



ανοικτά μαθήματα
opencourses

ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ Ι

Παντελής Δημήτριος
Τμήμα Μηχανολόγων Μηχανικών



ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

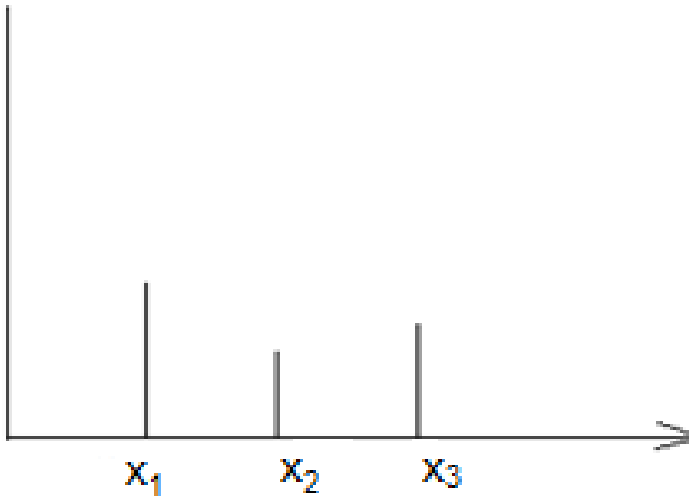
- Σε κάθε αποτέλεσμα του πειράματος αντιστοιχεί μία αριθμητική τιμή
- Μαθηματικός ορισμός: Τυχαία μεταβλητή X είναι συνάρτηση με πεδίο ορισμού το δειγματικό χώρο Ω και πεδίο τιμών το σύνολο των πραγματικών αριθμών

$$X: \Omega \rightarrow \mathbb{R}$$

- Για $\omega \in \Omega$, $X(\omega)$ είναι η τιμή που αντιστοιχεί στο αποτέλεσμα ω του πειράματος
- Παραδείγματα
 - αριθμός K σε 3 ρίψεις νομίσματος (0, 1, 2, 3)
 - άθροισμα δύο ρίψεων ζαριού (2, 3, ..., 12)
 - χρόνος παραγωγής προϊόντων (τιμή σε κάποιο συνεχές διάστημα)
- Συμβολισμός
 X : τυχαία μεταβλητή
 x : αριθμητική τιμή

ΔΙΑΚΡΙΤΕΣ ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

- Πεπερασμένο ή αριθμήσιμο σύνολο τιμών
- Συνάρτηση μάζας πιθανότητας (pmf)
 - πιθανότητα για κάθε τιμή της X



- σχήμα: πιθανότητα ως μάζα συγκεντρωμένη στην αντίστοιχη τιμή
- $p_X(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$
- $p_X(x) \geq 0, \sum_X p_X(x) = 1$

ΔΙΑΚΡΙΤΕΣ ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

- Παράδειγμα: $X =$ αριθμός Κ σε δύο ρίψεις νομίσματος
 $\Omega = \{ΚΚ, ΚΓ, ΓΚ, ΓΓ\}$

$$p_X(0) = P(X=0) = P(\Gamma\Gamma) = \frac{1}{4}$$

$$p_X(2) = P(X=2) = P(KK) = \frac{1}{4}$$

$$p_X(1) = P(X=1) = P(K\Gamma) + P(\Gamma K) = \frac{1}{2}$$

ΠΑΡΑΔΕΙΓΜΑΤΑ ΔΙΑΚΡΙΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- Πείραμα με δύο αποτελέσματα
“Επιτυχία” (E: πιθανότητα p)
“Αποτυχία” (A: πιθανότητα $1-p$)
- Τυχαία μεταβλητή Bernoulli
X: αριθμός επιτυχιών (0 ή 1)
– $p_X(1) = p, p_X(0) = 1 - p$
- Δυωνυμική τυχαία μεταβλητή
X: αριθμός επιτυχιών σε n ανεξάρτητες επαναλήψεις του πειράματος
– $p_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k=0,1,\dots,n$
- Γεωμετρική τυχαία μεταβλητή
X: αριθμός επαναλήψεων του πειράματος έως την πρώτη επιτυχία
– $X = k \Rightarrow k-1$ αποτυχίες και μία επιτυχία
– $p_X(k) = P(X = k) = (1 - p)^{k-1} p, \quad k=1,2,3,\dots$

ΠΑΡΑΔΕΙΓΜΑΤΑ ΔΙΑΚΡΙΤΩΝ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- Τυχαία μεταβλητή Poisson

