Πανεπιστήμιο Αιγαίου

Πιθανότητες και Στατιστική

Ελισάβετ Κωνσταντίνου

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Ζεύγη τυχαίων μεταβλητών

- X: το βάρος ενός ατόμου από κάποιο πληθυσμό
  συνάρτηση κατανομής $F_X(x) = P[X \leq x]$
- Y: το ύψος ενός ατόμου από τον ίδιο πληθυσμό
  σ.κ. $F_Y(y) = P[Y \leq y]$
- Πώς μπορούμε να υπολογίσουμε πιθανότητες όπως:
  - $P[X \leq x$ και $Y \leq y]$
  - $P[X \leq x$ και $Y > y]$
  - $P[X / Y^2 \leq 25]$

Ζεύγη τυχαίων μεταβλητών

- X: ξένες γλώσσες της μητέρας

n	0	1	2	3
P[X=n]	0.472	0.354	0.126	0.048

- E[X] = 0.750, σ_X = 0.853

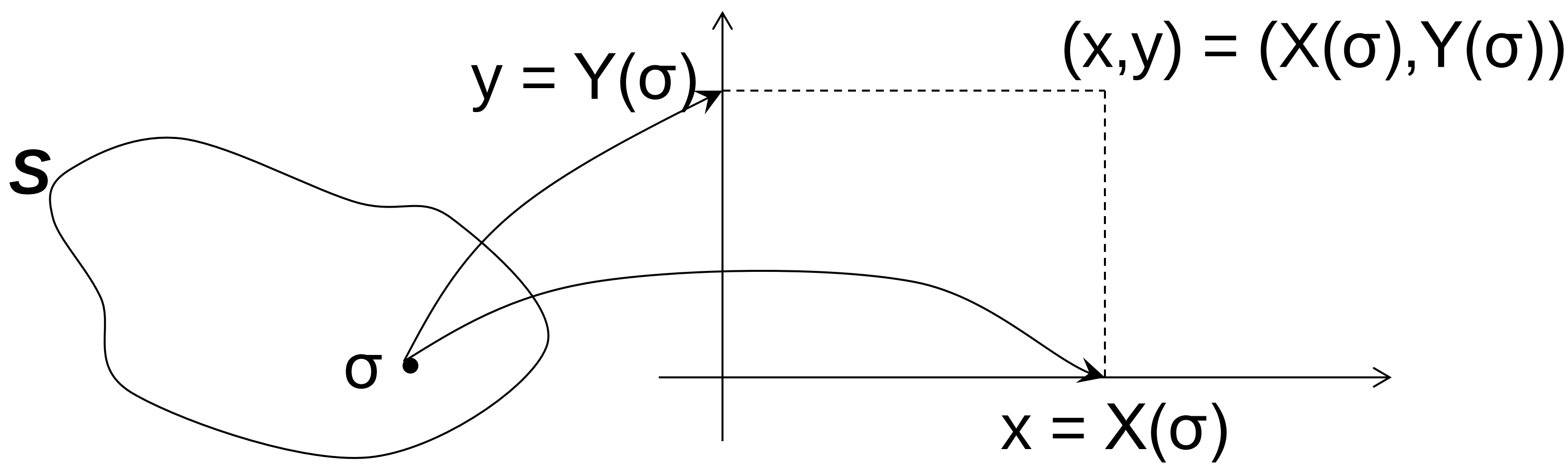
