Ασκήσεις Κεφαλαίου 3

1. Η κατανομή **Poisson** χρησιμοποιείται αντί της διωνυμικής όταν ο αριθμός m των επαναλήψεων των δοκιμών είναι μεγάλος και η πιθανότητα 'επιτυχίας' p σε κάθε δοκιμή είναι μικρή και τότε το γινόμενο $\lambda = mp$ ορίζει το μέσο αριθμό επιτυχιών. Η κατανομή **Poisson** χρησιμοποιείται επίσης για να περιγράψει την εμφάνιση πλήθους γεγονότων (επιτυχιών) σε ένα χρονικό διάστημα ή γενικότερα στο πεδίο αναφοράς, π.χ. αριθμός διακοπών σύνδεσης δικτύου σε μια μέρα, αριθμός καμμένων εικονοστοιχείων (pixel) σε μια οθόνη. Συμβολίζοντας X την τ.μ. του αριθμού εμφανίσεων των γεγονότων ενδιαφέροντος (επιτυχιών), η σππ της κατανομής Poisson είναι

$$f_X(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \tag{3.25}$$

όπου x είναι το πλήθος εμφανίσεων 'επιτυχιών', $x! = 1 \cdot 2 \cdots x$ είναι το παραγοντικό του x και λ η παράμετρος της κατανομής.

   (α΄) Έστω τυχαίο δείγμα ανεξάρτητων παρατηρήσεων $\{x_1, \ldots, x_n\}$ από κατανομή Poisson με άγνωστη παράμετρο λ. Δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας του λ είναι η δειγματική μέση τιμή.

   (β΄) Φτιάξτε μια συνάρτηση στο `matlab` που θα δημιουργεί M δείγματα μεγέθους n από κατανομή Poisson με δεδομένη παράμετρο λ και θα υπολογίζει τη δειγματική μέση τιμή για κάθε ένα από τα M δείγματα. Στη συνέχεια θα κάνει κατάλληλο ιστόγραμμα των M δειγματικών μέσων τιμών και θα δίνει ως έξοδο το μέσο όρο από τις M δειγματικές μέσες τιμές. Καλέστε τη συνάρτηση για διαφορετικούς συνδυασμούς των M, n και λ. Είναι πάντα (και σύμφωνα με το αποτέλεσμα στο υποερώτημα 1α΄) το κέντρο της κατανομής της δειγματικής μέσης τιμής (που περιγράφεται από το ιστόγραμμα) στην τιμή του λ;

   Βοήθεια (`matlab`*):* Για τη δημιουργία των τυχαίων αριθμών χρησιμοποίησε τη συνάρτηση `poissrnd`.

2. Στην ανάλυση αξιοπιστίας (reliability analysis) στη μηχανική εξετάζεται συχνά ο ρυθμός αποτυχίας (failure rate) που συνήθως συμβολίζεται με λ. Έστω ο χρόνος ζωής του συστήματος X και έστω ότι η διαδικασία που ορίζει το χρόνο ζωής του συστήματος δεν έχει μνήμη. Αυτό σημαίνει πως η πιθανότητα να εμφανιστεί αποτυχία σε λιγότερο από κάποιο χρόνο s δεν εξαρτάται από το χρόνο που λειτουργούσε το σύστημα ως τώρα t,

$P(X < t + s|t) = P(X < s)$. Τότε ο χρόνος ζωής X ακολουθεί εκθετική κατανομή με παράμετρο λ και σππ

$$f_X(x;\lambda) = \frac{\lambda^{-\lambda x}}{e}, \tag{3.26}$$

Επαναλάβετε τα ερωτήματα 1α΄ και 1β΄ για την εκθετική κατανομή.

Βοήθεια (`matlab`*):* Για τη δημιουργία των τυχαίων αριθμών χρησιμοποίησε τη συνάρτηση `exprnd`.

3. Σε συνέχεια της προηγούμενης άσκησης, προσομοιώστε το χρόνο ζωής n μηχανικών συστημάτων δημιουργώντας n τιμές από εκθετική κατανομή με μέσο χρόνο ζωής $1/\lambda = 15$ μήνες. Με βάση αυτό το δείγμα υπολογίστε το 95% παραμετρικό διάστημα εμπιστοσύνης για το μέσο χρόνο ζωής και εξετάστε αν περιέχεται σε αυτό η τιμή $1/\lambda = 15$.

