

Εισαγωγή στη στατιστική ανάλυση δεδομένων

Α. Η στατιστική περιγραφή της μέτρησης φυσικών μεγεθών

1. Ο στοχαστικός χαρακτήρας της μέτρησης

Στη διαδικασία μέτρησης ενός φυσικού μεγέθους η πραγματική του τιμή είναι άγνωστη. Μόνο μετά από εκτέλεση μεγάλου αριθμού μετρήσεων και με την προϋπόθεση ότι η κατάσταση του αντίστοιχου συστήματος παραμένει κατά προσέγγιση ίδια περιμένουμε η μέση τιμή των μετρήσεων να συγκλίνει στη πραγματική τιμή του παρατηρούμενου μεγέθους. Αυτή η μορφή άγνοιας που είναι θεμελιώδης αρχή για τη Πειραματική Φυσική έχει τις ρίζες της στην έλλειψη πληροφορίας ενός παρατηρητή για το σύστημα που μελετά και πηγάζει από δυο κυρίως παράγοντες:

- Το υπό μελέτη σύστημα είναι κατά κανόνα ανοικτό, δηλ. αλληλεπιδρά, έστω και πολύ ασθενικά, με το περιβάλλον του (η μετρητική συσκευή αποτελεί μέρος του περιβάλλοντος). Μια πλήρης γνώση αυτής της αλληλεπίδρασης είναι αδύνατη καθώς απαιτεί τη διαχείριση απείρου πλήθους βαθμών ελευθερίας που αντιστοιχούν στο περιβάλλον.
- Οι αλληλεπιδράσεις μεταξύ των βαθμών ελευθερίας που απαρτίζουν το σύστημα είναι συνήθως μη αρμονικές και έχουν σαν αποτέλεσμα την εξαιρετικά ευαίσθητη εξάρτηση της χρονικής εξέλιξής του από τις αρχικές συνθήκες. Η πλήρης γνώση των αρχικών συνθηκών είναι επίσης πρακτικά αδύνατη καθώς απαιτεί τη διαχείριση απείρου πλήθους αριθμητικών ψηφίων.

Έτσι η έλλειψη πληροφορίας, που εμφανίζεται σαν απροσπέλαστη αδυναμία του παρατηρητή, εισάγει την έννοια του τυχαίου στη διαδικασία της μέτρησης φυσικών μεγεθών και επιδέχεται μαθηματική περιγραφή με τη χρήση στατιστικών μεθόδων. Οι δύο παράγοντες που προαναφέρθηκαν αφορούν την εμφάνιση στοχαστικής συμπεριφοράς κυρίως σε κλασικά μακροσκοπικά συστήματα. Θα μπορούσε κανείς να ισχυριστεί ότι στο μικρόκοσμο οι δύο παράγοντες που αναφέρθηκαν πιο πάνω είναι δυνατόν να τεθούν υπό έλεγχο και επομένως δεν έχουν επίπτωση στη διαδικασία μέτρησης. Όμως στα μικροσκοπικά συστήματα η εμφάνιση στοχαστικών χαρακτηριστικών έχει ακόμη πιο θεμελιώδη προέλευση καθώς επάγεται από τη

πιθανοκρατική περιγραφή που επιβάλλει η κβαντική τους υπόσταση. Έτσι, σύμφωνα με τη κβαντική θεωρία τα περισσότερα φυσικά φαινόμενα έχουν στοχαστικό χαρακτήρα. Ένα τυπικό παράδειγμα είναι η διάσπαση ραδιενεργών πυρήνων. Τα ραδιενεργά υλικά διασπώνται σε χρόνους που εμφανίζονται ως τυχαίοι. Για κάθε ραδιενεργή ουσία όμως υπάρχει καθορισμένη πιθανότητα κάποιος πυρήνας να διασπασθεί σε δεδομένο χρονικό διάστημα. Αυτή η πιθανότητα προσδιορίζεται στα πλαίσια της Κβαντικής Μηχανικής και εξαρτάται μόνο από το είδος του πυρήνα δηλαδή είναι η ίδια για όλους τους πυρήνες αυτού του είδους. Δεν υπάρχει τρόπος να προβλέψει κανείς το χρόνο στον οποίο θα διασπαστεί ένας ραδιενεργός πυρήνας καθώς η διαδικασία αυτή είναι καθαρά στοχαστική. Όταν όμως διασπασθεί ένα μεγάλο πλήθος από όμοιους πυρήνες τότε μπορεί να καθορισθεί με ακρίβεια ένας μέσος ρυθμός διάσπασης που χαρακτηρίζει μονοσήμαντα το είδος τους. Αν επιλέξει κανείς να μετρήσει το ρυθμό διάσπασης παρατηρώντας τον αριθμό διασπάσεων σε προκαθορισμένο μικρό χρονικό διάστημα εύρους Δt αυτός θα παρουσιάζει διακυμάνσεις γύρω από τη μέση τιμή. Η πειραματική μελέτη των πυρηνικών διασπάσεων μιας ραδιενεργού πηγής εστιάζεται στη στατιστική περιγραφή των διακυμάνσεων αυτών. Για να εξοικειωθεί με τη μεθοδολογία που απαιτείται για μελέτη τέτοιου τύπου ο ασκούμενος φοιτητής, καλείται στην άσκηση 5 να προσδιορίσει την κατανομή του αριθμού διασπάσεων μιας ραδιενεργού πηγής όπως την καταγράφει ένας ανιχνευτής Geiger-Mueller.

2. Τρόπος περιγραφής και πειραματικής μελέτης στοχαστικών διακυμάνσεων

Όπως προαναφέραμε λοιπόν συχνά στην εργαστηριακή μελέτη φυσικών μεγεθών και διαδικασιών με στοχαστικά χαρακτηριστικά το ζητούμενο είναι ο προσδιορισμός κατάλληλα ορισμένων κατανομών. Έστω A ένα παρατηρήσιμο μέγεθος (π.χ. ο αριθμός διασπάσεων που καταγράφει ένας ανιχνευτής παρουσία μιας ραδιενεργού πηγής) που σε μια μέτρηση μπορεί να πάρει κάποιο φάσμα τιμών. Οι τιμές αυτές ονομάζονται ενδεχόμενα και σε μια πειραματική μελέτη αυτό που μπορεί να προσδιορίσει κανείς είναι η συχνότητα εμφάνισης κάθε ενδεχομένου. Αν θεωρήσουμε για απλότητα ότι το φυσικό μέγεθος A παίρνει διακριτό και πεπερασμένο πλήθος τιμών: a_1, a_2, \dots, a_M και ότι κατά την εκτέλεση ενός πειράματος μέτρησης του A βρίσκουμε το ενδεχόμενο i ($i=1, 2, \dots, M$), δηλ. την τιμή a_i , O_i φορές, τότε αν το σύνολο των μετρήσεων μας είναι N η πιθανότητα εμφάνισης της τιμής a_i εκτιμάται

από το πηλίκο $\frac{O_i}{N}$ που τείνει στην ακριβή τιμή της πιθανότητας p_i για $N \rightarrow \infty$. Αν η φυσική διαδικασία που μελετάμε στο πείραμα είναι στοχαστική τότε μια θεωρητική πρόβλεψη θα πρέπει να προκαθορίζει την τιμή των πιθανοτήτων αυτών. Αυτή τη πληροφορία θα μπορούσε π.χ. να παρέχει ένας υπολογισμός από πρώτες αρχές στα πλαίσια της Κβαντικής Μηχανικής. Εναλλακτικά, μπορεί κανείς, βασιζόμενος στις προσεγγίσεις των πιθανοτήτων p_i μέσω των μετρήσεων, να διατυπώσει ένα απλό μαθηματικό πρότυπο για να αναπαράγει τις τιμές αυτές με ανεκτή ακρίβεια. Σε οποιαδήποτε περίπτωση, ο πειραματικός φυσικός, κάνοντας την υπόθεση ότι το μαθηματικό αυτό πρότυπο (ή το αποτέλεσμα του κβαντικού υπολογισμού) ισχύει, καλείται να ελέγξει αν η υπόθεση αυτή είναι αποδεκτή ή όχι συγκρίνοντας με τις μετρήσεις του. Επειδή όμως οι μετρήσεις έχουν στατιστικές διακυμάνσεις αλλά και συστηματικά σφάλματα η αποδοχή ή όχι της υπόθεσης του δεν είναι απόλυτη αλλά γίνεται με κάποιο επίπεδο εμπιστοσύνης. Επομένως έρχεται αντιμέτωπος με τον ορισμό εννοιών που αρχικά εμφανίζονται να έχουν υποκειμενικό χαρακτήρα όπως π.χ. ικανοποιητική ακρίβεια ή επίπεδο εμπιστοσύνης. Για τον αυστηρό προσδιορισμό των εννοιών αυτών έχουν αναπτυχθεί τα κατάλληλα μεθοδολογικά εργαλεία που περιγράφονται στην επόμενη ενότητα.