- Y: ξένες γλώσσες του παιδιού

n	0	1	2	3
P[Y=n]	0.246	0.520	0.202	0.032

- E[Y] = 1.020, σ_Y = 0.759

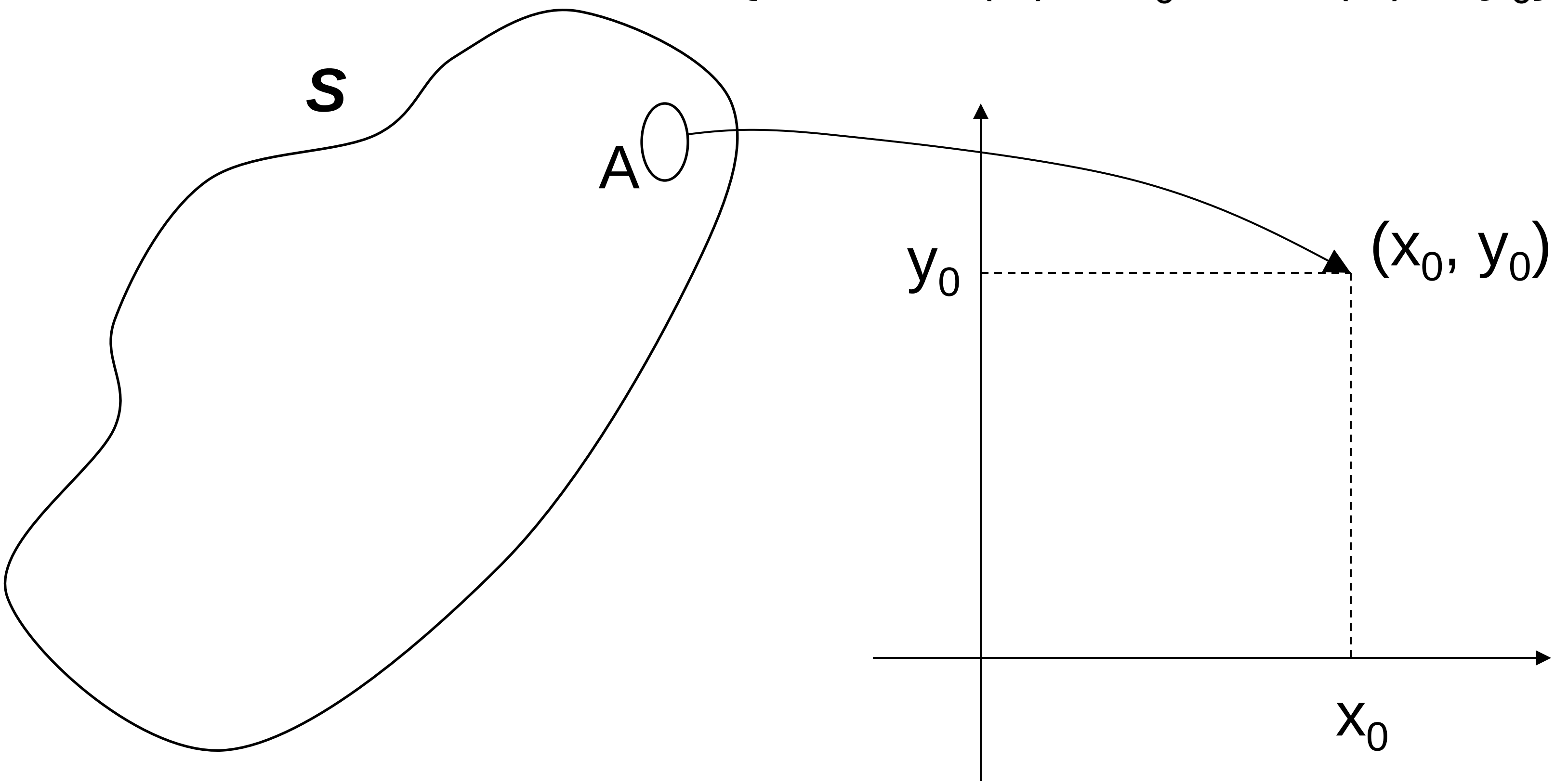
Ζεύγη τυχαίων μεταβλητών

- Εχουμε δύο τ.μ. X και Y ορισμένες στον ίδιο δ.χ. ***S***, δηλ. X: $\sigma \in \boldsymbol{S} \rightarrow X(\sigma) \in \mathbf{R}$ και Y: $\sigma \in \boldsymbol{S} \rightarrow Y(\sigma) \in \mathbf{R}$ (π.χ. το ύψος και το βάρος ενός ατόμου).
- Τότε το ζεύγος (X,Y) αντιστοιχεί σε κάθε σημείο του ***S*** ένα σημείο (x,y) στο επίπεδο: $(X,Y): \sigma \in \boldsymbol{S} \rightarrow (X(\sigma), Y(\sigma)) \in \mathbf{R}^2$.



