Προσομοίωση και Στοχαστικά μοντέλα
Πανεπιστημιακές παραδόσεις

Οικονομικό Πανεπιστήμιο Αθηνών
Τμήμα Στατιστικής

Πέτρος Δελλαπόρτας

Οκτώβριος 1994

Περιεχόμενα

Μέρος Α

ΘΕΩΡΙΑ

Κεφάλαιο 1

Εισαγωγή

1.1 Στοχαστική προσομοίωση

- Τί είναι προσομοίωση;

  Προσόμοιος: Σχεδόν όμοιος, παρεμφερής,...

  Simulate: Pretend to be, act like,...

- Τι είναι στοχαστική προσομοίωση;

  1. Η μίμηση της συμπεριφοράς ενός (περίπλοκου) μοντέλου πιθανότητας
  2. Η χρήση στοχαστικών τεχνικών για την μελέτη μη στοχαστικών μαθηματικών προβλημάτων

- Παραδείγματα

  1. – Συστήματα ουρών (μαγαζιά, νοσοκομεία, αεροπλάνα, δίκτυα υπολογιστών)
     – Επιδημίες (Ιλαρά, λύσσα, AIDS, αρρώστιες φυτών)
     – Πληθυσμοί ζώων
     – Κατανομή κάποιας παράξενης Στατιστικής συνάρτησης
     – Το πρόβλημα του Sylvester: 4 σημεία είναι τυχαία σε ένα κύκλο, ποιά είναι η πιθανότητα ότι η κυρτή περίμετρος είναι τρίγωνο;
     – Απόσταση της μικρότερης γειτονιάς: N σημεία τοποθετούνται τυχαία στο μοναδιαίο τετράγωνο (δηλ. ένα σημείο έχει συντεταγμένες x, y με $x, y \in [0, 1]$), και έστω d η μικρότερη απόσταση μεταξύ 2 σημείων. Ποιά η κατανομή του d; Πολλές εφαρμογές στην Αστρονομία, Οικολογία κλπ, ειδικά αν τα x,y προέρχονται από διμετάβλητη κανονική, Poisson κλπ.
     – 'Ερευνα για ασυμπτοτική συμπεριφορά
     – Κατανομή εκτιμήτριας μεγίστης πιθανοφάνειας

- Κεντρικό οριακό θεώρημα
- Ανθεκτικότητα: Ευαισθησία συμπερασματολογίας σε υποθέσεις

2. - Υπολογισμός περιπλόκων ολοκληρωμάτων (Αριθμητική ανάλυση)
   - Το πρόβλημα του ταξιδιώτη πωλητή: Υπάρχουν n πόλεις, και ένας πωλητής πρέπει να επισκεφθεί μία φορά μόνο την κάθε μία, με κάποια σειρά. Το κόστος μεταφοράς από την πόλη i στην πόλη j είναι $d(i, j)$. Ζητείται η οικονομικότερη διαδρομή.

1.2 Πιθανοθεωρητικό υπόβαθρο

Θεωρείται ότι οι παρακάτω έννοιες είναι γνωστές. Στην συνέχεια των σημειώσεων θα γίνεται αναφορά στις έννοιες αυτής της ενότητας, προϋποθέτοντας ότι ο αναγνώστης τις έχει πλήρως κατανοήσει.

- Τυχαία μεταβλητή X, συνάρτηση πυκνότητα πιθανότητας (σ.π.π.) $f_X(x)$.
- Αθροιστική συνάρτηση κατανομής (α.σ.κ.) $F_X(x)$
- Τυχαίες μεταβλητές $X_1, \ldots X_n$
  - Από κοινού πυκνότητα $f(\boldsymbol{x}) = f(x_1, \ldots, x_n)$
  - Περιθώριες κατανομές πιθανότητας
  - Δεσμευμένες κατανομές πιθανότητας
- Ροπές κ-τάξεως: $\mu_\kappa = E(X^\kappa), \kappa = 1, 2, \ldots$
- Ροπογεννήτριες: $M_X(t) = E(e^{tX})$
  - Ροπές με διαφοροποίηση και $t = 0$
  - Αθροίσματα τυχαίων μεταβλητών
    $Y = X_1 + X_2 + \ldots + X_n$
    $M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$
- Μετασχηματισμοί μιας τυχαίας μεταβλητής: $Y = g(X) \;\; (1-1)$
  $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$
  Αρα $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$
- Μετασχηματισμοί δύο τυχαίων μεταβλητών:
  $Y_1 = g_1(X_1, X_2)$, $Y_2 = g_2(X_1, X_2)$
  $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) \, |J|$

όπου J, η Ιακωβιανή ορίζουσα, δίνεται από την σχέση

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

1.3 Περιγραφή του μαθήματος-Αναφορές

Το μάθημα καλύπτει τα εξής θέματα:

- Τυχαίοι αριθμοί και η παραγωγή τους
- Απόκτηση κατανομών με μετασχηματισμό τυχαίων αριθμών
- 'Ελεγχος ποιότητας τυχαίων αριθμών
- Προσομοίωση στοχαστικών μοντέλων
- Εφαρμογές

Προτεινόμενες αναφορές είναι

- Byron J T Morgan (1984), *Elements of Simulation*; Chapman and Hall.
- Brian D Ripley (1987), *Stochastic Simulation*; John Wiley and Sons.
- John Dagpunar (1988), *Principles of Random Variate Generation*; Oxford University Press.

Κεφάλαιο 2

Παραγωγή ομοιόμορφων τυχαίων μεταβλητών

2.1 Αναγωγικές γεννήτριες (ψευδο-)τυχαίων αριθμών

Θεωρούμε την αναγωγή

$$x_{n+1} = (ax_n + b) \bmod m, \quad n = 0, 1, 2, \ldots, \tag{2.1}$$

όπου

- x_0: σπόρος (αρχική τιμή)
- m: μέτρο (modulus)
- b: βήμα (αύξηση)
- a: πολλαπλασιαστής

και τα x_0, m, b, a είναι θετικοί ακέραιοι, με $x_0, b, a \in \{0, 1, \ldots, m-1\}$. Έτσι ορίζεται μία κλάση "γεννητριών" αριθμών $x_0, x_1, \ldots$, όπου $x_i \in \{0, 1, \ldots, m-1\}$. Αν ζητούμε τιμές στο (0,1), τότε

$$u_n = \frac{x_n}{m} \Rightarrow u_n \in (0, 1)$$

για κάθε ακέραιο n.

ΙΔΕΑ: Αν m είναι πολύ μεγάλο, μπορούμε να βρούμε x_0, b, a τέτοια ώστε τα παραγόμενα $u_i, i = 1, 2, \ldots$ να "μοιάζουν" με παραγόμενους τυχαίους αριθμούς από την ομοιόμορφη κατανομή $U(0, 1)$;

Ειδικές περιπτώσεις

- Πολλαπλασιαστικές γεννήτριες: $b = 0$, $x_{n+1} = ax_n \bmod m$
- Μικτές γεννήτριες: $b \neq 0$, $x_{n+1} = (ax_n + b) \bmod m$

Παρατηρήσεις

1. Δεδομένων $x_0, x_1, ..., x_n$, μελλοντικές τιμές είναι εντελώς καθωρισμένες από την x_n.
2. Τα $x_0, x_1, ..., x_m$ δεν μπορεί να είναι διακριτά.
3. Υπάρχουν κάποια i, k τέτοια ώστε $x_i = x_{i+k}$
4. Τα $x_i, x_{i+1}, ..., x_{i+k-1}$ επαναλαμβάνονται.

Συνεπώς η αναγωγική γεννήτρια (2.1) παράγει μία περιοδική ακολουθία με περίοδο η κύκλο $k \leq m$. Μπορούμε να έχουμε $k = m$ θέτοντας, για παράδειγμα $a = 1, b = 1$ (οι παραγόμενοι αριθμοί όμως δεν είναι ΚΑΘΟΛΟΥ τυχαίοι!). Αν η γεννήτρια είναι πολλαπλασιαστική, τότε ισχύει $k \leq m - 1$ (προσέξτε το 0!).

Κριτήρια επιλογής x_0, a, b, m

- Εύκολες πράξεις: για παράδειγμα $m = 10^a$. Τότε

$$276623242 \bmod 10^5 = 23242$$

  Στον υπολογιστή, λόγω του δυαδικού συστήματος, $m = 2^p$ και έχουμε

$$1010001000101011101 \bmod 2^8 = 01011101$$

- Επιζητούμε μεγάλη περίοδο $k(\leq m)$. Το k εξαρτάται από την επιλόγή όλων των στοιχείων της γεννήτριας x_0, a, b, m. Στο ακόλουθο παράδειγμα ο σπόρος επηρεάζει το μήκος της περιόδου:

  $m = 2^4$, $a = 5$, $b = 4$: $x_{n+1} = 5x_n + 4 \bmod 16$

$x_0 = 0$	$0, 4, 8, 12, 0, \ldots$	$k = 4$
$x_0 = 1$	$1, 9, 1, 9, \ldots$	$k = 2$
$x_0 = 3$	$3, 3, \ldots$	$k = 1$

Μαθηματικές ιδιότητες γεννητριών

Θεώρημα 1 *Μια αναγωγική γεννήτρια της μορφής (2.1) έχει πλήρη περίοδο $k = m$ εάν και μόνο εάν*

1. *b και m δεν έχουν κοινούς παράγοντες εκτός του 1.*
2. *$a = 1 \bmod p$ για κάθε πρώτο παράγοντα p του m.*
3. *$a = 1 \bmod 4$ αν το 4 διαιρεί τον m.*

Πόρισμα 1 *Αν m είναι πρώτος, πλήρης περίοδος επιτυγχάνεται μόνο αν $\alpha = 1$.*

Θεώρημα 2 *Μία πολλαπλασιαστική γεννήτρια της μορφής (2.1) με* $m = 2^b \geq 16$ *έχει μέγιστη περίοδο* $m/4$ *εάν και μόνο εάν* $a \bmod 8 = 3$ *η* $a \bmod 8 = 5$*, και* x_0 *είναι περιττός αριθμός.*

Παράδειγμα 1 *Βάσει του θεωρήματος 1, αν* $m = 2^k$*, (3) συνεπάγεται ότι* $a = 4c + 1$ *για κάποιο θετικό ακέραιο* c*. Επιπλέον,* a *αυτής της μορφής ικανοποιεί επίσης την (2). Τέλος, η (1) ικανοποιείται εύκολα αν θέσουμε* b *έναν οποιοδήποτε περιττό αριθμό.*

Παράδειγμα 2 *Βάσει του θεωρήματος 2, μία κατάλληλη επιλογή του* a *είναι η* $a = 5^{2q+1}$, $q = 0, 1, 2, \ldots$*, γιατί*

$$5^{2q+1} \bmod 8 = (1+4)^{2q+1} \bmod 8 = 1 + 4(2q+1) \bmod 8 = 5 \bmod 8$$

Αρα πολλαπλασιαστικές γεννήτριες της μορφής $x_{n+1} = 5^{2q+1} x_n \bmod 2^\beta$ *έχουν περίοδο* $2^{\beta-2}$*. Τέτοια παραδείγματα γεννητριών που έχουν χρησιμοποιηθεί στο παρελθόν είναι*

a	β
5^{13}	$36, 39$
5^{17}	$40, 42, 43$

Τα ανωτέρα θεωρήματα δείχνουν τρόπους εύρεσης γεννητριών με μεγάλες περιόδους, αλλά μπορούμε να δεχθούμε σοβαρά μία αναγωγική γεννήτρια με πεπερασμένη περίοδο; Για να καταδείξουμε ότι οι ανωτέρω περίοδοι είναι ικανοποιητικές για πρακτικούς λόγους, ας υποθέσουμε ότι χρησιμοποιούμε, σύμφωνα με το τελευταίο παράδειγμα, $\beta = 43$, άρα η περίοδος είναι 2^{41}. Με παραγωγή 1000 αριθμών το δευτερόλεπτο, η ακολουθία δεν θα επαναληφθεί για > 63 χρόνια!

Μήπως η μεγάλη περίοδος δεν είναι αρκετή; Προφανώς δεν είναι το μόνο που χρειαζόμαστε, γιατί θα θέλαμε σίγουρα, για να επιτύχουμε τυχαιότητα, να αποφύγουμε ιδιότυπους "σχηματισμούς" συσχετίσεων μεταξύ συνεχομένων ανεξαρτήτων τυχαίων αριθμών. Ας δούμε ένα παράδειγμα όπου δημιουργείται ένα τέτοιο πρόβλημα:

Ας θεωρήσουμε την ακολουθία

$$x_{n+1} = 5x_n \bmod (m)$$

Τότε μπορεί να αποδειχθεί ότι

$$x_{n+1} = 5x_n - h_i m, \ h_i \in 0, 1, 2, 3, 4$$

Η ανωτέρω σχέση συνεπάγει ότι συνεχόμενες τιμές δίνουν καρτεσιανές συντεταγμένες σημείων που βρίσκονται πάνω σε μία από 5 γραμμές. Οσο μεγαλύτερο είναι το m, τόσο περισσότερο η ακολουθία των αριθμών θα παραμείνει πάνω σε μία από τις γραμμές πριν μετακινηθεί σε άλλη.

Παράδειγμα 3 *Για κάποιο μικρό* m*, π.χ.* $m = 11, x_0 = 1$*, έχουμε αλλαγές γραμμής κάθε φορά:*

$$x_1 = 5, x_2 = 3, x_3 = 4, x_4 = 9, x_5 = 1$$

Για κάποιο μεγάλο m*, π.χ.* $m = 1000, x_0 = 1$*, έχουμε*

$$x_1 = 5, x_2 = 25, x_3 = 125, x_4 = 625, x_5 = 125$$

Δηλαδή τα $x_1, \ldots, x_4$ *είναι όλα στην ευθεία* $x_{n+1} = 5x_n$*, ενώ μετά συνεχόμενες τιμές εναλλάσσονται μεταξύ* $x_{n+1} = 5x_n$ *και* $x_{n+1} = 5x_n - 3 \times (m = 1000) = 5x_n - 3000$

Παράδειγμα 4 *Η IBM χρησιμοποιούσε (IBM 360/370, PDP II) την γεννήτρια "RANDU", που δίνεται από την αναγωγική σχέση:*

$$x_{n+1} = (2^{16} + 3)x_n \bmod 2^{31}$$

Τότε, για ακέραιους c_1, c_2, c_3, c_4, έχουμε:

$$\begin{aligned}
x_{n+1} &= (2^{16} + 3)x_n + c_1 2^{31} \\
x_{n+2} &= [(2^{16} + 3)x_n + c_1 2^{31}](2^{16} + 3) + c_2 2^{31} \\
&= (2^{16} + 3)^2 x_n + c_1 2^{31}(2^{16} + 3) + c_2 2^{31} \\
&= 2^{32} x_n + 6 \cdot 2^{16} x_n + 9x_n + c_1 2^{31}(2^{16} + 3) + c_2 2^{31} \\
&= (6 \cdot 2^{16} + 9)x_n + 2^{31}(2x_n + c_1(2^{16} + 3) + c_2) \\
&= 6(2^{16} + 3)x_n - 9x_n + c_3 2^{31} \\
&= 6x_{n+1} - 9x_n + c_4 2^{31}
\end{aligned}$$

'Εστω $u_n = \frac{x_n}{2^{31}}$. Τότε

$$u_{n+2} = 6u_{n+1} - 9u_n + c_4$$

Από την παραπάνω σχέση, που δηλώνει ότι ο $u_{n+2} - 6u_{n+1} + 9u_n$ είναι ακέραιος, συνεπάγεται ότι συνεχόμενες τριάδες από υποθετικά τυχαίους αριθμούς βρίσκονται σε υπερεπίπεδα του μοναδιαίου κύβου. Αρα, αν ξέρουμε τα (u_{n-2}, u_{n-1}), to u_n δεν είναι τόσο τυχαίο!

2.2 Ελεγχοι τυχαίων αριθμών

2.2.1 Εισαγωγή

'Εχουμε δεί πως μπορούμε να κατασκευάσουμε ακολουθίες τυχαίων αριθμών με μεγάλη περίοδο, αλλά παράγουν αυτές οι ακολουθίες αριθμούς που "μοιάζουν" με πραγματώσεις μιας ανεξάρτητης ομοιόμορφης τυχαίας μεταβλητής; Υπάρχουν διάφορα είδη ελέγχων τυχαίων αριθμών που μπορούμε να χρησιμοποιήσουμε για να απαντήσουμε το παραπάνω ερώτημα. Εν γένει, διαφορετικοί έλεγχοι απαντούν σε διαφορετικά ερωτήματα, για παράδειγμα μπορεί να μας ενδιαφέρει ο έλεγχος "Πόσο συχνά εμφανίζονται ορισμένα ψηφία", και να αγνοήσουμε το ερώτημα "Ποιά η εξάρτηση μεταξύ συνεχόμενων ζευγαριών, τριάδων, τετράδων κλπ". Τότε, υπάρχει ο κίνδυνος να δεχθούμε μία μικτή αναγωγική γεννήτρια με $\alpha = 1, \beta = 1$ οπου τα ψηφία εμφανίζονται με σωστή συχνότητα, αλλά είναι (ΠΟΛΥ!) εξαρτημένα μεταξύ τους.

Παρατηρήσεις

1. Λόγω της φύσης των τυχαίων αριθμών, μερικές ακολουθίες θα απορριφθούν σίγουρα από κάποιον έλεγχο ομοιομορφίας.