3. Το κριτήριο χ^2 και η μέθοδος προσαρμογής

Το κύριο στατιστικό εργαλείο για τον πειραματικό έλεγχο μιας θεωρητικής πρόβλεψης/υπόθεσης είναι το **κριτήριο χ^2** . Οι ακόλουθες σκέψεις μας βοηθούν να κατανοήσουμε τον ορισμό του. Ας υποθέσουμε αρχικά ότι με βάση μια θεωρητική υπόθεση μπορούμε να υπολογίσουμε τις αναμενόμενες συχνότητες e_1, e_2, \dots, e_M κάποιων πιθανών ενδεχομένων E_1, E_2, \dots, E_M . Εάν τις συχνότητες αυτές προσπαθήσουμε να τις προσδιορίσουμε πειραματικά, οι τιμές O_1, O_2, \dots, O_M που θα προκύψουν (με αντίστοιχα σφάλματα δO_i), συνήθως διαφέρουν απ' αυτές που περιμέναμε. Έτσι μας ενδιαφέρει να κρίνουμε εάν η διαφορά αυτή είναι **σημαντική**.

Για να γίνει αυτό θεωρούμε ως μέτρο της διαφοράς ανάμεσα στη θεωρία και το πείραμα την τυχαία μεταβλητή

$$\chi^2 = \sum_{k=1}^M \frac{(O_k - e_k)^2}{e_k} \quad (1)$$

όπου M είναι το σύνολο των διαφορετικών ενδεχομένων. Προφανώς θα ισχύει:

$$\sum_{k=1}^M e_k = \sum_k O_k = N_{ολ}$$

όπου $N_{ολ}$ είναι το άθροισμα των συχνοτήτων, δηλαδή ο συνολικός αριθμός των μετρήσεων.

Η μεταβλητή (1), όταν είναι $N_{ολ} \gg O_k$, $e_k \gg 1$, ακολουθεί την λεγόμενη κατανομή χ^2 με ν βαθμούς ελευθερίας:

$$f_\nu(\chi^2) = \frac{1}{2^{\nu/2} \Gamma(\nu/2)} (\chi^2)^{(\nu/2)-1} e^{-\chi^2/2} \quad (2)$$

Οι βαθμοί ελευθερίας της κατανομής είναι όσοι και οι ανεξάρτητοι όροι του αθροίσματος (1): $M - (I+k)$ όπου k είναι το πλήθος των παραμέτρων (αν υπάρχουν) της θεωρητικής κατανομής που προσδιορίζονται πειραματικά.

Το κριτήριο χ^2 , το κριτήριο, δηλαδή, με βάση το οποίο θα αποφασίσουμε εάν είναι σημαντική η διαφορά ανάμεσα στη θεωρία και το πείραμα, τίθεται ως εξής: Εάν, πραγματοποιώντας το πείραμα, υπολογίσουμε τιμή της χ^2 μεγαλύτερη από κάποια κρίσιμη τιμή (έστω χ_p^2) τότε απορρίπτουμε την υπόθεσή μας (στην αντίθετη περίπτωση η υπόθεση δεν απορρίπτεται). Η πιθανότητα να υπερβούμε αυτήν την κρίσιμη τιμή:

$$\tilde{P}(\chi_\nu^2 > \chi_p^2) = \int_0^{\chi_p^2} dt f_\nu(t) \equiv 1 - \tilde{p} \quad (3)$$

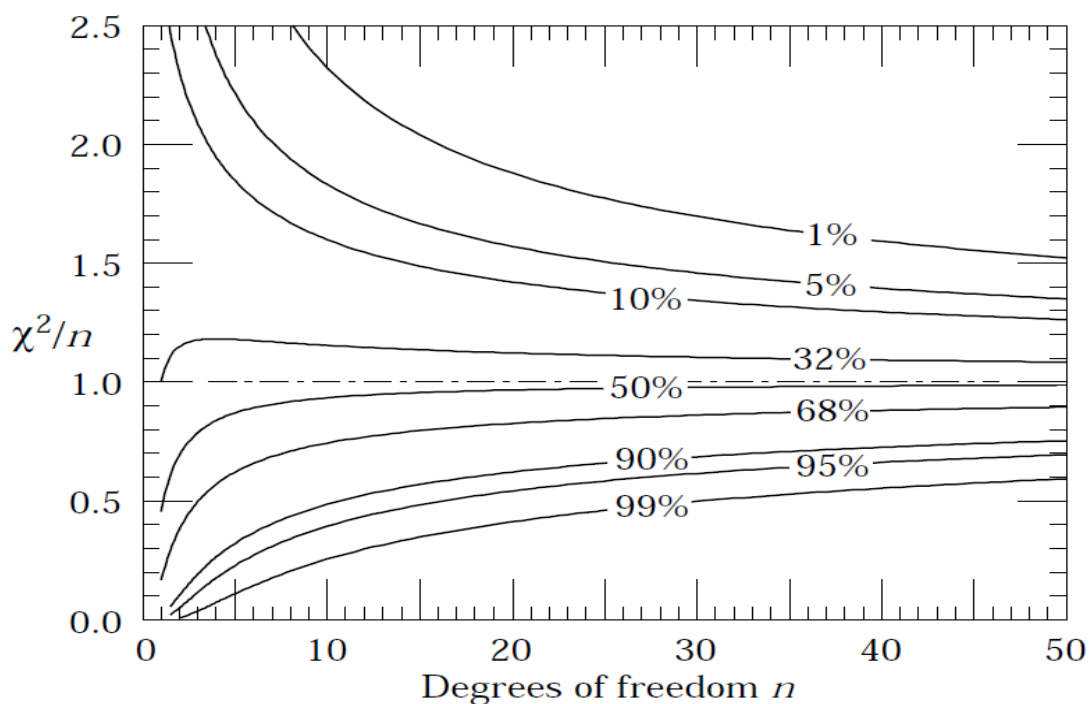
ορίζει και την πιθανότητα να κάνουμε λάθος απορρίπτοντας την υπόθεσή μας ή, ισοδύναμα, αντιπροσωπεύει το σφάλμα που δεχόμαστε να κάνουμε προκειμένου να την απορρίψουμε. Λέγεται **επίπεδο σημαντικότητας** ενώ το **επίπεδο εμπιστοσύνης** εκφράζει την εμπιστοσύνη που έχουμε στην απόρριψη που κάνουμε. Αν θεωρήσουμε ότι $|O_k - e_k| \approx \sqrt{e_k}$ (στατιστικό σφάλμα στην ιδανική περίπτωση άπειρων μετρήσεων) τότε είναι προφανές ότι $\chi^2 \approx M$ και το χ^2 ανά βαθμό ελευθερίας θα είναι $\chi^2 / dof \approx 1$. Από το κατωτέρω διάγραμμα προκύπτει σε αυτή τη περίπτωση ότι το επίπεδο εμπιστοσύνης απόρριψης είναι $\approx 50\%$. Αυτό είναι ένα πολύ συνηθισμένο σενάριο για τη τιμή του χ^2 . Για να απορρίψει ένας πειραματικός μια θεωρητική υπόθεση θα πρέπει το αντίστοιχο επίπεδο εμπιστοσύνης απόρριψης να είναι μεγαλύτερο ή ίσο από 95%. Να σημειωθεί ότι τόσο οι πολύ μεγάλες όσο

και οι πολύ μικρές τιμές του χ^2 έχουν μικρή πιθανότητα εμφάνισης γι' αυτό και το κριτήριο είναι καλό να εφαρμόζεται αμφίπλευρα.

Το κριτήριο χ^2 μπορεί να χρησιμοποιηθεί και για την εκτίμηση παραμέτρων προσαρμογής. Ο ορισμός (1) σε αυτή τη περίπτωση δεν είναι ο καταλληλότερος και είναι δόκιμο να γραφεί ως:

$$\chi^2 = \sum_{k=1}^M \left(\frac{O_k - e_k}{\delta O_k} \right)^2 \quad (4)$$

όπου δO_k είναι τα σφάλματα στη μέτρηση των O_k . Στην περίπτωση της προσαρμογής οι παράμετροι που καθορίζουν τις θεωρητικές συχνότητες e_k δεν υπολογίζονται από τις πειραματικές μετρήσεις αλλά προσδιορίζονται έτσι ώστε για δεδομένα O_k και δO_k να ελαχιστοποιείται το άθροισμα (4).



Παρατήρηση: Άλλη μια χρήση του κριτηρίου χ^2 είναι ο έλεγχος (δηλ. προσδιορισμός συστηματικού σφάλματος) μετρητικών συσκευών. Έστω ότι μια θεωρητική υπόθεση έχει ήδη ελεγχθεί πειραματικά από μια σειρά πειραμάτων μεγάλης ακρίβειας έτσι ώστε να θεωρείται ότι ισχύει με σχεδόν μηδενικό επίπεδο εμπιστοσύνης απόρριψης. Φαντασθείτε τώρα ότι εκτελείτε στο εργαστήριό σας ένα πείραμα ελέγχου αυτής της θεωρητικής υπόθεσης χρησιμοποιώντας το κριτήριο χ^2 . Είναι φανερό ότι η εύρεση μεγάλης τιμής του χ^2 θα πρέπει να αναχθεί σε δυσλειτουργία της μετρητικής συσκευής αν φυσικά έχει κανείς εξασφαλίσει τη σωστή εκτέλεση του πειράματος και την φυσιολογική κατάσταση της ραδιενεργού πηγής.

