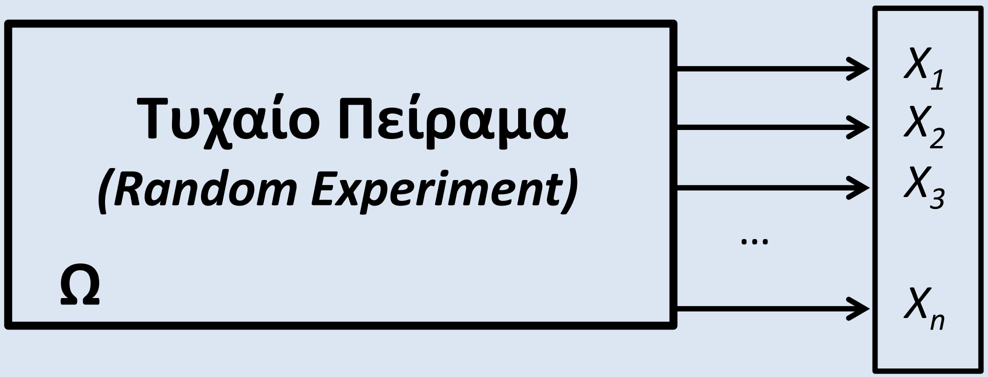
Πολυδιάστατες τυχαίες μεταβλητές
(*multivariate random variables*)

Πολυδιάστατες μεταβλητές

- **Ποσοτικά χαρακτηριστικά** που σχετίζονται με τη διαδικασία εκτέλεσης ενός πειράματος.



Πολυμεταβλητές
ή
Διάνυσμα τυχαίων μεταβλητών

- **Παράδειγμα:**
  - Πειραματική **διαδικασία:** τυχαία επιλογή ατόμου
  - Μετράμε πολλές **τυχαίες μεταβλητές (χαρακτηριστικά):**
    - Ηλικία ($\boldsymbol{X_1}$)
    - Βάρος ($\boldsymbol{X_2}$)
    - Ύψος ($\boldsymbol{X_3}$)
    - ….

Πολυδιάστατες μεταβλητές

Επικεντρωνόμαστε στα εξής:

- **από κοινού συμπεριφορά** όλων των μεταβλητών (χαρακτηριστικών) **ως μία διανυσματική οντότητα** εκφρασμένη σε έναν πολυδιάστατο χώρο,
- **ατομική συμπεριφορά** μιας μεταβλητής εκφρασμένη πάνω σε έναν άξονα του πολυδιάστατου χώρου,
- **συσχετίσεις** μεταξύ των μεταβλητών ώστε να βρίσκουμε τον **βαθμό εξάρτησής** τους και να μελετούμε ένα υποσύνολο των μεταβλητών όταν γνωρίζουμε τις τιμές των υπολοίπων (**δεσμευμένη** έκφραση).

Επέκταση σε πολλές διαστάσεις

Επεκτείνουμε τη βασική θεωρία μίας μεταβλητής εισάγοντας

- ✓ **από-κοινού κατανομή (*joint distribution*)**: κατανομή του συνόλου **όλων** των ***n*** μεταβλητών
- ✓ **περιθώρια κατανομή (*marginal distribution*)**: κατανομή **υποσυνόλου** των μεταβλητών
- ✓ **δεσμευμένη ή υπό συνθήκη κατανομή (*conditional distribution*)**: κατανομή υποσυνόλου των μεταβλητών όταν οι υπόλοιπες έχουν **σταθερή τιμή** (είναι **γνωστές**)

- Επικεντρωνόμαστε στις **δύο διαστάσεις *(X,Y)*,** και γενικεύουμε για την περίπτωση πολλών διαστάσεων $\boldsymbol{X=(X_1, X_2,\ldots, X_n)}$.

