ΚΕΦΑΛΑΙΟ 11

ΤΥΧΑΙΟΙ ΑΡΙΘΜΟΙ - ΠΡΟΣΟΜΟΙΩΣΗ

ΤΥΧΑΙΟΙ ΑΡΙΘΜΟΙ

Θα εισαγάγουμε την έννοια του *τυχαίου αριθμού* με ένα παράδειγμα.

Παράδειγμα: Θεωρούμε μια τυχαία μεταβλητή με συνάρτηση πιθανότητας η οποία σε κάθε ένα από τους αριθμούς 0.0000, 0.0001, 0.0002, …, 0.9999 αντιστοιχίζει την πιθανότητα 1/10000 (η κατανομή αυτή λέγεται *διακριτή ομοιόμορφη*). Μπορούμε να κατασκευάσουμε ένα πίνακα με τέτοιους τετραψήφιους αριθμούς επαναλαμβάνοντας το τυχαίο πείραμα που εκφράζεται μέσω της παραπάνω τυχαίας μεταβλητής ένα μεγάλο αριθμό φορών, ανεξαρτήτων μεταξύ τους. Η συλλογή των αριθμών που παίρνουμε με τον τρόπο αυτό (ή ισοδύναμα οι παρατηρηθείσες τιμές της τυχαίας μεταβλητής) μπορούν να αποτελέσουν ένα πίνακα τετραψήφιων τυχαίων αριθμών (βλέπε τον σχετικό πίνακα του παραρτήματος).

Γενικά, ας θεωρήσουμε το σύνολο των αριθμών που αντιπροσωπεύουν τις παρατηρηθείσες τιμές μιας τυχαίας μεταβλητής για ένα μεγάλο αριθμό ανεξάρτητων δοκιμών. Έστω ότι η τυχαία μεταβλητή έχει μια κατανομή πιθανότητας που αντιστοιχίζει την πιθανότητα 10^{-k} σε κάθε ένα από τους αριθμούς $1/10^k$, $2/10^k$, …, $(10^k-1)/10^k$. Οι παρατηρήσεις στις οποίες καταλήγουμε μπορούν να θεωρηθούν ως ένας πίνακας k-ψηφίων τυχαίων αριθμών. Διαλέγοντας μια κατάλληλα μεγάλη τιμή για το k, μπορούμε να κάνουμε το σύνολο των τυχαίων αριθμών που αντιστοιχούν σ’ αυτή την τυχαία μεταβλητή να πλησιάσει όσο θέλουμε κοντά σε ένα σύνολο αριθμών που έχουν επιλεγεί τυχαία από το διάστημα (0, 1).

Η κατασκευή ενός συνόλου τυχαίων αριθμών δεν είναι και τόσο εύκολη. Στα πρώτα βήματα της ανάπτυξης της Στατιστικής, χρησιμοποιήθηκαν μερικές πολύπλοκες και ευφυείς μέθοδοι για την κατασκευή πινάκων τυχαίων αριθμών. Στην εποχή μας, με την διάδοση των ηλεκτρονικών υπολογιστών, υπάρχουν διαθέσιμα πακέτα με προγράμματα που παράγουν αριθμούς με τα χαρακτηριστικά των τυχαίων αριθμών. Για τους τυχαίους αυτούς αριθμούς που παράγονται από τον ηλεκτρονικό υπολογιστή χρησιμοποιούμε συχνά τον όρο ψευδο-τυχαίοι αριθμοί. Ο λόγος είναι ότι τα προγράμματα που παράγουν τους αριθμούς αυτούς είναι τέτοια που, αν ο αρχικός αριθμός (ο αριθμός με τον οποίο ξεκινάμε) είναι γνωστός, όλοι οι άλλοι αριθμοί της ακολουθίας μπορούν να καθορισθούν με απλές αριθμητικές πράξεις. Οι αριθμοί αυτοί, παρά τον ντετερμινιστικό τρόπο κατασκευής τους, συμπεριφέρονται σαν να είχαν κατασκευασθεί τυχαία.

Οι τυχαίοι αριθμοί μπορούν να χρησιμοποιηθούν για να προσομοιώσουν πολλά φυσικά φαινόμενα. Χαρακτηριστικό είναι το ακόλουθο απλό παράδειγμα.