- $p_X(k) = P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k=0,1,2,\dots$

- Προσεγγίζει ικανοποιητικά τη διωνυμική τυχαία μεταβλητή για μεγάλο n , μικρό p και $np=\lambda$

- Παράδειγμα : 100 συσκευές, πιθανότητα βλάβης 0,01 για κάθε μία

$P(5 \text{ παθαίνουν βλάβη}) = ;$

$X =$ αριθμός μηχανών που παθαίνουν βλάβη

$X:$ διωνυμική με $n=100, p=0,01$

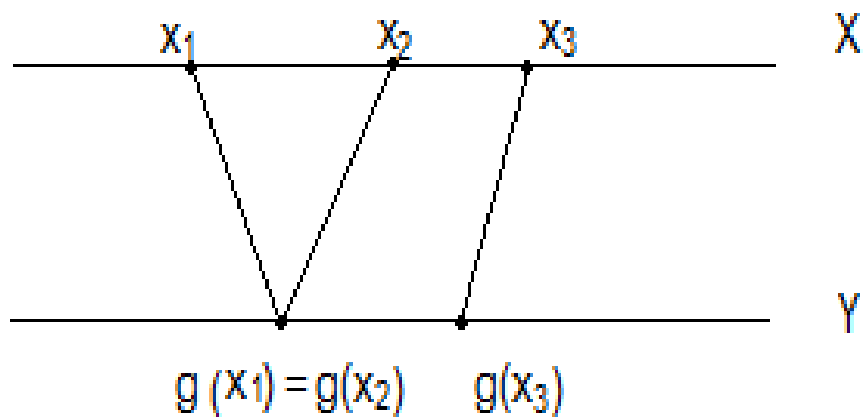
$$P(X = 5) = \binom{100}{5} \cdot 0,01^5 \cdot (1-0,01)^{95} = 0,0029$$

Προσεγγιστικά, $X:$ Poisson με $\lambda=100 \cdot 0,01=1$

$$P(X = 5) = \frac{e^{-1} \cdot 1^5}{5!} = 0,00306$$

ΣΥΝΑΡΤΗΣΕΙΣ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- $Y = g(X)$



- X : τυχαία μεταβλητή με γνωστή pmf $p_X(x)$
- Ζητούμενο: pmf της Y , $p_Y(y)$
- $p_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{x:g(x)=y} p_X(x)$

ΣΥΝΑΡΤΗΣΕΙΣ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- Παράδειγμα

$$p_X(x) = \frac{1}{9}, \quad x = -4, -3, -2, -1, 0, 1, 2, 3, 4$$

$$Y = |X|$$

$$p_Y(0) = p_X(0) = \frac{1}{9}$$

$$\text{Για } k=1,2,3,4, \quad p_Y(k) = p_X(k) + p_X(-k) = \frac{2}{9}$$

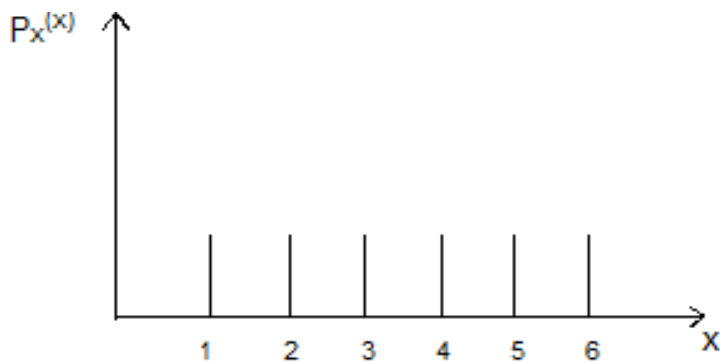
ΜΕΣΗ ΤΙΜΗ

- Ορισμός: $E(X) = \sum_x x \cdot p_X(x)$
- Ερμηνείες
 - Μέσος όρος αποτελεσμάτων σε μεγάλο αριθμό επαναλήψεων του πειράματος
 - Κέντρο βάρους της συνάρτησης μάζας πιθανότητας

- Παράδειγμα

X : αποτέλεσμα ρίψης ζαριού

$$p_X(x) = \frac{1}{6}, \quad x=1,2,3,4,5,6$$



$$E(X) = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6}$$

ΜΕΣΗ ΤΙΜΗ

- Λόγω συμμετρίας το κέντρο βάρους βρίσκεται στο μέσον του διαστήματος $[1,6]$: $\frac{1+6}{2} = 3,5$