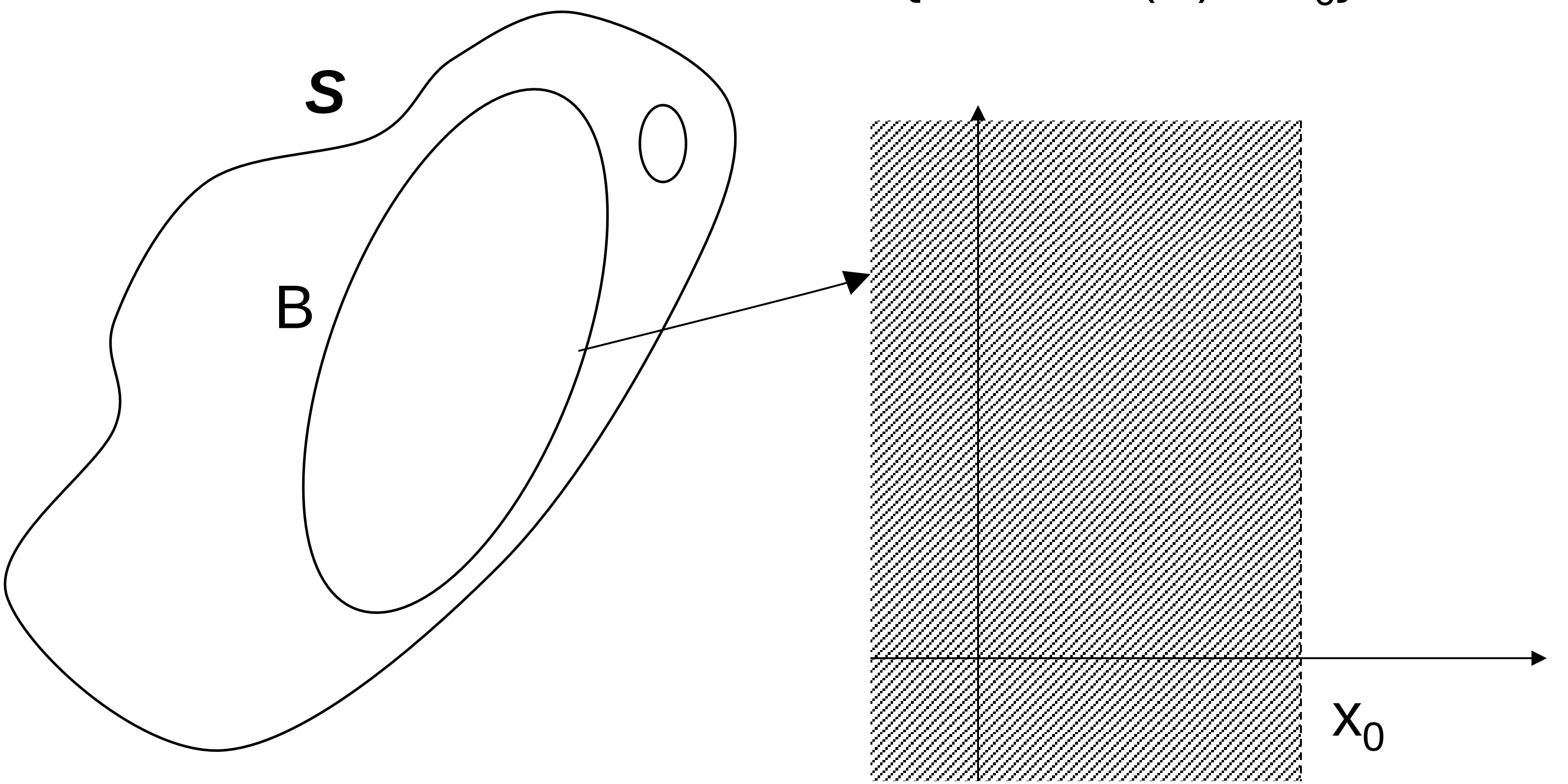
Ενδεχόμενα που ορίζονται από δύο τ.μ.

$A = \{\sigma \in \boldsymbol{S} \;/\; X(\sigma) = x_0 \text{ και } Y(\sigma) = y_0\}$



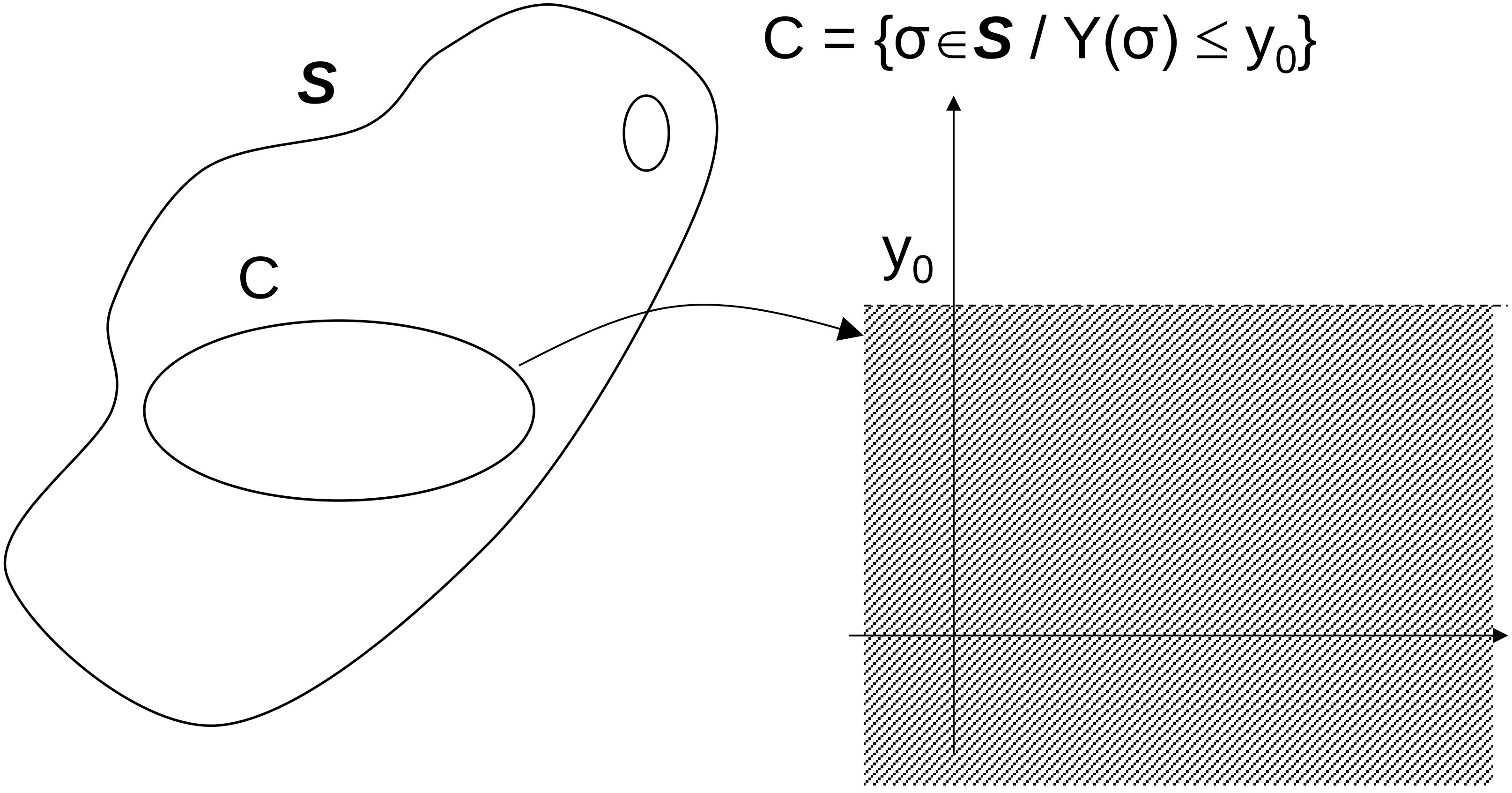
Ενδεχόμενα που ορίζονται από δύο τ.μ.