2. Μπορεί σε μία πλήρη περίοδο κάποιος έλεγχος να δείχνει ότι οι ιδιότητες της ακολουθίας είναι καλές, αλλά αυτο να μην συμβαίνει στην ίδια ακολουθία αν χρησιμοποιήσουμε κάποιο υποσύνολο της περιόδου.

3. Υπάρχουν θεωρητικοί έλεγχοι για έλεγχο ανεξαρτησίας συνεχόμενων κ-άδων για μικρά κ, πχ δες B. Ripley (1987), Stochastic Simulation, John Wiley and Sons, NY.

4. Οι έλεγχοι τυχαίων αριθμών είναι επίσης χρήσιμοι σαν τρόποι ελέγχου ότι η γεννήτρια έχει προγραμματισθεί/υλοποιηθεί επιτυχώς.

2.2.2 'Ελεγχοι για τυχαία ψηφία

Ας υποθέσουμε ότι έχουμε n αριθμούς που παρήχθησαν από κάποια γεννήτρια τυχαίων αριθμών.

Επανάληψη: χ^2 *έλεγχος καλής προσαρμογής:* 'Εστω ότι αποτελέσματα ανήκουν σε N πιθανές κατηγορίες, και είναι στατιστικά ανεξάρτητα. Επίσης, έστω ότι υπάρχουν παρατηρούμενοι αριθμοί $\Pi_1, \Pi_2, \ldots, \Pi_N$, με $\Sigma\Pi_i = n$. Η (μηδενική) υπόθεση για παραγωγή δεδομένων μας επιτρέπει, βάσει του μοντέλου που υποθέτουμε ότι παράγει τα δεδομένα, τον υπολογισμό αναμενόμενων τιμών σε κάθε κατηγορία $A_1, A_2, \ldots, A_N$, με $\sum_{i=1}^N A_i = n$. Τότε, η στατιστική συνάρτηση ελέγχου είναι $\chi = \sum_{i=1}^N \frac{(\Pi_i - A_i)^2}{A_i}$.

Διαισθητικά, αν η στατιστική συνάρτηση είναι μεγάλη, υπάρχουν υποψίες για την μηδενική υπόθεση. Αν είναι πολύ μικρή, υπάρχουν πάλι υποψίες (!), τα αναμενόμενα αποτελέσματα είναι πολύ καλά. Συνεπώς, κατάλληλοι έλεγχοι είναι οι δίπλευροι. Ας χρησιμοποιήσουμε το γνώριμο (αντι-)παράδειγμα αναγωγικής μικτής γεννήτριας με $\alpha = 1, \beta = 1$. Η μηδενική υπόθεση ικανοποιείται γιατί υπάρχει ο ακριβής αριθμός ψηφίων ανά κατηγορία σε μία πλήρη περίοδο, και αν χρησιμοποιήσουμε μονόπλευρο έλεγχο θα δεχθούμε (λανθασμένα) την μηδενική υπόθεση.

Κάτω από την μηδενική υπόθεση, και καθώς $n \to \infty$, ισχύει ότι $\chi \sim \chi^2_{N-1}$. Οι N κατηγορίες είναι αυθαίρετες, αλλά ένας γενικός κανόνας που πρέπει να προσεχθεί είναι ότι $A_i \geq 5$.

Σημείωση: Ασφαλώς αυτοί οι έλεγχοι θα αποτύχουν περιστασιακά κατά τύχη. Σε σοβαρές έρευνες είναι καλή ιδέα να δοκιμάσουμε ελέγχους σε μεγάλο αριθμό μη αλληλοεπικαλυπτομένων ακολουθιών.

2.2.3 'Ελεγχος συχνότητας

Εξετάζεται αν τα 10 ψηφία εμφανίζονται με ίδια συχνότητα. 'Εχουμε $N = 10$, παρατηρούμενες τιμές $\Pi_0, \Pi 1, \ldots, \Pi_9$ και αναμενόμενες τιμές $A_0 = n/10, A_1 = n/10, \ldots, A_9 = n/10$. Συγκρίνουμε το

$$\sum_{i=o}^{9} \frac{(\Pi_i - n/10)^2}{n/10}$$

με την περιοχή ουράς της χ^2_9 κατανομής. Σημειώνουμε ότι ακόμα και αν δεν απορρίψουμε την μηδενική υπόθεση, αυτός ο έλεγχος δεν περιέχει *καμμία* πληροφορία για την δομή της εξάρτησης. Τέλος, ας δούμε πως συμπεριφέρεται αυτός ο έλεγχος στα δεκαδικά ψηφία του $e = 2.71828\ldots$. Για τα πρώτα 2000 ψηφία έχουμε $\chi = 1.06 < \chi^2_9(.001)$, άρα υπάρχουν ισχυρές ενδείξεις για να απορρίψουμε την μηδενική υπόθεση (το χ είναι πολύ μικρό!). Αν δούμε τα πρώτα 10000 ψηφία, τότε $\chi = 8.61 \simeq E(\chi^2_9) = 9$ που είναι ότι περίπου περιμένουμε σε δείγμα τυχαίων αριθμών.

2.2.4 Σειριακός έλεγχος

'Εστω ότι n_{jk} δηλώνει πόσες φορές το ψηφίο j ακολουθείται από το ψηφίο k. Τότε $N = 100$ και οι παρατηρούμενες και αναμενόμενες τιμές είναι:

$$\begin{array}{ll} \Pi_i & n_{00}, n_{01}, \ldots, n_{09}, n_{10}, n_{11}, \ldots, n_{98}, n99 \\ A_i & n/100, n/100, \ldots, n/100, n/100, n/100, \ldots, n/100, n/100 \end{array}$$

Για να διατηρήσουμε την απαραίτητη προυπόθεση της ανεξαρτησίας για να χρησιμοποιήσουμε τον έλεγχο χ^2, πρέπει να έχουμε μη αλληλοκαλυπτόμενες ακολουθίες ζευγών αριθμών. Συνεπώς, $\sum_{j=0}^{9} \sum_{k=0}^{9} n_{jk} = n$ και σε αυτήν τη περίπτωση ο συνολικός αριθμός των ελεγχομένων ψηφίων είναι $2n$. Για τον στατιστικό έλεγχο συγκρίνουμε

$$\sum_{j=0}^{9} \sum_{k=0}^{9} \frac{(n_{jk} - n/100)^2}{n/100}$$

με την περιοχή ουράς της χ^2_{99}.

Ο σειριακός έλεγχος απαιτεί πολύ κόπο, ενώ έλεγχος εξάρτησης μεγαλύτερης τάξης (τριάδων, τετράδων κλπ) απαιτεί πολύ προσπάθεια και κρίνεται μάλλον ακατάλληλος.

2.2.5 'Ελεγχος χάσματος

Εξετάζονται πιθαναί σχηματισμοί στην ακολουθία τυχαίων αριθμών, ως εξής: Διαλέγουμε ένα ψηφίο, έστω 3. Η ελεγχόμενη ακολουθία θα είναι

$$x_1, x_2, x_3, 3, x_5, \ldots, x_{k+5}, 3, \ldots$$

Αν η ακολουθία είναι ομοιόμορφη και ανεξάρτητη, η κατανομή χάσματος είναι γεωμετρική με παράμετρο $p = 0.1$:

Μέγεθος χάσματος	0	1	2, ...
Πιθανότητα	$p_0 = \frac{1}{10}$	$p_1 = \frac{9}{10}\frac{1}{10}$	$p_2 = (\frac{9}{10})^2\frac{1}{10}, \ldots$

Για παράδειγμα, ας υποθέσουμε ότι κατασκευάζουμε 7 διαφορετικές κατηγορίες:

Χάσματα	0	1	2	3	4	5	> 5
Π_i	Π_0	Π_1	Π_2	Π_3	Π_4	Π_5	Π_+
A_i	np_0	np_1	np_2	np_3	np_4	np_5	$n(1 - \sum_{i=0}^{5} p_i)$

Ο σειριακός έλεγχος συγκρίνει την στατιστική συνάρτηση

$$\sum_i \frac{(\Pi_i - np_i)^2}{np_i}$$

με την περιοχή ουράς της χ_6^2

2.2.6 'Ελεγχος 'POKER'

Χωρίζουμε την ακολουθία τυχαίων αριθμών σε συνεχόμενες ομάδες των 5:

$$| \; x_1, \ldots, x_5 \; || \; x_6, \ldots, x_{10} \; | \ldots$$

Κάθε ομάδα μπορεί να έχει έναν από τούς επόμενους σχεδιασμούς:

5 διακεκριμένα	Ολα διαφορετικά	αβγδε
4 διακεκριμένα	'Ενα ζευγάρι	ααγδε
3 διακεκριμένα	Δύο ζευγάρια	ααββγ
3 διακεκριμένα	Μία τριπλή	αααβγ
2 διακεκριμένα	Φούλ	αααββ
2 διακεκριμένα	Καρρέ	ααααβ
1 διακεκριμένο	Πέντε όμοια	ααααα

Αν θεωρήσουμε πέντε διαφορετικές κατηγορίες 'r διακεκριμένα ψηφία', $r = 1, ..., 5$, σχηματίζουμε $N = 5$ κατηγορίες και έχουμε ότι

$$Pr[r \text{ διακεκριμένα}] = \frac{10(10-1)\ldots(10-r+1)}{10^5} S(5, r)$$

όπου $S(k, r)$ είναι οι αριθμοί Sterling δευτέρου είδους, που δηλώνουν τον αριθμό των τρόπων που μπορούμε να χωρίσουμε k αντικείμενα σε r μέρη. Για παράδειγμα $S(k, 1) = S(k, k) = 1$.

Οπως και πριν, παρατηρούμε τα Π_i, υπολογίζουμε τα $A_i = n \times Pr[r = i]$, όπου n ο αριθμός των ομάδων, και χρησιμοποιούμε τον χ^2 έλεγχο καλής προσαρμογής.

2.2.7 Έλεγχος μεταθέσεων

Γι αυτόν τον έλεγχο ομαδοποιούμε τα ψηφία σε ομάδες μήκους k:

$$\mid x_1, \ldots, x_k \mid\mid x_{k+1}, \ldots, x_{2k} \mid, \ldots$$

Υπάρχουν $k!$ πιθανές διατάξεις σε κάθε ομάδα, και έστω n ο αριθμός των ομάδων. Αν η ακολουθία παράγει ανεξάρτητα και ομοιόμορφα, όλες οι $k!$ διατάξεις είναι ισοπίθανες. Κατασκευάζουμε πάλι τις αναμενόμενες και παρατηρούμενες τιμές

$$\begin{array}{ccccc} \textit{Διατάξεις} & 1, & 2, & \ldots & k! \\ \Pi_i & \Pi_1 & \Pi_2 & \ldots & \Pi_{k!} \\ A_i & \frac{n}{k!} & \frac{n}{k!} & \ldots & \frac{n}{k!} \end{array}$$

και χρησιμοποιούμε τον χ^2 έλεγχο καλής καλής προσαρμογής.

2.3 Μέθοδοι παραγωγής τυχαίων μεταβλητών

2.3.1 Εισαγωγή

Υποθέτουμε ότι έχουμε μία ακολουθία ανεξαρτήτων $U(0,1)$ ποσοτήτων, και επιθυμούμε να παράγουμε ένα ανεξάρτητο δείγμα από τυχαίες μεταβλητές που ακολουθούν μονοδιάστατες κατανομές, διακριτές ή συνεχείς. Θα περιγράψουμε την μέθοδο της αντιστροφής, την μέθοδο της απόρριψης και την μέθοδο της σύνθεσης. Ολες οι μέθοδοι θα επιδειχθούν με αναφορές σε γνωστές κατανομές, και θα κλείσουμε αυτήν την ενότητα με τις μεθόδους Box-Muller και Polar Marsaglia ειδικά σχεδιασμένες για παραγωγή από κανονική κατανομή.

2.3.2 Η μέθοδος της αντιστροφής

Ας υποθέσουμε ότι θέλουμε να παράγουμε μία συνεχή τυχαία μεταβλητή X με αθροιστική συνάρτηση κατανομής $F(x)$, και επίσης ας υποθέσουμε ότι η αντίστροφη συνάρτηση, $F^{-1}(u)$, είναι καλά ορισμένη για $0 \leq u \leq 1$. Θα αποδείξουμε ότι αν U είναι μία τυχαία μεταβλητή (τ.μ.) που ακολουθεί την ομοιόμορφη $U(0,1)$ κατανομή, τότε η $X = F^{-1}(U)$ έχει την ζητούμενη κατανομή. Πράγματι, επειδή $Pr(U \leq x) = x$, έχουμε

$$X = F^{-1}(U) \Rightarrow Pr(X \leq x) = Pr(F^{-1}(U) \leq x) = Pr(U \leq F(x)) = F(x)$$

Αλλά διακριτή F^{-1} δεν υπάρχει, οπότε αν η τ.μ. από την οποία θέλουμε να παράγουμε ανεξάρτητο δείγμα είναι διακριτή, η παραπάνω μεθοδολογία πρέπει να τροποποιηθεί: ορίζουμε

$$F^{-}(u) = \min\{x : F(x) \geq u\} \tag{2.2}$$

Για κάποιο u που παρήχθη από την $U(0,1)$, η $F^{-}(u)$ παίρνει ως τιμή το μικρότερο από εκείνα τα x για τα οποία ισχύει η σχέση $F(x) \geq u$. Συνεπώς η μέθοδος της αντιστροφής ισχύει εν γένει. Επί πλέον, αν θέλουμε ανεξάρτητο δείγμα από την X, ξεκινούμε με ανεξάρτητο δείγμα από την $U(0,1)$

Παράδειγμα 5 *Ας υποθέσουμε ότι θέλουμε να προσομοιώσουμε μία τ.μ. X με κατανομή*

$$X = \begin{cases} 1 & \text{με πιθανότητα } p \\ 0 & \text{με πιθανότητα 1-}p \end{cases}$$

Η διαίσθησή μας θα μας προέτρεπε στην εξής διαδικασία: Παράγουμε $U \sim U(0,1)$, και αν $U \geq 1-p$ θέτουμε $X = 1$. Αλλιώς θέτουμε $X = 0$. Αυτή είναι και η προτεινόμενη μέθοδος σύμφωνα με την (2.2).

Παράδειγμα 6 *Έστω ότι η τ.μ. X ακολουθεί την εκθετική κατανομή με παράμετρο λ, σ.π.π. $f(x)$, και α.σ.κ. $F_X(x)$:*

$$X \sim Exp(\lambda),\ f(x) = \lambda e^{-\lambda x},\ F_X(x) = 1 - e^{-\lambda x}, \quad x, \lambda > 0$$

Για να προσομοιώσουμε την X, θέτουμε $X = F^{-1}(U)$, δηλαδή έχουμε

$$U = 1 - e^{-\lambda X} \Rightarrow X = -\lambda^{-1} log(1-U)$$

Αλλά αν $U \sim U(0,1) \Rightarrow (1-U) \sim U(0,1)$, όπως είναι διαισθητικά προφανές και αποδεικνύεται εύκολα με μετασχηματισμό μεταβλητών $U \to 1-U$. Αρα ο ζητούμενος αλγόριθμος είναι

1. *Παράγουμε $U \sim U(0,1)$*
2. *Θέτουμε $X = -\lambda^{-1} log(U)$*
3. *Τότε $X \sim Exp(\lambda)$*

Παράδειγμα 7 *'Εστω ότι η τ.μ. X ακολουθεί την κατανομή Cauchy με σ.π.π.*

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \; x \in \Re$$

Τότε

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \frac{1}{2} + \pi^{-1} \varepsilon\varphi^{-1} x \Rightarrow$$
$$\Rightarrow F^{-1}(u) = \varepsilon\varphi\left(\pi(u - \frac{1}{2})\right)$$

και ο αλγόριθμος προσομοίωσης της X είναι

1. *Παράγουμε $U \sim U(0,1)$*
2. *Θέτουμε $X = \varepsilon\varphi\left(\pi(U - \frac{1}{2})\right)$*
3. *Τότε $X \sim Cauchy$*

Παράδειγμα 8 *'Εστω ότι θέλουμε να παράγουμε τιμές από μία τ.μ. X που ακολουθεί την τυποποιημένη κανονική κατανομή. Τότε η α.σ.κ. δινεται από*

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{2\pi} e^{-2s^2} ds \Rightarrow F_X(x) = \,?$$

Δηλαδή η α.σ.κ. δεν είναι δυνατόν να αντισταφεί. Η απάιτηση της ύπαρξης της αντιστρόφου της α.σ.κ. είναι λοιπόν ΣΗΜΑΝΤΙΚΟΣ ΠΕΡΙΟΡΙΣΜΟΣ *της μεθόδου της αντιστροφης.*

2.3.3 Μέθοδος της απόρριψης

Πρόβλημα: Θέλουμε να παράγουμε X από σ.π.π $f = f_X$, αλλά είναι δύσκολο. Ταυτόχρονα, είναι εύκολο να παράγω Y από σ.π.π. g. Ας υποθέσουμε ότι οι f, g είναι τέτοιες ώστε να υπάρχει $M > 0$ με $\frac{f}{g} \leq M < \infty$.

ΒΗΜΑ 1:

1. Παράγουμε Y από g
2. Για κάποια συνάρτηση $h(.)$ με τιμές στο $[0,1]$, δεδομένης $Y=y$, θέτουμε $X=y$ με πιθανότητα $h(y)$. Αλλιώς, επιστρέφουμε στο 1.