Διακριτές δυσδιάστατες τυχαίες μεταβλητές

- Έστω **δύο τυχαίες μεταβλητές(*X,Y*)** (*δυσδιάστατο διάνυσμα*) με τιμές από ένα **αριθμήσιμο σύνολο** με **ζεύγη** τιμών:

$$\Omega_{X,Y} = \{(\boldsymbol{x_j}, \boldsymbol{y_k}), j = 1, \ldots, n_x \;\; k = 1, \ldots, n_y\}$$

- Δηλαδή, πλήθος ζευγών: $n_x \times n_y$

Διακριτή τυχαία μεταβλητή (*υπενθύμιση*)

- Η συνάρτηση κατανομής πιθανότητας ***F(x)=P(X ≤ x)*** είναι **μη φθίνουσα**, **δεξιά συνεχής** και **κλιμακωτή** με **άλματα** στις τιμές της μεταβλητής.

- **Η συνάρτηση πυκνότητας (ή μάζας) πιθανότητας (σ.π.π.)** είναι το σύνολο πιθανοτήτων των τιμών μιας διακριτής τ.μ. X :

$$f(x) = P\left(X = x\right) \geq 0 \qquad \forall x \in \Omega_X = \left\{x_1, x_2, \ldots, x_n\right\}$$

Ισχύει: $\sum_{i=1}^{n} f\left(x_i\right) = 1$

από-κοινού συνάρτηση πυκνότητας

- **Η από-κοινού συνάρτηση πυκνότητας (ή μάζας) πιθανότητας (*joint density*)** των ***(X,Y)*** εκφράζει τις **από-κοινού πιθανότητες** όλων των δυνατών **ζευγών τιμών** (***x_j***, ***y_k***) των δύο μεταβλητών:

$$f_{X,Y}(x_j, y_k) = P(X = x_j \cap Y = y_k) = \\ = P(X = x_j, Y = y_k)$$

Ισχύει ότι: $$\sum_{j=1}^{n_x}\sum_{k=1}^{n_y} f_{X,Y}(x_j, y_k) = 1$$

Υπολογισμός πιθανότητας ενδεχομένων

- Ένα **ενδεχόμενο** ***A*** περιλαμβάνει τιμές (x_j, y_k) των δύο μεταβλητών (μία ή περισσότερες).
- Η **πιθανότητα ενός ενδεχομένου** ***A*** εκφράζεται ως το άθροισμα πιθανοτήτων όλων των ζευγών τιμών που περιλαμβάνονται στο ***A***, δηλ.

$$P(A) = \sum_{(x_j, y_k) \in A} P\left(X = x_j, Y = y_k\right) = $$

$$= \sum_{(x_j, y_k) \in A} f_{X,Y}\left(x_j, y_k\right)$$

Περιθώρια κατανομή (*marginal distribution*)

- Ενώ η **από-κοινού κατανομή** μας δίνει πληροφορίες για την από-κοινού συμπεριφορά των δύο μεταβλητών *X* και *Y*, πολλές φορές μας ενδιαφέρει **ξεχωριστά** η συμπεριφορά **μιας** εκ των δύο μεταβλητών.

- Ορίζεται η **περιθώρια κατανομή (*marginal distribution*) μιας** μεταβλητής,
- ή κατ' επέκταση **ενός υποσυνόλου** των ***n*** τυχαίων μεταβλητών του διανύσματος.

Περιθώρια κατανομή (*marginal distribution*)

- Έστω 2 τ.μ. ***X, Y*** με από-κοινού σ.π.π. $f_{X,Y}(x,y)$

$$\Omega_{X,Y} = \{(x_j, y_k), \ j = 1, \ldots, n_x \ \ k = 1, \ldots, n_y\}$$

- Δηλαδή το πλήθος τιμών είναι $n_x \times n_y$.

