

Algorithm Analysis for Random Quran Question Generation

Submitted By:

1. Mahmoud Wagih Mohamed El-Sayed - SEC 7
2. Mrawan Nabil Mohamed Motawe - SEC 8
3. Malak Mahmoud Metwally El-Gizawy - SEC 8
4. Nour Saad Abd El-Fattah Al-Adrosy - SEC 9
5. Yahya Adel Mohamed - SEC 9

1. Problem Identification

1.1 Problem Definition

The project addresses the challenge of selecting N testing points (Ayahs) randomly from a specific range of the Holy Quran (e.g., a specific Juz, Surah, or the entire Quran) for revision purposes (Muraja'ah). A key requirement for an effective "Hifz" revision session is that questions should not be clustered in a single area but must be fairly distributed across the requested range to ensure comprehensive coverage.

1.2 Mathematical Formulation

To formalize the problem, let the Quranic range be represented as a discrete set of indices corresponding to Ayahs:

$$\mathcal{A} = \{a_1, a_2, \dots, a_M\}$$

Where M is the total number of Ayahs in the selected range (e.g., $M = 6236$ for the whole Quran based on Madani Mushaf). We aim to select a subset Q subset \mathcal{A} such that $|Q| = N$. The challenge is to maximize the **Uniformity** of the distribution of Q over \mathcal{A} to avoid clustering.

1.3 Relevance & Complexity

This problem finds immediate relevance in educational technology for Quran memorizers. The complexity lies in ensuring a **Deterministic Spread** of random variables using real data. The problem compares a probabilistic coverage model (Naive) against a guaranteed coverage model (Stratified), highlighting the trade-off between implementation simplicity and execution efficiency ($O(N^2)$ vs $O(N)$).

2. Algorithm Development

2.1 Naive Algorithm (Blind Random Sampling)

The naive approach treats the selected Quranic range as a single pool. It selects \$N\$ indices using a uniform random generator. Uniquely, it utilizes **Real Quran Data** to map Juz to Ayahs but relies on an inefficient **Linear Search** check against all previously selected questions to ensure uniqueness. Finally, it **Sorts** the generated list before displaying, adding extra computational overhead.

Pseudocode:

```
ALGORITHM Generate_Naive_Questions
// Inputs:
// - Start_Juz, End_Juz: The range of Quran parts to test on.
// - N: Total number of questions required.
// - Verses_Per_Q: Length of each question (recitation scope).

INPUTS: Start_Juz, End_Juz, N, Verses_Per_Q
OUTPUT: Final_List_of_Questions

// 1. Convert Juz Range to Actual Ayah Indices (Real Quran Data)
Start_Global_Ayah = GET_REAL_START_AYAH(Start_Juz)
End_Global_Ayah = GET_REAL_END_AYAH(End_Juz)

// 2. Initialize Empty List
Result_List = []

// 3. Generation Phase
WHILE LENGTH(Result_List) < N DO

    // STEP A: Blind Random Selection
    Candidate_Ayah = RANDOM_INTEGER(Start_Global_Ayah, End_Global_Ayah)

    // STEP B: Linear Search for Duplicates (O(N^2) Bottleneck)
    Is_Duplicate = FALSE
    FOR EACH existing_item IN Result_List DO
        IF existing_item == Candidate_Ayah THEN
            Is_Duplicate = TRUE
            BREAK
        END IF
    END FOR
```

```

// STEP C: Add to List if Unique
IF Is_Duplicate IS FALSE THEN
    APPEND Candidate_Ayah TO Result_List
END IF

END WHILE

// 4. Sorting Phase (Extra Overhead O(N log N))
// Required because random selection is unordered
SORT(Result_List)

// 5. Return Final List
RETURN Result_List

```

END ALGORITHM

2.2 Optimized Algorithm (Stratified Random Sampling)

The optimized approach divides the actual Ayah count of the selected range into **N** equal partitions (strata). The algorithm forces the selection of exactly one question from each stratum.

Key Advantage (Natural Sorting):

Unlike the Naive approach, this algorithm generates questions in a strictly ascending order because it iterates through the strata sequentially (from Segment 0 to Segment **N-1**). This produces a Naturally Sorted list without requiring any post-generation sorting step, reducing computational complexity significantly.

Pseudocode:

```

ALGORITHM Generate_Stratified_Questions
// Inputs:
// - Start_Juz, End_Juz: The range of Quran parts to test on.
// - N: Total number of questions required.
// - Verses_Per_Q: Length of each question (recitation scope).