Παράδειγμα: Έστω Χ ο αριθμός των Κ που παρατηρούμε όταν στρίψουμε πέντε φορές τέσσερα αμερόληπτα νομίσματα, ανεξάρτητα το ένα από το άλλο και στην τύχη. (Όπως είναι γνωστό, $X \sim b(4, 1/2)$). Για να προσομοιώσουμε το πείραμα, ας θεωρήσουμε πέντε σύνολα από τέσσερις τυχαίους αριθμούς το καθένα. Έστω X_i ο αριθμός των Κ στην i επανάληψη του πειράματος, i=1,2,3,4,5. Εάν αντιστοιχίσουμε στο ενδεχόμενο Κ τους τυχαίους αριθμούς 0.0000, 0.0001, …, 0.4999 και στο ενδεχόμενο Γ τους τυχαίους αριθμούς 0.5000, 0.5001, …, 0.9999 για i=1,2,3,4,5, η παρατηρηθείσα τιμή του X_i είναι ο αριθμός των τυχαίων αριθμών που είναι μικρότεροι ή ίσοι από το 0.4999 στο i σύνολο των τετραψήφιων τυχαίων αριθμών. Για παράδειγμα, ας θεωρήσουμε τα ακόλουθα πέντε σύνολα τυχαίων αριθμών.

0.5734	0.2395	0.7258	0.0832	0.1117
0.4420	0.8102	0.4562	0.5983	0.8509
0.0024	0.9033	0.3228	0.9644	0.7107
0.0127	0.5661	0.5261	0.8162	0.3004

Για αυτές τις ομάδες των τυχαίων αριθμών, θα έχουμε $X_1=3$, $X_2=1$, $X_3=2$, $X_4=1$, $X_5=2$.

Το προηγούμενο παράδειγμα αποτελεί και μια μέθοδο προσομοίωσης παρατηρήσεων από μια διωνυμική κατανομή με παραμέτρους n και p. Υπάρχει όμως μια γενική μέθοδος για την προσομοίωση παρατηρήσεων από μια οποιαδήποτε διακριτή τυχαία μεταβλητή X. Η μέθοδος αυτή έχει ως εξής:

Προσομοίωση Διακριτών Τυχαίων Μεταβλητών

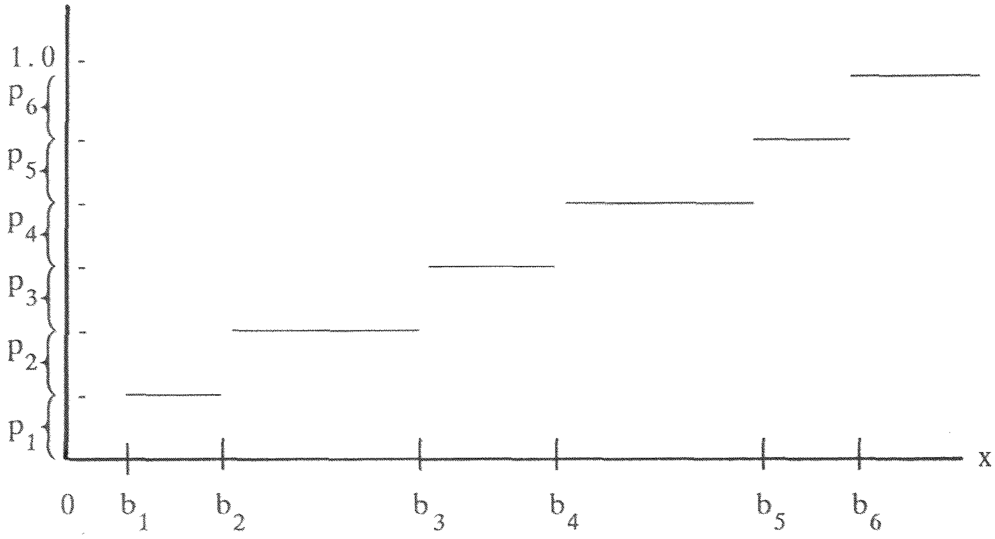
Έστω X μια διακριτή τυχαία μεταβλητή με πεδίο τιμών $R = \{b_1, b_2, b_3, \ldots\}$. Για ευκολία, υποθέτουμε ότι η X παίρνει έξι τιμές και έστω ότι $0 < b_1 < b_2 < \ldots < b_6$. Έστω $p_i = P(X=b_i)$, $i = 1,2,\ldots,6$. Για να προσομοιώσουμε μια παρατήρηση της X, παράγουμε ένα τυχαίο αριθμό $Y = y$ όπου Y ακολουθεί την ομοιόμορφη κατανομή στο [0,1]. Εάν $y \leq F(b_1) = p_1$, τότε θεωρούμε ως τιμή της X την b_1 ($X= b_1$). Αν $F(b_1)<y\leq F(b_2)$ (δηλαδή, ισοδύναμα, αν $p_1< y \leq p_1+ p_2$), τότε θεωρούμε ως τιμή της X την b_2 ($X= b_2$). Συνεχίζουμε με τον ίδιο τρόπο και αν $F(b_5) < y \leq F(b_6)$ (ή ισοδύναμα $p_1+ p_2+\ldots+p_5 < y \leq p_1+ p_2+\ldots+p_6$), θεωρούμε ως τιμή της X την b_6 ($X= b_6$). Έτσι, για παράδειγμα, $P(p_1<Y\leq p_1+ p_2) = F(b_2) - F(b_1) = p_1+ p_2- p_1= p_2$ που είναι και η σωστή πιθανότητα για το ενδεχόμενο $\{X = b_2\}$. Γενικά,