$B = \{\sigma \in \boldsymbol{S} \ / \ X(\sigma) \le x_0\}$



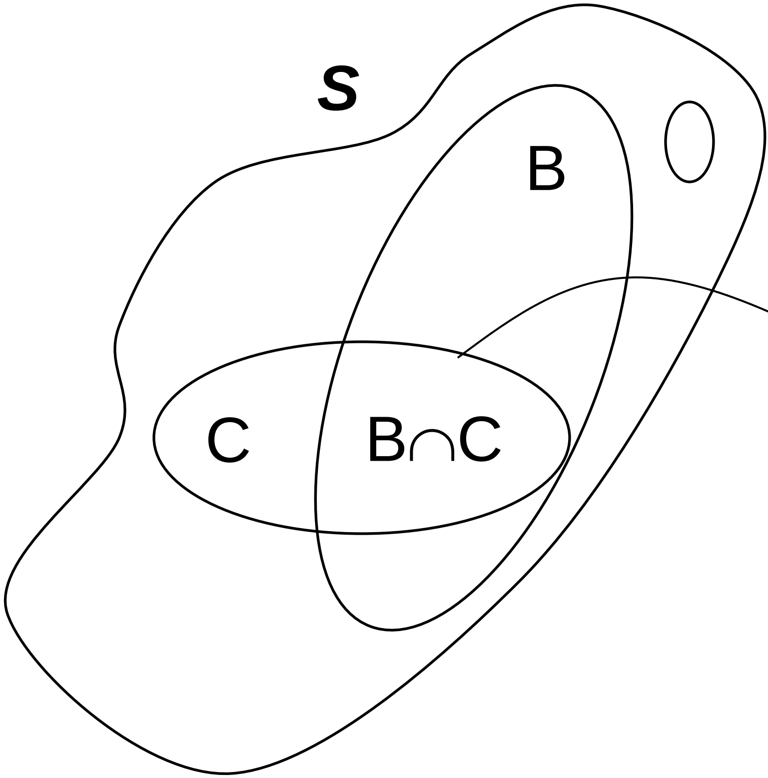
Ενδεχόμενα που ορίζονται από δύο τ.μ.

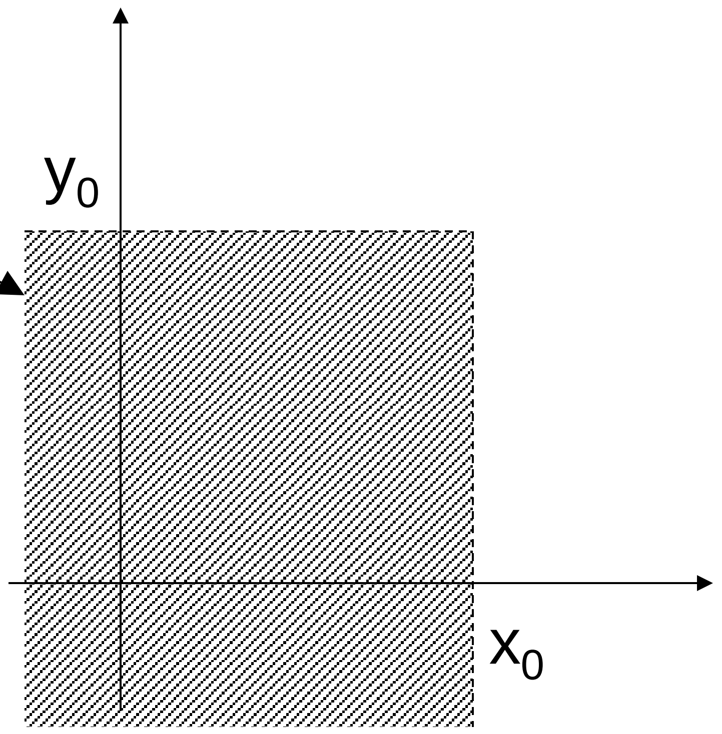
$C = \{\sigma \in \boldsymbol{S} \,/\, Y(\sigma) \le y_0\}$



Ενδεχόμενα που ορίζονται από δύο τ.μ.