- Η **περιθώρια κατανομή** του ***X*** καθορίζεται μέσω του προσδιορισμού της συνάρτησης πυκνότητας (ή μάζας) πιθανότητας της τυχαίας μεταβλητής:

$$f_X(x_j) = P(X = x_j)$$

$$\Omega_X = \{x_j, \quad j = 1, \ldots, n_x\}$$

Περιθώρια κατανομή (*marginal distribution*)

- **Εύρεση της περιθώριας κατανομής**

ολική πιθανότητα

$$f_X(\boldsymbol{x_j}) = P(X = \boldsymbol{x_j}) = P(X = \boldsymbol{x_j} \cap Y = '\boldsymbol{οτιδηποτε}') =$$

$$= P\left(\{X = \boldsymbol{x_j} \cap Y = \boldsymbol{y_1}\} \cup \{X = \boldsymbol{x_j} \cap Y = \boldsymbol{y_2}\} \cup \cdots \cup \{X = \boldsymbol{x_j} \cap Y = \boldsymbol{y_{n_y}}\}\right) =$$

$$= P(X = \boldsymbol{x_j}, Y = \boldsymbol{y_1}) + P(X = \boldsymbol{x_j}, Y = \boldsymbol{y_2}) + \cdots + P\left(X = \boldsymbol{x_j}, Y = \boldsymbol{y_{n_y}}\right) =$$

$$= \sum_{\boldsymbol{k}=1}^{n_y} f_{X,Y}(\boldsymbol{x_j}, \boldsymbol{y_k})$$

$$\Omega_X = \{\boldsymbol{x_j}, j = 1, \dots, n_x\}$$

- φυσικά ισχύει ότι:

$$\sum_{j=1}^{n_x} f_X(\boldsymbol{x_j}) = \sum_{j=1}^{n_x} \sum_{k=1}^{n_y} f_{X,Y}(\boldsymbol{x_j}, \boldsymbol{y_k}) = 1$$

- Παρόμοια, η **περιθώρια κατανομή** της μεταβλητής ***Y*** υπολογίζεται ως:

$$f_Y(\boldsymbol{y_k}) = P(Y = \boldsymbol{y_k}) = P(X = '\boldsymbol{οτιδηποτε}' \cap Y = \boldsymbol{y_k}) =$$

$$= P\left(\{X = \boldsymbol{x_1} \cap Y = \boldsymbol{y_k}\} \cup \{X = \boldsymbol{x_2} \cap Y = \boldsymbol{y_k}\} \cup \cdots \cup \left\{X = \boldsymbol{x_{n_x}} \cap Y = \boldsymbol{y_k}\right\}\right) =$$

$$= \sum_{j=1}^{n_x} f_{X,Y}\left(\boldsymbol{x_j}, \boldsymbol{y_k}\right)$$

$$\Omega_Y = \left\{\boldsymbol{y_k},\ k = 1, \dots, n_y\right\}$$

- ισχύει ότι: $$\sum_{k=1}^{n_y} f_Y(\boldsymbol{y_k}) = 1$$

Δεσμευμένη (υπό-συνθήκη) κατανομή (*conditional distribution*)

- Θέλουμε να βρούμε την κατανομή μιας μεταβλητής (***X***) όταν η τιμή της άλλης έχει σταθερή τιμή (***Y=y***).

- Έστω 2 τυχαίες μεταβλητές *X*, *Y* με από-κοινού σ.π.π. $f_{XY}(x,y)$ και περιθώριες $f_X(x)$, $f_Y(y)$

- Ορίζεται έτσι η **δεσμευμένη κατανομή** ***X |Y=y***

Δεσμευμένη κατανομή (*conditional distribution*)

- Δεσμευμένη κατανομή ***X|Y=y*** με σ.π.π.