Τι ιδιότητες έχει αυτός ο αλγόριθμος;

$$P\left[(Y \leq x)\bigcap(X \text{ παίρνει την τιμή της } Y\right] = \int_{-\infty}^{x} h(y)g(y)dy$$

$$\Rightarrow P\left[X \text{ παίρνει την τιμή της } Y\right] = \int_{-\infty}^{\infty} h(y)g(y)dy$$

Αρα

$$P\left[Y \leq x \mid X \text{ παίρνει την τιμή της } Y\right] = \frac{\int_{-\infty}^{x} h(y)g(y)dy}{\int_{-\infty}^{\infty} h(y)g(y)dy}$$

Τώρα αν διαλέξουμε $h(y)=\frac{f(y)}{Mg(y)}$;

$$\Rightarrow \text{Η σ.π.π. των ''δεκτών X''} = \frac{\frac{f}{Mg}g}{\int \frac{f}{Mg}g} = \frac{f}{\int f} = f \quad (\text{ότι ζητάμε!})$$

Αρα

ΑΛΓΟΡΙΘΜΟΣ

1. Παράγουμε $Y \to Y=y$ από $g(y)$
2. Παράγουμε $U \sim U(0,1) \to U=u$
3. AN $u \leq \frac{f(y)}{Mg(y)}$ TOTE $X=y$
   ΑΛΛΙΩΣ επιστρέφουμε στο 1.

Ο αλγόριθμος δίνει ένα τρόπο παραγωγής τυχαία μέσα στην περιοχή κάτω από την καμπύλη Mg. Δεχόμαστε τα σημεία που πέφτουν κάτω από την καμπύλη f. Αυτό φαίνεται στην συνθήκη $u \leq f(y)/(Mg(y))$ ή $Mg(y)u \leq f(y)$, όπου $Mg(y)u$ είναι τυχαίο σημείο κάτω από την Mg. Λόγω της παραπάνω διαισθητικής περιγραφής, η Mg ονομάζεται *φάκελλος* της f. Πόσες προσπάθειες χρειαζόμαστε μέχρι να δεχθούμε την τιμή $X=y$ στο βήμα 3;

$$P\left[\text{δεχόμαστε X}\right] = \int_{-\infty}^{\infty} h(y)g(y)dy = \int_{-\infty}^{\infty} \frac{f(y)}{Mg(y)}g(y)dy = \frac{1}{M}\int f = \frac{1}{M}$$

$$\Rightarrow E\left[\text{προσπάθειες να πάρουμε μία τιμή της X}\right] = M$$

Σημειώσεις

1. Η σταθερά M σχετίζεται με την επιλογή της g. Οσο πιό πολύ μιμείται η g το σχήμα της f τόσο πιό μικρό είναι το Μ, άρα πιό αποδοτική είναι η διαδικασία. Η επιλογή της g, όμως, θα χαρακτηριζόταν καλύτερα σαν *τέχνη!*

2. Σχεδόν πάντοτε μπορούμε να βρούμε μία σ.π.π. g και μία σταθερά Μ που να ικανοποιούν την συνθήκη $f \leq Mg$, εκτός ίσως όταν η f είναι μη φραγμένη ή οι ουρές της είναι πολύ παχείς (πχ $f \equiv Cauchy$, $g \equiv Κανονική$. Τότε $f/g \not\leq M\ \forall x$.

3. Οσον αφορά την αποδοτικότητα του αλγόριθμου, πρέπει να βρούμε μία g να μιμείται την f. Χρειαζόμαστε

   (a) Να καταλάβουμε το σχήμα της f με λεπτομέρεια

   (b) Να βρούμε ένα τρόπο να παράγουμε από την g.

Παράδειγμα 9 *Ας υποθέσουμε ότι θέλουμε να παράγουμε τιμές από την*

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, \quad x \geq 0$$

και χρησιμοποιούμε σαν σ.π.π. που θα παράγει τις αρχικές τιμές την $g(x) = e^{-x}$. Ο 'φάκελλος' που θα χρησιμοποιήσουμε είναι η $Mg(x)$, για κατάλληλο Μ. Καθώς $x \to \infty$, η f συγκλίνει στο 0 με πιο γρήγορο ρυθμό από ότι η g, και συνεπώς, αν η Mg τέμνει την f σε ένα σημείο, αυτό θα σημαίνει αυτόματα ότι $Mg(x) \geq f(x)\ \forall x$. 'Ενας τρόπος να υπολογίσουμε το Μ είναι να βρούμε την συνθήκη κατά την οποία η εξίσωση

$$Me^{-x} = \sqrt{\frac{2}{\pi}} e^{\frac{-x^2}{2}}$$

έχει δύο ρίζες. Είναι προφανές, ότι καμμία ρίζα ισοδυναμεί με κανένα σημείο επαφής, και δύο ρίζες με δύο σημεία τομής. Η παραπάνω εξίσωση καταλήγει στην

$$x^2 - 2x + 2\ln\left(M\sqrt{\frac{\pi}{2}}\right) = 0$$

και ίσες ρίζες έχουμε αν $M^2\pi = 2e \Leftrightarrow M = \sqrt{\frac{2e}{\pi}}$. Συνεπώς ο αλγόριθμος προσομοίωσης της $f(x)$ είναι:

1. *Παράγουμε X από την σ.π.π e^{-x}, για $x \geq 0$. Είναι γνωστό ότι αυτό επιτυγχάνεται με την μέθοδο της αντιστροφής, θέτοντας $X = -\ln U_1$ όπου $U_1 \sim U(0,1)$.*

2. *Για $U_2 \sim U(0,1)$, θέτουμε $Y = MU_2 e^{-x} = MU_2 U_1$*

3. *Δεχόμαστε X αν και μόνο αν*

$$Y < \sqrt{\frac{2}{\pi}} e^{-X^2/2} \Leftrightarrow MU_1U_2 < \sqrt{\frac{2}{\pi}} e^{-X^2/2} \Leftrightarrow U_1U_2 < e^{-\frac{1}{2}\left(1+X^2\right)}$$

Αρα τελικά ο αλγόριθμος δεν περιλαμβάνει (άμεσα) το M. Σημειώνουμε εδώ ότι η $f(x)$ είναι η μισή κανονική τυποποιημένη κατανομή. Αν θέλουμε να προσομοιώσουμε από την κανονική κατανομή $\tilde{X}$, μπορούμε να παράγουμε μία τιμή από την $U(0,1)$, και ελέγχοντας αν είναι μικρότερη η μεγαλύτερη από το $1/2$, να δίνουμε τιμές αντίστοιχα $\tilde{X} = -X$ ή $\tilde{X} = X$.

Εδώ, χρησιμοποιούμε ουσιαστικά την μέθοδο της σύνθεσης, που θα συναντήσουμε παρακάτω.

Επέκταση

'Εστω ότι θέλουμε να παράγουμε X από σ.π.π. $f^*/\int f^*$ ή, γενικώτερα, έστω ότι θέλουμε X από σ.π.π. $\propto f^*$. Ακολουθούμε ακριβώς την ίδια διαδικασία, θέτοντας $h = f^*/Mg$. Η διαφορά είναι ότι

$$P\,[\text{δεχόμαστε X}] = \int_{-\infty}^{\infty} hg = \frac{\int f^* g}{Mg} = \frac{\int f^*}{M}$$

Για παράδειγμα

$$h = \frac{f^*}{Mg}, \text{ όπου } M = \sup_x \frac{f^*(x)}{g(x)}$$

Παράδειγμα 10 *Οι σ.π.π που συναντώνται στην Μπευζιανή Στατιστική, βασίζονται στο θεώρημα του Bayes:*

$$f(x \mid z) = \frac{f(z \mid x)f(x)}{\int f(z \mid x)f(x)dx} \propto f(z \mid x)f(x)$$

'Ενα (σπουδαίο) πρόβλημα που συναντάται είναι η παραγωγή ανεξάρτητου τυχαίου δείγματος από την a posteriori κατανομή $f(x \mid z)$, αλλά συνήθως το ολοκλήρωμα στον παρονομαστή (σταθερά κανονικοποίησης) είναι δύσκολο να υπολογισθεί αναλυτικά, και η a posteriori είναι διαθέσιμη μόνο σαν $f^* = f(z \mid x)f(x)$

Παράδειγμα 11 *'Εστω ότι θέλουμε να παράγουμε $X \sim Beta(a,b)$ με σ.π.π.*

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \propto x^{a-1}(1-x)^{b-1} \equiv f^*(x), \quad x \in (0,1)$$

'Εστω ότι $a > 1, b > 1$. Είναι ευκολότερο να δουλέψουμε με την f^* παρά με την f, και ας δούμε πως υλοποιείται η μέθοδος της απόρριψης:

1. *Επιλέγουμε να προσομοιώσουμε από την f^**
2. *Επιλογή της g; Εύκολη λύση: $g \equiv U(0,1)$*
3. *Εξεύρεση M: $M = \max f^*/g$.*

$$\frac{f^*(x)}{g(x)} = x^{a-1}(1-x)^{b-1}$$

$$\frac{d}{dx}\left(\frac{f^*(x)}{g(x)}\right) = 0 \Leftrightarrow \hat{x} = \frac{a-1}{a+b-2}$$

$$\Rightarrow \frac{f^*}{g} \leq \frac{(a-1)^{a-1}(b-1)^{b-1}}{(a+b-2)^{a+b-2}} = M \Leftrightarrow$$
$$M = \left(\frac{a-1}{a+b-2}\right)^{a-1}\left(\frac{b-1}{a+b-2}\right)^{b-1}.$$

2.3.4 Μέθοδος της σύνθεσης

Σε πολλές περιπτώσεις, η σ.π.π $f(x)$ που θέλουμε να προσομοιώσουμε μπορεί να γραφεί σαν σταθμικό άθροισμα ή *μίξη* k άλλων σ.π.π.:

$$f = \pi_1 f_1 + \pi_2 f_2 + \ldots \pi_k f_k$$

όπου $\pi_i \geq 0$, $\sum \pi_i = 1$, και οι f_i, $i = 1, 2, \ldots, k$ είναι σ.π.π. και λέγονται *πυκνότητες μίξης*. Τα π_i μπορούν να θεωρηθούν σαν συνάρτηση μάζας πιθανότητας για μιά διακριτή τ.μ. Y, με $Pr(Y = i) = \pi_i$. Στο πρώτο στάδιο της μεθόδου της σύνθεσης παράγεται μία τιμή της Y, έστω $Y = j$, χρησιμοποιώντας τα π_i. Το δεύτερο στάδιο αποτελείται από παραγωγή μιάς τιμής της X_j από την σ.π.π. f_j. 'Εχουμε ότι

$$\begin{aligned}
Pr(X \leq x) &= \pi_1 Pr\left(X \leq x \mid \text{δείγμα από } f_1\right) + \\
&\quad\ \pi_2 Pr\left(X \leq x \mid \text{δείγμα από } f_2\right) + \\
&\quad\ \vdots \\
&\quad\ \pi_k Pr\left(X \leq x \mid \text{δείγμα από } f_k\right) = \\
&= \pi_1 \int_{-\infty}^{x} f_1(y)dy + \ldots + \pi_k \int_{-\infty}^{x} f_k(y)dy = \\
&= \int_{-\infty}^{x} f(y)dy.
\end{aligned}$$

Η ιδέα εδώ είναι ότι μπορεί η προσομοίωση των f_i να είναι πολύ ευκολότερη από την παραγωγή της f. Επίσης, οι μίξεις σ.π.π. συναντώνται συχνά σε (ρεαλιστικά) μοντέλα στην ψυχολογία, ιατρική, κλπ (πχ δες Titterington, D.M., Smith A.F.M. και Makov, U.E. (1985): *Statistical Analysis of Finite Mixture Distributions*, New York, Wiley).

Παράδειγμα 12 *'Εστω ότι θέλουμε να προσομοιώσουμε την τ.μ. X που ακολουθεί την κατανομή βήτα με παραμέτρους a, b, με $a = b < 1$. Η σ.π.π. είναι*

$$f(x) = \frac{\Gamma(2a)}{\Gamma(a)\Gamma(a)} x^{a-1}(1-x)^{a-1} = \frac{1}{2} f_1(x) + \frac{1}{2} f_2(x), \ x \in (0,1)$$

όπου

$$f_1(x) = \begin{cases} \frac{2\Gamma(2a)}{\Gamma(a)\Gamma(a)} x^{a-1}(1-x)^{a-1}, & x \in (0, 1/2) \\ 0 & x \in (1/2, 1) \end{cases}$$

και

$$f_2(x) = \begin{cases} \frac{2\Gamma(2a)}{\Gamma(a)\Gamma(a)} x^{a-1}(1-x)^{a-1}, & x \in (1/2, 1) \\ 0 & x \in (0, 1/2) \end{cases}$$

Για την μέθοδο της σύνθεσης απαιτείται η παραγωγή των f_1 και f_2, και χρησιμοποιούμε γι αυτό την μέθοδο της απόρριψης. Η εύρεση του 'φάκελλου' των f_1 και f_2 δεν είναι εύκολη, και εδώ θα αναφέρουμε ένα παράδειγμα επιλογής της g_1, ενώ η g_2 μπορεί να επιλεγεί παρόμοια. Έστω

$$g_1(x) = \begin{cases} a2^a x^{a-1} & x \in (0, 1/2) \\ 0 & x \in (1/2, 1) \end{cases}$$

Παράγουμε από την g_1 με την μέθοδο της αντιστροφής:

$$G_1(x) = \int_0^x a2^a t^{a-1} dt = \begin{cases} (2x)^a & x \in (0, 1/2) \\ 0 & x < 0 \\ 1 & x > 1/2 \end{cases} \quad \Rightarrow G^{-1}(u) = \frac{1}{2} u^{1/a}$$

Συνεπώς ο αλγόριθμος παραγωγής από την f, αποτελείται αρχικά από την επιλογή της f_1 ή f_2 με πιθανότητα $1/2$, και στην συνέχεια παραγωγή από πχ την f_1 με την μέθοδο της απόρριψης.

2.3.5 Μέθοδοι Box-Müller και Polar-Marsaglia

Οι μέθοδοι Box-Müller και Polar-Marsgalia προσομοιώνουν την τυποποιημένη κανονική κατανομή χρησιμοποιώντας έναν $1 \leftrightarrow 1$ μετασχηματισμό δύο ομοιόμορφα κατανεμημένων $U(0, 1)$ τ.μ. Η διαδικασία Box-Müller προχωρεί ως εξής:

Έστω δύο τυποποιημένες κανονικές τ.μ. $N_1(0, 1)$ και $N_2(0, 1)$, και ας υποθέσουμε ότι παράγουν δύο τιμές αντίστοιχα N_1 και N_2. Έτσι, ορίζεται στον διδιάστατο χώρο ένα σημείο με καρτεσινές συντεταγμένες, και αν θέλουμε να περάσουμε σε πολικές συντεταγμένες, τότε γράφουμε

$$N_1 = R\sigma\upsilon\nu\Theta$$

$$N_2 = R\eta\mu\Theta$$

και όπως έχει ήδη αποδειχθεί (Φρον. ασκήσεις 1) R και Θ είναι ανεξάρτητες τ.μ. με

$$\Theta \sim U(0, 2\pi)$$

$$R^2 = N_1^2 + N_2^2 \sim \chi_2^2 \equiv Exp(1/2)$$

Επί πλέον, για να παράγουμε Θ, απλά παράγουμε $U_2 \sim U(0, 1)$ και θέτουμε $\Theta = 2\pi U_2$, και για να παράγουμε R θέτουμε $R = (-2 \ln U_1)^{1/2}$ όπου $U_1 \sim U(0, 1)$. Συνεπώς, η μέθοδος Box-Müller παράγει πολικές συντεταγμένες και μετά μετασχηματίζει τα R, Θ στις καρτεσιανές συντεταγμένες N_1 και N_2. Για κάθε ζευγάρι τιμών κανονικής κατανομής, απαιτούνται δύο ανεξάρτητες τιμές U_1 και U_2 και επίσης ο υπολογισμός δύο τριγωνομετρικών συναρτήσεων. Αν επιθυμούμε δείγμα μεγάλου μεγέθους από την κανονική κατανομή, οι τριγωνομετρικές συναρτήσεις μπορεί να επιβαρύνουν αισθητά την αποδοτικότητα του αλγόριθμου, και συνιστάται η παρακάτω τροποποίηση, συνήθως ονομαζόμενη Polar-Marsaglia μέθοδος.