$B \cap C = \{\sigma \in \boldsymbol{S} \ / \ X(\sigma) \leq x_0 \text{ και } Y(\sigma) \leq y_0\}$





Συναρτήσεις κατανομής

- Ορίζουμε τις εξής συναρτήσεις κατανομής:

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

  $F_Y(x) = P(Y \leq y)$

  $F(x, y) = P(X \leq x \text{ και } Y \leq y) = P(\{X \leq x\} \cap \{Y \leq y\})$
- Οι δύο πρώτες λέγονται «περιθωριακές» (marginal) σ.κ., δίνουν την πιθανότητα μιας ημιευθείας των πραγματικών αριθμών και είναι οι σ.κ. που ξέραμε μέχρι τώρα.
- Η $F(x, y)$ λέγεται «από κοινού» (joint) σ.κ. των X και Y και μας δίνει την πιθανότητα ενός τεταρτημόριου του επιπέδου.

Ανεξάρτητες τ.μ.

- ΟΡΙΣΜΟΣ: Δύο τ.μ. ορισμένες στον ίδιο δ.χ. ***S*** λέγονται ανεξάρτητες αν τα ενδεχόμενα $\{X \leq x\}$ και $\{Y \leq y\}$ είναι ανεξάρτητα για κάθε x και y,

  δηλ. αν:

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

  ή ισοδύναμα:

  $$F(x, y) = F_X(x)\, F_Y(y)$$

Ιδιότητες της από κοινού σ.κ.

$0 \leq F(x, y) \leq 1$ για κάθε $x, y \in \mathbf{R}$

Αν $x_1 \leq x_2$ και $y_1 \leq y_2$, τότε $F(x_1, y_1) \leq F(x_2, y_2)$

$$\lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$$

$$\lim_{x \to -\infty} F(x, y) = 0$$

$$\lim_{y \to -\infty} F(x, y) = 0$$

Ιδιότητες της από κοινού σ.κ. (2)

$$\lim_{y \to +\infty} F(x,y) = F_X(x) \qquad \lim_{x \to +\infty} F(x,y) = F_Y(y)$$

Παρατήρηση: Από την από κοινού σ.κ. μπορούμε να βρούμε τις περιθωριακές, αλλά όχι αντίστροφα.

Η σ.κ. είναι συνεχής από δεξιά:

$$\lim_{x \to \alpha+} F(x,y) = F(\alpha, y) \qquad \lim_{y \to \beta+} F(x,y) = F(x, \beta)$$

Για κάθε a, b, c, d με $a \leq b$ και $c \leq d$:

$F(a, c) – F(a, d) – F(b, c) + F(b, d) \geq 0$

Η παραπάνω ποσότητα είναι η $P(a < X \leq b, c < Y \leq d)$

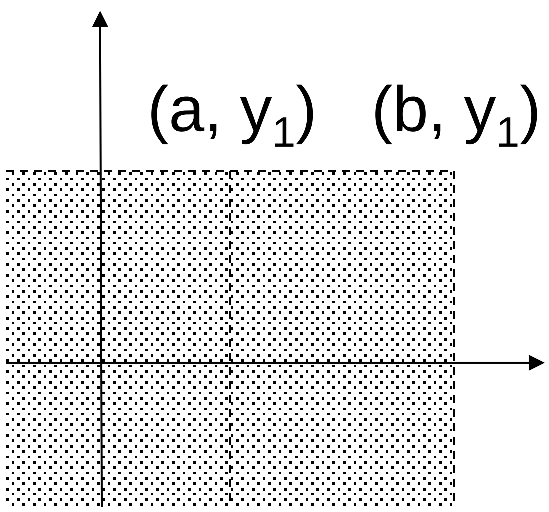
Υπολογισμός πιθανοτήτων από την σ.κ.

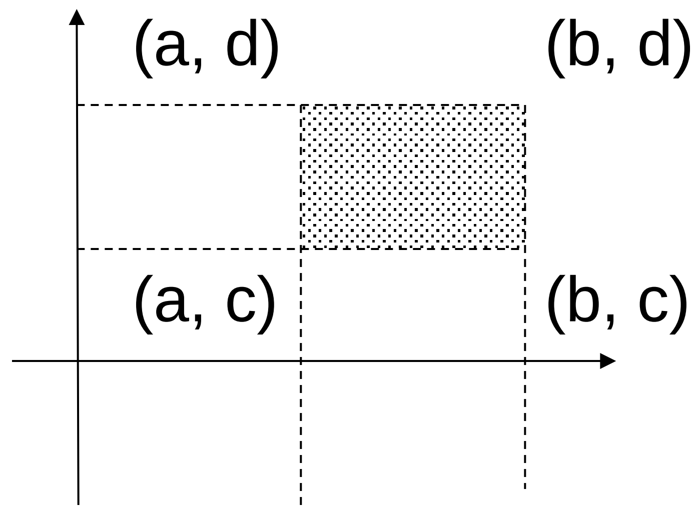
$P(X \leq x_1, Y \leq y_1) = F(x_1, y_1)$