$$f_{X|Y}(x\,|\,y) = P(X = x\,|\,Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- Έτσι, η δεσμευμένη του *X* ισούται με το πηλίκο της από-κοινού προς την περιθώρια του *Y* (*δέσμευσης*)
- Παρόμοια, **η δεσμευμένη σ.π.π. του *Y | X=x***

$$f_{Y|X}(y\,|\,x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

- Χρήσιμοι τύποι:

$$f_{XY}(x,y) = f_{X|Y}(x \mid y) f_Y(y)$$

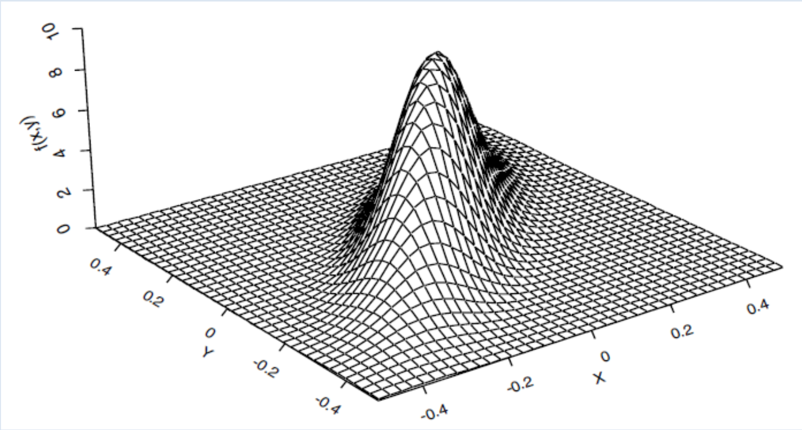
ή

$$f_{XY}(x,y) = f_{Y|X}(y \mid x) f_X(x)$$

- Η **από-κοινού** σ.π.π. είναι το **γινόμενο** της **δεσμευμένης** με την **περιθώρια** σ.π.π.

Συνεχείς δυσδιάστατες τυχαίες μεταβλητές

- Δύο τυχαίες μεταβλητές *X*, *Y* είναι **από-κοινού συνεχείς** όταν υπάρχει μία **μη-αρνητική** συνάρτηση $f_{X,Y}(x,y) \geq 0$ που ονομάζεται **από-κοινού συνάρτηση πυκνότητας πιθανότητας** για την οποία ισχύει ότι:

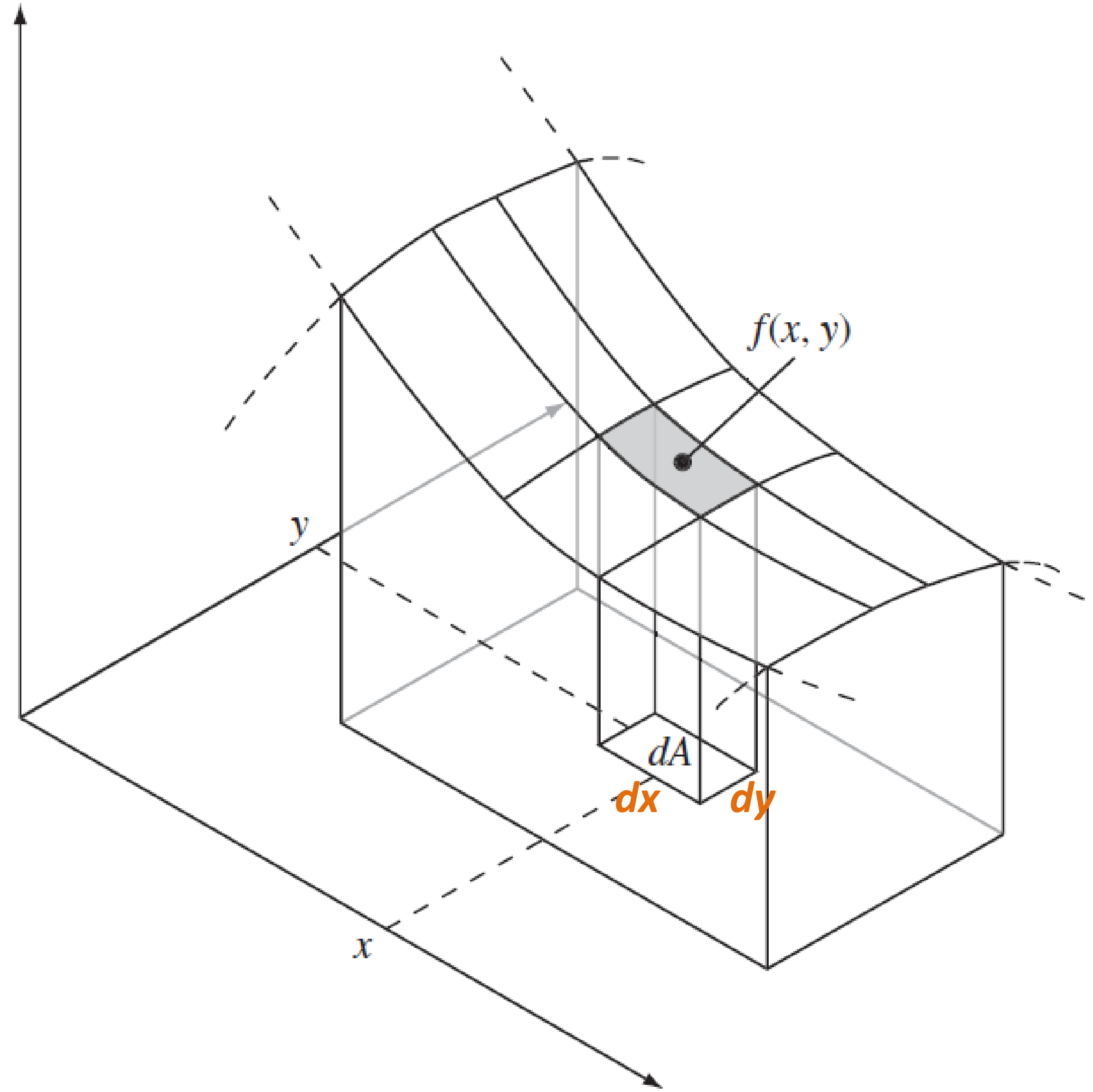


$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

Παρατηρήσεις

- Καθώς υπάρχουν **άπειρα σημεία** (x,y) η **μάζα πιθανότητας** που υπάρχει στο καθένα από αυτά είναι **μηδέν**.
- Στον συνεχή χώρο η συνάρτηση $f_{X,Y}(x,y) \geq 0$ **δεν παριστάνει πιθανότητα**.
- **Προσεγγιστικά ισχύει**:

$$P(X = x, Y = y) =$$
$$= P(x \leq X < x + dx, y \leq Y < y + dy) =$$
$$= f_{X,Y}(x,y)dxdy$$