Η κεντρική ιδέα βασίζεται στην κατασκευή ημιτόνων και συνημιτόνων ομοιόμορφα κατανεμημένων γωνιών χωρίς πρώτα να προσομοιωθούν οι γωνίες (!). Αυτό επιτυγχάνεται με την μέθοδο της απόρριψης, ως ακολούθως: Παράγουμε ανεξάρτητα $U_1, U_2 \sim U(0, 1)$ και θέτουμε

$$V_i = 2U_i - 1 \Rightarrow V_i \sim U(-1, 1), \quad i = 1, 2$$

Αρα το ζευγάρι (V_1, V_2) αντιστοιχεί σε ένα τυχαίο σημείο στο τετράγωνο μήκους 2 με κέντρο το $(0,0)$. Συνεχίζουμε να παράγουμε τέτοια ζευγάρια

$$(V_1, V_2), (V_2, V_3), \ldots, (V_i, V_{i+1}, \ldots$$

έως ότου $V_i^2 + V_{i+1}^2 < 1$, δηλαδή έως ότου το σημείο ανήκει στον μοναδιαίο κύκλο με κέντρο το $(0,0)$. Τότε θέτουμε

$$R^2 = V_1^2 + V_2^2$$

$$\varepsilon\varphi\Theta = \frac{V_2}{V_1}$$

και μπορεί να αποδειχθεί ότι

$$R^2 \sim U(0,1) \quad \text{και} \quad \Theta \sim U(0, 2\pi)$$

όπου τα R^2 και Θ είναι ανεξάρτητες τ.μ. Συνεπώς το ζευγάρι (R, Θ) είναι ότι απαιτείται για την μέθοδο Box Müller, και μπορούμε να γράψουμε

$$\eta\mu\Theta = V_2(V_1^2 + V_2^2)^{-\frac{1}{2}}$$

$$\sigma\upsilon\nu\Theta = V_1(V_1^2 + V_2^2)^{-\frac{1}{2}}.$$

Αρα ένα ζευγάρι ανεξαρτήτων $N(0,1)$ τ.μ., N_1 και N_2, δίνεται από

$$N_1 = \left(-2\ln(R^2)\right)^{\frac{1}{2}} V_2(V_1^2 + V_2^2)^{-\frac{1}{2}}$$

$$N_2 = \left(-2\ln(R^2)\right)^{\frac{1}{2}} V_1(V_1^2 + V_2^2)^{-\frac{1}{2}}$$

ή

$$N_1 = \left(-2\ln(V_1^2 + V_2^2)\right)^{\frac{1}{2}} V_2(V_1^2 + V_2^2)^{-\frac{1}{2}}$$

$$N_2 = \left(-2\ln(V_1^2 + V_2^2)\right)^{\frac{1}{2}} V_1(V_1^2 + V_2^2)^{-\frac{1}{2}}$$

καταλήγοντας στις

$$N_1 = V_2 \left(\frac{-2\ln W}{W}\right)^{\frac{1}{2}}$$

$$N_2 = V_1 \left(\frac{-2\ln W}{W}\right)^{\frac{1}{2}}$$

όπου $W = V_1^2 + V_2^2$. Η ιδέα της χρήσης της μεθόδου της απόρριψης για να αποφύγουμε τον υπολογισμό των τριγωνομετρικών συναρτήσεων ίσως φαίνεται παράξενη, αλλά παρατηρείστε ότι τα ζευγάρια (V_1, V_2) απορρίπτονται μόνο με πιθανότητα $1 - \pi/4$.

2.3.6 Παραγωγή τυχαίων μεταβλητών από διακριτές κατανομές

Μέθοδος της αντιστροφής

Έχουμε ήδη δεί ότι διακριτές κατανομές παράγονται εύκολα με την μέθοδο της αντιστροφής. Έστω $Pr(X = i) = p_i$, και $P_i = \sum_{j \leq i} p_j$, για $i = 1, 2, \ldots, n$. Τότε

$$F^{-1}(u) = \min\{x \mid F_X(x) \geq u\} = i, \text{ αν } P_{i-1} < u \leq P_i, \ P_0 = 0$$

και ο αλγόριθμος είναι

1. $U \sim U(0,1) \to U = u$
2. Έλεγχος $P_i \leq u$
   ΑΝ ΝΑΙ $i \to i+1$
   ΑΝ ΟΧΙ $X = i$

Αλλά ο παραπάνω αλγόριθμος παρουσιάζει το μειονέκτημα ότι αρχίζει από αριστερά και απαιτεί για την επιστροφή $X = i$, i συγκρίσεις. Συνεπώς

$$E(\text{αριθμός συγκρίσεων}) = \sum_i i p_i = E(X)$$

και ο αλγόριθμος, ειδικά για μεγάλο n, μπορεί να είναι πολύ αργός.

Αν η σ.π.π. είναι γνωστό ότι είναι περίπου συμμετρική και με μία κορυφή, τότε ένας εναλλακτικός αλγόριθμος θα άρχιζε από την "μέση", και θα επαναλάμβανε την διαδικασία ερευνώντας αριστερά-δεξιά:

1. Θέτουμε $L = 0, R = n$
2. Παράγουμε $U \sim U(0,1), \to U = u$
3. Θέτουμε $i = \frac{L+R}{2}$
4. ΑΝ $u > P_i$ θέτουμε $L = i$
   ΑΝ $u \leq P_i$ θέτουμε $R = i$
   UNTIL $L \geq R - 1$
5. Θέτουμε $X = i$

Μέθοδος της σύνθεσης

Ας υποθέσουμε ότι θέλουμε να προσομοιώσουμε την $X \sim Bin(3, 1/3)$. Υπάρχουν δύο προφανείς μέθοδοι:

1. Παράγουμε ανεξάρτητες τιμές u_1, u_2, u_3 από την ομοιόμορφη $U(0,1)$, και μετρούμε πόσες από αυτές ανήκουν στο διάστημα $(0, 1/3)$. Εν γένει, για την $B(n, 1/3)$ απαιτούνται n παραγωγές της $U(0,1)$.

2. Χρησιμοποιούμε την μέθοδο της αντιστροφής

Θα δούμε τώρα μία άλλη μέθοδο: ας σχηματίσουμε τις πιθανότητες με ακρίβεια τριών δεκαδικών.

i	p_i
0	0.2 9 6
1	0.4 4 5
2	0.2 2 2
3	0.0 3 7
Αθροισμα:	81820

'Εχουμε

$$p_0 = 0.296 = 0.8 \times \frac{2}{8} + 0.18 \times \frac{9}{18} + 0.02 \times \frac{6}{20}$$
$$p_1 = 0.445 = 0.8 \times \frac{4}{8} + 0.18 \times \frac{4}{18} + 0.02 \times \frac{5}{20}$$
$$p_2 = 0.222 = 0.8 \times \frac{2}{8} + 0.18 \times \frac{2}{18} + 0.02 \times \frac{2}{20}$$
$$p_3 = 0.037 = 0.8 \times \frac{0}{8} + 0.18 \times \frac{3}{18} + 0.02 \times \frac{7}{20}$$

Δηλαδή $p_i = 0.8f_{i1} + 0.18f_{i2} + 0.02f_{i3}$, όπου

$$(f_{01}, f_{11}, f_{21}, f_{31}) = \left\{\frac{2}{8}, \frac{4}{8}, \frac{2}{8}, 0\right\}$$

$$(f_{02}, f_{12}, f_{22}, f_{32}) = \left\{\frac{9}{18}, \frac{4}{18}, \frac{2}{18}, \frac{3}{18}\right\}$$

$$(f_{03}, f_{13}, f_{23}, f_{33}) = \left\{\frac{6}{20}, \frac{5}{20}, \frac{2}{20}, \frac{7}{20}\right\}$$

Αρα ο αλγόριθμος είναι:

1. Παράγουμε j^* από

j	π_j
1	0.8
2	0.18
3	0.02

2. Δεδομένης της τιμής j^*, παράγω i^* από

i	f_{ij^*}
0	f_{0j^*}
1	f_{1j^*}
2	f_{2j^*}
3	f_{3j^*}

3. Θέτουμε $X = i^* \Rightarrow X \sim Bin(3, 1/3)$

Βελτιώνεται η αποδοτικότητα με τον παραπάνω αλγόριθμο; είναι μεν πιό γρήγορος, αλλά έχει το μειονέκτημα της κατασκευής των π_j, f_{ij}. Η παραπάνω όμως ιδέα έχει "καλλιεργηθεί" και έτσι ο παραπάνω αλγόριθμος αποτελεί ένα παράδειγμα της μεθόδου "Alias". Αν και δεν θα ασχοληθούμε εδώ με αυτή την μέθοδο, θα δούμε απλώς πως διυλίζεται ο παραπάνω αλγόριθμος. Το πρόβλημα παραμένει προσομοίωση της δυωνυμικής $Bin(3, 1/3)$, και αρχίζουμε πάλι με την κατασκευή της σ.π.π.:

i	p_i
0	8/27
1	12/27
2	6/27
3	6/27

και κατασκευάζουμε (με κάποιο τρόπο) την ακόλουθη σύνθεση:

$$\frac{8}{27} = \frac{1}{4}\left(\frac{9}{27} + 0 + 0 + \frac{23}{27}\right)$$
$$\frac{12}{27} = \frac{1}{4}\left(\frac{18}{27} + \frac{9}{27} + \frac{21}{27} + 0\right)$$
$$\frac{6}{27} = \frac{1}{4}\left(0 + \frac{8}{27} + \frac{6}{27} + 0\right)$$
$$\frac{1}{27} = \frac{1}{4}\left(0 + 0 + 0 + \frac{4}{27}\right)$$

και όπως και προηγουμένως, $p_i = \sum \pi_j f_{ij}$, $\pi_j = \frac{1}{4}$, τα f_{ij} δίνονται από

$$(f_{01}, f_{11}, f_{21}, f_{31}) = \left\{\frac{9}{27}, \frac{18}{27}, 0, 0\right\}$$
$$(f_{02}, f_{12}, f_{22}, f_{32}) = \left\{0, \frac{9}{27}, \frac{18}{27}, 0\right\}$$
$$(f_{03}, f_{13}, f_{23}, f_{33}) = \left\{0, \frac{21}{27}, \frac{6}{27}, 0\right\}$$
$$(f_{04}, f_{14}, f_{24}, f_{34}) = \left\{\frac{23}{27}, 0, 0, \frac{4}{27}\right\},$$

και ο αλγόριθμος είναι

1. Διαλέγουμε f (ή π_j) με πιθανότητα 1/4
2. Διαλέγουμε $i = 0, 1, 2$ ή 3 με πιθανότητα που δίνεται από τις τιμές f_{ij}, για δεδομένο j.
3. Θέτουμε $X = i \Rightarrow X \sim Bin(3, 1/3)$

Σημειώνουμε εδώ ότι όλες οι διακριτές σ.π.π. f είναι κατανομές 2 σημείων, άρα πολύ εύκολες για παραγωγή. Επίσης, $\pi_j = 1/4$ και σε αυτόν τον αλγόριθμο δεν χρειάζεται να κατασκευάσουμε τα π_j. Ο αλγόριθμος "Alias" κατασκευάζει πάντα τέτοιου είδους αποσύνθεση της διακριτής κατανομής που θέλουμε να προσομοιώσουμε, και είναι πολύ αποδοτικός (γρήγορος).

Κεφάλαιο 3

Τεχνικές ελάττωσης διασποράς και ολοκλήρωσης Monte Carlo

3.1 Monte Carlo ολοκλήρωση

Ας υποθέσουμε ότι η τιμή του ολοκληρώματος

$$\theta = \int \phi(x) f(x) dx \tag{3.1}$$

είναι άγνωστη, και έστω ότι η $f(x)$ είναι σ.π.π. Για παράδειγμα, αν ή $\phi(x)$ παίρνει τιμές x, x^r, $I_{(x<c)}$ το θ δίνει αντίστοιχα τον μέσο, τις ροπές r-τάξης, και την αθροιστική συνάρτηση κατανομής της f. Περιπτώσεις όπου το ολοκλήρωμα της (3.1) είναι αναλυτικά αδύνατο να υπολογισθεί, είναι πχ όταν το $\phi(x)f(x)$ ή το διάστημα της ολοκλήρωσης είναι πολύ περίπλοκα.

Το ανωτέρο πρόβλημα είναι μη-πιθανοθεωρητικό, αλλά η λύση του μπορεί να βασιστεί σε ιδέα παρμένη από την θεωρία πιθανοτήτων, ως εξής:

Ας υποθέσουμε ότι παράγουμε τυχαίο δείγμα $X_1, X_2, \ldots, X_n$ από την κατανομή με σ.π.π. f. Θέτουμε

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} \phi(X_i)$$

Τότε

$$\begin{aligned} E(\hat{\theta}) &= n^{-1} \sum_{i=1}^{n} E\left(\phi(X_i)\right) = n^{-1} n E\left(\phi(X)\right) = \\ &= E\left(\phi(X)\right) = \int \phi(x) f(x) dx = \theta \end{aligned}$$

$$V(\hat{\theta}) = n^{-1} V\left(\phi(X)\right) = n^{-1} \int \left[\phi(x) - \theta\right]^2 f(x) dx = \frac{c}{n} \tag{3.2}$$

για κάποια σταθερά c. 'Αρα το τυπικό σφάλμα της εκτιμήτριας $\hat{\theta}$ είναι ανάλογο του $1/\sqrt{n}$ και η $\hat{\theta}$ είναι συνεπής εκτιμήτρια της θ. Για ευκολία μας, σε αυτές τις σημειώσεις, θα ονομάζουμε την παραπάνω μέθοδο '$\phi - f$' μέθοδο.

Είναι σαφές ότι διαφορετικές επιλογές f και ϕ καταλήγουν σε διαφορετικές σταθερές c στην (3.2), και το πρόβλημα είναι να βρούμε κατάλληλες επιλογές που να επιτυγχάνουν όσο το δυνατόν μικρότερο c. Ας δούμε μερικά παραδείγματα επιλογών '$\phi - f'$, όταν ζητείται η τιμή του ολοκληρώματος

$$\theta = \int_2^\infty \frac{1}{\pi}\frac{1}{1+x^2}dx, \tag{3.3}$$

που ισούται με $P(X > 2)$ όταν $X \sim Cauchy$:

1.

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \Re$$

$$\phi(x) = I_{(x>2)} = \begin{cases} 1, x > 2 \\ 0, x \leq 2 \end{cases}$$

2. Το θ μπορεί να είναι $\theta = \frac{1}{2}P(|x| > 2)$

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \Re$$

$$\phi(x) = I_{(|x|>2)} = \begin{cases} 1, |x| > 2 \\ 0, |x| \leq 2 \end{cases}$$

3.

$$1 - 2\theta = \int_{-2}^2 \frac{1}{\pi(1+x^2)}dx = 2\int_0^2 \frac{1}{\pi(1+x^2)}dx$$

$$f(x) = \frac{1}{2}, x \in (0,2)$$

$$\phi(x) = \frac{2}{\pi(1+x^2)}$$

4. Αν $y = x^{-1}$, τότε

$$\theta = \int_2^\infty \frac{1}{\pi(1+x^2)}dx = \int_0^{1/2} \frac{y^{-2}}{\pi(1+y^{-2})}dy = \int_0^{1/2} \frac{1}{\pi(1+y^2)}dy.$$

$$f(x) = 2, x \in (0, 1/2)$$

$$\phi(x) = \frac{1}{2\pi(1+x^2)}$$

Στην συνέχεια, και για να επιτραπεί κάποια σύγκριση της '$\phi - f'$ μεθόδου με μία πιο απλούστερη τεχνική ολοκλήρωσης, θα εισάγουμε την μέθοδο ολοκλήρωσης 'hit and miss'. 'Εστω ότι $\phi(x)$, $x \in (a,b)$ είναι μία φραγμένη συνάρτηση με $0 \leq \phi(x) \leq c$, και ας υποθέσουμε ότι ζητείται η τιμή ψ της περιοχής κάτω από την καμπύλη ϕ, δηλαδή

$$\psi = \int_a^b \phi(x)dx = (b-a)\int_a^b \phi(x)f(x)dx, \quad f(x) = \frac{1}{b-a}, \quad x \in (a,b)$$

Η μέθοδος 'hit and miss'

Ο αλγόριθμος 'hit and miss' παράγει τυχαία σημεία στο παραλληλόγραμμο που ορίζεται από τα σημεία $(a,0),(b,0),(a,c),(b,c)$ και εκτιμά το ψ από την συχνότητα των σημείων που πέφτουν κάτω από την καμπύλη $\phi(x)$. Πιο συγκεκριμένα, έστω $U_i \sim U(a,b)$ και $V_i \sim U(0,c)$, $i=1,2,\ldots,n$· δημιουργούμε έτσι ένα ομοιόμορφο τυχαίο δείγμα μεγέθους n στο παραλληλόγραμο που περικλείει την ϕ. Ορίζουμε

$$\tilde{\psi} = \underbrace{c(b-a)}_{\substack{\text{Εμβαδόν}\\ \text{του}\\ \text{παρ/μου}}} \overbrace{\frac{1}{n}\sum_{i=1}^{n} I_{(V_i \leq \phi(U_i))}}^{\substack{\text{συχνότητα σημείων}\\ \text{κάτω από την } \phi(x)}}$$

$$E(\tilde{\psi}) = c(b-a)\frac{1}{n}nE\left[I_{(V\leq\phi(U))}\right] = c(b-a)P\left(V \leq \phi(U)\right) = \psi$$

$$\begin{aligned} V(\tilde{\psi}) &= c^2(b-a)^2\frac{1}{n^2}n\left[E(I^2)-E^2(I)\right], \quad I \sim Bernoulli \\ &= c^2(b-a)^2\frac{1}{n}\left\{P\left(V\leq\phi(U)\right)\left[1-P\left(V\leq\phi(U)\right)\right]\right\} \\ &= \frac{c^2(b-a)^2}{n}\left[\frac{\psi}{c(b-a)}\left(1-\frac{\psi}{c(b-a)}\right)\right] = \frac{1}{n}\psi\left[c(b-a)-\psi\right] \end{aligned}$$

Η μέθοδος '$\phi - f'$

$$\hat{\psi} = (b-a)\frac{1}{n}\sum_{i=1}^{n}\phi(X_i), \quad X_i \sim U(a,b)$$

$$\begin{aligned} E(\hat{\psi}) &= \psi \\ V(\hat{\psi}) &= (b-a)^2\frac{1}{n^2}nV\left(\phi(X)\right) \\ &= \frac{(b-a)^2}{n}\left[E(\phi^2)-E^2(\phi)\right], \\ &= \frac{(b-a)^2}{n}\left\{\int_a^b \frac{\phi^2(x)}{b-a}dx - \left[\int_a^b \frac{\phi(x)}{b-a}dx\right]^2\right\} \\ &= \frac{(b-a)^2}{n}\left\{\int_a^b \frac{\phi^2(x)}{b-a}dx - \frac{\psi^2}{(b-a)^2}\right\} \end{aligned}$$

Θα αποδείξουμε τώρα ότι $V(\hat{\psi}) \leq V(\tilde{\psi})$. Γι αυτό, παρατηρώντας ότι $c\phi(x) \geq \phi^2(x)$, έχουμε

$$\begin{aligned} V(\hat{\psi}) &\leq \frac{(b-a)^2}{n}\left\{c\int_a^b \frac{\phi(x)}{b-a}dx - \frac{\psi^2}{(b-a)^2}\right\} = \\ &= \frac{1}{n}\left[c(b-a)\psi - \psi^2\right] = V(\tilde{\psi}) \end{aligned}$$

Συνεπώς, η ‘hit and miss’ μέθοδος είναι χειρότερη από απλοικό Monte-Carlo!