```

INPUTS: Start_Juz, End_Juz, N, Verses_Per_Q
OUTPUT: Final_List_of_Questions

```

// 1. Convert Juz Range to Actual Ayah Indices (Real Quran Data)
Start_Global_Ayah = GET_REAL_START_AYAH(Start_Juz)

```

```

End_Global_Ayah = GET_REAL_END_AYAH(End_Juz)

// 2. Calculate Partition Size (Stratum)
Total_Range = End_Global_Ayah - Start_Global_Ayah + 1
Step_Size = Total_Range / N

// 3. Initialize Empty List
Result_List = []

// 4. Main Generation Loop (Iterates exactly N times)
FOR i FROM 0 TO N-1 DO

    // STEP A: Define Strict Segment Boundaries
    Segment_Start = Start_Global_Ayah + (i * Step_Size)
    Segment_End = Segment_Start + Step_Size - 1

    IF Segment_End > End_Global_Ayah THEN
        Segment_End = End_Global_Ayah
    END IF

    // STEP B: Stratified Random Selection
    Candidate_Ayah = RANDOM_INTEGER(Segment_Start, Segment_End)

    // STEP C: Add to List (Naturally Sorted)
    APPEND Candidate_Ayah TO Result_List

END FOR

// 5. Return the Final List
RETURN Result_List

END ALGORITHM

```

3. Implementation Details

Both algorithms now use **Real Quran Data** via a lookup table JUZ_START_INDICES to map Juz numbers to exact Ayah indices (e.g., Juz 30 starts at Ayah 5673).

1. **Naive Implementation:** Uses mt19937 for high-quality randomness but suffers from an inefficient $O(N^2)$ duplicate check logic, followed by an $O(N \log N)$ sorting step.
2. **Optimized Implementation:** Divides the real Ayah range into segments. It guarantees $O(N)$ performance and zero overlaps. The output is **Naturally Sorted** by design, saving processing power.

4. Theoretical Analysis

4.1 Complexity Analysis

- **Naive Algorithm:**
 - **Time Complexity:** $O(N^2)$ due to linear collision checking, plus $O(N \log N)$ for the explicit sorting phase. The dominant term is $O(N^2)$.
 - **Space Complexity:** $O(N)$.
- **Stratified Algorithm:**
 - **Time Complexity:** $O(N)$. It iterates exactly N times. No searching or sorting is required because the segmentation guarantees order and uniqueness by design.
 - **Space Complexity:** $O(N)$.

4.2 Quality Analysis

- **Naive:** High clustering risk. It might select Ayah 50 and then Ayah 51, leading to redundant testing.
- **Stratified:** Guaranteed coverage. If you ask for 30 questions from the whole Quran, it mathematically guarantees exactly one question per Juz-sized segment.

5. Empirical Analysis & Visualization

5.1 Performance Metrics

We tested both algorithms with varying input sizes (N) on the full Quran range (6236 Ayahs).

Input Size (N)	Naive Time ($O(N^2)$)	Stratified Time ($O(N)$)	Observation
10	~0.005 ms	~0.003 ms	Similar for small N
100	~0.500 ms	~0.018 ms	Naive starts slowing down (Sorting adds overhead)
1000	~40.00 ms	~0.145 ms	Significant Gap (Naive is ~270x slower)
Large N	Degrades rapidly	Linear scaling	Naive becomes unusable

5.2 Visualization of Distribution (Simulation)

Naive Sampling (Clustered & Overlapping):

[x x x x] <-- Uneven Spread
Start End

Note: Gaps are left untested, while some areas are tested twice.

Stratified Sampling (Uniform & Fair):

[x | x | x | x | x] <-- Even Spread
Seg 1 Seg 2 Seg 3 Seg 4 Seg 5

Note: Every segment is tested exactly once.

6. Results Comparison & Summary

6.1 Discrepancies

The empirical results confirm that the **Naive algorithm is significantly slower** and less reliable.

- **Reason:** The nested loop for collision checking ($O(N^2)$) plus the sorting overhead ($O(N \log N)$) makes it much heavier than the Stratified method ($O(N)$).
- **Verdict:** The Stratified algorithm is strictly superior for creating balanced Quranic exams.

6.2 Comparison Summary

Aspect	Naive Sampling	Stratified Sampling	Winner
Time Complexity	$O(N^2) + O(N \log N)$	$O(N)$	Stratified
Space Complexity	$O(N)$	$O(N)$	Draw
Risk of Clustering	High	None	Stratified
Sorting Strategy	Explicit (Slow)	Natural (Free)	Stratified