$P(p_1+ p_2+\ldots+p_{i-1} < Y \leq p_1+ p_2+\ldots+p_{i-1}+p_i) = P(X=b_i)$, $i =1,2,\ldots,6$

δηλαδή, τα ενδεχόμενα

$(p_1+ p_2+\ldots+p_{i-1} < Y \leq p_1+ p_2+\ldots+p_{i-1}+p_i)$ και $\{X=b_i\}$, $i = 1,2,\ldots,6$ είναι ισοπίθανα. Επομένως, η διαδικασία αυτή κατασκευάζει τις επιθυμητές πιθανότητες για τα ενδεχόμενα $\{X=b_1\}$, $\{X=b_2\}$, …, $\{X=b_6\}$. Το σχήμα που ακολουθεί βοηθά στην κατανόηση της μεθόδου.

Σχήμα



Σημείωση: Παρότι η μέθοδος αυτή ισχύει για οποιαδήποτε διακριτή τυχαία μεταβλητή, δεν είναι πάντοτε και η καλύτερη που μπορεί να χρησιμοποιήσει κανείς. Για παράδειγμα, για την διωνυμική κατανομή που προσομοιώσαμε στο προηγούμενο παράδειγμα, η γενική μέθοδος είναι ασφαλώς περισσότερο χρονοβόρα από την μέθοδο που χρησιμοποιήσαμε.

Στην συνέχεια, χρησιμοποιούμε το συνεχές ανάλογο της παραπάνω μεθόδου για την προσομοίωση συνεχών τυχαίων μεταβλητών.

Προσομοίωση Συνεχών Τυχαίων Μεταβλητών

Θεώρημα: Έστω X μια συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής F(x). Υποθέτουμε ότι η F(x) είναι γνησίως αύξουσα. Ορίζουμε την τυχαία μεταβλητή Y = F(x). Η τυχαία μεταβλητή Y έχει συνάρτηση κατανομής $F_y(y) = y$, $0<y<1$. (Όπως είναι γνωστό, η κατανομή αυτή ονομάζεται ομοιόμορφη κατανομή στο (0,1)).

Απόδειξη: Η συνάρτηση κατανομής της Y είναι

$$P(Y \le y) = P(F(X) \le y),\ 0<y<1$$

Είναι όμως δεδομένο ότι η F(x) είναι γνησίως αύξουσα. Επομένως, το ενδεχόμενο $\{F(x)\leq y)$ είναι ισοδύναμο με το ενδεχόμενο $\{X\leq F^{-1}(y)\}$. Συνεπώς,

$$\begin{aligned} P(Y\leq y) &= P(X\leq F^{-1}(y)) \\ &= F(F^{-1}(y)) \\ &= y, \qquad 0<y<1 \end{aligned}$$

Δηλαδή, η τυχαία μεταβλητή Y έχει την συνάρτηση κατανομής μιας τυχαίας μεταβλητής που κατανέμεται ομοιόμορφα στο (0,1).

Παράδειγμα: Έστω η τυχαία μεταβλητή X με συνάρτηση πυκνότητας πιθανότητας

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

(η κατανομή αυτή ονομάζεται λογιστική κατανομή).
Τότε,

$$F(x) = \int_{-\infty}^{x} \frac{e^{-t}}{(1+e^{-t})^2} dt = \int_{-\infty}^{x} \frac{d(1+e^{-t})}{(1+e^{-t})^2} = \frac{1}{1+e^{-x}}, \qquad \infty < x < \infty$$

Επομένως, η τυχαία μεταβλητή

$$Y = \frac{1}{1+e^{-X}}$$

έχει συνάρτηση πυκνότητας πιθανότητας

$$g(y) = 1, \qquad 0<y<1$$

Το αντίστροφο του θεωρήματος αυτού, όπως διατυπώνεται στην συνέχεια, μπορεί να χρησιμοποιηθεί για την προσομοίωση τυχαίων μεταβλητών.

