Ερώτηση 1

Από το Α.Μ. μου επιλέγονται A = 02 και B = 40. Πρώτα πρέπει να ταξινομήσουμε κατά αύξουσα σειρά τους ακεραίους:

02, 13, 14, 18, 23, 27, 29, 32, 36, 40, 41, 44, 46, 55, 60, 61, 62, 64, 65, 73, 75, 78, 80, 84, 87, 93, 96, 99

Για μέγεθος 4 έχουμε τις διαμερίσεις:

- Bin 1: 02, 13, 14, 18
- Bin 2: 23, 27, 29, 32
- Bin 3: 36, 40, 41, 44
- Bin 4: 46, 55, 60, 61
- Bin 5: 62, 64, 65, 73
- Bin 6: 75, 78, 80, 84
- Bin 7: 87, 93, 96, 99

α) bin means: Για κάθε διαμέριση υπολογίζουμε το μέσο όρο και αντικαθιστούμε όλα τα στοιχεία με τη τιμή του. Οι μέσοι όροι υπολογίζονται ως:

- Bin 1: 11.75
- Bin 2: 27.75
- Bin 3: 40.25
- Bin 4: 55.5
- Bin 5: 66
- Bin 6: 79.25
- Bin 7: 93.75

Καθώς όμως τα δεδομένα μας είναι ακέραιοι, πρέπει και οι διαμερίσεις να περιέχουν ακεραίους, επομένως στρογγυλοποιούμε και έχουμε:

- Bin 1: 12, 12, 12, 12
- Bin 2: 28, 28, 28, 28
- Bin 3: 40, 40, 40, 40
- Bin 4: 56, 56, 56, 56
- Bin 5: 66, 66, 66, 66
- Bin 6: 79, 79, 79, 79
- Bin 7: 94, 94, 94, 94

β) bin boundaries: Για κάθε διαμέριση υπολογίζουμε το ανώτατο και κατώτατο όριο και αναθέτουμε σε κάθε στοιχείο το όριο στο οποίο βρίσκεται πιο κοντά. Τα όρια υπολογίζονται ως:

- Bin 1: 02, 18
- Bin 2: 23, 32
- Bin 3: 36, 44
- Bin 4: 46, 61
- Bin 5: 62, 73
- Bin 6: 75, 84
- Bin 7: 87, 99

Επομένως οι διαμερίσεις γίνονται:

- Bin 1: 02, 18, 18, 18
- Bin 2: 23, 32, 32, 32
- Bin 3: 36, 44, 44, 44
- Bin 4: 46, 61, 61, 61
- Bin 5: 62, 62, 62, 73
- Bin 6: 75, 75, 84, 84
- Bin 7: 87, 99, 99, 99

Ερώτηση 2

Από το Α.Μ. μου επιλέγονται A = 4 και B = 7 (Τελευταίο ψηφίο είναι το 0). Επομένως έχουμε τις δύο hash functions:

1. $h_1(x) = (4x + 11) mod 20$
2. $h_2(x) = (7x + 2) mod 20$

Έχουμε δηλαδή $w = 2$ hash functions και επίσης μέγεθος φίλτρου $m = 20$. Δεν γνωρίζουμε το πλήθος n των στοιχείων που έχουν παρέλθει.

1. Υπολογίζουμε τη τιμή των hash functions για $y = 8$:

$h_1(8) = 43 mod 20 = 3$
$h_2(8) = 58 mod 20 = 18$

και βρίσκουμε από το state:

$s_3(t) = 1$
$s_{18}(t) = 1$

Αφού και τα δύο bits είναι 1 υπάρχει καλή πιθανότητα το 8 να έχει περάσει ήδη.

2. Η πιθανότητα να έχουμε False Positive (F) δίνεται από την εξίσωση 1. Επομένως η πιθανότητα να πέρασε (P) δίνεται από την εξίσωση 2.

$$F = \left(1 - e^{-\frac{nw}{m}}\right)^w \tag{1}$$

$$P = 1 - F \tag{2}$$

Καθώς δεν γνωρίζουμε το n, το καλύτερο που μπορούμε να κάνουμε είναι να υπολογίσουμε το πλήθος N των ξεχωριστών στοιχείων που έχουν περάσει από το φίλτρο, το οποίο μπορούμε να το προσεγγίσουμε χρησιμοποιόντας την εξίσωση 3. Σημειώνεται πως m_0 είναι το πλήθος των 0 στο φίλτρο.

$$N \approx \frac{m \ln \frac{m}{m_0}}{w} \tag{3}$$

Στο φίλτρο υπάρχουν $m_0 = 11$ μηδενικά. Έτσι έχουμε $n = N \approx 6$.

Επομένως, από την (1) λαμβάνουμε $F = 0.2$.

Τελικά από την (2) λαμβάνουμε $P = 0.8$.

3. Υπολογίζουμε τη τιμή των hash functions για $x = 13$:

$h_1(8) = 63 mod 20 = 3$
$h_2(8) = 93 mod 20 = 13$

και βρίσκουμε από το state:

$s_3(t) = 1$
$s_{13}(t) = 0$

Επομένως έχουμε ένα collision για το 3^o bit, οπότε αυξάνεται η πιθανότητα να έχουμε False Positive, ενώ το 13^o bit θα γίνει 1. Έτσι έχουμε:

$s(t+1) = [1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0]$

4. Θα χρησιμοποιήσουμε την εξίσωση 3 με το καινούριο state. Έχουμε πλέον $m_0 = 10$ μηδενικά και λαμβάνουμε $n = N \approx 7$ ξεχωριστά στοιχεία. Μπορούμε να πούμε, λοιπόν, πως έχουν περάσει **τουλάχιστον** 7 ξεχωριστά στοιχεία με μεγάλη βεβαιότητα, αλλά δεν έχουμε κάποια πληροφορία για τον συνολικό αριθμό των στοιχείων που έχουν περάσει.

Ερώτηση 3

Επιλέγουμε x = 3 καθώς το Α.Μ λήγει σε 0. Έτσι έχουμε το πίνακα μετάβασης της εξίσωσης 4.

$$P = \begin{bmatrix} 0.3 & 0.7 \\ 0.3 & 0.7 \end{bmatrix} \tag{4}$$

Αρχικοποίηση (c):

- $\alpha_1(1) = \Pi(1) \times \theta_1(c) = 0$
- $\alpha_1(2) = \Pi(2) \times \theta_2(c) = 0.5$

Αναδρομή:

Για t = 2 (b):

- $\alpha_2(1) = \sum_{i=1}^{2}(\alpha_1(i) \times P_{\mathrm{i},1}) \times \theta_1(b) = 0.075$
- $\alpha_2(2) = \sum_{i=1}^{2}(\alpha_1(i) \times P_{\mathrm{i},2}) \times \theta_2(b) = 0$

Για t = 3 (a):

- $\alpha_3(1) = \sum_{i=1}^{2}(\alpha_2(i) \times P_{\mathrm{i},1}) \times \theta_1(a) = 0.01125$
- $\alpha_3(2) = \sum_{i=1}^{2}(\alpha_2(i) \times P_{\mathrm{i},2}) \times \theta_2(a) = 0.02625$

Τερματισμός:

$P(cba) = \sum_{i=1}^{2} \alpha_3(i) = 0.0375$