3.2 Δειγματοληψία σπουδαιότητος

Η δειγματοληψία σπουδαιότητος (importance sampling) στηρίζεται στην ίδια ιδέα με την '$\phi - f$' μέθοδο, αλλά, ενώ η επιλογή των ϕ και f που να επιτυγχάνει όσο το δυνατόν μικρότερη διασπορά της εκτιμήτριας μπορεί να είναι πολύ δύσκολη, ή και αδύνατη, εδώ προτείνεται επιλογή μιας συνάρτησης που μιμείται, όσο το δυνατόν περισσότερο, το σχήμα της συνάρτησης προς ολοκλήρωση. Ας δούμε πως επιτυγχάνεται αυτό.

'Εστω $\theta = \int h(x)dx$ το ζητούμενο ολοκλήρωμα, και έστω g μία σ.π.π. 'Εχουμε, ξεκινώντας όπως στην '$\phi - f$' μέθοδο,

$$\theta = \int \phi(x)f(x)dx = \int \phi(x)\frac{f(x)}{g(x)}g(x)dx = \int \psi(x)g(x)dx \, (= E_g [\psi(x)])$$

Τότε, μπορούμε να διαλέξουμε μιας εκ των δύο μεθοδολογιών:

1. Παράγουμε δείγμα $X_1, X_2, \ldots, X_n$ από την f, και εκτιμούμε το θ με την

$$\hat{\theta}_f = \frac{1}{n}\sum_i^n \phi(X_i)$$

2. Παράγουμε δείγμα $X_1, X_2, \ldots, X_n$ από την g, και εκτιμούμε το θ με την

$$\hat{\theta}_g = \frac{1}{n}\sum_i^n \psi(X_i)$$

Η πρώτη είναι η γνωστή '$\phi - f$' μέθοδος, ας δούμε τις ιδιότητες της δεύτερης:

$$\begin{aligned}
E(\hat{\theta}_g) &= \frac{1}{n}\sum E_g\left(\psi(X_i)\right) = \theta \\
V(\hat{\theta}_g) &= E\left(\hat{\theta}_g(X) - \theta\right)^2 \\
&= \frac{1}{n}\int \left[\frac{\phi(x)f(x)}{g(x)} - \theta\right]^2 g(x)dx \\
&= \frac{1}{n}\int \left[\frac{h(x)}{g(x)} - \theta\right]^2 g(x)dx \\
&= \frac{1}{n}V_g\left(\frac{h}{g}\right)
\end{aligned}$$

'Αρα, η διασπορά της $\hat{\theta}_g$ μηδενίζεται αν $h(x)/g(x)$ ισούται με θ, και επειδή αυτό προυποθέτει a priori γνώση του θ, προσπαθούμε να επιτύχουμε ισότητα του $h(x)/g(x)$ με κάποια σταθερά, ή, ισοδύναμα, προσπαθούμε να βρούμε μία g που να μιμείται όσο το δυνατόν περισσότερο το σχήμα της h. Η συνάρτηση $g(x)$ καλείται συχνά συνάρτηση σπουδαιότητας (importance function) και η διαδικασία παραγωγής X_i ως προς g καλείται δειγματοληψία σπουδαιότητος. Ας σημειώσουμε εδώ, ότι η επιλογή των σ.π.π. f ή g στην '$\phi - f'$ ή δειγματοληψία σπουδαιότητος αντίστοιχα, είναι μεν το κλειδί της επιτυχίας των μεθόδων, αλλά επιφυλάσσει κινδύνους· για παράδειγμα, μπορεί να κερδίσουμε σε αποδοτικότητα διαλέγοντας μία g που μιμείται το σχήμα της h, αλλά ταυτόχρονα να χάσουμε επειδή παραγωγή τιμών από την g να είναι πολύ δαπανηρή. Θα ξανατονίσουμε λοιπόν, ότι συνταγές για καλή επιλογή των σ.π.π. δεν υπάρχουν και η όλη διαδικασία χαρακτηρίζεται μάλλον "τέχνη".

Παράδειγμα 13 *Ας υποθέσουμε ότι θέλουμε να υπολογίσουμε το ολοκλήρωμα* (3.3)*:*

$$\theta = \int_2^\infty \frac{1}{\pi}\frac{1}{1+x^2}dx$$

Επιλέγουμε

$$g(x) = \frac{2}{x^2},\ x \in (2,\infty)$$

που μιμείται το σχήμα της ποσότητας πρός ολοκλήρωση στο $(2,\infty)$.
'Εχουμε

$$G(x) = \int_2^x \frac{2}{t^2}dt = 1 - \frac{2}{x}$$

και μπορούμε να παράγουμε τιμές από την g με την μέθοδο της αντιστροφής (προφανώς $G(\infty) = 1 \Rightarrow g$ είναι σ.π.π.). Επίσης, αν $U \sim U(0,1) \Rightarrow 1-U \sim U(0,1)$, συνεπώς $X = 2/U$ είναι τ.μ. με σ.π.π. g.

$$\begin{aligned} \int \psi(x)g(x)dx &= \int_2^\infty = \frac{x^2}{2\pi(1+x^2)}\frac{2}{x^2}dx = E_X\left(\psi(X)\right) = \\ &= E_{1/X}\left(\psi(\frac{1}{X})\right) = E_{U(0,1/2)}\left(\frac{1}{2\pi(1+x^2)}\right), \\ &\quad \text{γιατί } \frac{1}{X} \sim U(0,1/2) \end{aligned}$$

'Αρα καταλήξαμε στην τέταρτη περίπτωση της ενότητας 3.1.

3.3 Αντίθετες τυχαίες μεταβλητές

Ας υποθέσουμε ότι $\hat{\theta}_1(X), \hat{\theta}_2(X)$ είναι αμερόληπτες εκτιμήτριες του θ με διασπορές αντίστοιχα $V(\hat{\theta}_1), V(\hat{\theta}_2)$. Έχουμε

$$E\left(\frac{1}{2}\left(\hat{\theta}_1(X)+\hat{\theta}_2(X)\right)\right)=\theta$$

$$V\left(\frac{1}{2}\left(\hat{\theta}_1+\hat{\theta}_2\right)\right)=\frac{1}{4}V(\hat{\theta}_1)+\frac{1}{4}V(\hat{\theta}_2)+\frac{1}{2}Cov(\hat{\theta}_1,\hat{\theta}_2)$$

Αν υποθέσουμε ότι $V(\hat{\theta}_1)=V(\hat{\theta}_2)$. Τότε

$$\begin{aligned} V\left(\frac{1}{2}\left(\hat{\theta}_1+\hat{\theta}_2\right)\right) &= \frac{1}{2}V(\hat{\theta}_1)+\frac{1}{2}Cov(\hat{\theta}_1,\hat{\theta}_2) \\ &= \frac{1}{2}V(\hat{\theta}_1)\left[1+\frac{Cov(\hat{\theta}_1,\hat{\theta}_2)}{\sqrt{V(\hat{\theta}_1)V(\hat{\theta}_2)}}\right] \\ &= \frac{1}{2}V(\hat{\theta}_1)\left[1+Corr(\hat{\theta}_1,\hat{\theta}_2)\right] \end{aligned}$$

Συνεπώς αν η $Corr(\hat{\theta}_1,\hat{\theta}_2)$ είναι μεγάλη και αρνητική,

$$V\left[\frac{1}{2}(\hat{\theta}_1+\hat{\theta}_2)\right]\ll V(\hat{\theta}_1)$$

Παράδειγμα 14 *Έστω*

$$\theta=\int_0^1\sqrt{1-x^2}dx \qquad \left(=\frac{\pi}{4}\right)$$

και ας δούμε πως θα χρησιμοποιούσαμε διάφορες εκτιμήτριες και την βελτίωση που επιτυγχάνουν οι αντίθετες μεταβλητές:

1. *Η 'hit and miss' μέθοδος παράγει ομοιόμορφα σημεία στο μοναδιαίο τετράγωνο που περικλείει την $\sqrt{1-x^2}$, και εκτιμά το θ από την συχνότητα των σημείων που παράγονται μέσα στην προς ολοκλήρωση συνάρτηση:*

$$n\tilde{\theta}\sim Bin\left(n,\frac{\pi}{4}\right)\Rightarrow V(\tilde{\theta})=\frac{\frac{\pi}{4}\left(1-\frac{\pi}{4}\right)}{n}\approx\frac{0.182}{n}$$

Σημειώνουμε εδώ ότι για την παραγωγή n τιμών για την εκτίμηση του θ χρειάζονται $2n$ τιμές της ομοιόμορφης $U(0,1)$.

2. *Η '$\phi-f$' μέθοδος, με $\phi(x)=\sqrt{1-x^2}, f(x)\sim U(0,1)$, κατασκευάζει την εκτιμήτρια*

$$\hat{\theta}=\frac{1}{n}\sum_i\phi(x_i)$$

και η διασπορά είναι

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{n}\int_0^1 (\phi-\theta)^2 dx = \frac{1}{n}\left[\int_0^1 \phi^2(x)dx - \theta^2\right] \\ &= \frac{1}{n}\left[\left(1-\frac{1}{3}\right)-\left(\frac{\pi}{4}\right)^2\right] = \frac{0.0498}{n} \end{aligned}$$

και αν λάβουμε υπόψιν την προηγούμενη παρατήρηση, τότε, επειδή για την '$\phi - f'$ μέθοδο χρειάζονται n τιμές από την $U(0,1)$, θα γράψουμε ότι

$$V(\hat{\theta}) = \frac{0.0249}{n}$$

3. *Αν $U \sim U(0,1)$ τότε η $1-U$ είναι η αντίθετη τ.μ. της U. Ας χρησιμοποιήσουμε δύο φορές την '$\phi - f'$ μέθοδο:*

$$\phi(u) = \sqrt{1-u^2} = \sqrt{1-(1-u)^2}$$

Τότε

$$\theta^* = \frac{1}{n}\sum_i \frac{1}{2}\left[\sqrt{1-u_i^2} + \sqrt{1-(1-u_i)^2}\right]$$

$$\begin{aligned} nV(\theta^*) &= \frac{1}{2}\left\{V\left(\sqrt{1-U^2}\right) + Cov\left(\sqrt{1-U^2}, \sqrt{1-(1-U)^2}\right)\right\} \\ &= \frac{1}{2}\left(\frac{2}{3}-\frac{\pi^2}{16}\right) + \\ &\quad + \frac{1}{2}\left\{E\left(\sqrt{(U+1)(U)(U-1)(U-2)}\right) - \frac{\pi^2}{16}\right\} \end{aligned}$$

Μπορεί να αποδειχθεί ότι

$$V(\theta^\star) = \frac{0.0052}{n}$$

και συγκρινόμενη με την 'hit and miss' μέθοδο

$$Var(\theta^\star) = \frac{0.0026}{n}$$

Τέλος, για λόγους πληρότητας,δίνουμε το παρακάτω θεώρημα:

Θεώρημα 3 *'Εστω g μία μονότονη συνάρτηση στο $(0,1)$ και $U \sim U(0,1)$, τότε $Corr\left(g(U), g(1-U)\right) < 0$*

Απόδειξη: 'Εστω, χωρίς βλάβη της γενικότητας, ότι η g είναι αύξουσα. 'Εστω ακόμη, ότι $\theta = E\left(g(U)\right) = E\left(g(1-U)\right)$. Θέτουμε

$$inf(u : g(u) > \theta) = 1 - t$$

'Εχουμε

$$\begin{aligned} Cov\left\{g(U), g(1-U)\right\} &= E\left\{(g(U)-\theta)(g(1-U)-\theta)\right\} \\ &= E\left\{g(U)(g(1-U)-\theta)\right\} \\ &= \int_0^1 g(u)\left[g(1-u)-\theta\right] du \\ &= \int_0^t g(u)\left[g(1-u)-\theta\right] du + \\ &\quad + \int_t^1 g(u)\left[g(1-u)-\theta\right] du \\ &< g(t)\int_0^1 \left[g(1-u)-\theta\right] du \\ &< 0 \end{aligned}$$

3.4 Τυχαίες μεταβλητές ελέγχου

Θα εξετάσουμε εδώ μία άλλη μέθοδο εκτίμησης της τιμής ενός ολοκληρώματος που μπορεί να επιτύχει ελάττωση της διασποράς της εκτιμήτριας. 'Οπως θα φανεί, υπάρχει σαφής σχέση της μεθόδου με την παλινδρόμηση. Το πρόβλημα είναι πάλι η εκτίμηση του θ, και έστω

$$\theta = E(Z) = E\left[\phi(X)\right]$$

Ας υποθέσουμε ότι υπάρχει $W = \psi(X)$ με $E(W)$ γνωστή και W συσχετισμένη με την Z. Τότε, με δεδομένο δείγμα $X_1, \ldots, X_n$, έχουμε

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left[Z_i - (W_i - E(W_i))\right] = \frac{1}{n}\sum_{i=1}^{n}\left[\phi(x_i) - \left[\psi(x_i) - E\left(\psi(x_i)\right)\right]\right]$$

Τότε

$$\begin{aligned} E(\hat{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} E\left(\phi(X)\right) - \left[E\left(\psi(x_i)\right) - E\left(\psi(x_i)\right)\right] \\ &= \frac{1}{n} \times nE\left(\phi(X)\right) = \theta \end{aligned}$$

Αλλά $V(\hat{\theta})$ περιλαμβάνει την $-Cov(Z,W)$, άρα, εάν η $Corr(Z,W)$ είναι μεγάλη τότε επιτυγχάνεται μικρή $V(\hat{\theta})$.

Προεκτείνοντας τα παραπάνω, έστω ότι

$$W_j = \psi_j(X),\ j = 1, 2, \ldots, p, \qquad \text{με } E(W_j) \text{ γνωστή}$$

Τότε, για κάποια $\beta_1, \beta_2, \ldots, \beta_p$

$$\hat{\theta} = \frac{1}{n}\left\{\sum_{i=1}^{n}\phi(x_i) - \beta_1\left(W_1 - E(W_1)\right) - \cdots - \beta_p\left(W_p - E(W_p)\right)\right\}$$

και έχουμε

$$\begin{aligned} E(\hat{\theta}) &= \theta \\ V(\hat{\theta}) &= \frac{1}{n^2} n \left\{ V\left[\phi(X_i) - \beta_1 \left(W_1 - E(W_1) \right) - \cdots - \beta_p \left(W_p - E(W_p) \right) \right] \right\} \\ &= \frac{1}{n} \Big\{ E\left\{ \left[\phi(X_i) - \beta_1 \left(W_1 - E(W_1) \right) - \cdots - \beta_p \left(W_p - E(W_p) \right) \right] \right\}^2 \\ & \quad - E^2 \left\{ \left[\phi(X_i) - \beta_1 \left(W_1 - E(W_1) \right) - \cdots - \beta_p \left(W_p - E(W_p) \right) \right] \right\} \Big\} \end{aligned}$$

Άρα η ελαχιστοποίηση της $V(\hat{\theta})$ ως προς τα $\beta_i,\ i = 1, \ldots, p$ για δεδομένα W επιτυγχάνεται με την ελαχιστοποίηση της

$$E\left\{ \left[Z - \beta_1 \left(W_1 - E(W_1) \right) - \cdots - \beta_p \left(W_p - E(W_p) \right) \right] \right\}^2 \tag{3.4}$$

που είναι η αναμενόμενη τετραγωνική απόσταση της Z από τον γραμμικό συνδιασμό των W, και είναι προφανής η σύνδεση με την πολλαπλή γραμμική παλινδρόμιση.

Στην πράξη, ξεκινάμε επιλέγοντας μερικές W, και από κάποια pilot study παράγουμε δείγμα $z_j, w_{1j}, \ldots, w_{pj}$, όπου $j = 1, \ldots, N$, και επιλέγουμε $\beta_i,\ i = 1, \ldots, p$ που ελαχιστοποιούν το άθροισμα $\sum_{j=1}^{N} (z_j - \sum_{i=1}^{p} \beta_i w_{ij})^2$.