Θεώρημα: Έστω ότι η συνάρτηση F(·) έχει τις ιδιότητες της συνάρτησης κατανομής μιας συνεχούς τυχαίας μεταβλητής με F(a)=0 και F(b)=1. Υποθέτουμε επίσης ότι η F(·) είναι γνησίως αύξουσα στο διάστημα (a,b), όπου a και b μπορούν να πάρουν τις τιμές $-\infty$ και $+\infty$ αντίστοιχα. Έστω Y μια άλλη τυχαία μεταβλητή που ακολουθεί την ομοιόμορφη κατανομή στο (0,1). Τότε, η τυχαία μεταβλητή X που οριζεται ως

$$X = F^{-1}(Y)$$

είναι μίας συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής F(x).

Απόδειξη: Η συνάρτηση κατανομής της X είναι

$$F(x) = P(X \leq x) = P(F^{-1}(Y) \leq x)$$

Επειδή όμως η $F(\cdot)$ είναι γνησίως αύξουσα, το ενδεχόμενο $\{F^{-1}(Y) \leq x\}$ είναι ισοδύναμο με το ενδεχόμενο $\{Y \leq F(x)\}$ και επομένως

$$P(X \leq x) = P(Y \leq F(x))$$

Έχουμε όμως υποθέσει ότι η Y είναι ομοιόμορφη στο (0,1).
Δηλαδή,

$$P(Y \leq y) = y \quad \text{για} \quad 0 < y < 1$$

Συνεπώς,

$$P(X \leq x) = F(x), \qquad 0 < F(x) < 1$$

Άρα η F(x) είναι η συνάρτηση κατανομής της X.

Παράδειγμα: Να προσομοιωθεί μια τυχαία μεταβλητή X η οποία ακολουθεί την κατανομή με συνάρτηση πυκνότητας πιθανότητας

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

Λύση: Όπως αποδείχθηκε προηγουμένως, η συνάρτηση κατανομής της X δίνεται από τον τύπο.

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

Είναι προφανές ότι η $F(\cdot)$ πληροί τις υποθέσεις του θεωρήματος. Η τυχαία μεταβλητή Y=F(X) ακολουθεί την ομοιόμορφη κατανομή στο (0,1). Τότε, σύμφωνα με το θεώρημα που μόλις αποδείξαμε, αν πάρουμε ένα τυχαίο αριθμό y από τον πίνακα των τυχαίων αριθμών (βλέπε παράρτημα), η τιμή

$$x = \ln\frac{y}{1-y}$$

παριστάνει μια παρατήρηση της τυχαίας μεταβλητής X.

Παράδειγμα: Να παραχθεί τυχαίο δείγμα πέντε παρατηρήσεων από την κατανομή $F(x) = 1\text{-}e^{-x}$, $0 \leq x < \infty$ (δηλαδή, από την εκθετική κατανομή με παράμετρο θ=1).

Λύση: Είναι προφανές ότι η F(x) είναι απολύτως αύξουσα στο $[0,+\infty)$ με $F(+\infty) =1$, $F(0) = 0$. Σύμφωνα με το θεώρημα, η τυχαία μεταβλητή

$$Y = F(X) =1\text{-}e^{-X}$$

ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (0,1). Τότε, αν

$$y_1, y_2, y_3, y_4, y_5$$

είναι πέντε τυχαίοι αριθμοί, οι τιμές

$$x_i = -\ln(1-y_i)\,, \quad i = 1,2,\ldots,5$$

αποτελούν ένα τυχαίο δείγμα πέντε παρατηρήσεων από την κατανομή F(x). Στην συγκεκριμένη περίπτωση, έστω

0.9189, 0.0693, 0.0570, 0.5956, 0.6017

πέντε τυχαίοι αριθμοί από τον πίνακα τυχαίων αριθμών. Τότε, ένα τυχαίο δείγμα πέντε παρατηρήσεων από την υπό εξέταση κατανομή είναι το

$$x_1 = 2.5127,\ x_2 = 0.0718,\ x_3 = 0.0587,\ x_4 =0.8439,\ x_5=0.9206$$

Σημείωση: Παρότι οι αποδείξεις των δύο προηγουμένων θεωρημάτων προϋποθέτουν ότι η F(x) είναι απολύτως αύξουσα, τα συμπεράσματα των θεωρημάτων αυτών ισχύουν και χωρίς αυτόν τον περιορισμό.