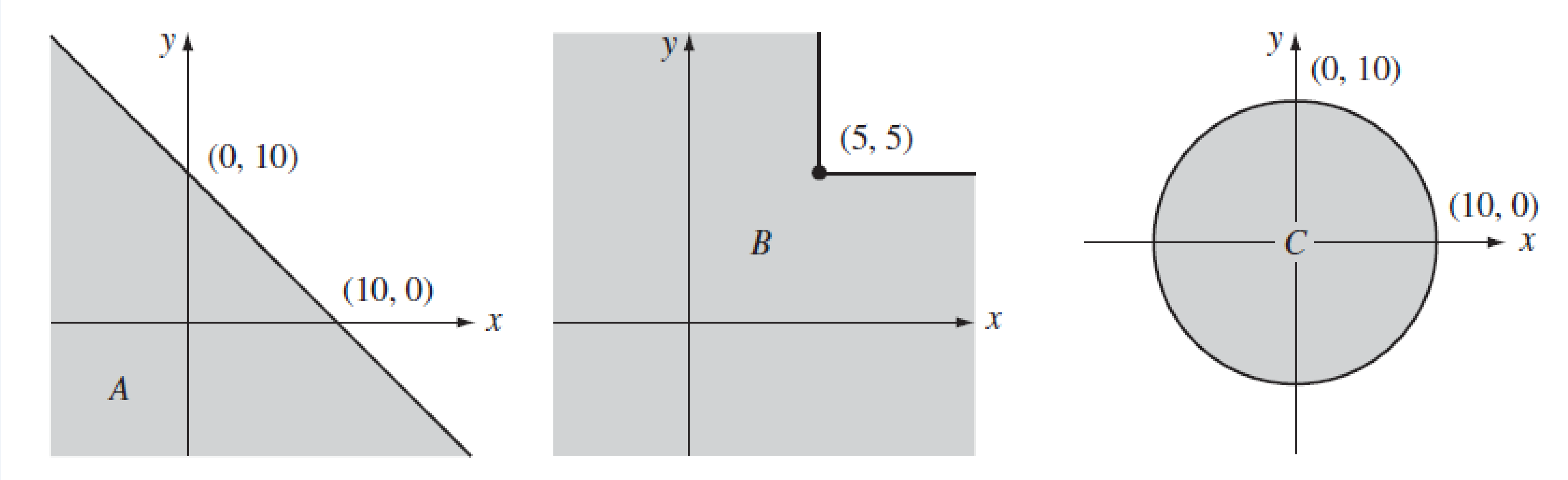


$$P(\boldsymbol{X}=x,\boldsymbol{Y}=y)=f_{\boldsymbol{X},\boldsymbol{Y}}(x,y)\boldsymbol{dxdy}$$

Υπολογισμός πιθανοτήτων στο συνεχή χώρο

- Τα **ενδεχόμενα** ορίζονται ως **περιοχές** του x-y επιπέδου. Παραδείγματα:

$$A = \{X + Y \leq 10\}$$
$$B = \{\min(X, Y) \leq 5\} = \{X \leq 5\} \cup \{Y \leq 5\}$$
$$C = \{X^2 + Y^2 \leq 100\}.$$



Υπολογισμός πιθανοτήτων στο συνεχή χώρο

- Τότε, η **πιθανότητα** ενός ενδεχομένου υπολογίζεται από το **διπλό ολοκλήρωμα** **της** **από-κοινού** **συνάρτησης πυκνότητας των 2 μεταβλητών** στην περιοχή ολοκλήρωσης που ορίζει το ενδεχόμενο:

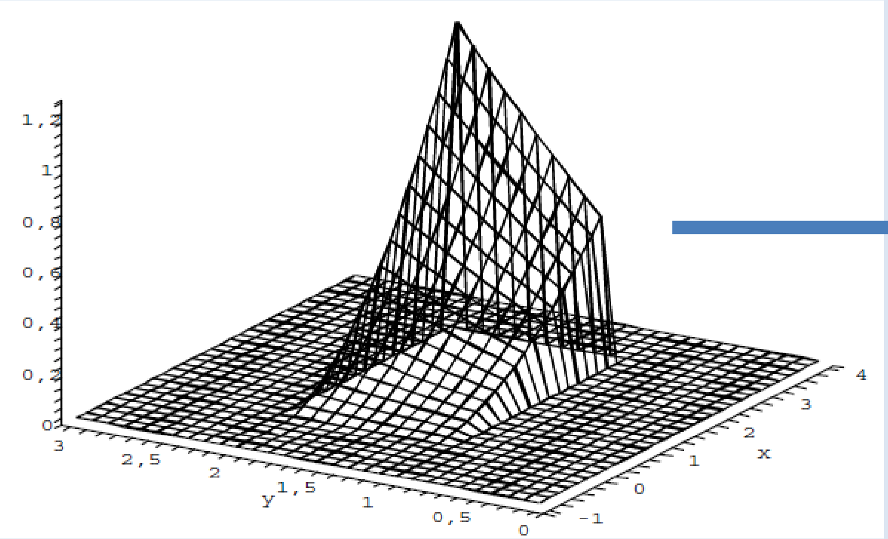
$$P\big((X,Y) \in \boldsymbol{A}\big) = \iint\limits_{\boldsymbol{A}} f_{X,Y}(x,y)\,dx\,dy$$

- Δηλαδή ισούται με τον **όγκο** κάτω από την **σ.π.π.** $\boldsymbol{f_{X,Y}(x,y)}$ στην περιοχή ολοκλήρωσης του ενδεχομένου ***A***.