Παράδειγμα 15 *Επανερχόμαστε στο γνώριμο ολοκλήρωμα (3.3):*

$$\theta = P(X > 2) = \int_2^{\infty} \frac{1}{\pi} \frac{1}{1 + x^2} dx$$

όπου $X \sim$ Cauchy. Σύμφωνα με την τέταρτη μέθοδο της 3.1, το ολοκλήρωμα (3.3) καταλήγει στο

$$\int_0^{1/2} \frac{1}{\pi(1 + y^2)} dy$$

με

$$f(x) = 2,\ x \in (0, \frac{1}{2}), \qquad \phi(x) = \frac{1}{2\pi(1 + x^2)}.$$

Θεωρούμε τις εκτιμήτριες, για $X \sim U(0, 1/2)$,

$$a\left(X^2 - E(X^2) \right) + b\left(X^4 - E(X^4) \right)$$

ή

$$a\left(X^2 - 1/12) \right) + b\left(X^4 - 1/80 \right)$$

σαν

$$\beta_1 \left(W_1 - E(W_1) \right) - \beta_2 \left(W_2 - E(W_2) \right).$$

Στην συνέχεια, παράγουμε $X_1, \ldots, X_N$, υπολογίζουμε τα X_i^2, X_i^4 και $\phi(X_i)$, και ελαχιστοποιούμε το άθροισμα

$$\min_{a,b} \sum_{i=1}^{N} \left[\phi(x_i) - a\left(x_i^2 - \frac{1}{12} \right) - b\left(x_i^4 - \frac{1}{80} \right) \right]^2$$

Χρησιμοποιώντας τα αποτελέσματα κάποιας συγκεκριμένης pilot study παίρνουμε $\hat{\alpha} = 0.312$, $\hat{\beta} = -0.233$, *και συνεπώς η* (3.4) *εκτιμείται από το*

$$\frac{\sum_{i=1}^{N} \left[\phi(x_i) - \hat{a}\left(x_i^2 - \frac{1}{12}\right) - \hat{b}\left(x_i^4 - \frac{1}{80}\right)\right]^2}{N} \approx 1.1 \times 10^{-9}$$

Συνεπώς $V(\hat{\theta}) \approx \frac{1.1\times 10^{-9}}{n}$ *που είναι εκπληκτικά μικρή, ένα δείγμα μεγέθους* $n = 1$ *είναι αρκετό!*

3.5 4 τρόποι εκτίμησης του ολοκληρώματος (3.3)

Θα κλείσουμε το κεφάλαιο επανερχόμενοι στους τέσσερεις τρόπους της ενότητας 3.1, και εξετάζοντας τις διασπορές κάθε μιάς των εκτιμητριών. Υπενθυμίζουμε ότι $\theta = P(X > 2)$, $X \sim Cauchy$ και είναι απλό να δούμε ότι $\theta = 1 - F(2) = 1/2 - \pi^{-1}\varepsilon\varphi 2 = 0.1476$

1. Παράγουμε n τιμές από την Cauchy, και έστω $\hat{\theta}$ το ποσοστό των τιμών που που είναι μεγαλύτερες του 2. Τότε

$$n\hat{\theta} \sim Bin(n, \theta) \Rightarrow Var(\hat{\theta}) = \frac{\theta(1-\theta)}{n} \approx \frac{0.126}{n}$$

2.

$$2n\hat{\theta} \sim Bin(n, 2\theta) \Rightarrow Var(\hat{\theta}) = \frac{2\theta(1-2\theta)}{4n} \simeq \frac{0.052}{n}$$

δηλαδή έχουμε μείωση κατά 2.4 φορές.

3.

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{n}\int [\phi(x) - \theta]^2 f(x)dx \\ &= \frac{1}{2n}\int_0^2 \left(\frac{2}{\pi(1+X^2)} - 0.1476\right)^2 dx = \frac{0.028}{n} \end{aligned}$$

και επιτυγχάνουμε περαιτέρω μείωση κατά 2 φορές.

4.

$$V(\hat{\theta}) = \frac{2}{n}\int_0^{1/2} \left(\frac{1}{2\pi(1+X^2)} - 0.1476\right)^2 dx = \frac{0.00093}{n}$$

και με την τελευταία ”έξυπνη” δειγματοληψία σπουδαιότητας επιτυγχάνουμε περαιτέρω μείωση κατα 300 φορές.

Κεφάλαιο 4

Παραγωγή εξαρτημένων τυχαίων μεταβλητών

4.1 Διατεταγμένο δείγμα

Σε πολλές εφαρμογές χρειάζεται, αντί να δουλεύουμε με *ανεξάρτητο* δείγμα $X_1, X_2, \ldots, X_n$ να χρησιμοποιούμε αντ' αυτού το *εξαρτημένο* διατεταγμένο δείγμα $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Εκ πρώτης όψεως, δεν υπάρχει πρόβλημα: απλά παράγουμε $X_1, X_2, \ldots, X_n$ και το διατάσσουμε χρησιμοποιώντας έναν από τους γνωστούς αλγόριθμους διάταξης. Αλλά, ειδικά για μεγάλα n, αυτοί οι αλγόριθμοι είναι υπολογιστικά πολύ δαπανηροί.

Μια ιδέα είναι να χρησιμοποιήσουμε $X_{(i)} = F^{-1}(U_{(i)})$ όπου $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ είναι διατεταγμένο δείγμα από την $U(0,1)$. Το πρόβλημα τώρα μετατίθεται στην παραγωγή διατεταγμένων τιμών από την ομοιόμορφη κατανομή $U(0,1)$, και θα εξετάσουμε παρακάτω τεχνικές παραγωγής $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$.

4.1.1 Μέθοδος της ακολουθίας

- Παράγουμε ανεξάρτητο δείγμα $U_1, U_2, \ldots, U_n$ από την $U(0,1)$.
- Ορίζουμε $U_{(n)} = U_n^{1/n}$

  $\vdots$

  $U_{(k)} = U_{(k+1)} \times U_k^{1/k}, \quad k = n-1, n-2, \ldots, 1$

  $\vdots$
- *Ισχυρισμός:* $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ *είναι* διατεταγμένο δείγμα από την $U(0,1)$.

Απόδειξη: Ας θεωρήσουμε το max {n ανεξάρτητων τιμών από την $U(0,1)$}. Τότε

$$\begin{aligned} \{\text{α.σ.κ. του μέγιστου στην τιμή } y &= P(max \leq y)\} \\ &= P(\text{όλα } \leq y) \end{aligned}$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} P(\text{ κάθε ένα } \leq y) \\
&= y^n
\end{aligned}
$$

'Αρα μπορούμε να παράγουμε $U_{(n)}$ με την μέθοδο της αντιστροφής: θέτουμε $U_{(n)}$ την n-οστή ρίζα μιάς τιμής από την $U(0,1)$. 'Ετσι έχουμε αποδείξει, ότι η $U_{(n)} = U_n^{1/n}$ είναι μία τ.μ. με κατανομή $\max\{U_1, \ldots, U_n\}$. Στην συνέχεια, παρατηρούμε ότι το πρόβλημα μετατίθεται στην εύρεση διατεταγμένου δείγματος μεγέθους n-1 ομοιόμορφα κατανεμημένου στο $(0, U_{(n)})$. Συνεπώς, το αποτέλεσμα έπεται με επαγωγή.

4.1.2 Εκθετικά διαστήματα

'Εστω $X_1, X_2, \ldots, X_{n+1}$ ανεξάρτητο δείγμα από την εκθετική κατανομή με σ.π.π. e^{-x}, $x > 0$. Ορίζουμε

$$S_k = X_1 + X_2 + \cdots + X_k \qquad k = 1, 2, \ldots, n+1$$

Μπορεί να αποδειχθεί, ότι αν θέσουμε

$$U_{(k)} = \frac{S_k}{S_{n+1}}, \qquad k = 1, 2, \ldots, n$$

Τότε $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ είναι διατεταγμένο δείγμα από την $U(0,1)$. Για μία διαισθητική απόδειξη, σκεφθείτε ότι σε μία ανέλιξη Poisson οι χρόνοι μεταξύ γεγονότων είναι ανεξάρτητες τ.μ. από την εκθετική σ.π.π., οι τιμές S_k δίνουν τούς αθροιστικούς χρόνους κάθε γεγονότος από τον χρόνο 0, και διαιρώντας με S_{n+1} παίρνουμε τιμές από το διάστημα $(0,1)$.

4.2 Πολυμεταβλητή κανονική κατανομή

Έστω ότι

$$\boldsymbol{X} = (X_1, \ldots, X_n) \sim MVN(\boldsymbol{\mu}, V)$$

όπου συμβολίζουμε με MVN την πολυδιάστατη κανονική κατανομή με σ.π.π.

$$f_X(\boldsymbol{X}) = (2\pi)^{-n/2}|V|^{-1/2}exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T V^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}$$

και

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T,\ V_{n\times n} = (V_{ij})$$

$$E(X_i) = \mu_i,\ V_{ii} = V(X_i),\ V_{ij} = Cov(X_i, X_j)$$

Προσομοίωση της $\boldsymbol{X}$ επιτυγχάνεται με μετασχηματισμό της σε τυποποιημένη πολυδιάστατη κανονική κατανομή. Θέτουμε

$$\boldsymbol{Y} = L^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$$

όπου L είναι ορθογώνιος πίνακας τέτοιος ώστε $LL^T = V$. Τότε ισχύει

$$\boldsymbol{Y} \sim MVN(\mathbf{o}, \boldsymbol{I}_{n\times n}) \Rightarrow Y_i \sim N(0,1),\ i = 1, 2, \ldots, n$$

Συνεπώς παραγωγή του $\boldsymbol{Y}$είναι εύκολη. Αν τώρα θέσουμε $\boldsymbol{X} = \boldsymbol{\mu} + L\boldsymbol{Y}$, παίρνουμε τιμές από την $\boldsymbol{X}$. Η μόνη απαίτηση λοιπόν είναι η εύρεση του πίνακα L που είναι απλή· δες, για παράδειγμα, την μέθοδο Cholesky.

4.3 Ανέλιξη Poisson

Σε μία ανέλιξη Poisson σε χρόνο έντασης λ, ισχύουν τα εξής (σημαντικά) αποτελέσματα:

1. Οι χρόνοι μεταξύ γεγονότων είναι ανεξάρτητες τ.μ. από την εκθετική σ.π.π. $\lambda e^{-\lambda x}$, $x > 0$.

2. Ο αριθμός των γεγονότων σε ένα χρονικό διάστημα μήκους t ακολουθεί την κατανομή Poisson με παράμετρο λt.

Το δεύτερο αποτέλεσμα δηλώνει ότι για να προσομοιώσουμε μία τ.μ. με κατανομή Poisson με παράμετρο λ, αρκεί να προσομοιώσουμε μία διαδικασία Poisson με παράμετρο λ και να μετρήσουμε τον αριθμό των γεγονότων που συμβαίνουν σε χρονικό διάστημα μοναδιαίου μήκους. Σύμφωνα με το πρώτο αποτέλεσμα, η διαδικασία Poisson μπορεί να προσομοιωθεί παράγοντας ανεξάρτητες τιμές από μία εκθετική κατανομή με παράμετρο λ, έστω $E_1, E_2, \ldots, E_k, \ldots$, και ελέγχοντας πότε $E_1 + E_2 + \cdots + E_k + E_{k+1} > 1$. Τότε, k είναι μία ανεξάρτητη τιμή από την κατανομή Poisson με παράμετρο λ.

4.4 Αλυσίδες Markov

Έστω μία αλυσίδα Markov που ορίζεται από τις τυχαίες μεταβλητές

$$X_1, X_2, \ldots, X_n, \ldots$$

που παίρνουν τιμές (καταστάσεις του συστήματος) $S_1, S_2, \ldots, S_k$. Ο πίνακας μεταβάσεως της αλυσίδας ορίζεται ως $P = (p_{ij})$, όπου

$$P(X_n = S_j \mid X_{n-1} = S_i) = p_{ij}, \quad \sum_{j=1}^{k} p_{ij} = 1$$

και ορίζει *πλήρως* την εξέλιξη της ανέλιξης, αν είναι γνωστή η αρχική κατανομή $(p_1^{(0)}, \ldots, p_k^{(0)})$ με $P(X_0 = S_j) = p_j^{(0)}$. Κάθε γραμμή του πίνακα P είναι μία διακριτή κατανομή στο $\{S_1, S_2, \ldots, S_k\}$.

Η προσομοίωση της αλυσίδας Markov είναι εύκολη:

1. Παράγουμε μία τιμή για την X_0 από την $(p_1^{(0)}, \ldots, p_k^{(0)})$.

2. Δεδομένης μίας τιμής $X_n = S_i$, παράγουμε X_{n+1} από την $(p_{i1}, \ldots, p_{ik})$ (δηλ. την i-οστή γραμμή του πίνακα P).

4.4.1 Ιδιότητες της αλυσίδας Markov

Θα ασχοληθούμε με *ομογενείς* αλυσίδες Markov δηλ. αλυσίδες τέτοιες ώστε για τυχόν m, η πιθανότητα μετάβασης από μία κατάσταση σε μία άλλη μετά από χρόνο m είναι η ίδια ανεξαρτήτου του χρόνου εκκίνησης:

$$P(X_{n+m} = S_j \mid X_n = S_i) = P(X_m = S_j \mid X_0 = S_i) \equiv p_{ij}^{(m)}$$

Εξισώσεις Chapman-Kolmogorov

Οι τιμές p_{ij} του πίνακα μεταβάσεως P πληρούν τις λεγόμενες εξισώσεις Chapman-Kolmogorov:

$$p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}$$

Θεώρημα

Έστω

$$\boldsymbol{P}^{(\circ)} = (p_1^{(0)}, \ldots, p_k^{(0)}), \quad p_j^{(0)} = P(X_0 = S_j)$$

και

$$\boldsymbol{P}^{(n)} = (p_1^{(n)}, \ldots, p_k^{(n)}), \quad p_j^{(n)} = P(X_n = S_j)$$

οι απόλυτες πιθανότητες των καταστάσεων στον χρόνο 0 και n αντίστοιχα. Τότε

$$\boldsymbol{P}^{(n)} = \boldsymbol{P}^{(0)} P^n$$

Απόδειξη:

$$p_k^{(n)} = \sum_j p_j^{(n-1)} p_{jk} \Rightarrow \boldsymbol{P}^{(n)} = \boldsymbol{P}^{(n-1)} P$$

και μετά από επαναλήψεις βασισμένες στην παραπάνω σχέση

$$\boldsymbol{P}^{(n)} = \boldsymbol{P}^{(n-2)} PP = \boldsymbol{P}^{(n-2)} P^2 = \cdots = \boldsymbol{P}^{(0)} P^n$$

Πόρισμα: Η $p_{jk}^{(n)}$ δίνεται από το j,k στοιχείο του P^n.

Υπενθυμίζουμε ότι μία αλυσίδα Markov ονομάζεται *ανάγωγη* αν όλες οι καταστάσεις της επικοινωνούν μεταξύ τους, και λέγεται *απεριοδική* εάν δεν υπάρχουν περιοδικές καταστάσεις, δηλαδή καταστάσεις στις οποίες η επάνοδος επιτυγχάνεται μόνο σε χρόνους $N, 2N, \ldots$. Επίσης, μία κατάσταση λέγεται *εργοδική* εάν είναι απεριοδική, *έμμονη*, (δηλαδή το σύστημα επιστρέφει σε αυτή τουλάχιστον μία φορά με πιθανότητα 1), και *θετική* (δηλαδή έχει μέσο χρόνο επανόδου $< \infty$).

Στην συνέχεια, θα εστιάσουμε την προσοχή μας στην *κατανομή ισορροπίας* μιας αλυσίδας Markov

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$$

που ορίζεται σαν το όριο

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix} = \boldsymbol{1}\boldsymbol{\pi}$$

Το παραπάνω όριο υπάρχει εάν και μόνο εάν η αλυσίδα Markov είναι ανάγωγη και έχει μόνο εργοδικές καταστάσεις· είναι επίσης ανεξάρτητο του $\boldsymbol{P}^{(0)}$, γιατί

$$\forall\ \boldsymbol{P}^{(0)},\ \lim_{n\to\infty} \boldsymbol{P}^{(n)} = \lim_{n\to\infty} \left\{ \boldsymbol{P}^{(0)} P^n \right\} = \boldsymbol{P}^{(0)} \boldsymbol{1}\boldsymbol{\pi} = \boldsymbol{\pi}$$

Σημειώσεις: Αν $\boldsymbol{P}^{(0)} = \boldsymbol{\pi}$ τότε $\boldsymbol{P}^{(n)} = \boldsymbol{\pi}\ \forall\ n$. Επίσης, αν υπάρχει η $\boldsymbol{\pi}$ τότε $\boldsymbol{\pi} = \boldsymbol{\pi} P$. Η συνθήκη που εξασφαλίζει ότι $\boldsymbol{\pi} = \boldsymbol{\pi} P$ είναι η

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall\ j \neq i. \tag{4.1}$$

Αυτό συμβαίνει επειδή

$$(\boldsymbol{\pi} P)_j = \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

Έχουμε ήδη δείξει παραπάνω, πως, αν γνωρίζουμε τα $\boldsymbol{P}^{(0)}, P$, μπορούμε να παράγουμε μία αλυσίδα Markov η οποία *μπορεί* να έχει κατανομή ισορροπίας. Ας υποθέσουμε τώρα ότι θέλουμε να παράγουμε τιμές από μία αλυσίδα Markov που έχει μία συγκεκριμένη κατανομή ισορροπίας $\boldsymbol{\pi}$. Η στρατηγική μας θα είναι η ακόλουθη:

1. Διαλέγουμε οποιαδήποτε αλυσίδα Markov με πίνακα μεταβιβάσεως P που να εκπληρώνει την (4.1)

2. Προσομοιώνουμε τιμές από την παραπάνω αλυσίδα Markov για "πολύ χρόνο".

3. Οι τιμές που παράγονται από τα X_n για μεγάλα n, είναι "σαν" τιμές παραγόμενες από την $\boldsymbol{\pi}$.