B. Η προσομοίωση στοχαστικών διαδικασιών

1. Η μέθοδος προσομοίωσης Monte-Carlo

Η πρόοδος της επιστήμης της πληροφορίας και η τεχνολογική της εξέλιξη μέσω της δημιουργίας υπολογιστικών διατάξεων άνοιξε το δρόμο για ένα εναλλακτικό τρόπο ελέγχου μιας θεωρητικής υπόθεσης μέσω της λεγομένης διαδικασίας **προσομοίωσης**. Στη διαδικασία αυτή δημιουργείται αλγόριθμος που περιέχει όλες τις θεωρητικές παραδοχές που αφορούν τη μελέτη ιδιοτήτων ενός πολύπλοκου φυσικού συστήματος. Εφαρμόζεται όταν η μελέτη αυτή δεν επιδέχεται αναλυτικό ή αριθμητικό χειρισμό λόγω έλλειψης πληροφορίας και επομένως ύπαρξης στοχαστικών χαρακτηριστικών (όπως αναφέραμε στην εισαγωγή). Με τον αλγόριθμο προσομοίωσης γίνεται ο υπολογισμός ενός σχετικά μεγάλου αριθμού περιπτώσεων που αφορούν διαφορετικές υλοποιήσεις των στοχαστικών χαρακτηριστικών, από τις οποίες, με στατιστικές μεθόδους, εξάγονται οι ζητούμενες πληροφορίες για τις ιδιότητες του συστήματος. Στην εποχή μας η μέθοδος προσομοίωσης είναι το βασικό εργαλείο κάθε ερευνητή. Κυριαρχούν δύο κατηγορίες αλγορίθμων προσομοίωσης: (i) η **μοριακή δυναμική** που ενδείκνυται για τη μελέτη προβλημάτων που αφορούν **τη δυναμική** δηλ. **τη χρονική εξέλιξη** μετρήσιμων **ιδιοτήτων** ενός πολύπλοκου συστήματος και (ii) η μέθοδος **Monte Carlo** που εφαρμόζεται για τη μελέτη **στατιστικών ιδιοτήτων**

πολύπλοκων συστημάτων. Στη παρούσα άσκηση θα εξοικειωθούμε με τη μέθοδο Monte Carlo.

Η μέθοδος Monte Carlo αναπτύχθηκε από τους von Neuman, Ulam και Metropolis στο τέλος του Β' Παγκοσμίου Πολέμου για τη μελέτη της διάχυσης νετρονίων σε ύλη που μπορεί να υποστεί σχάση [1]. Το όνομα "Monte Carlo" προέρχεται από το ότι αυτή η μέθοδος χρησιμοποιεί τυχαίους αριθμούς, όμοιους με αυτούς που προκύπτουν στη διάρκεια ενός παιχνιδιού ρουλέτας. Για μια αξιόπιστη περιγραφή των χαρακτηριστικών ενός φυσικού συστήματος χρειάζεται μεγάλο πλήθος (τυπικά $\sim 10^{10}$) από τυχαίους αριθμούς και η παραγωγή τους με φυσικές διαδικασίες είναι χρονοβόρα και ακριβή. Έτσι επιστρατεύονται οι ηλεκτρονικοί υπολογιστές για αυτό το σκοπό. Όμως σε ένα ηλεκτρονικό υπολογιστή η παραγωγή των τυχαίων αριθμών θα γίνει μέσω ενός αλγορίθμου και επομένως οι παραγόμενοι αριθμοί δεν μπορεί να είναι πραγματικά τυχαίοι. Ο σωστός όρος για τον χαρακτηρισμό αυτών των αριθμών είναι ο όρος: "ψευδοτυχαίοι αριθμοί". Καθώς όμως στις προσομοιώσεις χρησιμοποιούνται σχεδόν αποκλειστικά τέτοιου τύπου αριθμοί έχει επικρατήσει η ονομασία "τυχαίοι" και για τους αριθμούς που παράγονται μέσω κατάλληλων αλγορίθμων. Εκτενέστερη συζήτηση για τον τρόπο υλοποίησης τυχαίων αριθμών σε ένα σύγχρονο ψηφιακό υπολογιστικό σύστημα καθώς και παρουσίαση κάποιων ποιοτικών χαρακτηριστικών τους γίνεται στο Παράρτημα Α. Ο ακριβής αλγόριθμος εφαρμογής της μεθόδου Monte Carlo σε προσομοίωση πολύπλοκων φυσικών διαδικασιών εξαρτάται από το υπό μελέτη πρόβλημα [2].

2. Η Monte-Carlo (ή στοχαστική) ολοκλήρωση

Συχνά στη μελέτη πολύπλοκων συστημάτων δεν αρκεί η προσομοίωση μιας κατανομής αλλά χρειάζεται και ο ακριβής υπολογισμός μέσων τιμών. Έτσι στην ουσία πρέπει κανείς να επινοήσει μεθόδους αριθμητικής ολοκλήρωσης, εν γένει σε πολυδιάστατους χώρους, με χρήση τυχαίων αριθμών. Ο τρόπος αυτός υπολογισμού ολοκληρωμάτων καλείται **στοχαστική ολοκλήρωση** και στη συνέχεια δίνουμε μια πολύ συνοπτική περιγραφή της.

Ας θεωρήσουμε για απλότητα το ολοκλήρωμα:

$$I = \int_0^1 f(x) dx ,$$

μίας συνάρτησης $f(x)$ στο διάστημα $[0,1]$. Εάν το x θεωρηθεί τυχαία μεταβλητή που ακολουθεί την ομοιόμορφη κατανομή τότε το ολοκλήρωμα είναι απλά η μέση τιμή της συνάρτησης της τυχαίας μεταβλητής:

$$I = E[f(x)] .$$

Επομένως είναι δυνατόν να γίνει **στατιστικός υπολογισμός** της μέσης τιμής, άρα και του ολοκληρώματος. Για αυτό το σκοπό λαμβάνεται ένα δείγμα στοιχείων της τυχαίας μεταβλητής, οπότε η στατιστική εκτίμηση του ολοκληρώματος δίνεται από την έκφραση:

$$\langle f(x) \rangle = \frac{1}{N} \sum_{i=1}^N f(x_i) = \frac{1}{N} \sum_{j=1}^k n_j f(x_j) \quad (5)$$

Όπου n_j (με $\sum_{j=1}^k n_j = N$) η συχνότητα εμφάνισης της κάθε τιμής της στοχαστικής μεταβλητής x . Επειδή τα στοιχεία του δείγματος πρέπει να είναι τυχαία, ανεξάρτητα μεταξύ τους και ισοπίθανα, το δείγμα αποτελείται από **τυχαίους αριθμούς οι οποίοι ακολουθούν την ομοιόμορφη κατανομή**.

Με βάση τον νόμο των μεγάλων αριθμών πράγματι, $\langle f(x) \rangle \rightarrow E[f(x)]$ όταν $N \rightarrow \infty$. Επί πλέον εάν θεωρήσουμε την $\langle f(x) \rangle$ ως τυχαία μεταβλητή, σύμφωνα με το κεντρικό οριακό θεώρημα (βλέπε Παράρτημα Β) αυτή ακολουθεί κανονική κατανομή με μέση τιμή την μέση τιμή της συνάρτησης, $E[f(x)]$, και διασπορά $\frac{\sigma(f)}{\sqrt{N}}$, όπου $\sigma(f)$ η διασπορά της συνάρτησης $f(x)$. Επίσης από το ίδιο θεώρημα συνάγεται ότι

$$P \left[-a \frac{\sigma(f)}{\sqrt{N}} \leq \langle f(x) \rangle - E[f(x)] \leq b \frac{\sigma(f)}{\sqrt{N}} \right] = \frac{1}{\sqrt{2\pi}} \int_{-a}^b e^{-\frac{t^2}{2}} dt \quad (6)$$

Από την ανωτέρω σχέση προκύπτει και το σφάλμα του υπολογισμού του ολοκληρώματος. Ας σημειωθεί ότι το σφάλμα αυτό είναι **στατιστικό**. Τα πολλαπλάσια της διασποράς δεν αποτελούν άνω η κάτω φράγμα του ολοκληρώματος αλλά χρησιμοποιούνται για τον καθορισμό της πιθανότητας να απέχει η στατιστική εκτίμηση από την πραγματική τιμή το αντίστοιχο πολλαπλάσιο της διασποράς. Για παράδειγμα εάν $a = b = 3$, η τιμή του ολοκληρώματος είναι **0,997**. Δηλαδή η στατιστική εκτίμηση απέχει από την πραγματική τιμή λιγότερο από $3 \frac{\sigma(f)}{\sqrt{N}}$, με πιθανότητα **0.997**.