ΙΔΙΟΤΗΤΕΣ ΜΕΣΗΣ ΤΙΜΗΣ

- X τυχαία μεταβλητή με γνωστή pmf, $Y = g(X)$
- $E(Y) = \sum_y y \cdot p_Y(y)$
 - Εύρεση $p_Y(y)$ δεν είναι απαραίτητη
 - $E(Y) = E(g(X)) = \sum_x g(x) \cdot p_X(x)$
- Παράδειγμα
 V : ταχύτητα (km/h), $P(V=5) = 0,6$, $P(V=30) = 0,4$
 T : χρόνος για απόσταση 2 km
$$E(T) = E\left(\frac{2}{V}\right) = \frac{2}{5} \cdot 0,6 + \frac{2}{30} \cdot 0,4 = \frac{4}{15}$$
- Γενικά, $E(g(X)) \neq g(E(X))$
 - Εξαίρεση: γραμμικές συναρτήσεις
$$E(\alpha X + \beta) = \alpha E(X) + \beta$$

ΔΙΑΣΠΟΡΑ

- $\text{var}(X) = E\left[\left[X - E(X)\right]^2\right]$
 - Μέτρο δικύμανσης των τιμών της X
- Ιδιότητες
 - $\text{var}(X) \geq 0$
 - $\text{var}(\alpha X + \beta) = \alpha^2 \text{var}(X)$

Η σταθερά β μετατοπίζει τις τιμές της τυχαίας μεταβλητής χωρίς να αλλάζει τη διακύμανσή τους
- Για ευκολότερο υπολογισμό
$$\text{var}(X) = E(X^2) - [E(X)]^2$$
- Τυπική απόκλιση
$$\sigma_X = \sqrt{\text{var}(X)}$$

ΔΙΑΣΠΟΡΑ

- Παράδειγμα

X : αποτέλεσμα ρίψης ζαριού

$$E(X) = \frac{7}{2}$$

$$E(X^2) = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) \cdot \frac{1}{6} = \frac{91}{6}$$

$$\text{var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

ΜΕΣΗ ΤΙΜΗ ΚΑΙ ΔΙΑΣΠΟΡΑ BERNOULLI, POISSON

- Bernoulli

$$p_X(0) = 1 - p, \quad p_X(1) = p$$

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$E(X^2) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

$$\text{var}(X) = p - p^2 = p \cdot (1 - p)$$

- Poisson

$$p_X(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

$$E(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda \cdot (\lambda + 1)$$

$$\text{var}(X) = \lambda \cdot (\lambda + 1) - \lambda^2 = \lambda$$

ΑΠΟ ΚΟΙΝΟΥ ΣΥΝΑΡΤΗΣΗ ΜΑΖΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

- Ταυτόχρονη μελέτη δύο ή περισσότερων τυχαίων μεταβλητών

- Για δύο τυχαίες μεταβλητές X, Y

$$p_{X,Y}(x, y) = P(X = x \text{ και } Y = y)$$

- $\sum_{x,y} p_{X,Y}(x, y) = 1$

- $p_X(x) = P(X = x) = \sum_y P(X = x \text{ και } Y = y) = \sum_y p_{X,Y}(x, y)$

- $p_Y(y) = \sum_x p_{X,Y}(x, y)$

ΑΠΟ ΚΟΙΝΟΥ ΣΥΝΑΡΤΗΣΗ ΜΑΖΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

- Παράδειγμα

Y				
4	0	1/20	1/20	1/20
3	1/20	2/20	3/20	1/20
2	1/20	2/20	3/20	1/20
1	1/20	1/20	1/20	0
	1	2	3	4
				X

$$p_X(2) = p_{X,Y}(2,1) + p_{X,Y}(2,2) + p_{X,Y}(2,3) + p_{X,Y}(2,4) = \frac{6}{20}$$

$$p_Y(4) = p_{X,Y}(1,4) + p_{X,Y}(2,4) + p_{X,Y}(3,4) + p_{X,Y}(4,4) = \frac{3}{20}$$

ΣΥΝΑΡΤΗΣΕΙΣ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- $Z = g(X, Y)$
 - $p_Z(z) = P(g(X, Y) = z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y)$
 - $E[g(X, Y)] = \sum_x \sum_y g(x, y) \cdot p_{X,Y}(x, y)$
 - Γενικά $E[g(X, Y)] \neq g(E(X), E(Y))$
 - Εξαίρεση: γραμμικές συναρτήσεις
 $E(\alpha X + \beta Y + \gamma) = \alpha E(X) + \beta E(Y) + \gamma$