Ας σταματήσουμε λίγο, και ας δούμε τις επιπτώσεις της παραπάνω στρατηγικής: Έστω ότι $\boldsymbol{\pi}$ είναι μία διακριτή κατανομή στο $\{S_1, \ldots, S_k\}$, και θέλουμε να προσομοιώσουμε τυχαίες μεταβλητές με κατανομή $\boldsymbol{\pi}$. Θα μπορούσαμε ασφαλώς να χρησιμοποιήσουμε την μέθοδο της αντιστροφής, αλλά αν το k είναι πολύ μεγάλο; Χρησιμοποιώντας την παραπάνω στρατηγική, μπορούμε απλά να παράγουμε μία αλυσίδα Markov με κατανομή ισορροπίας $\boldsymbol{\pi}$. Το πρόβλημα απλά μετατίθεται τώρα στην επιλογή του πίνακα P.

4.4.2 Ο αλγόριθμος Metropolis

Επιλέγουμε κάποιον συμμετρικό πίνακα μεταβιβάσεως Q, και έστω q_i η i-οστή γραμμή. Αν βρισκόμαστε στην κατάσταση S_i, επιλέγουμε S_j χρησιμοποιώντας την κατανομή q_i, δηλαδή με πιθανότητα q_{ij}. Στον αλγόριθμο Metropolis μετακινούμαστε στο S_j με πιθανότητα $\min\{1, \pi_j/\pi_i\}$, αλλιώς παραμένουμε στο S_i. Άρα

$$p_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} q_{ij}, \quad i \neq j$$

$$p_{ii} = q_{ii} + \sum_{j \neq i} \max\left\{0, 1 - \frac{\pi_j}{\pi_i}\right\} q_{ij}$$

Τότε $P = (p_{ij})$ είναι ένας πίνακας μεταβιβάσεως, και η (4.1) ισχύει γιατί

$$\pi_i p_{ij} = \min\{\pi_i, \pi_j\} q_{ij} = \min\{\pi_j, \pi_i\} q_{ji} = \pi_j p_{ji}$$

και συνεπώς η $\boldsymbol{\pi}$ είναι η κατανομή ισορροπίας της αλυσίδας Markov με πίνακα μεταβιβάσεως P.

4.4.3 Εφαρμογή στο πρόβλημα του ταξιδιώτη πωλητή

Στο κεφάλαιο 1 είχαμε αναφέρει το παράδειγμα του ταξιδιώτη πωλητή σαν ένα πρόβλημα όπου οι μέθοδοι προσομοίωσης καλούνται να δώσουν λύση. Ας το θυμηθούμε: Υπάρχουν n πόλεις, και ένας πωλητής πρέπει να επισκεφθεί μία φορά μόνο την κάθε μία, με κάποια σειρά. Το κόστος μεταφοράς από την πόλη i στην πόλη j είναι $d(i,j)$. Ζητείται η οικονομικότερη διαδρομή.

Οι πιθανές διαδρομές του πωλητή είναι $n!$ και έστω $c(\boldsymbol{x})$ το κόστος της διαδρομής $\boldsymbol{x}$, όπου $\boldsymbol{x} \in \{x_1, x_2, \ldots, x_{n!}\}$. Θέλουμε να ελαχιστοποιήσουμε, ως προς $\boldsymbol{x}$, το κόστος $c(\boldsymbol{x})$, και ας σκεφθούμε ότι άμεση έρευνα μπορεί να είναι αδύνατη: για παράδειγμα, $n = 100$.

Τέχνασμα: Ορίζουμε

$$p_\lambda(\boldsymbol{x}) = \frac{e^{-\lambda c(\boldsymbol{x})}}{\sum_{\boldsymbol{x}} e^{-\lambda c(\boldsymbol{x})}} = ke^{-\lambda c(\boldsymbol{x})}.$$

Τότε $p_\lambda(\boldsymbol{x})$ είναι μία σ.π.π. στο $\{x_1, x_2, \ldots, x_{n!}\}$.

Σημείωση: Για μεγάλο λ, τα $\boldsymbol{x}$ που αντιστοιχούν σε μεγάλα $c(\boldsymbol{x})$ έχουν πολύ μικρή $p_\lambda(\boldsymbol{x})$. Συνεπώς, μόνο $\boldsymbol{x}$ με πολύ μικρή $c(\boldsymbol{x})$ έχουν μη αμελητέα $p_\lambda(\boldsymbol{x})$. Καθώς $\lambda \to \infty$, μόνο τα $\boldsymbol{x}$ που ελαχιστοποιούν την $c(x)$ λαμβάνουν μή μηδενική πιθανότητα. Στην οριακή τιμή $\lambda = \infty$ παίρνουμε μία κατανομή με "ακίδες" στα $\boldsymbol{x}$ που ελαχιστοποιεί την $c(x)$. ·Αρα τα $\boldsymbol{x}$ που παράγονται από την $p_\lambda(\boldsymbol{x})$ για πολύ μεγάλο λ είναι τα $\boldsymbol{x}$ που ελαχιστοποιούν την $c(x)$.

Το πρόβλημα έχει λοιπόν μεταφερθεί στην προσομοίωση τιμών από την $p_\lambda(\boldsymbol{x})$: αρκεί να χρησιμοποιήσουμε τον αλγόριθμο Metropolis με κατανομή ισορροπίας p_λ.

Γενικώτερα, ας υποθέσουμε ότι θέλουμε να "εξερευνήσουμε" κάποια $f(\boldsymbol{x})$, που μπορεί να έχει πολλές διαστάσεις ώστε να μην είναι δυνατόν να απεικονισθεί, ή να είναι "κρυμμένη" σαν κώδικας υπολογιστή. Για παράδειγμα, σε μία διαδικασία παραγωγής, $f(\boldsymbol{x})$ μπορεί να είναι η έξοδος της διαδικασίας, πχ εσοδεία ή ποιότητα, και το $\boldsymbol{x}$ μπορεί να είναι (ίσως εκατοντάδες) μεταβλητές ελέγχου. Εδώ, σε αντίθεση με το προηγούμενο παράδειγμα όπου το ζητούμενο ήταν η ελαχιστοποίηση κάποιας συνάρτησης $c(\boldsymbol{x})$, ζητείται η μεγιστοποίηση της $f(\boldsymbol{x})$.

4.4.4 Αλγόριθμος του δειγματολήπτη Gibbs

Έστω η από κοινού σ.π.π. $p(X_1, X_2, \ldots, X_N)$ όπου ο αριθμός N είναι πολύ μεγάλος, και οι τ.μ. X_i, $i = 1, \ldots, N$ είναι διακριτές. Μπορούμε να γράψουμε μία σειρά από δεσμευμένες πιθανότητες:

$$\begin{array}{l} p(X_1 \mid X_2, X_3, \ldots, X_N) \\ p(X_2 \mid X_1, X_3, \ldots, X_N) \\ \vdots \\ p(X_N \mid X_1, X_2, \ldots, X_{N-1}) \end{array}$$

Ο αλγόριθμος του δειγματολήπτη Gibbs προχωρεί ως εξής:

1. Διαλέγουμε τυχαίες αρχικές τιμές $X_1^{(0)}, \ldots, X_N^{(0)}$

2. Παράγουμε τιμές

$$\begin{array}{rcl} X_1 = X_1^{(1)} & \text{από} & p(X_1 \mid X_2^{(0)}, \ldots, X_N^{(0)}) \\ X_2 = X_2^{(1)} & \text{από} & p(X_2 \mid X_1^{(1)}, X_3^{(0)}, \ldots, X_N^{(0)}) \\ X_3 = X_3^{(1)} & \text{από} & p(X_3 \mid X_1^{(1)}, X_2^{(1)}, X_4^{(0)}, \ldots, X_N^{(0)}) \\ & \vdots & \\ X_N = X_N^{(1)} & \text{από} & p(X_N \mid X_1^{(1)}, X_2^{(1)}, \ldots, X_{N-1}^{(1)}) \end{array}$$

3. Μετά από ένα κύκλο, περάσαμε από τις $\begin{pmatrix} X_1^{(0)} \\ X_2^{(0)} \\ \vdots \\ X_N^{(0)} \end{pmatrix}$ στις $\begin{pmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_N^{(1)} \end{pmatrix}$

4. Επαναλαμβάνουμε την παραπάνω διαδικασία t φορές, και λαμβάνουμε τις τιμές των X_i μετά από t κύκλους: $\begin{pmatrix} X_1^{(t)} \\ X_2^{(t)} \\ \vdots \\ X_N^{(t)} \end{pmatrix}$

5. Τι συμβαίνει όταν $t \to \infty$; Μπορεί να αποδειχθεί ότι καθώς $t \to \infty$ οι τιμές $\begin{pmatrix} X_1^{(t)} \\ X_2^{(t)} \\ \vdots \\ X_N^{(t)} \end{pmatrix}$ είναι "σαν" τιμές παραγόμενες από την $p(X_1,\ldots,X_N)$.

4.5 Τυχαία πεδία Markov

Θεωρούμε τα pixels μιας εικόνας, ή τα τμήματα οικοπέδου κάποιας έκτασης στην οποία γίνοναι καλλιέργειες, ή οποιαδήποτε άλλο πλέγμα (grid) κάποιας επιφάνειας. Το ενδιαφέρον μας βρίσκεται στο "αποτέλεσμα" του r τετραγώνου του πλέγματος, έστω X_r $, r = 1, 2, \ldots, N$. Για τα παραπάνω παραδείγματα, τα αποτελέσματα μπορεί να είναι το χρώμα των pixels ή η εσοδεία της συγκομιδής του χωραφιού.

Το πρόβλημα είναι η εύρεση κατάλληλου μοντέλου, ή ο προσδιορισμός της $p(X_1, \ldots, X_N)$. Μπορούμε να θεωρήσουμε τις δεσμευμένες πιθανότητες

$$P(X_r \mid X_s,\ s \neq r) = P(X_r \mid X_s), \quad \textit{όπου } s \textit{ ανήκει στην γειτονιά του } r)$$

Μία τέτοια χωρική διαδικασία καλείται τυχαίο πεδίο Markov.

Παράδειγμα 16 *Έστω μία εικόνα μεγέθους $M \times M$ με πιθανές τιμές των pixels άσπρο ή μαύρο. Ας υποθέσουμε ότι*

$$P(X_r = \textit{μαύρο} \mid \textit{υπόλοιπα}) = P(X_r = \textit{μαύρο} \mid \textit{γείτονες}) = \frac{e^{\eta}}{1 + e^{\eta}}$$

όπου $\eta = \beta \times$(αριθμός μαύρων γειτόνων − αριθμός άσπρων γειτόνων), και β είναι μία θετική παράμετρος. Μπορεί να αποδειχθεί ότι

$$p(X_1, \ldots, X_N) = c \times exp\{\beta \sum I_{(\text{ίδιο χρώμα})}\}$$

όπου το άθροισμα λαμβάνεται για όλα τα ζευγάρια που είναι γείτονες.

Πως προσομοιώνουμε τιμές από την $p(X_1, \ldots, X_N)$; Ας θυμηθούμε τον δειγματολήπτη Gibbs: Παράγουμε επαναληπτικά από όλες τις δεσμευμένες πιθανότητες, που στα τυχαία πεδία Markov μετασχηματίζονται σε πολύ απλές μαθηματικές σχέσεις· για παράδειγμα

$$p(X_i \mid \textit{υπόλοιπα}) = p(X_i \mid \textit{γείτονες})$$

Μπορούσαμε ακόμη να χρησιμοποιήσουμε τον αλγόριθμο Metropolis, κατασκευάζοντας μία αλυσίδα Markov με κατανομή ισορροπίας

$$\boldsymbol{\pi} = \{p(X_1, \ldots, X_N) \textit{ για όλες τις πιθανές "εικόνες"}\}$$

Το παραπάνω μοντέλο αναπαριστά την αληθινή εικόνα. Στην πραγματικότητα, η επεξεργασία εικόνας γίνεται ως εξής:

$$\begin{array}{ccc} \textit{Παρατηρούμενη εικόνα} = & \textit{Αληθινή εικόνα} & + \textit{ θόρυβος} \\ \downarrow & \downarrow & \\ \textit{Δεδομένα } (\Delta) & \textit{Άγνωστες τιμές των pixels } (A) & \end{array}$$

Στην Μπεϋζιανή Στατιστική, η ανωτέρω σχέση μπορεί να αναπαρασταθεί σύμφωνα με το Θεώρημα του Bayes ως εξής:

$$\begin{array}{ccc} P(A \mid \Delta) & \propto P(\Delta \mid A) & P(A) \\ \downarrow & \downarrow & \downarrow \\ \textit{Η πιο πιθανή εικόνα δίνεται} & \textit{διαδικασία} & \textit{μοντέλο της εικόνας} \\ \textit{από max a posteriori} & \textit{θορύβου} & \textit{(πχ τυχαίο πεδίο Markov)} \end{array}$$

Για παράδειγμα, έστω ότι σε κάποια θέση i, j *παρατηρούμε* y_{ij} *ενώ η αληθινή τιμή είναι* μ_{ij}. *Τότε*

$$P(\Delta|A) \propto exp\left\{-\frac{1}{2\sigma^2}\sum_i\sum_j(y_{ij}-\mu_{ij})^2\right\}$$

$$P(A) \propto exp\left\{\beta \sum I_{(\text{ιδιοχρώμα})}\right\}$$

όπου το άθροισμα λαμβάνεται για όλα τα ζευγάρια που είναι γείτονες, και

$$P(A|\Delta) \propto P(\Delta|A)P(A) = exp\{\text{περίπλοκη συνάρτηση των } \mu_{ij}\}$$

Για να "εξερευνήσουμε" την $P(A|\Delta)$, *παράγουμε τιμές χρησιμοποιώντας τεχνικές Gibbs ή Metropolis.*

Τέλος, αν μόνο ενδιαφερόμαστε για το $\max P(A|\Delta)$ *μπορούμε να αντικαταστήσουμε την σ.π.π.* $e^{-f(x)}$ *με σ.π.π.* $\propto e^{-\lambda f(x)}$ *για κάποιο μεγάλο* λ *(simulated annealing).*

4.6 Γιατι συγκλίνει ο δειγματολήπτης Gibbs;

Δεν είναι προφανές γιατί ο δειγματολήπτης Gibbs, μετά από πολλές επαναλήψεις, προσομοιώνει τιμές που μπορεί να θεωρηθεί ότι προέρχονται από την από κοινού σ.π.π., ή, με άλλα λόγια, ότι ο δειγματολήπτης Gibbs συγκλίνει. Αυτό οφείλεται στην στην Μαρκοβιανή φύση των επαναλήψεων, που θα αναπτύξουμε εδώ για μία απλή περίπτωση μιάς 2×2 πολυωνυμικής κατανομής.

'Εστω ότι X και Y είναι η κάθε μία τ.μ. Bernoulli με από κοινού σ.π.π.

$$\begin{bmatrix} f_{X,Y}(0,0) & f_{X,Y}(1,0) \\ f_{X,Y}(0,1) & f_{X,Y}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

όπου $p_i \geq 0,\ p_1 + p_2 + p_3 + p_4 = 1$. Η περιθώρια κατανομή της X είναι

$$f_X = [f_X(0) \quad f_X(1)] = [p_1 + p_3 \quad p_2 + p_4]$$

δηλαδή κατανομή Bernoulli με πιθανότητα επιτυχίας $p_2 + p_4$.

Οι δεσμευμένες πιθανότητες μπορούν να γραφούν, υπό μορφή πινάκων,

$$A_{Y|X} = \begin{bmatrix} f_{Y|X}(0 \mid 0) & f_{Y|X}(1 \mid 0) \\ f_{Y|X}(0 \mid 1) & f_{Y|X}(1 \mid 1) \end{bmatrix} = \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix}$$

και

$$A_{X|Y} = \begin{bmatrix} f_{X|Y}(0 \mid 0) & f_{X|Y}(1 \mid 0) \\ f_{X|Y}(0 \mid 1) & f_{X|Y}(1 \mid 1) \end{bmatrix} = \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix}$$

Ο δειγματολήπτης Gibbs είναι μία μέθοδος παραγωγής ενός δείγματος από την $f_{X,Y}$ με έμεσο τρόπο, παράγοντας δηλαδή από τις δεσμευμένες κατανομές $f_{X|Y}$ και $f_{Y|X}$. Σύμφωνα με τον αλγόριθμο της ενότητας 4.4.4 δίνουμε μία αρχική τιμή Y_0, και παράγοντας εναλλάξ από τις $f_{X|Y}$ και $f_{Y|X}$ παίρνουμε μία *ακολουθία τ.μ.*

$$Y_0, X_0, Y_1, X_1, Y_2, X_2, \ldots, Y_t, X_t, \ldots \tag{4.2}$$

Στο παράδειγμά μας η παραπάνω ακολουθία αποτελείται από 0 και 1. Οι πίνακες $A_{X|Y}$ και $A_{Y|X}$ μπορούν να θεωρηθούν σαν πίνακες μεταβιβάσεως που δίνουν την πιθανότητα μεταβιβάσεως από την κατάσταση x στην y και αντίστροφα, δηλαδή

$$P(X = x \mid Y = y) = \textit{Πιθανότητα μεταβιβάσεως } x \to y$$

Αν μόνο ενδιαφερόμαστε στην παραγωγή περιθωρίου κατανομής από την X, θεωρούμε τις τιμές της X από την ακολουθία (4.2):

$$X_0, X_1, X_2, \ldots$$

και μπορούμε να θεωρήσουμε αυτή την ακολουθία σαν αλυσίδα Markov, γράφοντας, για το πρώτο βήμα,

$$P(X_1 = x_1 \mid X_0 = x_0) = \sum_y P(X_1 = x_1 \mid Y_1 = y) P(Y_1 = y \mid X_0 = x_0)$$

ή, σε μορφή πίνακα, ο πίνακας μεταβιβάσεως δίνεται από

$$A_{X|X} = A_{Y|X} A_{X|Y}$$

και μπορούμε να υπολογίσουμε την κατανομή οποιουδήποτε X_k στην ακολουθία:

$$P(X_k = x_k \mid X_0 = x_0) = A_{X|X}^k$$

Επί πλέον, αν συμβολίσουμε την περιθώρια κατανομή του X_k με

$$f_k = [f_k(0) \quad f_k(1)]$$

τότε, για οποιοδήποτε k,

$$f_k = f_0 A_{X|X}^k = \left(f_0 A_{X|X}^{k-1}\right) A_{X|X} = f_{k-1} A_{X|X} \tag{4.3}$$

Είναι γνωστό, (δες, γιά παράδειγμα Feller, W: *An introduction to Probability Theory and its Applications*, Volume 1. New York, John Wiley) ότι εάν ο πίνακας $A_{X|X}$ έχει όλα τα στοιχεία του θετικά, τότε η παραπάνω σχέση συνεπάγεται ότι για οποιαδήποτε αρχική κατανομή f_0, και καθώς $k \to \infty$, η f_k συγκλίνει στην *μοναδική* κατανομή f που ικανοποιεί την (4.3):

$$\begin{aligned} f_X A_{X|X} &= [p_1 + p_3 \;\; p_2 + p_4] \times \begin{bmatrix} \frac{p_1}{p_1+p_3} & \frac{p_3}{p_1+p_3} \\ \frac{p_2}{p_2+p_4} & \frac{p_4}{p_2+p_4} \end{bmatrix} \times \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} \\ &= [p_1 + p_2 \;\; p_3 + p_4] \times \begin{bmatrix} \frac{p_1}{p_1+p_2} & \frac{p_2}{p_1+p_2} \\ \frac{p_3}{p_3+p_4} & \frac{p_4}{p_3+p_4} \end{bmatrix} \\ &= [p_1 + p_3 \;\; p_2 + p_4] \\ &= f_X. \end{aligned}$$

Συνεπώς, καθώς $k \to \infty$, η κατανομή της X_k πλησιάζει περισσότερο στην f_X, και αν σταματήσουμε την επαναληπτική διαδικασία σε ένα αρκετά μεγάλο k, μπορούμε να υποθέσουμε ότι η κατανομή του X_k προσεγγίζει την f_X.