Βέβαια ο υπολογισμός της διασποράς $\sigma(f)$ της συνάρτησης f είναι το ίδιο δύσκολος με τον υπολογισμό του ολοκληρώματος. Για τον λόγο χρησιμοποιείται η Monte – Carlo εκτίμηση της διασποράς:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N [f(x_i) - \langle f(x) \rangle]^2 = \frac{1}{N-1} \sum_{i=1}^N f^2(x_i) - \frac{N}{N-1} \langle f(x) \rangle^2 \quad (7)$$

οπότε το σφάλμα καθορίζεται από την έκφραση:

$$\varepsilon_N = \sqrt{s^2 / N}$$

Παράδειγμα 1: Το ολοκλήρωμα

$$I = \int_0^1 f(x) dx$$

μπορεί να προσεγγιστεί από την τιμή

$$I_N = \sum_{i=1}^N f(x_i) / N$$

με τις τιμές x_i να ακολουθούν ομοιόμορφη κατανομή στο διάστημα $[0, 1)$. Το ολοκλήρωμα π.χ.:

$$I = \int_0^1 e^{-x^2} dx$$

το οποίο δεν επιδέχεται εύκολο αναλυτικό υπολογισμό μπορεί να προσεγγιστεί από

το άθροισμα: $\frac{1}{N} \sum_{i=1}^N e^{-x_i^2}$ όπου τα x_i είναι τυχαίοι αριθμοί ομοιόμορφα κατανεμημένοι

στο $[0,1)$. Μπορείτε να το επιβεβαιώσετε αυτό χρησιμοποιώντας τους 10 αριθμούς:

0.0000, 0.0085, 0.6014, 0.8916, 0.9680, 0.1897, 0.5150, 0.3980, 0.2629, 0.7435

Η ακριβής τιμή του ολοκληρώματος είναι $I=0.747$.

Παράδειγμα 2: Το ολοκλήρωμα: $I_{ab} = \int_a^b f(x) dx$ μπορεί να υπολογιστεί με

παρόμοιο τρόπο αρκεί να κάνει κανείς τον γραμμικό μετασχηματισμό:

$$x = a + (b - a) t \quad (8\alpha)$$

οπότε θα γίνει:

$$I_{ab} = (b-a) \int_0^1 f(a + (b-a)t) dt \quad (8\beta)$$

και το t θα είναι τυχαία μεταβλητή ομοιόμορφα κατανεμημένη στο $[0,1)$.

Παράδειγμα 3: Το πλεονέκτημα της στοχαστικής ολοκλήρωσης είναι ότι γενικεύεται με πολύ απλό τρόπο σε πολυδιάστατους χώρους. Αρκεί να ορίσει κανείς τυχαία σημεία που κατανέμονται ομοιόμορφα σε ένα μοναδιαίο υπερκύβο της αντίστοιχης διάστασης.

Έτσι για παράδειγμα το πολυδιάστατο ολοκλήρωμα της μορφής:

$$I = \int_0^1 dx_1 \int_0^1 dx_2 \dots \int_0^1 dx_n f(x_1, x_2, \dots, x_n)$$

μπορεί να εκτιμηθεί από το άθροισμα:

$$S_N = \frac{1}{N} \sum_{i=1}^N f(r_1^{(i)}, r_2^{(i)}, \dots, r_n^{(i)})$$

όπου τα $r_j^{(m)}$ ($m = 1, 2, \dots, N$ και $j = 1, 2, \dots, n$) είναι τυχαίοι αριθμοί ομοιόμορφα κατανεμημένοι στο $[0,1)$. Με άλλα λόγια ο υπολογισμός του ολοκληρώματος γίνεται χρησιμοποιώντας n -άδες τυχαίων αριθμών με ομοιόμορφη κατανομή στον υπερκύβο n διαστάσεων με πλευρά 1 και πρώτη κορυφή στο $(0,0,\dots,0)$. Είναι σχετικά απλό να δείξει κανείς ότι το σφάλμα του υπολογισμού αυτού είναι ανάλογο του $\frac{1}{\sqrt{N}}$

ανεξάρτητα της διάστασης n του ολοκληρώματος. Εν γένει σε ένα πολλαπλό ολοκλήρωμα, αν γίνουν αναλυτικά μερικές ολοκληρώσεις, αυτό θα έχει σαν αποτέλεσμα την αύξηση της ακρίβειας της προσέγγισης. Σε άλλες περιπτώσεις, ιδιαίτερα αν το διάστημα $[a, b]$ δεν είναι φραγμένο, άλλοι μετασχηματισμοί είναι πιο κατάλληλοι για τη μετατροπή σε ένα ολοκλήρωμα της μορφής (8β).

Παράρτημα Α

Παράδειγμα αλγορίθμου γέννησης τυχαίων αριθμών:

Η επαναληπτική σχέση:

$$x_{v+1} = px_v + q \cdot \text{Mod } \Gamma \quad (12)$$

παρέχει με κατάλληλη εκλογή των p , q , Γ τυχαίους αριθμούς με ομοιόμορφη κατανομή στο διάστημα $[0, \Gamma-1]$. Με διαίρεση δια του Γ τα x_v δίνουν τυχαίους αριθμούς με ομοιόμορφη κατανομή στο διάστημα $[0, 1)$.

Οι τιμές των παραμέτρων p , q και Γ επιλέγονται έτσι ώστε να εξασφαλίσει κανείς ότι τα διαδοχικά x_v βρίσκονται στο $[0, \Gamma-1]$, δεν είναι συσχετισμένα μεταξύ τους και δεν ακολουθούν κάποια περιοδική δομή. Εμπειρικά μια καλή επιλογή είναι η ακόλουθη:

- p ακέραιος τέτοιος ώστε $p \text{ Mod } 8 = 5$ και $\sqrt{\Gamma} < p < \Gamma - \sqrt{\Gamma}$.
- q περιττός ακέραιος τέτοιος ώστε $q/\Gamma=0.211$.
- $\Gamma = 2^{m-1}$ όπου m ο μέγιστος αριθμός bits μιας “λέξης” που αποθηκεύεται στη μνήμη του επεξεργαστή. Για Intel 386 και κάτω $m=16$ ενώ για Intel 486 και πάνω $m=32$.

Συνήθως στις προσομοιώσεις χρησιμοποιεί κανείς ψευδοτυχαίους αριθμούς ομοιόμορφα κατανεμημένους στο $[0,1)$. Θα συζητήσουμε λίγο αργότερα για το πως μπορεί κανείς από μια ακολουθία τέτοιων αριθμών να παράγει ένα σύνολο ψευδοτυχαίων αριθμών που περιγράφονται από άλλη κατανομή. Φυσικά υπάρχουν πολλοί αλγόριθμοι που μπορούν να παράγουν ψευδοτυχαίους ομοιόμορφα κατανεμημένους στο $[0,1)$. Για να αποφανθεί κανείς ποιος από αυτούς τους αλγόριθμους είναι καλύτερος χρειάζονται κάποια τεστ δηλ. κριτήρια ελέγχου της ποιότητάς τους. Το απλούστερο ίσως από αυτά τα τεστ είναι να επιχειρήσει κανείς με τους ψευδοτυχαίους αριθμούς που παράγονται από μια γεννήτρια να γεμίσει ένα απλό κυβικό πλέγμα με L^3 κορυφές. Για το σκοπό αυτό πρέπει να υπολογιστεί το ποσοστό κατάληψης του πλέγματος. Ορίζουμε λοιπόν σε κάθε κορυφή τον αριθμό $n(k_1, k_2, k_3)$ με $k_i=1, 2, \dots, L$ και $i=1, 2, 3$ που έχει αρχικά μηδενική τιμή ($n=0$) για όλες τις

κορυφές του πλέγματος. Χρησιμοποιώντας τρεις ψευδοτυχαίους x_1, x_2, x_3 ομοιόμορφα καταναμημένους στο $[0,1)$ υπολογίζουμε τις συντεταγμένες της κορυφής του πλέγματος που θα καταληφθεί, από τις σχέσεις: $k_i = 1 + x_i L$ με $k_i = 1, 2, \dots, L$ και $i=1, 2, 3$. Επαναλαμβάνουμε τη διαδικασία $t L^3$ φορές με t της τάξης του 10 και καταμετρούμε τον αριθμό των κορυφών με μη μηδενική τιμή του $n(k_1, k_2, k_3)$. Θεωρητικά περιμένουμε ο αριθμός των άδειων κορυφών να μειώνεται με εκθετικό τρόπο $\sim \exp(-t)$. Έτσι ένας καλός αλγόριθμος θα πρέπει να μην αφήνει άδειες κορυφές για $L=20$. Αν εφαρμόσει κανείς τον αλγόριθμο που προτάθηκε προηγουμένως για $L=20$ και $t = 10$ θα βρει ότι οι άδειες κορυφές είναι περίπου 2000 (από σύνολο 8000) κάτι που υποδεικνύει ότι σ' αυτόν τον αλγόριθμο υπάρχουν συσχετίσεις και χρειάζεται περαιτέρω βελτίωση.

Ας δούμε στη συνέχεια πως μπορεί να παράγει κανείς τυχαίους αριθμούς που ακολουθούν μια αυθαίρετη κατανομή $p(x)$. Θα παρουσιάσουμε δύο μεθόδους που είναι και οι πλέον διαδεδομένες.

