ΕΠΑΝΑΛΗΨΗ ΒΑΣΙΚΩΝ ΕΝΝΟΙΩΝ ΣΤΑΤΙΣΤΙΚΗΣ

Στα πλαίσια του προπτυχιακού μαθήματος Χρονικές σειρές Τμήμα μαθηματικών ΑΠΘ

Διδάσκουσα: Αγγελική Παπάνα

Μονοδιάστατες τυχαίες μεταβλητές

Τυχαία μεταβλητή είναι κάθε πραγματική συνάρτηση, η οποία αντιστοιχεί κάθε στοιχειώδες αποτέλεσμα του πειράματος τύχης (το οποίο είναι στοιχείο του δειγματικού χώρου του πειράματος τύχης) σε έναν πραγματικό αριθμό.

Είναι μια συνάρτηση με πεδίο ορισμού τον δειγματικό χώρο του πειράματος και πεδίο τιμών την πραγματική ευθεία R.

Η **πιθανότητα** η τυχαία μεταβλητή X να παίρνει την τιμή x συμβολίζεται με $f(x) = P(X = x)$.

Η τυχαία μεταβλητή X λέγεται αν το σύνολο τιμών της είναι πεπερασμένο ή άπειρα μετρήσιμο.

Παράδειγμα

Υποθέτουμε ότι ένα φοιτητής γράφει εξετάσεις στο μάθημα Χρονικές Σειρές. Το αποτέλεσμα των εξετάσεων δεν είναι γνωστό εξ' αρχής.

Ο δειγματικός χώρος του πειράματος αποτελείται από δύο ενδεχόμενα, Π={προάγεται} και Α={δεν προάγεται}.

Ορίζουμε την εξής τυχαία μεταβλητή:

$$X = \begin{cases} 1, \text{αν το ενδεχόμενο είναι Π} \\ 0, \text{αν το ενδεχόμενο είναι Α} \end{cases}$$

ΙΔΙΟΤΗΤΕΣ ΤΗΣ ΔΙΑΚΡΙΤΗΣ ΜΕΤΑΒΛΗΤΗΣ

Η συνάρτηση $f(x) = P(X = x)$ λέγεται συνάρτηση πιθανότητας της διακριτής τυχαίας μεταβλητής X και έχει τις ακόλουθες ιδιότητες:

1. $P(X = x) = f(x) \geq 0$
2. $f(x_1) + f(x_2) + \cdots + f(x_n) = 1$, όπου n είναι το πλήθος των τιμών της X

Η κατανομή πιθανότητας ή απλώς κατανομή της διακριτής τυχαίας μεταβλητής X, αποτυπώνεται σε έναν πίνακα που περιέχει τις τιμές της τυχαίας μεταβλητής X και τις αντίστοιχες πιθανότητες με τις οποίες η τυχαία μεταβλητή X παίρνει τις τιμές αυτές.

ΑΘΡΟΙΣΤΙΚΗ ΣΥΝΑΡΤΗΣΗ ΚΑΤΑΝΟΜΗΣ ΤΗΣ ΔΙΑΚΡΙΤΗΣ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ X

Η συνάρτηση κατανομής ή αθροιστική συνάρτηση κατανομής $\mathrm{F}(x)$ της διακριτής τυχαίας μεταβλητής X ορίζεται ως εξής:

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

$$F(x) = \sum_{k=0}^{x} P(X = k)$$

ΜΕΣΗ ΤΙΜΗ ΤΗΣ ΔΙΑΚΡΙΤΗΣ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ Χ

$$E(X) = \sum_x xP(X = x) = \sum_x xf(x) = \mu$$

ΔΙΑΚΥΜΑΝΣΗ ΤΗΣ ΔΙΑΚΡΙΤΗΣ ΤΥΧΑΙΑΣ ΜΕΤΑΒΛΗΤΗΣ Χ

$$\mathrm{V}ar(X) = \sum_x (X - \mu)^2 f(x) = E[(X - \mu)^2] = \sigma^2$$

ΚΟΙΝΗ ΠΙΘΑΝΟΤΗΤΑ ΔΥΟ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Σε πολλά προβλήματα χρειάζεται να ορίσουμε την συνδυασμένη μεταβλητότητα δύο τυχαίων μεταβλητών X και Υ, δηλαδή την **κοινή κατανομή πιθανότητας** τους.

Έστω η διακριτή τυχαία μεταβλητή X με δυνατές διακεκριμένες τιμές $x_1, \dots x_n$ και Y με δυνατές διακεκριμένες τιμές $x_1, \dots x_m$, αντίστοιχα.

Η **κοινή συνάρτηση μάζας πιθανότητας** $f_{XY}(x, y)$ ορίζεται για κάθε ζεύγος δυνατών τιμών (x_i, y_i) ως

$$f_{XY}(x_i, y_i) = P(X = x_i, Y = y_i)$$

και η **κοινή αθροιστική συνάρτηση κατανομής** ορίζεται ως

$$F_{XY}(x_i, y_i) = P(X \leq x_i, Y \leq y_i) = \sum_{x \leq x_i} \sum_{y \leq y_i} f_{XY}(x_i, y_i)$$

ΠΕΡΙΘΩΡΙΑ ΣΥΝΑΡΤΗΣΗ ΠΙΘΑΝΟΤΗΤΑΣ

Έστω δύο τυχαίες μεταβλητές X και Υ, και έστω $f_{XY}(x, y)$ η **κοινή κατανομή πιθανότητας** τους. Οι συναρτήσεις $f_X(x)$, $f_Y(y)$ λέγονται περιθώριες συναρτήσεις κατανομής (marginal distribution functions) των X και Υ, αντίστοιχα.

