

## Ασκήσεις Προσομοίωσης - Σειρά 3

### Α. Μπουρνέτας, ΠΜΣ 'Βιοστατιστική'

**ΠΡΟΒΛΗΜΑ 1.** Έστω  $X$  συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = x + \frac{1}{2}, \quad 0 \leq x \leq 1.$$

(α) Να περιγραφεί γεννήτρια τυχαίων αριθμών από αυτή την κατανομή χρησιμοποιώντας τη μέθοδο αντίστροφου μετασχηματισμού.

(β) Να περιγραφεί γεννήτρια τυχαίων αριθμών από αυτή την κατανομή χρησιμοποιώντας τη μέθοδο απόρριψης.

(γ) Να υλοποιηθούν οι παραπάνω γεννήτριες και να δημιουργηθούν 1000 παρατηρήσεις από την  $X$  μέσω της κάθε μιας από αυτές. Να ελεγχθεί αν τα δύο δείγματα που δημιουργήθηκαν προέρχονται από την ίδια κατανομή χρησιμοποιώντας τον έλεγχο Kolmogorov-Smirnov.

**ΠΡΟΒΛΗΜΑ 2.** Μια ασφαλιστική εταιρεία έχει 500 ασφαλισμένους καθέναν από τους οποίους έχει πιθανότητα 5% να ζητήσει μια αποζημίωση στη διάρκεια της επόμενης χρονιάς. Υποθέστε ότι το ποσό μιας αποζημίωσης είναι τυχαία μεταβλητή που ακολουθεί εκθετική κατανομή με μέση τιμή 600 ευρώ. Να εκτιμήσετε μέσω προσομοίωσης την πιθανότητα το συνολικό ποσό αποζημιώσεων που θα πρέπει να πληρώσει η εταιρεία την επόμενη χρονιά να υπερβαίνει τις 35000 ευρώ.

**ΠΡΟΒΛΗΜΑ 3.** Η ηλικία των ατόμων ενός πληθυσμού ακολουθεί κατανομή  $Gamma(5, 0.1)$ . Η αρτηριακή πίεση ενός ατόμου ηλικίας  $x$  ακολουθεί κανονική κατανομή με μέση τιμή  $9 + x/12$  και τυπική απόκλιση  $0.04x$ .

(α) Να περιγραφεί γεννήτρια τυχαίων αριθμών για την αρτηριακή πίεση ενός τυχαίου ατόμου αυτού του πληθυσμού.

(β) Αν η αρτηριακή πίεση άνω του 15 θεωρείται υπερβολική, να εκτιμηθεί το μέσο επίπεδο υπέρβασης από τη φυσιολογική τιμή : (i) για όλο τον πληθυσμό και (ii) για άτομα άνω των 65 ετών.

**ΠΡΟΒΛΗΜΑ 4.** Έστω ένα μοντέλο παλινδρόμησης με εξαρτημένη μεταβλητή  $Y$  και ανεξάρτητη  $X$ , όπου η δεσμευμένη κατανομή της  $Y$  δεδομένου ότι  $X = x$  είναι εκθετική με μέση τιμή  $0.4x + 2$ . Υποθέτουμε ότι αυτή η σχέση δεν είναι γνωστή, και γίνεται μια στατιστική μελέτη για τη διερεύνηση της συσχέτισης μεταξύ των δύο μεταβλητών. Για τη μελέτη συγκεντρώνεται ένα δείγμα μεγέθους 100 στο οποίο υπάρχουν οι τιμές της  $X = 1, 2, \dots, 10$ , από 10 φορές η καθεμιά. Ο αναλυτής δε γνωρίζει την πραγματική κατανομή των καταλοίπων. Υποθέτει ότι αυτή είναι κανονική με σταθερή διασπορά, έτσι ώστε να μπορεί να εφαρμόσει το μοντέλο γραμμικής παλινδρόμησης. Μετά τον υπολογισμό των εκτιμήσεων ελαχίστων τετραγώνων γίνεται υπολογισμός των καταλοίπων (χωρίς κανονικοποίηση) και έλεγχος κανονικότητας αυτών μέσω του ελέγχου Shapiro-Wilk.

(α) Χρησιμοποιώντας προσομοίωση για τις τιμές της  $Y$  στο δοσμένο δείγμα, να εκτιμηθεί η μέση απόκλιση των εκτιμήσεων ελαχίστων τετραγώνων  $\hat{\beta}_0, \hat{\beta}_1$  από τις αληθινές τιμές 2 και 0.4, αντίστοιχα.

(β) Να εκτιμηθεί η πιθανότητα απόρριψης του ελέγχου κανονικότητας.

**ΠΡΟΒΛΗΜΑ 5.** Έστω το ολοκλήρωμα

$$A = \int_0^{\infty} \frac{\sqrt{x+4}}{x^3+1} dx.$$

(α) Να περιγράψετε μια μέθοδο εκτίμησης της τιμής του  $A$  με τη μέθοδο προσομοίωσης Monte Carlo.

(β) Να υπολογίσετε τον απαιτούμενο αριθμό παρατηρήσεων έτσι ώστε η τιμή του  $A$  να εκτιμηθεί με διάστημα εμπιστοσύνης 95% και ακρίβεια  $\pm 0.01$ .

(γ) Να βρεθεί το παραπάνω Δ.Ε.

**ΠΡΟΒΛΗΜΑ 6.** Μια βιολογική μέτρηση  $X$  είναι τυχαία μεταβλητή που περιγράφεται από το μοντέλο  $X = (1 - U^2)e^U$ , όπου η  $U$  ακολουθεί ομοιόμορφη κατανομή στο διάστημα  $(0, 1)$ . Έστω  $\theta = E(X)$  η μέση τιμή της  $X$ .

(α) Να εκτιμηθεί το  $\theta$  με ένα διάστημα εμπιστοσύνης 95% βασισμένο σε 1000 προσομοιωμένες παρατηρήσεις της  $X$ .

(α) Να εκτιμηθεί το  $\theta$  με ένα διάστημα εμπιστοσύνης 95% βασισμένο σε 1000 προσομοιωμένες παρατηρήσεις της  $X$ , χρησιμοποιώντας δειγματοληψία σπουδαιότητας με κατανομή σπουδαιότητας για τη  $U$  την  $g(u) = 2u, 0 \leq u \leq 1$ . Υπάρχει βελτίωση στην ακρίβεια της εκτίμησης με αυτή την κατανομή σπουδαιότητας, και αν ναι, κατά ποιο ποσοστό;

**ΠΡΟΒΛΗΜΑ 7.** Έστω το ζεύγος τυχαίων μεταβλητών  $(X, Y)$  με από κοινού συνάρτηση κατανομής πιθανότητας  $f(x, y) = kx, 0 \leq x \leq y \leq 1$ .

(α) Να περιγραφεί και να υλοποιηθεί η δημιουργία ανεξάρτητων (προσεγγιστικά) παρατηρήσεων από την παραπάνω κατανομή χρησιμοποιώντας τη μέθοδο Gibbs sampler.

(β) Να εκτιμηθούν η μέση τιμή και η διασπορά των  $X, Y$  όπως επίσης και ο συντελεστής συσχέτισής τους.