CSL 866, Problem Sheet 1

1. Consider the trivial algorithm for selecting the minimum element among n elements. We scan the elements given in an array $A[]$ and keep track of the minimum elements among the elements scanned so far - call it the *temprary* minimum. At the i-th step, we either upadate the minimum or we retain it. If the elements are presented to us in a random order, what is the expected number of times that we update the temporary minimum.

2. Complete the analysis of the sampling based point location data structure discussed in class where we want the subproblem size to be decreasing by a constant factor.
Can you obtain a better bound on the space by choosing a larger sample and combining with the expected bound on sum of subproblem sizes ? Choosing a larger sample reduces the number of levels in the data structure.

3. **Crossing number** A planar graph $G = (V, E)$ satisfies the Euler's relation

$$|E| - |V| + |F| = 2 \text{ where F is number of faces of a planar embedding}$$

(i) If the graph is not planar then any embedding will have edge crossings - denote this by $\mathcal{C}$. Verify using simple arguments that $\mathcal{C} \geq |E| - 3|V| + 6$.
Note that a traingulated graph has the maximum number of edges.
(ii) Obtain a superior bound on $\mathcal{C}$ using the probabilistic method, by sampling vertices with probability p and optimising p to get the best bound.

4. Consider the following simple sorting algorithm for point uniformly distributed in the interval $(0, 1)$. Divide the interval $[0, 1]$ into subintervals of appropriate size the correct subinterval. For example if intervals have size 0.1, then 0.65 belongs to the 7th sub-interval. After bucketing, we sort the elements within a subinterval using some simple $O(n^2)$ algorithm. Then output the sorted list by concatenating the sorted lists. Show the the expected time for sorting in $O(n)$.
Hint: Let X_i be the random variable denoting the number of elements in subinterval i. Then bound $E[X_i^2]$.

5. **Approximate half-plane range query** Given a set of points S of n on the plane, we can construct a simple data structure for half-plane range query by choosing a random sample $R \subset S$, such that for any query half plane ℓ we will compute the number of points in $\ell \cap R$, say n' and report the answer $n' \cdot \frac{n}{|R|}$.
Do a rigorous analysis of what guarantee you can provide for this (approximate) answer depending on $|R|$.

6. Calculate an upper bound expression for the term $E[|\Pi^k(R)|]$ where R is a random sample chosen by picking each element with probability $\frac{r}{n}$ from n objects. Here $k \geq 0$ is some fixed small integer constant.
Interpret the notations as per definitions given in the lecture.

7. The set of ranges $\mathcal{R}_p = \{r \in \mathcal{R} | p \in r\}$ The *dual range space* of a range space $S = (X, \mathcal{R})$ is the space $S^* = (\mathcal{R}, X^*)$ where $X^* = \{\mathcal{R}_p | p \in X\}$.
(i) Show that $(S^*)^* = S$.
(ii) Show that the VC dimension of the dual of a range space with VC dimension δ is bounded by $2^{\delta+1}$.
Hint You may find the incidence matrix representation of a range space more helpful for this purpose.