$P(a < X \leq b, Y \leq y_1) = F(b, y_1) – F(a, y_1)$

$P(a < X \leq b, c < Y \leq d) = F(a, c) – F(a, d) – F(b, c) + F(b, d)$

$P(a < X \leq b, Y > y_1) = F_X(b) – F_X(a) – F(b, y_1) + F(a, y_1)$





Παράδειγμα

- X: χρόνος επόμενης βλάβης από αστοχία υλικού
- Y: χρόνος επόμενης βλάβης από λάθος χειρισμού
- Από κοινού σ.κ.:

  $F(x, y) = 1 - e^{-x/24} - e^{-y/12} + e^{-x/24-y/12}$ για $x, y \geq 0$
- Πιθανότητα να έχουμε βλάβη και των δύο τύπων το επόμενο 24ωρο:

  $P(X < 24, Y < 24) = F(24, 24) = 0.5466$
- Πιθανότητα να μην έχουμε βλάβη και των δύο τύπων το επόμενο 24ωρο (μπορεί να έχουμε μόνο X ή Y):

  $P(X > 24$ ή/και $Y > 24) = 1 - F(24, 24) = 0.4534$

Παράδειγμα

- Πιθανότητα να μην έχουμε καμμία βλάβη το επόμενο 24ωρο, αλλά να έχουμε βλάβη και των δύο τύπων το μεθεπόμενο 24ωρο:

  $P(24 < X < 48 \text{ και } 24 < Y < 48)$
  $= F(48, 48) – F(24, 48) – F(48, 24) – F(24, 24) = 0.0272$

- Πιθανότητα να μην έχουμε κανενός είδους βλάβη το επόμενο 24ωρο:

  $P(X > 24, X > 24)$
  $= 1 – F_X(24) – F_Y(24) + F(24, 24) = 0.0498$

Παράδειγμα

- Οι τ.μ. X και Y είναι ανεξάρτητες;

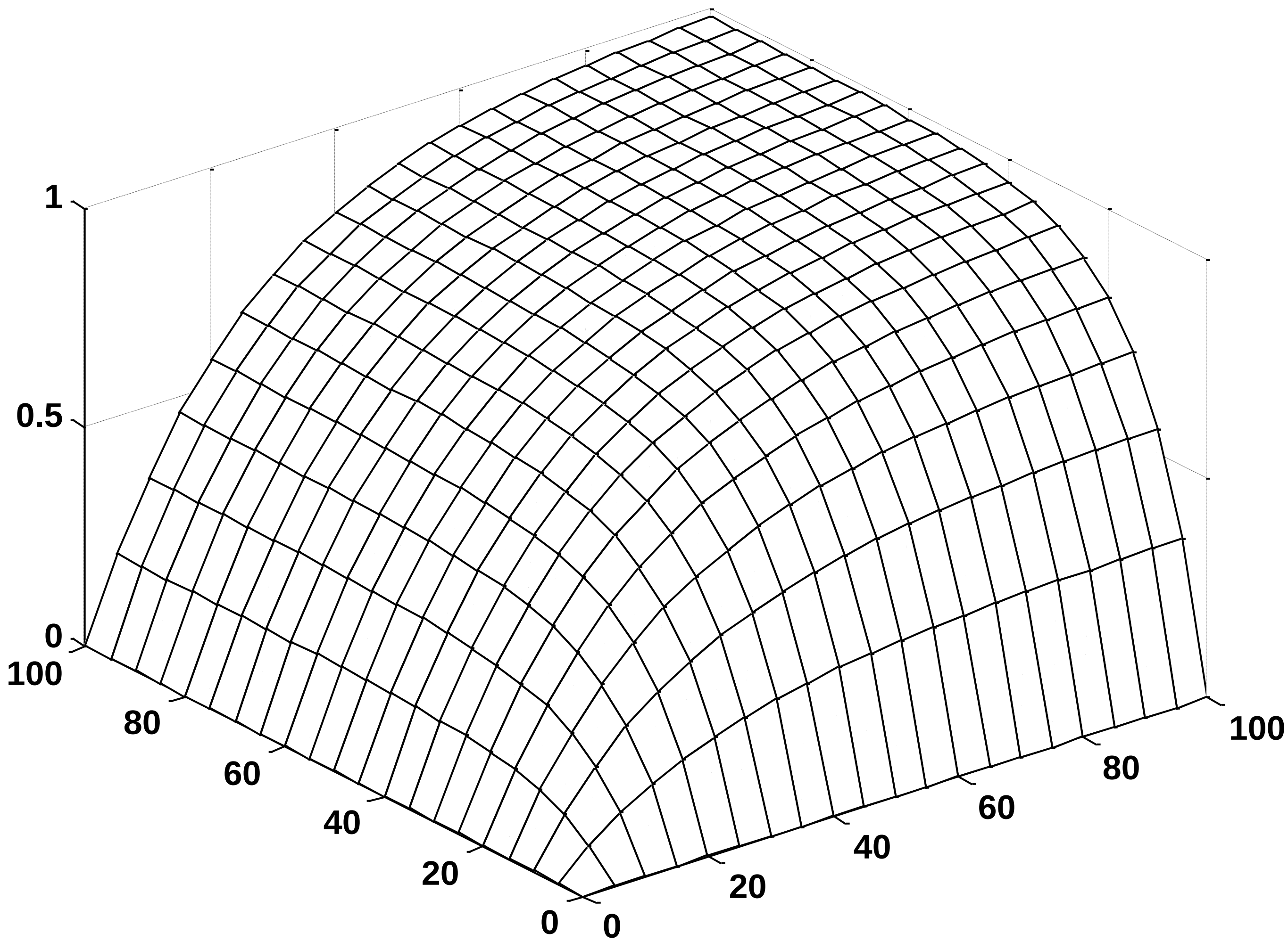
  $F(x, y) = 1 - e^{-x/24} - e^{-y/12} + e^{-x/24-y/12}$ για $x, y \geq 0$

