

## Πρώτη Σειρά Ασκήσεων

Αυτό είναι το πρώτο μέρος της πρώτης σειράς ασκήσεων. Η προθεσμία για την παράδοση αυτού του κομματιού είναι στις 30 Οκτωβρίου, στο ξεκίνημα του μαθήματος. Κάνετε turn-in τον κώδικα σας, και παραδώστε τις υπόλοιπες ερωτήσεις είτε ηλεκτρονικά, είτε σε χαρτί. Συστήνεται να παραδώσετε τις ερωτήσεις νωρίτερα, Παρασκευή 26 Οκτωβρίου, μιας και τότε θα ανακοινωθεί το δεύτερο μέρος της πρώτης άσκησης. Για καθυστερημένες υποβολές ισχύει η πολιτική στην σελίδα του μαθήματος. Οι λεπτομέρειες για το turn-in είναι στη σελίδα του μαθήματος.

### Άσκηση 1 (Reservoir Sampling)

Στην τάξη περιγράψαμε τον αλγόριθμο Reservoir Sampling για τη δειγματοληψία ενός αντικειμένου από ένα ρεύμα  $N$  αντικειμένων. Σε αυτή την άσκηση θα πρέπει να τροποποιήσετε τον αλγόριθμο ώστε να κάνει δειγματοληψία  $K$  αντικειμένων.

1. Περιγράψετε τον αλγόριθμο που διαλέγει ένα ομοιόμορφο δείγμα  $K$  αντικειμένων από ένα ρεύμα  $N$  αντικειμένων. Ο αλγόριθμος σας θα πρέπει να δουλεύει με ένα μόνο πέρασμα στα δεδομένα διαβάζοντας τα αντικείμενα ένα-ένα, χωρίς προηγούμενη γνώση του μεγέθους του ρεύματος, και να χρησιμοποιεί  $O(K)$  μνήμη (υποθέστε ότι το μέγεθος του κάθε αντικειμένου είναι σταθερό).  
(Υπόδειξη: Σε ένα ομοιόμορφα τυχαίο δείγμα, κάθε στοιχείο έχει πιθανότητα  $K/N$  να εμφανιστεί στο δείγμα.)
2. Αποδείξτε ότι ο αλγόριθμος σας παράγει ένα ομοιόμορφα τυχαίο δείγμα, δηλαδή, για κάθε  $i, 1 \leq i \leq N$ , το  $i$ -οστό στοιχείο έχει πιθανότητα  $K/N$  να εμφανιστεί στο δείγμα.  
(Υπόδειξη: Ποια είναι η πιθανότητα αν επιλεγεί το  $i$ -οστό στοιχείο, να αντικατασταθεί αργότερα από το  $j$ -οστό, για  $j > i$ ?)
3. Γράψτε ένα πρόγραμμα που υλοποιεί τον αλγόριθμο σας (σε όποια γλώσσα θέλετε). Το πρόγραμμα σας θα πρέπει να παράγει ένα δείγμα με  $K$  τυχαίες γραμμές ενός αρχείου εισόδου. Χρησιμοποιήστε το standard input για είσοδο και το standard output για έξοδο. Το πρόγραμμα σας θα πρέπει να τρέχει ως εξής: `sample <inputfile >outputfile`.

### Άσκηση 2 (Δειγματοληψία ακμών)

Έχουμε ένα γράφο  $G$  με  $N$  κορυφές και  $M$  ακμές, και βαθμό  $d_x$  για κάθε κόμβο  $x$ . Θέλουμε να διαλέξουμε μια ακμή ομοιόμορφα τυχαία από το γράφο (δηλαδή κάθε ακμή θα πρέπει να έχει πιθανότητα  $1/M$  να διαλεχτεί).

Θεωρήστε τον ακόλουθο αλγόριθμο για την δειγματοληψία μιας ακμής:

- Διαλέγουμε ένα κόμβο ομοιόμορφα τυχαία (δηλ., με πιθανότητα  $1/N$ ) από το σύνολο των κόμβων.
- Διαλέγουμε ομοιόμορφα τυχαία μία από τις ακμές με άκρο τον επιλεγμένο κόμβο  $x$  (δηλ., με πιθανότητα  $1/d_x$ ), και επιστρέφουμε αυτή την ακμή.

Απαντήστε τις ακόλουθες ερωτήσεις:

1. Υπολογίστε την πιθανότητα να επιλέξω την ακμή  $(x, y)$  ως συνάρτηση του αριθμού των κόμβων  $N$ , και των βαθμών  $d_x, d_y$  των κόμβων  $x$  και  $y$  αντίστοιχα.
2. Δώστε ένα αντιπαράδειγμα που δείχνει ότι ο αλγόριθμος δεν παράγει ένα τυχαίο ομοιόμορφο δείγμα (δηλ., **δεν** ισχύει ότι η πιθανότητα κάθε ακμής να επιλεγεί είναι  $1/M$ ).
3. Τροποποιήστε ένα από τα βήματα του παραπάνω αλγορίθμου ώστε να παράγει ένα ομοιόμορφα τυχαίο δείγμα.  
(**Υπόδειξη:** Ποια είναι η σχέση μεταξύ του  $M$  και των βαθμών των κόμβων?)