Homework Exercises for Part I of the Course

Deadline: 8 May

3-1 The definitions of ε-nets and ε-approximations can be enriched through addition of weights. Let (X, R) be a finite range space and let μ be a probability distribution on X (i.e., for every $x \in X$ we have $0 \le \mu(x) \le 1$ and $\sum_{x \in X} \mu(x) = 1$); for a range $r \in R$, $\mu(r)$ is defined as $\sum_{p \in r} \mu(p)$. Here, an ε-approximation is a *multi-set* $A \subset X$ such that for every $r \in R$

$$|\frac{|A \cap r|}{|A|} - \frac{\mu(r)}{\mu(X)}| \le \varepsilon.$$

Note that in a multi-set an element may appear more than once and the multiple elements are preserved in intersection (i.e., $\{2,2\} \cap \{2\} = \{2,2\}$). Given the above new definition of ε-approximation, prove the following.

A. Let $X_1, \ldots, X_t$ be disjoint subsets of X and A_i be an ε-approximation for $(X_i, R_{|X_i})$. Show that $A_1 \cup A_2 \ldots \cup A_t$ is an ε-approximation for $(Y, R_{|Y})$ in which $Y = \cup_{i=1}^{t} X_i$.

B. Use the original ε-approximation theorem to prove the existence of an ε-approximation of size $O(\varepsilon^{-2} \log \varepsilon^{-1})$ with respect to any probability distribution (hint: duplicate points to simulate weights).

3-2 Consider the following problem which is inspired by the convex hull computation in three dimensions.

Let P be a set of n points in $\mathbb{R}^2$ and assume there are h disjoint but unknown triangles that cover all the points in P such that each triangle covers at least one point. The problem that you need to solve is to discover all the triangles using the following operations:

(op. find) You select a subset $S \subset P$ and by this operation you obtain all the triangles that cover at least one point of S. The running time of this operation is $O(|P| \log |S|)$.

(op. del) You select one point $p \in P$ and permanently delete it. The running time of this operation is $O(1)$.

(op. organize) You select a set T of triangles and you find all the points that are covered by T. The running time of this operation is $O(|P| \log |T|)$.

Assuming that you have access all the operations defined above, answer the following questions.

A. Show that (op. find) easily results in an $O(n \log n)$-time algorithm for the problem.

B. Assume you know the value of h and furthermore, all the triangles cover asymptotically the same number of points. Describe an $O(n \log h)$-time algorithm for the problem.

C. Answer part B. but assuming that you do not know the value of h. Describe an $O(n \log h)$-time algorithm for the problem (hint: you might want to search for the value of h starting from small cases).

D. (hard) Let t_i be the number of points covered by the ith triangle. Describe an algorithm for the problem with running time of $O\left(\sum_{i=1}^{n} t_i \log(n/t_i)\right)$.

E. Define $\text{Entropy}(P) := \sum_{i=1}^{n} t_i \log(n/t_i)$. Show that $\text{Entropy}(P) = O(n \log h)$. Build an instance of the problem with $\text{Entropy}(P) = o(n \log h)$.

3-3 Consider a set P of n points in the plane. Since every three points in P create a triangle, there are $\binom{n}{3}$ triangles created by the points of P. The simplicial depth of a point q in the plane is the number of such triangles that contain q. Describe a data structure that can approximate the simplicial depth of any query point q with an additive error of εn^3 for any parameter $0 < \varepsilon$. The space and the query time of your data structure should be independent of n.

3-4 Consider two P and Q in the plane each containing an even number of points. A ham-sandwich-cut is a line that contains $|P|/2$ points of P and $|Q|/2$ points of Q on each side. Such ham-sandwich-cuts are known to exist in the plane. Using this, prove that for any planar point set of size n, there exists a weak $\frac{5}{8}$-net of size two (hint: you can start by dividing the point set into two sets of size $n/4$ and $3n/4$ using a vertical line). For simplicity, assume n is a multiple of 8.