Μέθοδοι παραγωγής τυχαίων αριθμών με κατανομή $p(x)$, x στο $[a,b)$

- (i) **Μέθοδος αντιστροφής:** αν είναι δυνατόν να υπολογιστεί το ολοκλήρωμα $F(x) = \int_a^x p(z) dz$ αναλυτικά τότε αποδεικνύεται εύκολα ότι η μεταβλητή $x = F^{-1}(\xi)$ ακολουθεί την $p(x)$ υπό την προϋπόθεση ότι η ξ ακολουθεί την ομοιόμορφη κατανομή στο $[0,1)$.
- (ii) **Μέθοδος απόρριψης:** έστω w η μέγιστη τιμή της $p(x)$ στο $[a,b)$. Επιλέγουμε δύο τυχαίους αριθμούς r_1 και r_2 ομοιόμορφα καταναμημένους στο $[0,1)$. Μετασχηματίζουμε τον r_1 σύμφωνα με τη σχέση: $x = a + (b-a) r_1$ έτσι ώστε ο τυχαίος αριθμός x να είναι ομοιόμορφα καταναμημένος στο $[a,b)$. Υπολογίζουμε κατόπιν τον λόγο $p(x)/w$. Αν ισχύει $r_2 < \frac{p(x)}{w}$ τότε η τιμή x γίνεται αποδεκτή ενώ στην αντίθετη περίπτωση απορρίπτεται. Το σύνολο των αποδεκτών τιμών του x σε ένα μεγάλο πλήθος επαναλήψεων της διαδικασίας αυτής, ακολουθεί την κατανομή $p(x)$.

Παράρτημα Β

Έστω N τυχαίες μεταβλητές X_1, X_2, \dots, X_N που περιγράφονται όλες από την ίδια κατανομή. Για την κατανομή αυτή υποθέτουμε μόνο ότι χαρακτηρίζεται από πεπερασμένη μέση τιμή μ και διασπορά σ^2 . Το **κεντρικό οριακό θεώρημα** αφορά την κατανομή του αθροίσματος αυτών των N μεταβλητών και αποτελεί το δεύτερο θεμελιώδες θεώρημα της θεωρίας πιθανοτήτων (μετά από το θεώρημα των μεγάλων αριθμών). Πιο συγκεκριμένα προβλέπει ότι το άθροισμα $S_N = \sum_{i=1}^N X_i$ ορίζει μια νέα

στοχαστική μεταβλητή $Z_N = \frac{S_N - N\mu}{\sigma\sqrt{N}}$ η οποία στο όριο $N \rightarrow \infty$ ακολουθεί τη

καθιερωμένη κανονική κατανομή $N(0,1)$ δηλαδή μια κατανομή Gauss με μέση τιμή 0 και διασπορά 1. Σύμφωνα με το θεώρημα αυτό λοιπόν το άθροισμα S_N θα ακολουθεί προσεγγιστικά και αυτό μια κατανομή Gauss με μέση τιμή $N\mu$ και διασπορά $\sigma\sqrt{N}$.

Μια γενίκευση του κεντρικού οριακού θεωρήματος που διατυπώθηκε από τον Lyapunov προβλέπει ότι ακόμη και αν κάθε μεταβλητή X_i στο άθροισμα ακολουθεί διαφορετική κατανομή με μέση τιμή μ_i και διασπορά σ_i^2 τότε η κανονικοποιημένη

μεταβλητή $Z_N = \frac{S_N - m_N}{s_N}$ όπου $m_N = \sum_{i=1}^N \mu_i$ και $s_N^2 = \sum_{i=1}^N \sigma_i^2$ στο όριο

$N \rightarrow \infty$ ακολουθεί και αυτή τη καθιερωμένη κανονική κατανομή $N(0,1)$!

Βιβλιογραφία

[1] J. Von Neumann, and S. Ulam, *Random ergodic theorems*, Bull. Am. Math. Soc. **51**, 660 (1945); N. Metropolis, and S. Ulam, *The Monte Carlo Method*, J. Am. Stat. Ass. **44**, 335 (1949); J. Von Neumann, *Various techniques used in connection with random digits*, US Nat. Bur. Stand. Appl. Math. Ser. **12**, 36 (1951).

[2] M. E. J. Newman, and G. T. Barkema, “*Monte Carlo methods in Statistical Physics*”, Cambridge University Press, 1999.

Στατιστική μελέτη ραδιενεργών διασπάσεων και εφαρμογή της μεθόδου προσομοίωσης Monte Carlo

Σκοπός της άσκησης: Μελέτη των στατιστικών διακυμάνσεων του ρυθμού διάσπασης μιας σταθερής ραδιενεργού πηγής και εξοικείωση με τη μέθοδο προσομοίωσης Monte Carlo.

1. Εισαγωγή

Η παρούσα άσκηση έχει σαν αντικείμενο αρχικά τη πειραματική καταγραφή των διακυμάνσεων του ρυθμού διάσπασης μιας ραδιενεργού πηγής και στη συνέχεια τη στατιστική τους μελέτη.

2. Θεωρητικό υπόβαθρο

2.1 Ραδιενεργές διασπάσεις

Έστω ότι η πιθανότητα διάσπασης ενός ραδιενεργού πυρήνα συγκεκριμένου είδους σε χρονικό διάστημα Δt είναι γνωστή και ίση με $\lambda \Delta t$ όπου λ είναι ο μέσος ρυθμός διάσπασης που χαρακτηρίζει αυτό το είδος των πυρήνων και καλείται **σταθερά διάσπασης**. Αντίστοιχα η πιθανότητα να μην διασπαστεί ένας πυρήνας στο διάστημα Δt θα είναι $1 - \lambda \Delta t$. Αν θεωρήσουμε ότι σε μια συλλογή από N πυρήνες κάθε πυρήνας διασπάται ανεξάρτητα από τους άλλους τότε περιμένει κανείς η πιθανότητα να έχουν διασπασθεί n πυρήνες στο διάστημα Δt να δίνεται από την διωνυμική κατανομή:

$$P(n, \Delta t) = \frac{N!}{(N-n)!n!} (\lambda \Delta t)^n (1 - \lambda \Delta t)^{N-n} \quad (1)$$

Όταν $N \gg 1$ και $\lambda \Delta t \ll 1$ με $N\lambda\Delta t = \mu$ σταθερό η κατανομή αυτή τείνει στη κατανομή Poisson:

$$P(n, \Delta t) = \frac{\mu^n}{n!} e^{-\mu} \quad (2)$$

όπου $\mu = \lambda N \Delta t$ είναι ο μέσος αριθμός διασπάσεων στο διάστημα Δt . Όταν επιπλέον ισχύει $\mu \gg 1$ η κατανομή αυτή τείνει στη κανονική (Gaussian):

$$P(n, \Delta t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \quad (3)$$

με διασπορά $\sigma^2 = \mu$.

Ας θεωρήσουμε τώρα τη μεταβολή του πληθυσμού των ραδιενεργών πυρήνων στο χρονικό διάστημα $[t, t + \Delta t]$. Έστω $N(t)$ το πλήθος των αδιάσπαστων πυρήνων τη χρονική στιγμή t . Ο αριθμός των αδιάσπαστων πυρήνων τη χρονική στιγμή $t + \Delta t$ θα δίνεται από τη σχέση: $N(t + \Delta t) = N(t) - \mu$ όπου, όπως προαναφέραμε, $\mu = \lambda N(t) \Delta t$ είναι ο μέσος αριθμός πυρήνων που διασπάστηκαν στο διάστημα $[t, t + \Delta t]$. Θα ισχύει λοιπόν:

$$\frac{N(t + \Delta t) - N(t)}{\Delta t} = -\lambda N(t) \quad (4)$$

και παίρνοντας το όριο $\Delta t \rightarrow 0$ καταλήγουμε στη σχέση:

$$-\frac{dN}{dt} = \lambda N \quad (5)$$

όπου λN είναι η **ενεργότητα** της πηγής. Η (5) αποτελεί μια συνήθη διαφορική εξίσωση για το $N(t)$ με λύση την:

$$N(t) = N(0) e^{-\lambda t} \quad (6)$$

Συνήθως αντί της σταθεράς διάσπασης λ χρησιμοποιείται ο χρόνος ημιζωής $T_{1/2}$ ως η φυσική παράμετρος που χαρακτηρίζει την διαδικασία διάσπασης. Ορίζεται σαν το χρόνο υποδιπλασιασμού ενός αρχικού πληθυσμού πυρήνων:

$$T_{1/2} = \frac{\ln 2}{\lambda}$$

Είναι φανερό ότι η μελέτη των διακυμάνσεων του ρυθμού διάσπασης ραδιενεργών πυρήνων μπορεί να υλοποιηθεί με δυο τρόπους

1. Είτε κρατώντας το διάστημα Δt σταθερό να προσδιορίσει κανείς τις διακυμάνσεις του αριθμού διασπάσεων σε αυτό το διάστημα
2. Είτε βρίσκοντας τις διακυμάνσεις των χρονικών διαστημάτων στα οποία υλοποιείται προκαθορισμένος αριθμός διασπάσεων.

Στη παρούσα άσκηση θα επιλεγεί ο πρώτος τρόπος δηλ. θα μετρηθεί ο αριθμός διασπάσεων σε προκαθορισμένο σταθερό διάστημα Δt .