- Παράδειγμα: $Z = X + 2Y$

Y				
4	0	1/20	1/20	1/20
3	1/20	2/20	3/20	1/20
2	1/20	2/20	3/20	1/20
1	1/20	1/20	1/20	0
	1	2	3	4 X

ΣΥΝΑΡΤΗΣΕΙΣ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

(συνέχεια παραδείγματος)

$$E(Z) = E(X) + 2E(Y)$$

$$p_X(1) = \frac{3}{20}, \quad p_X(2) = \frac{6}{20}, \quad p_X(3) = \frac{8}{20}, \quad p_X(4) = \frac{3}{20}$$

$$E(X) = 1 \cdot \frac{3}{20} + 2 \cdot \frac{6}{20} + 3 \cdot \frac{8}{20} + 4 \cdot \frac{3}{20} = 2,55$$

$$p_Y(1) = \frac{3}{20}, \quad p_Y(2) = \frac{7}{20}, \quad p_Y(3) = \frac{7}{20}, \quad p_Y(4) = \frac{3}{20}$$

$$E(Y) = 1 \cdot \frac{3}{20} + 2 \cdot \frac{7}{20} + 3 \cdot \frac{7}{20} + 4 \cdot \frac{3}{20} = 2,5$$

$$E(Z) = 2,55 + 2 \cdot 2,5 = 7,55$$

ΠΟΛΛΕΣ ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

- Γενίκευση για περισσότερες από δύο τυχαίες μεταβλητές

- Για 3 τυχαίες μεταβλητές

$$p_{X,Y,Z}(x,y,z) = P(X=x \text{ και } Y=y \text{ και } Z=z)$$

$$p_{X,Y}(x,y) = \sum_z p_{X,Y,Z}(x,y,z)$$

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x,y,z)$$

$$E(g(X,Y,Z)) = \sum_x \sum_y \sum_z g(x,y,z) \cdot p_{X,Y,Z}(x,y,z)$$

- $E(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n) = \alpha_1 E(X_1) + \alpha_2 E(X_2) + \dots + \alpha_n E(X_n)$

- Μέση τιμή διατηρεί την γραμμικότητα της συνάρτησης

- Μέση τιμή διωνυμικής τυχαίας μεταβλητής

$$E(X) = \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k}$$

Πιο εύκολα, $X = X_1 + X_2 + \dots + X_n$

- $X_1 + X_2 + \dots + X_n$ Bernoulli

$$E(X) = E(X_1) + E(X_2) + \dots + E(X_n) = n \cdot p$$

ΔΕΣΜΕΥΜΕΝΗ ΣΥΝΑΡΤΗΣΗ ΜΑΖΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

- Δέσμευση από γεγονός A

$$p_{X|A}(x) = P(X = x | A) = \frac{P(\{X = x\} \cap A)}{P(A)}$$

- Δέσμευση από άλλη τυχαία μεταβλητή

$$p_{X|Y}(x|y) = P(X = x | Y = y) = \frac{P(X = x \text{ και } Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

- Πολλαπλασιαστικός κανόνας

$$p_{X,Y}(x,y) = p_Y(y) \cdot p_{X|Y}(x,y)$$

- Θεώρημα συνολικής πιθανότητας

$$p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_{X|Y}(x,y) \cdot p_Y(y)$$

ΔΕΣΜΕΥΜΕΝΗ ΣΥΝΑΡΤΗΣΗ ΜΑΖΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

- Δεσμευμένη μέση τιμή

$$E(X|A) = \sum x \cdot p_{X|A}(x)$$

$$E(X|Y=y) = \sum_x x \cdot p_{X|Y}(x|y)$$

- Θεώρημα συνολικής μέσης τιμής

– A_1, A_2, \dots, A_n : διαμέριση δειγματικού χώρου

$$E(X) = \sum_i E(X|A_i) \cdot P(A_i)$$

– $E(X) = \sum_y E(X|Y=y) \cdot p_Y(y)$

ΓΕΩΜΕΤΡΙΚΗ ΤΥΧΑΙΑ ΜΕΤΑΒΛΗΤΗ

- Μέση τιμή

- $E(X) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p$

- Με χρήση δεσμευμένων μέσων τιμών

$$\begin{aligned} E(X) &= E(X | X=1) \cdot P(X=1) + E(X | X > 1) \cdot P(X > 1) \\ &= 1 \cdot p + (1 + E(X)) \cdot (1-p) \Rightarrow E(X) = \frac{1}{p} \end{aligned}$$

