INNOPOLIS UNIVERSITY

Week 6. Problem set

1. Apply COUNTING-SORT [CLRS, §8.2] to the following input array where each column corresponds to one item with its numeric key and single-character satellite data:

2	6	2	7	6	6	3	4	1	6	2	0	7
U	W	A	M	E	S	E	A	O	O	R	Y	E

   You **must** demonstrate the final state of both auxiliary arrays used in the algorithm, as well as the final state of the output array.

2. Explain how BUCKET-SORT [CLRS, §8.4] can be used to efficiently sort the points distributed uniformly[1] in a unit disk $\{\langle x, y\rangle \mid x^2 + y^2 \leq 1\}$ by their distance[2] from the origin $\langle 0, 0\rangle$:

   (a) Explain the idea of splitting the points into buckets.

   (b) Show how to assign each point to a bucket in $\Theta(1)$ time.

   (c) For each bucket, compute the probability of a random point to be assigned to it. Ensure that the sum of probabilities is 1.

   (d) Argue that the average case time complexity for your algorithm is $\Theta(n)$, relying on the result for standard BUCKET-SORT [CLRS, §8.4].

3. Consider a modification of the MERGE-SORT algorithm [CLRS, §2.3] that stops recursion when the size of subarray becomes less than or equal to k ($k \leq n$). For arrays of size $\leq k$, the modified algorithm performs BUCKET-SORT [CLRS, §8.4] with a randomized QUICK-SORT [CLRS, §7.3] to sort each bucket. Answer the following questions about the modified algorithm. Provide justification (4–6 sentences) for each answer.

   (a) What is the worst case time complexity in terms of n and k?

   (b) What is the best case time complexity in terms of n and k?

   (c) What is the average[3] case time complexity in terms of n and k?

   The answer should be given using Θ-notation.
   Provide a **brief** justification for each case (1–2 sentences).

References

[CLRS] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C., 2022. *Introduction to algorithms, Fourth Edition.* MIT press.

[1] Meaning that for each point the probability of being in any region R inside the disk is directly proportional to the area of that region: $P(R) = \frac{\mathsf{area}(R)}{\pi}$.

[2] We consider the standard Euclidean distance $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

[3] assuming all elements in the input array are distinct and any initial order in the array is equally likely