2.2 Εφαρμόζοντας το κριτήριο χ^2 για έλεγχο της θεωρητικής υπόθεσης

Από την ανωτέρω περιγραφή προκύπτει ότι αν θεωρηθεί ότι η πιθανότητα διάσπασης ενός ραδιενεργού πυρήνα συγκεκριμένου είδους είναι πολύ μικρή και ότι σε μια συλλογή με πολύ μεγάλο αριθμό πυρήνων (αντιστρόφως ανάλογο της πιθανότητας διάσπασης) αυτού του είδους κάθε πυρήνας διασπάται ανεξάρτητα από τους άλλους με την ίδια πιθανότητα, τότε ο αριθμός διασπάσεων σε προκαθορισμένο χρονικό διάστημα θα είναι μια τυχαία μεταβλητή που θα κατανέμεται σύμφωνα με τη κατανομή Poisson (2). Αυτή η θεωρητική πρόβλεψη μπορεί να ελεγχθεί πειραματικά με μέτρηση των συχνοτήτων εμφάνισης των διαφόρων δυνατών τιμών του αριθμού διασπάσεων σε προκαθορισμένο χρονικό διάστημα Δt (διακυμάνσεις του αριθμού διασπάσεων) και χρήση του κριτηρίου χ^2 .

Όπως ήδη είπαμε, ο αριθμός διασπάσεων σε προκαθορισμένο χρονικό διάστημα Δt που θα καταγράφει η πειραματική μετρητική συσκευή (ανιχνευτής), αν ισχύουν οι θεωρητικές υποθέσεις μας, θα ακολουθεί την κατανομή (2) με μέσο αριθμό διασπάσεων μ . (Εδώ αγνοούμε το «νεκρό χρόνο» του συστήματος ή άλλως το t είναι «ζωντανός χρόνος» που ισούται με τον «πραγματικό χρόνο» μείον τον «νεκρό χρόνο» (βλέπε άσκηση για ανιχνευτή ακτινοβολιών).

Κατά συνέπεια αν ληφθούν $N_{ολ}$ καταγραφές ίσων χρονικών διαστημάτων, Δt , η σχέση:

$$e_k = N_{ολ} P(k) = N_{ολ} e^{-\mu} \mu^k / k! \quad (7)$$

θα παρέχει την αναμενόμενη συχνότητα καταγραφής k διασπάσεων. Έστω ότι οι αντίστοιχες συχνότητες που παρατηρήθηκαν είναι O_k . Για τον υπολογισμό του χ^2 να σημειώσουμε ότι:

- Η τιμή του μ δεν είναι συνήθως γνωστή και θα πρέπει να προσδιοριστεί πειραματικά. Αυτό θα γίνει είτε με ανεξάρτητη μέτρηση είτε από τη σχέση:

$$\mu = \frac{1}{N_{ολ}} \sum_k O_k \cdot k \quad (8)$$

Στη δεύτερη περίπτωση εισάγεται μία ακόμη δεσμευτική σχέση μεταξύ των παραμέτρων της χ^2 . Να σημειώσουμε ότι στον υπολογισμό του μ υπεισέρχεται ένα σφάλμα σ_μ το οποίο είναι της τάξεως $\sigma/\sqrt{N_{ολ}}$ όπου σ η τυπική απόκλιση της κατανομής (2).

- Είπαμε ότι για να ακολουθεί το χ^2 την κατανομή που περιγράφηκε στην εισαγωγή θα πρέπει οι συχνότητες να είναι αρκετά μεγάλοι αριθμοί. Πρακτικά αρκεί να είναι $e_k, O_k > 5$. Αν δε συμβαίνει αυτό **πρέπει** να συμπτύξουμε (αθροίζοντας) γειτονικές συχνότητες ώστε να προκύψουν νέες τιμές > 5 . Φυσικά, το πλήθος των τιμών k και, βέβαια, οι βαθμοί ελευθερίας περιορίζονται ανάλογα.

2.2 Εφαρμόζοντας τη μέθοδο προσομοίωσης Monte-Carlo

Στην παρούσα άσκηση επιχειρείται η **προσομοίωση της διαδικασίας διάσπασης ραδιενεργών πυρήνων**. Σ' αυτό το πρόβλημα ο αλγόριθμος προσομοίωσης είναι πολύ απλός: έστω ένα σύνολο από N ίδιους ραδιενεργούς πυρήνες. Όπως αναφέραμε στην ενότητα 2.1 η πιθανότητα διάσπασης ενός πυρήνα αυτού του είδους σε χρονικό διάστημα Δt είναι $\lambda \Delta t$ όπου το λ χαρακτηρίζει το είδος του πυρήνα και το Δt προκαθορίζεται. Για να προσομοιώσουμε τη διαδικασία διάσπασης θεωρούμε N τυχαίους αριθμούς ομοιόμορφα κατανεμημένους στο $[0,1)$. Κάθε τυχαίος αριθμός αντιστοιχεί σε έναν πυρήνα. Συγκρίνουμε κάθε έναν από τους N αριθμούς με τη πιθανότητα $\lambda \Delta t$. Αν ο τυχαίος αριθμός είναι μικρότερος από $\lambda \Delta t$ τότε θεωρούμε ότι ο αντίστοιχος πυρήνας διασπάστηκε στο διάστημα Δt (γιατί;). Έτσι βρίσκουμε το συνολικό αριθμό διασπασμένων πυρήνων στο διάστημα Δt . Επαναλαμβάνουμε την ίδια διαδικασία χρησιμοποιώντας N διαφορετικούς τυχαίους αριθμούς ομοιόμορφα κατανεμημένους στο $[0,1)$. Καταλήγουμε έτσι σε ένα νέο αριθμό διασπασμένων πυρήνων. Η διαδικασία επαναλαμβάνεται M φορές με $M \gg 1$. Μετά το πέρας της διαδικασίας κατασκευάζουμε ιστόγραμμα με τους παρατηρημένους αριθμούς διασπασμένων πυρήνων. Κανονικοποιώντας τις συχνότητες εμφάνισης των διαφόρων αριθμών διασπάσεων έτσι ώστε το εμβαδόν του ιστογράμματος να είναι 1 παίρνουμε τη κατανομή του αριθμού διασπάσεων η οποία για $M \rightarrow \infty$ θα τείνει προς τη

κατανομή Poisson εάν οι θεωρητικές μας υποθέσεις για ανεξαρτησία των πυρήνων και σταθερή πιθανότητα διάσπασης ισχύουν.

3. Όργανα

1. Ανιχνευτής
2. Καταμετρητής με προρρυθμιση χρόνου/αριθμού διασπάσεων
3. Ραδιενεργός πηγή
4. Ηλεκτρονικός υπολογιστής

4. Εκτέλεση της άσκησης

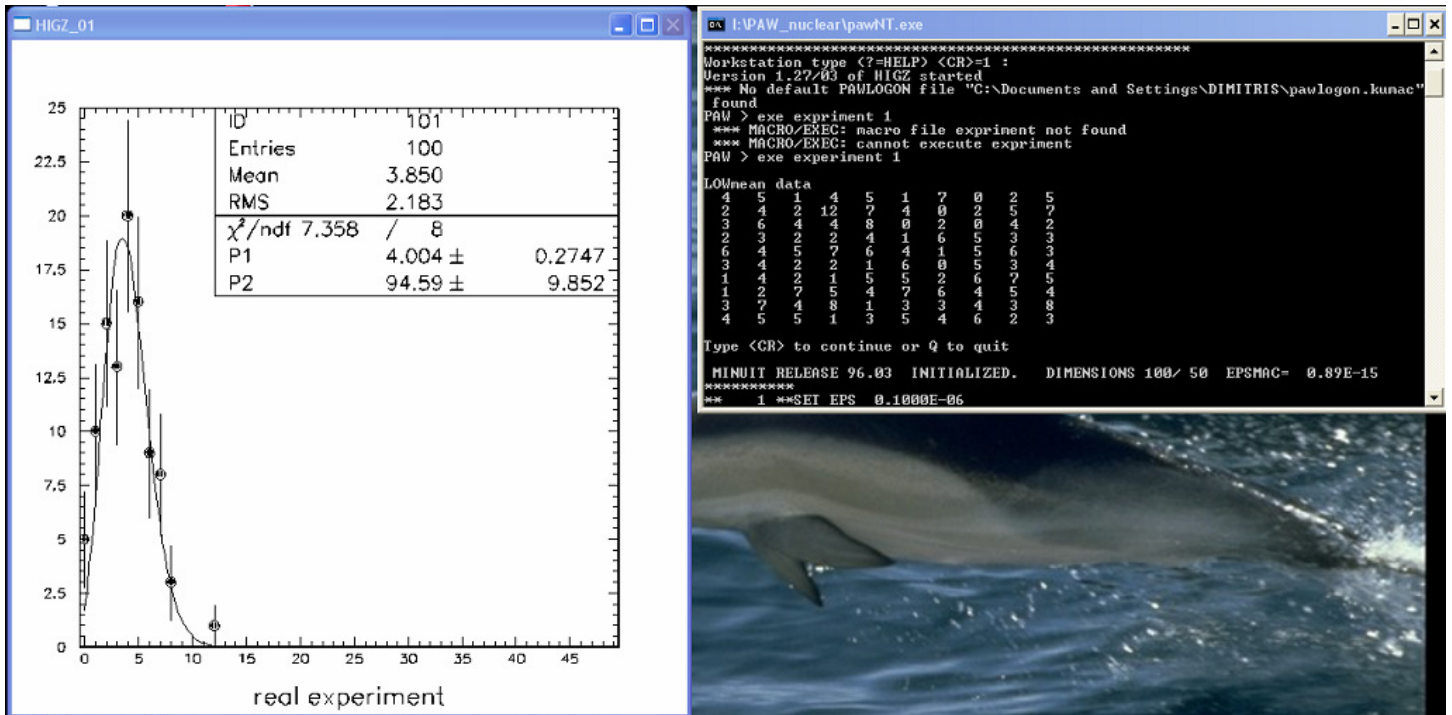
1. Αναγνωρίστε τα όργανα που θα χρησιμοποιήσετε. *Προσοχή στη χρήση της ραδιενεργής πηγής.*
2. Ρυθμίστε το προκαθορισμένο χρονικό διάστημα καταγραφής διασπάσεων στη τιμή $\Delta t=2$ sec. Καθορίστε τη θέση της πηγής έτσι ώστε σε αυτό το χρονικό διάστημα να καταγράφονται κατά μέσο όρο λιγότερες από 10 διασπάσεις (γιατί;).
3. Πάρτε 100 μετρήσεις του αριθμού διασπάσεων για τον έλεγχο χ^2 . Προσέξτε να μην μεταβάλλετε τη θέση της πηγής κατά τη διάρκεια των μετρήσεων (γιατί;). Γράψτε τις μετρήσεις σας απευθείας σε αρχείο απλού κειμένου και ονομάστε το lowmean.dat.

