COMP 480/580 Assignment 1 — Question 3 Solution
Inequalities and 5-independence in Linear Probing

Restatement

We must prove that linear probing has $O(1)$ expected search cost when the load factor $\alpha = 1/3$ and the hash family is 5-independent. We are given two hints:

1. Use the inequality
$$\Pr(|X - \mathbb{E}[X]| \geq a) \;\leq\; \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{a^4}.$$
2. 5-independence ensures that in the 4th-moment expansion of sums of indicator variables, products factorize so that cross terms can be controlled.

Step 1 — Interval analysis

For a dyadic interval (arc) of length 2^s, let
$$B_s = \sum_{i=1}^{N} X_i, \quad \mu_s = \mathbb{E}[B_s].$$

Here X_i is the indicator that key i lands in the arc, and $N = m/3$ inserted keys. Standard linear probing analysis shows the expected search cost is bounded by
$$\sum_{s=1}^{\lceil \log m \rceil} 2^s \; \Pr(B_s \geq 2\mu_s).$$

Step 2 — Apply the 4th-moment inequality (Hint 1)

Using $a = \mu_s$,
$$\Pr(B_s \geq 2\mu_s) \;=\; \Pr(B_s - \mu_s \geq \mu_s) \;\leq\; \frac{\mathbb{E}[(B_s - \mu_s)^4]}{\mu_s^4}.$$

Step 3 — Expanding the 4th central moment (Hint 2)

Write $B_s = \sum_{i=1}^{N} X_i$ with $\mathbb{E}[X_i] = p = 2^s/m$. Then $\mu_s = Np = \alpha 2^s$.

We expand
$$\mathbb{E}[(B_s - \mu_s)^4] = \mathbb{E}\Big[\Big(\sum_{i=1}^{N}(X_i - p)\Big)^4\Big].$$

Expanding gives sums of terms of the forms:

- **Single-variable terms:** $\mathbb{E}[(X_i - p)^4]$. Each is $O(p)$ since $X_i \in \{0, 1\}$. Total contribution $O(Np) = O(\mu_s)$.
- **Pairwise terms:** $\mathbb{E}[(X_i - p)^2(X_j - p)^2]$. Because of 5-independence, X_i and X_j are independent, so this factorizes: $\mathbb{E}[(X_i - p)^2] \cdot \mathbb{E}[(X_j - p)^2] = O(p^2)$. With $\binom{N}{2}$ such pairs, contribution $O(N^2 p^2) = O(\mu_s^2)$.
- **Mixed terms of degree 3 or 4 in distinct variables:** These vanish or reduce to lower-order contributions under 5-independence, since expectations factorize and centered indicators have mean 0.

Therefore

$$\mathbb{E}[(B_s - \mu_s)^4] = O(\mu_s + \mu_s^2).$$

Step 4 — Tail bound

Plugging into Step 2:

$$\Pr(B_s \geq 2\mu_s) \ \leq \ O\left(\frac{1}{\mu_s^3} + \frac{1}{\mu_s^2}\right).$$

For $\mu_s \geq 2$ we simplify to $\Pr(B_s \geq 2\mu_s) = O(1/\mu_s^2)$.

Step 5 — Summation

Since $\mu_s = \alpha 2^s = \frac{1}{3} 2^s$,

$$2^s \Pr(B_s \geq 2\mu_s) = O\left(\frac{2^s}{(2^s)^2}\right) = O\left(\frac{1}{2^s}\right).$$

Summing over $s = 1, \ldots, \log m$ gives a convergent geometric series, hence constant. For small s, there are only finitely many terms, each $O(1)$, so total contribution is constant.

Conclusion

By applying the 4th-moment inequality (Hint 1) and expanding the 4th central moment using 5-independence (Hint 2), we have shown that

$$\mathbb{E}[\text{search cost}] = O(1) \quad \text{when } \alpha = 1/3.$$