16, Given streaming data, design an algorithm to get approximate median of all previous data, use constant memory.

Box 1

P^2 Algorithm: To calculate the p-quantile of
$$\{x_1, \ldots, x_n\}$$

A. Initialization: Sort the first five observations
$\{x_1, x_2, x_3, x_4, x_5\}$ and set

Marker heights $\quad q_i \leftarrow x_{(i)}; \quad i = 1, \ldots, 5$

Marker positions $\quad n_i \leftarrow i; \quad i = 1, \ldots, 5$

Desired marker positions
$$n_1' \leftarrow 1; \quad n_2' \leftarrow 1 + 2p; \quad n_3' \leftarrow 1 + 4p;$$
$$n_4' \leftarrow 3 + 2p; \quad n_5' \leftarrow 5;$$

Note that n_i' are real variables, while n_i are integers.

To reduce CPU overhead, calculate and store the increment dn_i' in the desired marker positions:

$$dn_1' \leftarrow 0; \quad dn_2' \leftarrow p/2; \quad dn_3' \leftarrow p;$$
$$dn_4' \leftarrow (1 + p)/2; \quad dn_5' \leftarrow 1;$$

B. For each subsequent observation $x_j, j \geq 6$, perform the following:
1. Find cell k such that $q_k \leq x_j < q_{k+1}$ and adjust extreme values (q_1 and q_5) if necessary, that is,

    CASE of x_j

    $[x_j < q_1]$ $q_1 \leftarrow x_j; \; k \leftarrow 1;$

    $[q_1 \leq x_j < q_2]$ $k \leftarrow 1;$

    $[q_2 \leq x_j < q_3]$ $k \leftarrow 2;$

    $[q_3 \leq x_j < q_4]$ $k \leftarrow 3;$

    $[q_4 \leq x_j \leq q_5]$ $k \leftarrow 4;$

    $[q_5 < x_j]$ $q_5 \leftarrow x_j; \; k \leftarrow 4;$

    END CASE;

2. Increment positions of markers $k + 1$ through 5:

$$n_i \leftarrow n_i + 1 \quad i = k, \ldots, 5$$

Update desired positions for all markers:

$$n_i' \leftarrow n_i' + dn_i' \quad i = 1, \ldots, 5$$

3. Adjust heights of markers 2–4 if necessary:
FOR $i = 2$ **TO 4 DO**
  BEGIN
    $d_i \leftarrow n_i' - n_i$
    IF $\{(d_i \geq 1 \text{ and } n_{i+1} - n_i > 1) \text{ or}$
      $(d_i \leq -1 \text{ and } n_{i-1} - n_i < -1)\}$
    BEGIN
      $d_i \leftarrow \text{sign}(d_i)$
      Try adjusting q_i using P^2 formula:
      $q_i' \leftarrow q_i$ from parabolic formula

$$\textbf{IF } \{q_{i-1} < q_i' < q_{i+1}\}$$

$$\textbf{THEN } q_i \leftarrow q_i$$

ELSE use linear formula:

$$q_i \leftarrow q_i \text{ from linear formula;}$$

$$n_i \leftarrow n_i + d_i,$$

END IF;

END DO;

C. Return q_3 as the current estimate of p-quantile.