  $F_X(x) = F(x, +\infty) = 1 - e^{-x/24}$ για $x \geq 0$

  $F_Y(y) = F(+\infty, y) = 1 - e^{-y/12}$ για $y \geq 0$

  $F_X(x)\, F_Y(y) = F(x, y)$ για κάθε x, y, άρα είναι ανεξάρτητες

Παράδειγμα: γραφική παράσταση της σ.κ.



Διακριτές διδιάστατες τ.μ.

- X, Y διακριτές: από κοινού συνάρτηση πιθανότητας

  $p(i, j) = P(X = i, Y = j)$
- Ιδιότητες:

  $0 \leq p(i, j) \leq 1$

$$\sum_{i} \sum_{j} p(i, j) = 1$$

$$F(x, y) = \sum_{i \leq x} \sum_{j \leq y} p(i, j)$$

- X, Y ανεξάρτητες: $p(i, j) = p_X(i)\, p_Y(j)$

Παράδειγμα – από κοινού σ.π.

X (ξένες γλώσσες της μητέρας) \ Y (ξένες γλώσσες του παιδιού)	**0**	**1**	**2**	**3**
0	0.246	0.192	0.034	0.000
1	0.000	0.298	0.052	0.004
2	0.000	0.030	0.090	0.006
3	0.000	0.000	0.026	0.022

Επαληθεύουμε ότι $\sum_i \sum_j p(i,j) = 1$

Παράδειγμα – από κοινού α.σ.κ.

X (ξένες γλώσσες της μητέρας)	Y (ξένες γλώσσες του παιδιού)			
	0	**1**	**2**	**3**
0	0.246	0.438	0.472	0.472
1	0.246	0.736	0.822	0.826
2	0.246	0.766	0.942	0.952
3	0.246	0.766	0.968	1.000

Περιθωριακές κατανομές

- Από την από κοινού συνάρτηση πιθανότητας μπορούμε να υπολογίσουμε τις περιθωριακές των X, Y:

$$P(X = i) = p_X(i) = \sum_j p(i, j)$$

$$P(Y = j) = p_Y(j) = \sum_i p(i, j)$$

- Από τις περιθωριακές γενικά δεν μπορούμε να υπολογίσουμε την από κοινού, εκτός αν:
- Οι X, Y ανεξάρτητες: $p(i, j) = p_X(i)\, p_Y(j)$

Παράδειγμα – περιθωριακές σ.π.

X (ξένες γλώσσες της μητέρας) \ Y (ξένες γλώσσες του παιδιού)	**0**	**1**	**2**	**3**	$p_X(i)$↓
0	0.246	0.192	0.034	0.000	0.472
1	0.000	0.298	0.052	0.004	0.354
2	0.000	0.030	0.090	0.006	0.126
3	0.000	0.000	0.026	0.022	0.048
$p_Y(j)$→	0.246	0.520	0.202	0.032	1.000

Παράδειγμα – περιθωριακές σ.π.

n	0	1	2	3
$p_X(n)$	0.472	0.354	0.126	0.048
$p_Y(n)$	0.246	0.520	0.202	0.032

$E[X] = \Sigma n\, p_X(n) = 0.750$

$E[X^2] = \Sigma n^2\, p_X(n) = 1.290$

$Var[X] = E[X^2] – E[X]^2 = 0.728$

$\sigma_X = 0.853$

$E[Y] = 1.020$

$E[Y^2] = 1.616$

$Var[Y] = 0.576$

$\sigma_Y = 0.759$

Υπό συνθήκη κατανομές

- Υπό συνθήκη συνάρτηση πιθανότητας (conditional distribution) της X δεδομένου ότι Y = j:

$$p_{X|Y=j}(i,j) = P(X = i \mid Y = j) = \frac{P(X = i, Y = j)}{P(Y = j)} = \frac{p(i,j)}{p_Y(j)}$$

- Για δεδομένο j η $p_{X|Y=j}(i)$ είναι έγκυρη σ.π. (άθροισμα = 1)
- Εκφράζει την a posteriori πιθανότητα του X = i δεδομένου ότι Y = j.
- Αντίστοιχα η $p_X(i) = P(X = i)$ είναι η a priori πιθανότητα.
- Αντίστοιχα ορίζεται η $p_{Y|X=i}(j)$

Υπό συνθήκη μέσες τιμές

- Η υπό συνθήκη μέση τιμή και διασπορά της X δεδομένου ότι Y = j είναι αντίστοιχα:

$$E\{X \mid Y = j\} = \sum_i i\, p_{X|Y=j}(i, j)$$

$$Var\{X \mid Y = j\} = \sum_i (i - \mu)^2\, p_{X|Y=j}(i, j)$$

- και γενικά:

$$E\{g(X) \mid Y = j\} = \sum_i g(i)\, p_{X|Y=j}(i, j)$$

Παράδειγμα (συν.)