Κεφάλαιο 5

Ασκήσεις

5.1 Ασκήσεις 1

1. Να βρείτε την συνάρτηση πυκνότητας πιθανότητας της $Y = X^2$, όπου $X \sim N(0,1)$.

2. Αν $X \sim Exp(\lambda)$, ποιά είναι η σ.π.π. της $Y = X^2$ και ποιά της $Y = \sqrt{X}$;

3. Εάν $X_1, \ldots, X_n$ είναι ανεξάρτητες $N(0,1)$, να βρείτε την κατανομή της $Z = X_1^2 + \cdots + X_n^2$.

4. Εάν X και Y είναι ανεξάρτητες τ.μ., να βρείτε την κατανομή της $X+Y$ όταν:

   (a) $X \sim N(\mu_1, \sigma_1^2),\ Y \sim N(\mu_2, \sigma_2^2$

   (b) X, Y είναι *Poisson* με μέσους λ, μ, αντίστοιχα;

   (c) X και Y είναι εκθετικές με μέσους λ^{-1}, μ^{-1}, αντίστοιχα.

   (d) Τι συμβαίνει στην (iii) εάν $\lambda = \mu$; Προτείνατε και επαληθεύστε γενικεύσεις για τα αποτελέσματα που βρήκατε στην (i) και (ii) για τα αθροίσματα n ανεξαρτήτων τυχαίων μεταβλητών.

5. Να βρείτε την συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $X = -\lambda^{-1}\ln U$, όπου $U \sim U(0,1)$.

6. Ορίζουμε $\Theta = 2\pi U_1,\ R = (-2\ln U_2)^{1/2}$, όπου U_1, U_2 είναι ανεξάρτητες ομοιόμορφες κατανομές στο $(0,1)$. Θεωρώντας, στην σειρά, τούς μετασχηματισμούς

$$(U_1, U_2) \to (R, \Theta) \to (X, Y)$$

   αποδείξατε ότι

$$X = R\sigma\upsilon\nu\Theta,\ Y = R\eta\mu\Theta$$

   είναι ανεξάρτητες τυποποιημένες κανονικές κατανομές.

7. Θυμηθείτε την διαδικασία Poisson και ιδιαίτερα την σχέση μεταξύ Poisson αριθμό γεγονότων που συμβαίνουν σε εκθετικούς μεταξύ-γεγονότων χρόνους. Μπορείτε να εκμεταλλευθείτε την άσκηση 5 για να παράγετε, χρησιμοποιώντας μία ακολουθία ανεξαρτήτων ομοιόμορφων $U(0,1)$ ποσοτήτων, μία ακολουθία ανεξαρτήτων Poisson(λ) τυχαίων ποσότητων;

8. Εάν $U_1, \ldots, U_{12}$ είναι ανεξάρτητες ομοιόμορφες τυχαίες μεταβλητές στο $(0,1)$, να βρείτε ποιά είναι (προσεγγιστικά) η κατανομή της

$$X = \left[\sum_{i=1}^{12} U_i - 6 \right]$$

9. Εάν X έχει μία συνεχή συνάρτηση κατανομής F, αποδείξατε ότι $Y = F(X)$ έχει ομοιόμορφη κατανομή $U(0,1)$.

5.2 Ασκήσεις 2

1. Η μέθοδος του μέσου τετραγώνου για γέννηση τυχαίων αριθμών: Πάρτε ένα τετραψήφιο ακέραιο αριθμό και τετραγωνίστε τον: Αν ο καινούργιος ακέραιος έχει 8 ψηφία πάρτε τα μεσαία 4. Αν έχει λιγότερα από 8 ψηφία, προσθέστε στην αρχή μηδενικά ώστε να κατασκευάσετε 8 ψηφία, και πάρτε τα μεσαία 4. Αν, για παράδειγμα, αρχίσουμε με 2034, παίρνουμε

$$2034 \to 04137156 \to 1371 \to 01879641 \to 8796 \to 77369616 \to \cdots$$

Συνεπώς, η παραγόμενη ακολουθία τυχαίων αριθμών είναι:

$$1, 3, 7, 1, 8, 7, 9, 6, 3, 6, 9, 6, \ldots$$

   (a) Πειραματιστείτε με διάφορες αρχικές τιμές, μην σταματήσετε γρήγορα!
   (b) Τι συμβαίνει αν αρχίσετε με το 6600 ή 5500; (ή παρόμοιο;)
   (c) Εάν το (a) δεν σας έπεισε, δοκιμάστε 47 βήματα αρχίζοντας με το 8653.
   (d) Τι πιστεύετε για την παραπάνω μέθοδο γέννησης τυχαίων αριθμών;

2. Σχεδιάστε τις καρτεσιανές συντεταγμένες συνεχόμενων αριθμών για τις παρακάτω περιπτώσεις:

   (a) $m = 32, a = 45, b = 1$
   (b) $m = 32, a = 45, b = 7$
   (c) $m = 32, a = 129, b = 1$

Σχολιάστε την καταλληλότητα κάθε γεννήτριας.

3. Η γεννήτρια "Fibonacci" για ακέραιους $\{0, 1, \ldots, m-1\}$ δίνεται από

$$x_{n+1} = (x_n + x_{n-1}) \bmod m$$

με δεδομένα τα x_1, x_2.

   (a) Δοκιμάστε μερικές τιμές. Κάτι αξιοσημείωτο;
   (b) Πόσο συχνά συμβαίνει $x_{n-1} < x_{n+1} < x_n$;
   (c) Μπορείτε να βρείτε θεωρητική εξήγηση της παρατήρησής σας στο (b);
   (d) Αν η ακολουθία $x_1, x_2, \ldots$ ήταν ΠΡΑΓΜΑΤΙΚΑ ακολουθία ανεξαρτήτων ομοιόμορφων τυχαίων αριθμών στο $\{0, 1, \ldots, m-1\}$, αποδείξατε ότι

$$P(x_{n-1} < x_{n+1} < x_n) = \frac{1}{6}$$

   (e) Τι πιστεύετε για την γεννήτρια Fibonacci; Μπορείτε να την βελτιώσετε;

5.3 Ασκήσεις 3

1. Αποδείξατε ότι για μία αναγωγική γεννήτρια πλήρους περιόδου m, ο μέσος και η διασπορά των τιμών που παράγονται αν διαιρέσουμε κάθε αριθμό της ακολουθίας με το m, είναι αντίστοιχα
$$\frac{1}{2}\left(1-\frac{1}{m}\right) \quad \textit{και} \quad \frac{(1+1/m)}{12}$$

2. Αποδείξατε ότι αν U_1 και U_2 είναι ανεξάρτητες U(0,1) τυχαίες μεταβλητές, τότε το δεκαδικό μέρος του (U_1+U_2) είναι επίσης $U(0,1)$. Ακόμα αποδείξατε ότι αυτό το αποτέλεσμα επίσης ισχύει αν $U_1 \sim U(0,1)$ αλλά U_2 έχει οποιαδήποτε συνεχή κατανομή. Τέλος, αποδείξατε ότι αν U_1, U_2 και U_3 έχουν σχηματισθεί ανεξάρτητα από μικτές γεννήτριες με αντίστοιχα μήκη κύκλων c_1, c_2 και c_3, τότε μπορούμε να πάρουμε το δεκαδικό μέρος του $(U_1+U_2+U_3)$ σαν τιμή μιας ψευδοτυχαίας $U(0,1)$ τυχαίας μεταβλητής, και η προκύπτουσα ακολουθία από $(0,1)$ τιμές θα έχει μήκους κύκλου $c_1c_2c_3$ αν c_1, c_2 και c_3 είναι σχετικά πρώτοι (δηλαδή πρώτοι μεταξύ τους ανά δύο).

3. Σε μία αναγωγική γεννήτρια, αποδείξατε ότι αν $m = 10^k$ για κάποιο θετικό ακέραιο $k > 1$, τότε για να επιτύχουμε μήκος κύκλου ίσο με m, χρειαζόμαστε να θέσουμε $a = 20d + 1$, όπου d είναι θετικός ακέραιος.

4. Μία χαρακτηριστικά μη τυχαία πτυχή των αναγωγικών ψευδο-τυχαίων αριθμών είναι ότι κανένας αριθμός δεν εμφανίζεται δύο φορές σε ένα κύκλο. Προτείνατε μία απλή διαδικασία να ξεπεραστεί αυτό το μειονέκτημα.

5. Η αναγωγική γεννήτρια
$$x_{n+1} = 9x_n + 13 \bmod 32$$
έχει πλήρους μήκους περίοδο. Γράψτε την προκύπτουσα ακολουθία και εξετάστε πιθανούς συσχετισμούς. Για παράδειγμα, συγκρίνατε τους αριθμούς στο πρώτο μισό με αυτούς στο δεύτερο μισό, γράψτε τους αριθμούς σε δυαδική μορφή κλπ.

5.4 Ασκήσεις 4

1. Ποιό είναι το μήκος της περιόδου της αναγωγικής γεννήτριας με

$$x_0 = 5772156648$$

$$a = 3141592621$$

$$b = 2718281829$$

$$m = 10000000000$$

2. Είναι οι συνθήκες b=περιττός, $a = 1 \bmod 4$, επαρκείς για να εγγυηθούν πλήρη περίοδο m, όταν ο m είναι δύναμη του 2;

3. Εάν $m = 10^{\lambda}$, $\lambda \geq 2$ και b περιττός και όχι πολλαπλάσιο του 5, αποδείξατε ότι η αναγωγική γεννήτρια θα έχει περίοδο m εάν και μόνο εάν $a = 1 \bmod 20$

4. Αν $m = 2^{35} + 1$ $(= 3 \times 11 \times 43 \times 281 \times 86171)$, να βρείτε όλα τα a που ικανοποιούν τις συνθήκες του θεωρήματος 1. Σχολιάσατε την επιλογή του m.

5. Αν $m = 10^6 - 1$ $(3^3 \times 7 \times 11 \times 13 \times 37)$, να βρείτε όλα τα a που ικανοποιούν τις συνθήκες του θεωρήματος 1.

5.5 Ασκήσεις 5

1. Αποδείξατε ότι εάν $X \sim Exp(1)$, τότε $W = aX^{1/b}$ ακολουθεί την κατανομή Weibull με σ.π.π.
$$f_W(w) = ba^{-b}w^{b-1}exp\left[-(w/a)^b\right],\ w \in \Re^+,\ a, b \in \Re^+$$
Εξηγείστε πως μπορούμε να προσομοιώσουμε την τ.μ. Weibull χρησιμοποιώντας την μέθοδο της αντιστροφής.

2. Προσομοιώστε X-τιμές από πληθυσμό με α.σ.κ.
$$F(x) = \exp\{-\exp(-(x-a)/b)\}.$$
(Κατανομή Gumbel).

3. Προσομοιώστε X-τιμές από πληθυσμό με σ.π.π.
$$f(x) = \begin{cases} 4x & 0 \le 1/2 \\ 4(1-x) & 1/2 \le x \le 1 \end{cases}$$

4. Εάν X έχει σ.π.π.
$$f_X(x) = \lambda e^{-\lambda x}$$
βρείτε την κατανομή της $Y = \sqrt{2X}$. Προτείνατε μία μέθοδο παραγωγής Y-τιμών χρησιμοποιώντας τιμές από την ομοιόμορφη κατανομή.

5. Κατασκευάστε μέθοδο απόρριψης, με πιθανότητα αποδοχής μεγαλύτερης του 0.887 για προσομοίωση της B(2,2) κατανομής.

6. Θεωρείστε μία συνάρτηση-φάκελλο βασισμένη στο
$$g(x) = a^{-1}exp(-a^{-1}x),\ x \in \Re^+,$$
και κατασκευάστε μέθοδο απόρριψης για την προσομοίωση της Gamma(a,1) κατανομής, $a > 1$. Εάν $X \sim Gamma(a, 1)$, να βρείτε την κατανομή της $Y = b^{-1}X$ και έτσι προτείνατε μία μέθοδο για προσομοίωση της Gamma κατανομής με $a > 1$.

5.6 Ασκήσεις 6

1. Μία παραλαγή της μεθόδου της απόρριψης (squeezed rejection algorithm) για την προσομοίωση της Exp(1) κατανομής περικομμένης στο (0,2), μπορεί να γραφεί ως εξής:

   (a) Παράγουμε $Y \sim U(0,2),\ U \sim U(0,1)$

   (b) IF $U \leq e^{-a}(a+1-Y)$ GOTO (e)

   (c) IF $U > e^{-b}/(1-b+Y)$ GOTO (a)

   (d) IF $U > e^{-Y}$ GOTO (a)

   (e) Θέτουμε $X = Y$

   Αποδείξατε ότι η πιθανότητα επιτυχίας στο βήμα (b) είναι ίση με ae^{-a} εάν $a \geq 1$ και ίση με $1/4(a+1)^2e^{-a}$ εάν $a < 1$. 'Ετσι αποδείξατε ότι η καλύτερη επιλογή είναι $a = 1$.

2. Περιγράψτε πως θα προσομοιώνατε τ.μ. από την logistic σ.π.π.

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}, \quad (-\infty < x < \infty)$$

   χρησιμοποιώντας την μέθοδο της απόρριψης βασισμένη στον φάκελλο e^{-x} για $x \geq 0$.

3. 'Εστω ότι υπάρχει κάποιος αλγόριθμος για την παραγωγή τυχαίων δειγμάτων από την t_3 κατανομή με σ.π.π.

$$h(x) = \left\{3^{1/2}\frac{\pi}{2}(1+\frac{x^2}{3})^2\right\}^{-1}, \quad (-\infty < x < \infty).$$

   Δεδομένων τιμών από τον αλγόριθμο, δείξτε πως δείγματα από την τυποποιημένη κανονική κατανομή μπορούν να παραχθούν χρησιμοποιώντας την μέθοδο της απόρριψης και πιθανότητα αποδοχής $(9\sqrt{e})/(8\sqrt{1.5\pi})$.

4. Παράγετε 10 συνεχόμενες τιμές από την δυωνυμική (4,1/4) κατανομή χρησιμοποιώντας

   (a) άθροισμα ομοιόμορφων τ.μ.

   (b) μέθοδο της αντιστροφής

5. Παράγετε 5 συνεχόμενες τιμές από την Poisson (2) κατανομή χρησιμοποιώντας

   (a) γινόμενο ομοιόμορφων τ.μ.

   (b) μέθοδο της αντιστροφής

5.7 Ασκήσεις 7

1. Υπολογίστε το ολοκλήρωμα
$$I = \int_0^1 x^3 dx$$
χρησιμοποιώντας την μέθοδο hit-and-miss και την μέθοδο των αντιθέτων τ.μ. Ποιά η μείωση της διασποράς για δείγμα μεγέθους n;

2. Για την '$\phi - f$' μέθοδο εκτίμησης του
$$\int_0^1 \sqrt{1-x^2}dx,$$
με f ομοιόμορφη σ.π.π. στο $(0,1)$, πόσο μεγάλο πρέπει να είναι το n ώστε το 95% διάστημα εμπιστοσύνης για την τιμή του ολοκληρώματος να έχει πλάτος 0.01;

3. Εκτιμήστε το ολοκλήρωμα
$$\int_0^1 [x(x^2-1)(x-2)]^{1/2} dx$$
χρησιμοποιώντας την '$\phi - f$' μέθοδο.

4. Πειραματιστείτε με τεχνικές ελάττωσης διασποράς για την εκτίμηση του $P(X > 2.5)$, όπου $X \sim N(0,1)$.