   (α΄) Υπολογίστε $M = 1000$ δείγματα μεγέθους $n = 5$. Σε τι ποσοστό βρίσκεται ο πραγματικός μέσος χρόνος ζωής μέσα στο 95% διάστημα εμπιστοσύνης;

   (β΄) Κάνετε το ίδιο για $M = 1000$ αλλά $n = 100$. Διαφέρει το ποσοστό αυτό από το παραπάνω;

   Βοήθεια (`matlab`*):* Για τον υπολογισμό διαστήματος εμπιστοσύνης και ελέγχου για τη μέση τιμή με χρήση της κατανομής Student κάλεσε τη συνάρτηση `ttest`.

4. Η τάση διακοπής εναλλασσόμενου ρεύματος ενός μονωτικού υγρού δηλώνει τη διηλεκτρική ανθεκτικότητα του. Πήραμε τις παρακάτω παρατηρήσεις της τάσης διακοπής (kV) σε κάποιο κύκλωμα κάτω από ορισμένες συνθήκες.

41	46	47	47	48	50	50	50	50	50	50	50
48	50	50	50	50	50	50	50	52	52	53	55
50	50	50	50	52	52	53	53	53	53	53	57
52	52	53	53	53	53	53	53	54	54	55	68

   (α΄) Βρείτε 95% διάστημα εμπιστοσύνης για τη διασπορά της τάσης διακοπής του κυκλώματος.

   (β΄) Από παλιότερες μετρήσεις είχαμε βρει πως η τυπική απόκλιση της τάσης διακοπής παρόμοιου κυκλώματος ήταν περίπου 5 kV. Με βάση το δείγμα κάνετε έλεγχο για την υπόθεση πως αυτή είναι η τυπική απόκλιση της τάσης διακοπής.

(γ΄) Βρείτε 95% διάστημα εμπιστοσύνης για τη μέση τάση διακοπής του κυκλώματος.

(δ΄) Μπορούμε να αποκλείσουμε ότι η μέση τάση διακοπής είναι 52 kV;

(ε΄) Κάνετε έλεγχο $\mathcal{X}^2$ καλής προσαρμογής σε κανονική κατανομή και βρείτε την p-τιμή του ελέγχου.

Βοήθεια (`matlab`*):* Για τον υπολογισμό διαστήματος εμπιστοσύνης και ελέγχου για τη διασπορά με χρήση της κατανομής $\mathcal{X}^2$ κάλεσε τη συνάρτηση `vartest`. Για τον έλεγχο $\mathcal{X}^2$ καλής προσαρμογής κάλεσε τη συνάρτηση `chi2gof`.

5. Ο θερμοπίδακας Old Faithful στην Αμερική είναι από τους πιο γνωστούς θερμοπίδακες για το μέγεθος αλλά και την κανονικότητα των εξάρσεων του (eruptions)
(δες `http://en.wikipedia.org/wiki/Old_Faithful`).
Στο αρχείο δεδομένων `eruption.dat` στην ιστοσελίδα του μαθήματος δίνονται στην πρώτη και δεύτερη στήλη 298 μετρήσεις (σε λεπτά) του διαστήματος αναμονής (waiting time) και της διάρκειας του ξεσπάσματος (duration) για το 1989 και στην τρίτη στήλη 298 μετρήσεις του διαστήματος αναμονής έξαρσης για το 2006. Για κάθε ένα από τα τρία μετρούμενα μεγέθη κάνετε τα παρακάτω.

(α΄) Βρείτε 95% διάστημα εμπιστοσύνης για την τυπική απόκλιση του μεγέθους και ελέγξτε αν είναι 10′ για την αναμονή και 1′ για τη διάρκεια.

(β΄) Βρείτε 95% διάστημα εμπιστοσύνης για τη μέση τιμή του μεγέθους και ελέγξτε αν είναι 75′ για την αναμονή και 2.5′ για τη διάρκεια.

(γ΄) Κάνετε έλεγχο $\mathcal{X}^2$ καλής προσαρμογής σε κανονική κατανομή και βρείτε την p-τιμή του ελέγχου.

Με βάση τις 298 μετρήσεις για το χρόνο αναμονής και διάρκειας έξαρσης το 1989, εξετάστε αν μπορείτε να δεχθείτε τον παρακάτω ισχυρισμό (αντιγραφή από τη διεύθυνση της Wikipedia): ”With an error of 10 minutes, Old Faithful will erupt 65 minutes after an eruption lasting less than 2.5 minutes or 91 minutes after an eruption lasting more than 2.5 minutes.”