Προσοχή: Όλες οι διαδικασίες να εκτελούνται στο φάκελο: PAW_nuclear που βρίσκεται στην επιφάνεια εργασίας.

4. Μετακινήστε τη πηγή έτσι ώστε να καταγράφετε σε διάστημα 2 sec περισσότερες από 30 διασπάσεις κατά μέσο όρο. Πάρτε εκ νέου 100 μετρήσεις του αριθμού διασπάσεων προσέχοντας να μην μετακινηθεί η πηγή. Γράψτε τις μετρήσεις σας απευθείας σε αρχείο απλού κειμένου και ονομάστε το highmean.dat.
5. Με διπλό κλικ στο εικονίδιο του αρχείου “pawNT.exe” εισέρχεται στο περιβάλλον επεξεργασίας δεδομένων “paw”. Εκτελέστε το αρχείο “experiment.kumac” για τα δεδομένα lowmean.dat γράφοντας στην είσοδο εντολών “exe experiment 1” και πατώντας ακολούθως το πλήκτρο “Enter”.

Στην οθόνη του υπολογιστή σας θα εμφανισθεί το ιστόγραμμα των δεδομένων σας για μικρή τιμή του μέσου αριθμού διασπάσεων καθώς και οι μετρήσεις σας.

Εικόνα των παραθύρων που ανοίγει το PAW κατά την εκτέλεσή του. Το δεξιό παράθυρο είναι το παράθυρο που δίνονται οι εντολές και το αριστερό είναι το παράθυρο γραφικών. **Κατά την εκτέλεση των προγραμμάτων, το παράθυρο γραφικών δεν πρέπει να επικαλύπτεται από κανένα άλλο παράθυρο.**



Για παράδειγμα η εντολή για την εκτέλεση του προγράμματος experiment δίνεται στο παράθυρο εντολών όπως φαίνεται στην παρακάτω εικόνα:

```

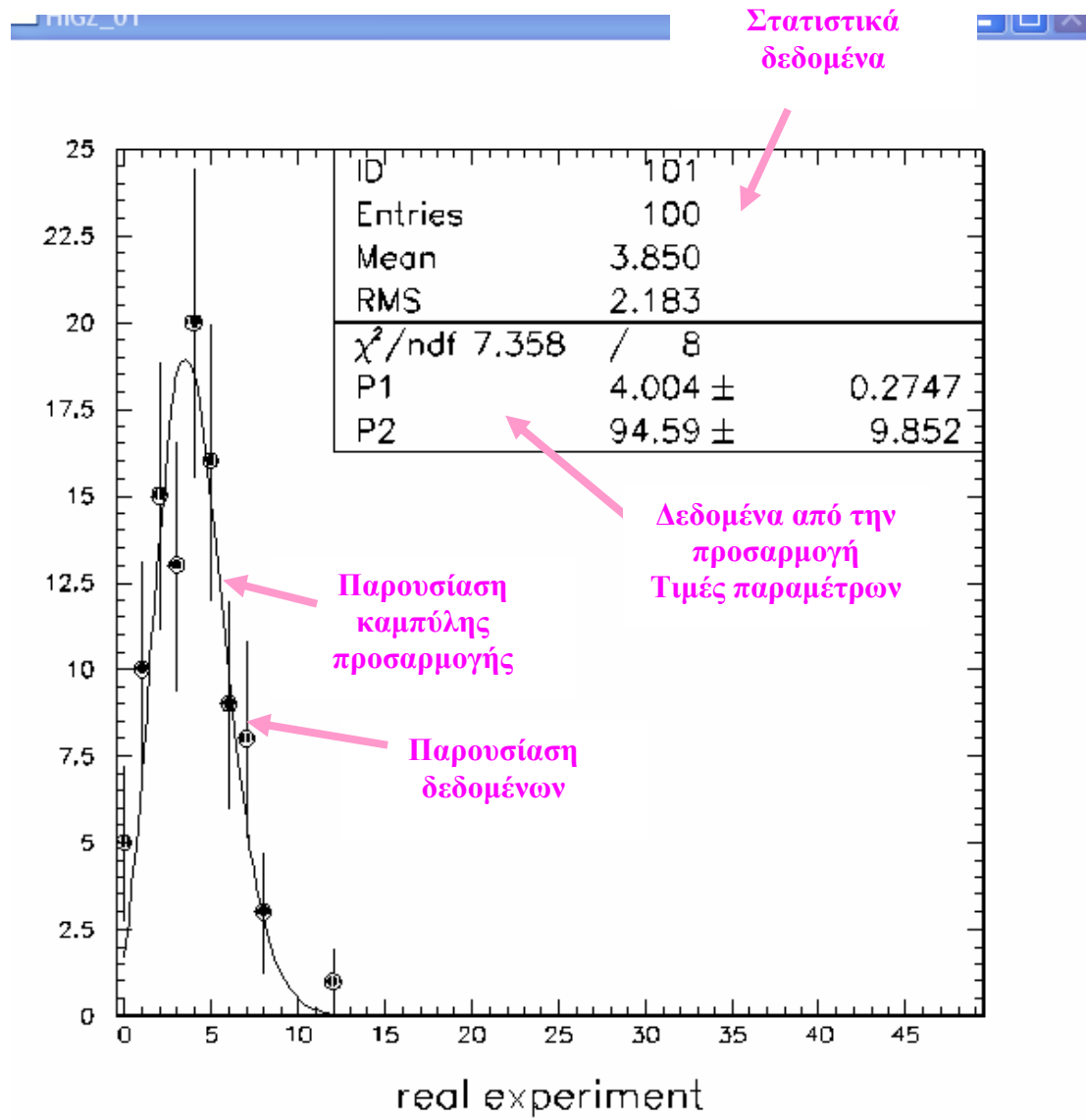
I:\PAW_nuclear\pawNT.exe
*****
WELCOME to PAW
*****
Version 2.12/22 13 June 2001
*****
Workstation type (?=HELP) <CR>=1 :
Version 1.27/03 of HIGZ started
*** No default PAWLOGON file "C:\Documents and Settings\DIMITRIS\pawlogon.kumac"
found
PAW > exe experiment 1

LOWmean data
4 5 1 4 5 1 7 0 2 5
2 4 2 12 7 4 0 2 5 7
3 6 4 4 8 0 2 0 4 2
2 3 2 2 4 1 6 5 3 3
6 4 5 7 6 4 1 5 6 3
3 4 2 2 1 6 0 5 3 4
1 4 2 1 5 5 2 6 7 5
1 2 7 5 4 7 6 4 5 4
3 7 4 8 1 3 3 4 3 8
4 5 5 1 3 5 4 6 2 3

Type <CR> to continue or Q to quit
  
```

Επακόλουθα στο ίδιο παράθυρο τυπώνονται τα δεδομένα από τη συγκεκριμένη μέτρηση.

Κατά την εκτέλεση του προγράμματος αυτού στο παράθυρο γραφικών εμφανίζεται η εικόνα που φαίνεται παρακάτω:



Επιβεβαιώστε με απ' ευθείας καταμέτρηση τις τιμές των διαφόρων συχνοτήτων εμφάνισης καθώς και τα αντίστοιχα σφάλματα.

6. Συγκρίνατε τη μέση τιμή του ιστογράμματος (είναι η τιμή της μεταβλητής "mean" στο εμφανιζόμενο πλαίσιο) με τη πειραματική τιμή του μ που βρίσκετε χρησιμοποιώντας τη σχέση (8).