- Πόσες ξένες γλώσσες μιλάει το παιδί, αν η μητέρα μιλάει 2 γλώσσες;

j	$p_{Y\|X=2}(j) = p(2, j) / p_X(2)$
0	0
1	0.030 / 0.126 = 0.238
2	0.090 / 0.126 = 0.714
3	0.006 / 0.126 = 0.048

Παράδειγμα (συν.)

- A priori μέση τιμή: $E[Y] = 1.020$

  (μέση τιμή των γλωσσσών που μιλούν τα παιδιά – σύνολο του πληθυσμού)
- A posteriori μέση τιμή: $E[Y \mid X = 2] = 1.810$

  (μέση τιμή των γλωσσσών που μιλούν τα παιδιά των μητέρων που γνωρίζουν δύο γλώσσες)
- Αντίστοιχα a priori: $Var[Y] = 0.576$
  a posteriori: $Var[Y \mid X = 2] = 0.249$

Παράδειγμα (συν.)

- Πόσες ξένες γλώσσες μιλούν οι μητέρες των παιδιών που γνωρίζουν 2 γλώσσες;

i	$p_{X\|Y=2}(i) = p(i, 2) / p_Y(2)$
0	0.034 / 0.202 = 0.168
1	0.052 / 0.202 = 0.257
2	0.090 / 0.202 = 0.446
3	0.026 / 0.202 = 0.129

$E[X \mid Y = 2] = 1.536$, $Var[X \mid Y = 2] = 0.843$

Ανεξάρτητες τ.μ.

- X, Y ανεξάρτητες $\Leftrightarrow$ $p(i, j) = p_X(i)\, p_Y(j)$
  $\Leftrightarrow$ $p_{X\,|\,Y}(i, j) = p(i, j) \,/\, p_Y(j) = p_X(i)$
  (εκ των υστέρων κατανομή = εκ των προτέρων)
- Πληροφορία που έχουμε για την Y δεν μας προσφέρει καμμία πληροφορία για την X.

Συνδιασπορά δύο τ.μ.

$$\begin{aligned}\mathrm{Cov}\{X, Y\} &= E\{(X - E\{X\})\,(Y - E\{Y\})\} \\ &= E\{XY\} - E\{X\}\,E\{Y\}\end{aligned}$$

- Φυσική σημασία: πόσο μεταβάλλονται **από κοινού** δύο τ.μ. γύρω από τις μέσες τιμές τους.
- Ισοδύναμα, πόσο ισχυρή είναι η εξάρτηση μιας τ.μ. από την άλλη.
- Μπορεί να είναι οποιοσδήποτε πραγματικός αριθμός (η διασπορά είναι θετική ή 0)
- Για κάθε X, Y: $\mathrm{Var}\{X + Y\} = \mathrm{Var}\{X\} + \mathrm{Var}\{Y\} + 2\mathrm{Cov}\{X, Y\}$
- Αν οι X, Y είναι ανεξάρτητες, $\mathrm{Cov}\{X, Y\} = 0$.

Συντελεστής συσχέτισης

$$\mathrm{Corr}\{X, Y\} = \frac{\mathrm{Cov}\{X, Y\}}{\sigma_X \sigma_Y}$$

- $-1 \leq \rho \leq 1$
- Για $\rho > 0$, υπάρχει θετική συσχέτιση των X, Y: μεγάλες τιμές της X αντιστοιχούν γενικά σε μεγάλες τιμές της Y.
- Για $\rho < 0$, υπάρχει αρνητική συσχέτιση των X, Y: μεγάλες τιμές της X αντιστοιχούν γενικά σε μικρές τιμές της Y.
- Για $\rho = 1$, $Y = \alpha X + \beta$, $\alpha > 0$
- Για $\rho = -1$, $Y = \alpha X + \beta$, $\alpha < 0$
- Για $\rho = 0$, οι X, Y είναι ασυσχέτιστες

Παράδειγμα (συν.)

- Ποιά σχέση υπάρχει μεταξύ της γλωσσομάθειας των παιδιών και της γλωσσομάθειας των μητέρων;

  $E[XY] = \Sigma\ \Sigma\ i\ j\ p(i, j) = 1.224$

  $Cov[X, Y] = E[XY] – E[X]\ E[Y] = 0.459$

  $\rho = Cov[X, Y] / (\sigma_X\ \sigma_Y) = 0.709$
- Υπάρχει ισχυρή θετική συσχέτιση μεταξύ X και Y, δηλ. η γλωσσομάθεια της μητέρας επηρεάζει σημαντικά το παιδί.