- Διασπορά

$$\begin{aligned} E(X^2) &= E(X^2 | X=1) \cdot P(X=1) + E(X^2 | X > 1) \cdot P(X > 1) \\ &= 1 \cdot p + E\left[(1+X)^2\right] \cdot (1-p) \Rightarrow E(X^2) = \frac{2}{p^2} - \frac{1}{p} \end{aligned}$$

$$\text{var}(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

ΑΝΕΞΑΡΤΗΣΙΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

- Από κοινού συνάρτηση μάζας πιθανότητας ισούται με το γινόμενο των επιμέρους συναρτήσεων μάζας πιθανότητας
 - π.χ. X, Y, Z ανεξάρτητες αν
$$p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z) \text{ για κάθε } x,y,z$$

- Παράδειγμα

Y	4	0	1/20	1/20	1/20
	3	1/20	2/20	3/20	1/20
	2	1/20	2/20	3/20	1/20
	1	1/20	1/20	1/20	0
		1	2	3	4
				X	

- Είναι οι X, Y ανεξάρτητες;
 - Όχι: $p_{X,Y}(2,4) = \frac{1}{20} \neq p_X(2) \cdot p_Y(4) = \frac{18}{400}$
 - Εναλλακτικά, $p_{X|Y}(1|4) = 0 \neq p_X(1)$

ΑΝΕΞΑΡΤΗΤΕΣ ΤΥΧΑΙΕΣ ΜΕΤΑΒΛΗΤΕΣ

- Αν X, Y ανεξάρτητες

- $E(XY) = E(X) \cdot E(Y)$

- $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$

- $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

- Παράδειγμα

X, Y ανεξάρτητες, $Z = X - 4Y$

$$\text{var}(Z) = \text{var}(X) + \text{var}(-4Y) = \text{var}(X) + (-4)^2 \text{var}(Y) = \text{var}(X) + 16 \text{var}(Y)$$

- Γενίκευση: Αν X_1, X_2, \dots, X_n ανεξάρτητες

$$\text{var}(X_1 + X_2 + \dots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n)$$

- Διασπορά διωνυμικής τυχαίας μεταβλητής

$$X = X_1 + X_2 + \dots + X_n$$

- X_1, X_2, \dots, X_n ανεξάρτητες Bernoulli

$$\text{var}(X) = \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n) = n \cdot p \cdot (1-p)$$

ΣΥΝΔΙΑΣΠΟΡΑ

- $\text{cov}(X, Y) = E\left[\left[X - E(X)\right] \cdot \left[Y - E(Y)\right]\right]$
 - Θετική: μεγάλες(μικρές) τιμές της X συνήθως αντιστοιχούν σε μεγάλες (μικρές) τιμές της Y
 - Αρνητική: μεγάλες(μικρές) τιμές της X συνήθως αντιστοιχούν σε μικρές(μεγάλες) τιμές της Y
 - Μηδενική: δεν υπάρχει τέτοιου είδους συσχέτιση, X, Y ασυσχέτιστες
- $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$
 - X, Y ανεξάρτητες $\Rightarrow \text{cov}(X, Y) = 0$
(Το αντίθετο δεν ισχύει απαραίτητα)

- Παράδειγμα: $p_{X,Y}(1,0) = p_{X,Y}(0,1) = p_{X,Y}(-1,0) = p_{X,Y}(0,-1) = \frac{1}{4}$

$$\left. \begin{array}{l} XY = 0 \text{ πάντα} \Rightarrow E(XY) = 0 \\ E(X) = E(Y) = 0 \end{array} \right\} \Rightarrow \text{cov}(X, Y) = 0$$

X, Y δεν είναι ανεξάρτητες διότι

$X \neq 0 \Rightarrow Y = 0$ (X δίνει πληροφορία για το Y)

ΣΥΝΔΙΑΣΠΟΡΑ

- Διασπορά αθροίσματος τυχαίων μεταβλητών

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) + \sum_{(i,j): i \neq j} \text{cov}(X_i, X_j)$$

ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ

- Αδιάστατη μορφή της συνδιασποράς

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

- $-1 \leq \rho \leq 1$
 - $|\rho| = 1 \Leftrightarrow Y = \alpha X + \beta$ (γραμμική σχέση)
 - $\rho = 0$: X, Y ασυσχέτιστες
- Παράδειγμα: ρίψη νομίσματος n φορές
 X : αριθμός Κ, Y : αριθμός Γ
 $\rho(X, Y) =$;
 $X + Y = n \Rightarrow Y = n - X$ (γραμμική σχέση)
Άρα $\rho(X, Y) = -1$ (αρνητική κλίση)