7. Ξαναπατήστε το πλήκτρο “Enter” για να προσαρμόσετε τη κατανομή Poisson στα δεδομένα σας. Βρείτε τη μέση τιμή της κατανομής Poisson που προσαρμόσατε και συγκρίνετέ την με τις τιμές του μ που βρήκατε στο 6. Σχολιάστε τα αποτελέσματά σας και προσδιορίστε το επίπεδο εμπιστοσύνης απόρριψης της θεωρητικής πρότασης.
8. Στο περιβάλλον “paw” εκτελέστε το αρχείο προσομοίωσης της διαδικασίας διάσπασης “simulation.kumac” γράφοντας:

“exe simulation argument1 argument2”

όπου argument1 είναι ο μέσος αριθμός κρούσεων (πραγματικός αριθμός) και argument2 είναι ο αριθμός δοκιμών (προσοχή πρέπει να είναι ακέραιος). Επιλέξτε 100 δοκιμές με μέσο αριθμό κρούσεων αυτόν που προκύπτει από την εξίσωση (8) για τη περίπτωση μικρού μέσου αριθμού διασπάσεων. Συγκρίνατε τα αποτελέσματά σας με τα αντίστοιχα πειραματικά δεδομένα, πατώντας κάθε φορά το πλήκτρο “Enter”. Ξαναπατώντας το πλήκτρο “Enter” προσαρμόζετε σε μια κατανομή Poisson στα αντίστοιχα δεδομένα προσομοίωσης. Καταγράψτε επίσης το σφάλμα για το μέσο αριθμό κρούσεων όπως αυτό προκύπτει από την προσαρμογή.

9. Ξαναεκτελέστε το αρχείο προσομοίωσης επιλέγοντας τώρα 10.000 δοκιμές. Πατώντας “Enter” προσαρμόστε στα δεδομένα της προσομοίωσης αρχικά μια κατανομή Poisson και κατόπιν ξαναπατώντας “Enter” μια κατανομή Gauss. Σημειώστε τη τιμή της μεταβλητής χ^2 σε κάθε προσαρμογή που επιχειρείτε. Τελειώνοντας τη διαδικασία θα έχετε καταγεγραμμένες 2 τιμές του χ^2 . Συγκρίνατε τις τιμές χ^2_{Poisson} με χ^2_{Gauss} και σχολιάστε τα αποτελέσματά σας προσδιορίζοντας τα επίπεδα εμπιστοσύνης απόρριψης σε κάθε περίπτωση. Επίσης καταγράψτε το σφάλμα του μέσου αριθμού κρούσεων για την προσαρμογή των δεδομένων προσομοίωσης με κατανομή Poisson. Συγκρίνατε τα σφάλματα που βρήκατε στη παρούσα προσαρμογή με αυτά που βρήκατε στο βήμα 8. Εξηγήστε τη διαφορά που βρίσκετε.
10. Επαναλάβετε τη διαδικασία που περιγράφεται στα σημεία 5-9 πιο πάνω για τα δεδομένα του αρχείου “highmean.dat”. Για να το κάνετε αυτό εκτελέστε το αρχείο “experiment.kumac” γράφοντας στην είσοδο εντολών “exe experiment

2” και πατώντας ακολούθως το πλήκτρο “Enter”. Στην οθόνη του υπολογιστή σας θα εμφανισθεί το ιστόγραμμα των δεδομένων σας για μεγάλη τιμή του μέσου αριθμού διασπάσεων καθώς και οι μετρήσεις σας. Επιβεβαιώστε με απ’ ευθείας καταμέτρηση τις τιμές των διαφόρων συχνοτήτων εμφάνισης καθώς και τα αντίστοιχα σφάλματα. Κατόπιν ακολουθείστε την ανάλογη διαδικασία με αυτή που περιγράφεται στα βήματα 5-9 για τα δεδομένα που αφορούσαν το μικρό μέσο αριθμό διασπάσεων. Τι παρατηρείτε για τις τιμές χ^2_{Poisson} και χ^2_{Gauss} που προκύπτουν από τη προσαρμογή των δεδομένων προσομοίωσης για 10000 δοκιμές στη περίπτωση μεγάλου μέσου αριθμού διασπάσεων; Συγκρίνατε αυτές τις τιμές με τις αντίστοιχες τιμές που βρήκατε στο βήμα 9 για τα δεδομένα με μικρό μέσο αριθμό διασπάσεων.

11. Παραμένοντας στο περιβάλλον “raw” εκτελέστε το αρχείο “central_limit.kumac” (“exe central”). Στην οθόνη του υπολογιστή εμφανίζονται 9 ιστογράμματα που περιγράφουν την κατανομή των τιμών 9 τυχαίων μεταβλητών $X_1, X_2, X_3, \dots, X_9$ με ομοιόμορφη κατανομή στο $[0,1)$. (Τα ιστογράμματα έχουν προκύψει από δειγματοληψία 10000 τιμών για κάθε μεταβλητή επιλέγοντας εύρος ιστού 0.01). Κάντε μια πρόβλεψη για τη μορφή που θα έχει η κατανομή του αθροίσματος X_1+X_2 . Πατώντας το πλήκτρο “Enter” θα εμφανιστεί στην οθόνη σας το ιστόγραμμα της τυχαίας μεταβλητής που αντιστοιχεί στο άθροισμα X_1+X_2 . Στη συνέχεια σκεφτείτε πως περιμένετε να μοιάζει η κατανομή του αθροίσματος:

$$\sum_{i=1}^3 X_i \text{ και ελέγξτε τη πρόβλεψή σας πατώντας πάλι το πλήκτρο “Enter”}.$$

Πατώντας εκ νέου το ίδιο πλήκτρο εμφανίζεται στην οθόνη σας η κατανομή του αθροίσματος: $\sum_{i=1}^5 X_i$. Τέλος πατώντας ακόμη μια φορά το “Enter” εμφανίζεται

στην οθόνη η κατανομή του αθροίσματος: $\sum_{i=1}^9 X_i$. Αν έχετε εκπλαγεί από το

αποτέλεσμα, θα βρείτε στο παράρτημα Β της εισαγωγής την εξήγηση για το τι συμβαίνει. Για περισσότερες εκπλήξεις μπορείτε να ξαναπατήσετε το πλήκτρο “Enter”. Τώρα εμφανίζονται στην οθόνη 9 ιστογράμματα που περιγράφουν την κατανομή των τιμών 9 τυχαίων μεταβλητών $X_1, X_2, X_3, \dots, X_9$. Όπως βλέπετε, κάθε μεταβλητή ακολουθεί τώρα διαφορετική κατανομή στο $[0,1)$. Ποια είναι

άραγε η κατανομή της μεταβλητής $Y = \sum_{i=1}^9 X_i$; Την απάντηση θα την δείτε πατώντας μια τελευταία φορά το “Enter” και την εξήγηση και πάλι στο παράρτημα Β της εισαγωγής.

5. Ερωτήσεις (πρέπει να απαντηθούν στις συνοδευτικές κόλλες κατά τη διάρκεια της άσκησης)

1. Περιγράψτε σύντομα πως λειτουργεί ο ανιχνευτής που χρησιμοποιείτε. Γιατί νομίζετε ότι οι διαδοχικές μετρήσεις σας μπορούν να θεωρηθούν ανεξάρτητες μεταξύ τους;
2. Δείξτε ότι για $N \gg 1$ και $\lambda \Delta t \ll 1$ με $N\lambda \Delta t = \text{σταθερό}$ η κατανομή (1) τείνει στη κατανομή (2). Θεωρείστε γνωστό το όριο: $\lim_{x \rightarrow \infty} (1 + \frac{1}{x})^x = e$.
3. Υπολογίστε για τη κατανομή Poisson $P_\mu(n) = \frac{\mu^n e^{-\mu}}{n!}$ τα μεγέθη: $\langle n \rangle$, $\langle n^2 \rangle$ και $\langle (n-\mu)^2 \rangle$
4. Έστω ότι ο μέσος ρυθμός καταγραφής ενός ανιχνευτή τυχαίων γεγονότων είναι 2.5 καταγραφές ανά δευτερόλεπτο. Ποια είναι η πιθανότητα να μην έχει παρατηρηθεί καταγραφή σε διάστημα ενός δευτερολέπτου; Ποιο είναι το πιθανότερο χρονικό διάστημα μεταξύ διαδοχικών καταγραφών;
5. Προσπαθήστε να εξηγήσετε τον αλγόριθμο αντιστροφής για την παραγωγή ψευδοτυχαίων αριθμών που ακολουθούν μια οποιαδήποτε κατανομή $p(x)$.
Εφαρμογή: Έστω X τυχαία μεταβλητή με ομοιόμορφη κατανομή στο $[0,1)$. Βρείτε μετασχηματισμό $Y=f(X)$ τέτοιοι ώστε η μεταβλητή Y να κατανέμεται εκθετικά στο $[0, \infty)$ εφαρμόζοντας τη μέθοδο αντιστροφής.
6. Εξηγείστε τον αλγόριθμο απόρριψης για την παραγωγή ψευδοτυχαίων αριθμών που ακολουθούν μια οποιαδήποτε κατανομή $p(x)$.
Εφαρμογή: Έστω N ζεύγη τυχαίων αριθμών (x_i, y_i) ομοιόμορφα κατανεμημένων στο $[0,1) \times [0,1)$. Από αυτά n ζεύγη πληρούν τη συνθήκη: $x_i^2 + y_i^2 \leq 1$.

Προσδιορίστε τη τιμή του π συναρτήσει των n, N . Βρείτε και το αντίστοιχο σφάλμα. Μπορείτε να διατυπώσετε τον υπολογισμό αυτό σαν ένα δισδιάστατο ολοκλήρωμα στο επίπεδο (x,y) . Ποια είναι η αντίστοιχη γεωμετρική ερμηνεία; Πως σχετίζεται η διαδικασία που ακολουθήσατε με τον αλγόριθμο απόρριψης;