Οι **περιθώριες κατανομές πιθανότητας** των X και Υ υπολογίζονται ως εξής:

$$f_X(x_i) = P(X = x_i) = \sum_{y_i} f_{XY}(x_i, y_i)$$

$$f_Y(y_i) = P(Y = y_i) = \sum_{x_i} f_{XY}(x_i, y_i)$$

Παράδειγμα

Στρίβουμε ένα αμερόληπτο νόμισμα τρεις φορές.

Ορίζουμε τις τυχαίες μεταβλητές X και Υ ως εξής:

$$X = \begin{cases} 0, \text{αν έχουμε Κ στην πρώτη δοκιμή} \\ 1, \text{αν έχουμε Γ στην πρώτη δοκιμή} \end{cases}$$ και

Υ = ο αριθμός των Κ στις τρεις δοκιμές

Επομένως: $x = 0, 1$ και $y = 0,1,2,3$ και π.χ. για

{X=0} = {ΚΚΚ, ΚΚΓ, ΚΓΚ, ΚΓΓ}, {X=1} = {ΓΚΚ, ΓΚΓ, ΓΓΚ, ΓΚΚ}

Για τις κατανομές πιθανότητας των X και Y έχουμε αντίστοιχα:

x	0	1
$f_X(x)$	½	½

y	0	1	2	3
$f_Y(y)$	1/8	3/8	3/8	1/8

Η από κοινού συνάρτηση πιθανότητας των X και Y είναι:

$X \backslash Y$	0	1	2	3	$f_X(x)$
0	0	1/8	2/8	1/8	½
1	1/8	2/8	1/8	0	½
$f_Y(y)$	1/8	3/8	3/8	1/8	1

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Ορισμός: Δύο τυχαίες μεταβλητές X και Y θα λέγονται ανεξάρτητες (independent) αν για κάθε ζεύγος συνόλων A και B πραγματικών αριθμών

$$P\{X \in A, Y \in B\} = P\{X \in A\}\, P\{Y \in B\}$$

Με άλλα λόγια, οι X και Y είναι ανεξάρτητες αν για όλα τα A και B τα ενδεχόμενα $E_A = \{X \in A\}$ και $E_B = \{Y \in B\}$ είναι ανεξάρτητα.

Ο παραπάνω ορισμός είναι ισοδύναμος με τη σχέση

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

Αν X και Y είναι δύο **ανεξάρτητες** τυχαίες μεταβλητές, τότε:

$$\mathrm{E(XY) = E(X)E(Y)}.$$

ΣΥΝΔΙΑΣΠΟΡΑ

Όταν μελετάμε δύο τυχαίων μεταβλητών X και Υ που δεν είναι ανεξάρτητες, έχει ενδιαφέρον να προσδιορίσουμε πόσο ισχυρά συσχετίζεται η μια με την άλλη.

Για αυτό ορίζουμε τη συνδιασπορά ή συνδιακύμανση (covariance) των τυχαίων μεταβλητών X και Υ:

$$\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{E}(\mathrm{X} - \mu_X)(\Upsilon - \mu_\Upsilon) = \mathrm{E}(\mathrm{XY}) - \mathrm{E}(\mathrm{X})\mathrm{E}(\Upsilon)$$

ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

Το μειονέκτημά της συνδιασποράς είναι ότι η τιμή της εξαρτάται από τις μονάδες μέτρησης των τ.μ. Χ και Υ. Για αυτό όταν θέλουμε να μετρήσουμε τη γραμμική συσχέτιση δύο τ.μ. Χ και Υ, χρησιμοποιούμε συνήθως το **συντελεστή συσχέτισης** (correlation coefficient)

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- $-1 \leq \rho \leq 1$
- Αν οι τ.μ. Χ και Υ είναι ανεξάρτητες είναι $\rho = 0$, αλλά $\rho = 0$ δε δηλώνει ότι οι Χ και Υ είναι ανεξάρτητες αλλά απλά ότι δεν είναι γραμμικά συσχετισμένες.
- $\rho = -1$ ή $\rho = 1$ αν και μόνο αν $Y = \alpha + \beta X$ για κάποιους αριθμούς α και β.

Ιδιότητες μέσης τιμής, διακύμανσης και συνδιασποράς

$\mathrm{E}(\alpha + \beta \mathrm{X}) = \alpha + \beta \mathrm{EX}$

$\mathrm{E}(\mathrm{X} - \mu)^2 = \mathrm{Var}(\mathrm{X})$

$\mathrm{Var}(X) = E(X^2) - (E\mathrm{X})^{\,2}$

$\mathrm{Var}(\alpha + \beta \mathrm{X}) = \beta^2 \mathrm{Var}(\mathrm{X})$

$\mathrm{Var}(\mathrm{X} + \mathrm{Y}) = \mathrm{Var}(\mathrm{X}) + \mathrm{Var}(\mathrm{Y})$ (για ανεξάρτητες μεταβλητές)

$\mathrm{Var}(\mathrm{X} + \mathrm{Y}) = \mathrm{Var}(\mathrm{X}) + \mathrm{Var}(\mathrm{Y}) + 2\mathrm{Cov}(\mathrm{X}, \mathrm{Y})$ (για εξαρτημένες μεταβλητές)

$\mathrm{Cov}(\mathrm{X}, \mathrm{X}) = \mathrm{Var}(\mathrm{X})$

$\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{EX\Upsilon} - \mathrm{EXE\Upsilon}$

$\mathrm{Cov}(\alpha + \beta \mathrm{X}, \gamma + \delta \mathrm{Y}) = \beta\delta \mathrm{Cov}(\mathrm{X}, \Upsilon)$