ. After processing X_1, X_2, we will obtain either nothing, or a bit Z_1.

(a) [3pts] Show that, if a bit is obtained, it is fair, i.e., $\mathbb{P}(Z_1 = 0|Z_1 \neq \lambda) = \mathbb{P}(Z_1 = 1|Z_1 \neq \lambda) = 1/2$.

In general we can process the X sequence in successive n-tuples via a function $f : \{1, 2, 3, 4, 5, 6\}^n \to \{0, 1\}^*$ where $\{0, 1\}^*$ denotes the set of all finite length binary sequences (including the empty string λ). [The case in (a) is the function where $f(3, 4) = 0$, $f(4, 3) = 1$, and $f(j, m) = \lambda$ for all other choices of j and m.] The function f is chosen such that $(Z_1, \ldots, Z_K) = f(X_1, \ldots, X_n)$ are i.i.d., and fair (here K may depend on $(X_1, \ldots, X_n)$).

(b) [3pts] Letting $H(X)$ denote the entropy of the (unknown) distribution $(p_1, p_2, \ldots, p_6)$, prove the following chain of (in)equalities.

$$\begin{aligned} nH(X) &= H(X_1, \ldots, X_n) \\ &\geq H(Z_1, \ldots, Z_K, K) \\ &= H(K) + H(Z_1 \ldots, Z_K|K) \\ &= H(K) + \mathbb{E}[K] \\ &\geq \mathbb{E}[K]. \end{aligned}$$

Consequently, on the average no more than $nH(X)$ fair bits can be obtained from $(X_1, \ldots, X_n)$.

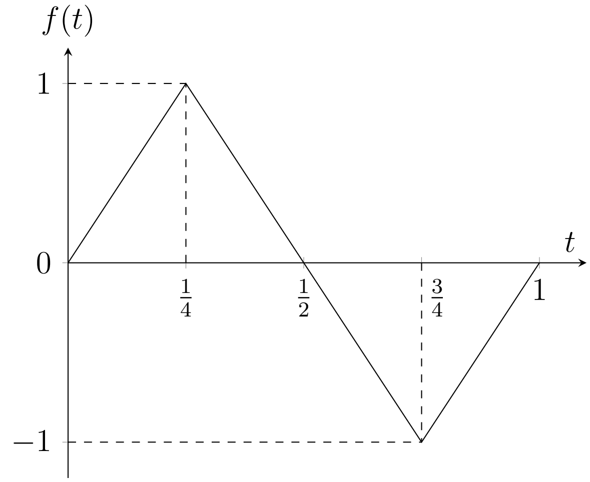
(c) [4pts] Describe how you would find a good f (with high $\mathbb{E}[K]$) for $n = 4$ which would work for any distribution $(p_1, p_2, ..., p_6)$.

(space for problem 2)

(space for problem 2)

(space for problem 2)

Problem 3 (Haar Wavelet)**.** [10pts] Consider the following function $f(t)$.



Let $\psi(t)$ be the Haar wavelet. (1 for $t \in [0, 1/2]$ and -1 for $t \in [1/2, 1]$).
As in the class we define $\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n)$, for $m, n \in \mathbb{Z}$.

(i) [5pts] Note that $f(t) = \sum_{m,n} a_{m,n}\psi_{m,n}(t)$. Find the scales $m \in \mathbb{Z}$ such that $\forall n, a_{m,n} = 0$.

(ii) [5pts] Let $f_{m^*}(t)$ be the projection of $f(t)$ to the space spanned by $\{\psi_{m,n} : m, n \in \mathbb{Z},\ m \geq m^*\}$ w.r.t. the standard L_2 norm. Find

$$\max_{t\in\mathbb{R}} |f_{m^*}(t) - f(t)|$$

as a function of $m^* \in \mathbb{Z}$.

[Hint : Try m^* equals to 0 and sketch $f_0(t)$.]

(space for problem 3)

(space for problem 3)

(space for problem 3)

Problem 4 (UCB With Geometric Intervals)**.** [10pts] Consider the following slight variant of the UCB algorithm. We have K arms. As in the lecture notes, assume that each of these K arms corresponds to a random variable which is 1-subgaussian. For the first K steps we sample each of these arms once. After these K first steps we have an interval of length 1, then an interval of length 2, then one of length 4, and so on. At the beginning of each such interval we choose the arm in the same manner as the UCB algorithm. More precisely, if t marks the beginning of a new interval then

$$A_t = \text{argmax}_k \hat{\mu}_k(t-1) + \sqrt{\frac{2\ln f(t)}{T_k(t-1)}},$$

where $f(g) = 1 + t\ln^2(t)$ as for the case we discussed in the course and where $T_k(t-1)$ denotes the number of times we have chosen arm k in the last $t-1$ steps. But unlike the standard UCB algorithm, for all other steps in this interval we keep the same arm. Why might we be interested in such an algorithm? One motivation is complexity. Computing which arm is best takes some effort. In this way we only have to compute the best arm a logarithmic (in the time horizon) number of times.

Recall that in the analysis of the original algorithm the key to the analysis was to find a good upper bound on $T_k(n)$ for $k > 1$, assuming that arm 1 is the optimum arm. In turn, we upper bounded the probability that we choose arm k at a particular point in time t by the probability that arm 1 had an empirical mean at least an ϵ below its true mean μ_1 and that the empirical mean of arm k was above $\mu_1 - \epsilon$. In formulae we had

$$T_k(n) = \sum_{t=1}^{n} \mathbb{1}_{\{A_t=k\}} \leq \sum_{t=1}^{n} \mathbb{1}_{\left\{\hat{\mu}_1(t-1)+\sqrt{\frac{2\ln f(t)}{T_1(t-1)}}\leq \mu_1-\epsilon\right\}} + \sum_{t=1}^{n} \mathbb{1}_{\left\{\hat{\mu}_k(t-1)+\sqrt{\frac{2\ln f(t)}{T_k(t-1)}}\geq \mu_1-\epsilon \,\wedge\, A_t=k\right\}} \tag{3}$$

Let us proceed in the same fashion. Let $n = K + 2^L - 1$. In words, we are at the end of the L-th interval, where $L \in \mathbb{N}$.

(i) [5pts] What is the expression equivalent to (3) for our case?

(ii) [5pts] Look at the first of the two terms on the right of (3) in your equivalent expression. Derive a suitable upper bound for this first term. If you do not have time for the whole derivation just write down the first few steps. These are the most crucial ones.

(space for problem 4)

(space for problem 4)

(space for problem 4)