Συνεχείς δυσδιάστατες τυχαίες μεταβλητές (συν.)

Παράδειγμα:

$$f(x,y) = \frac{2}{75}(2x^2y + xy^2) \quad \text{for } \boxed{0 \le x \le 3 \ \text{ and } \ 1 \le y \le 2,}$$

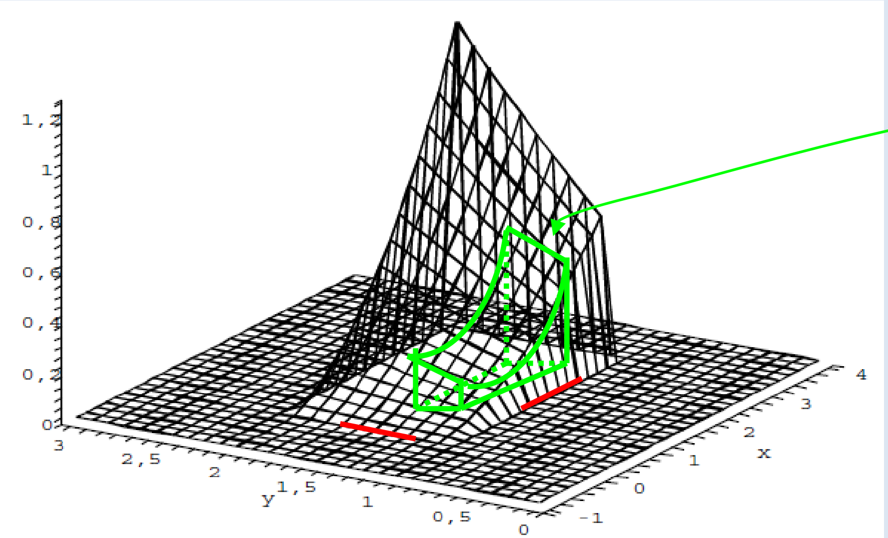


Ο **όγκος** της **περιοχής** κάτω από την συνάρτηση πυκνότητας πιθανότητας ***f(x, y)*** **είναι 1**

Συνεχείς δυσδιάστατες τυχαίες μεταβλητές (συν.)

Παράδειγμα:

$$f(x,y) = \frac{2}{75}\left(2x^2y + xy^2\right) \quad \text{for } 0 \le x \le 3 \text{ and } 1 \le y \le 2,$$



A

$$\begin{aligned} \mathrm{P}\left(1 \le X \le 2, \frac{4}{3} \le Y \le \frac{5}{3}\right) &= \int_1^2 \int_{\frac{4}{3}}^{\frac{5}{3}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \frac{2}{75}\int_1^2 \left(\int_{\frac{4}{3}}^{\frac{5}{3}} (2x^2y + xy^2)\,\mathrm{d}y\right)\mathrm{d}x \\ &= \frac{2}{75}\int_1^2 \left(x^2 + \frac{61}{81}x\right)\mathrm{d}x \;=\; \frac{187}{2025}. \end{aligned}$$

Εύρεση περιθώριας κατανομής (*marginal distribution*)

- Η **περιθώρια κατανομή** μιας **συνεχής** τυχαίας μεταβλητής (π.χ. *X*) προκύπτει **ολοκληρώνοντας την από-κοινού** συνάρτηση πυκνότητας ως προς την άλλη μεταβλητή (π.χ. *Y*):

$$f_X(\boldsymbol{x}) = \int_{-\infty}^{\infty} f_{X,Y}(\boldsymbol{x}, \boldsymbol{y'})d\boldsymbol{y'}$$

και

$$f_Y(\boldsymbol{y}) = \int_{-\infty}^{\infty} f_{X,Y}(\boldsymbol{x'}, \boldsymbol{y})d\boldsymbol{x'}$$

Δεσμευμένη (ή υπό-συνθήκη) κατανομή (*conditional distribution*)

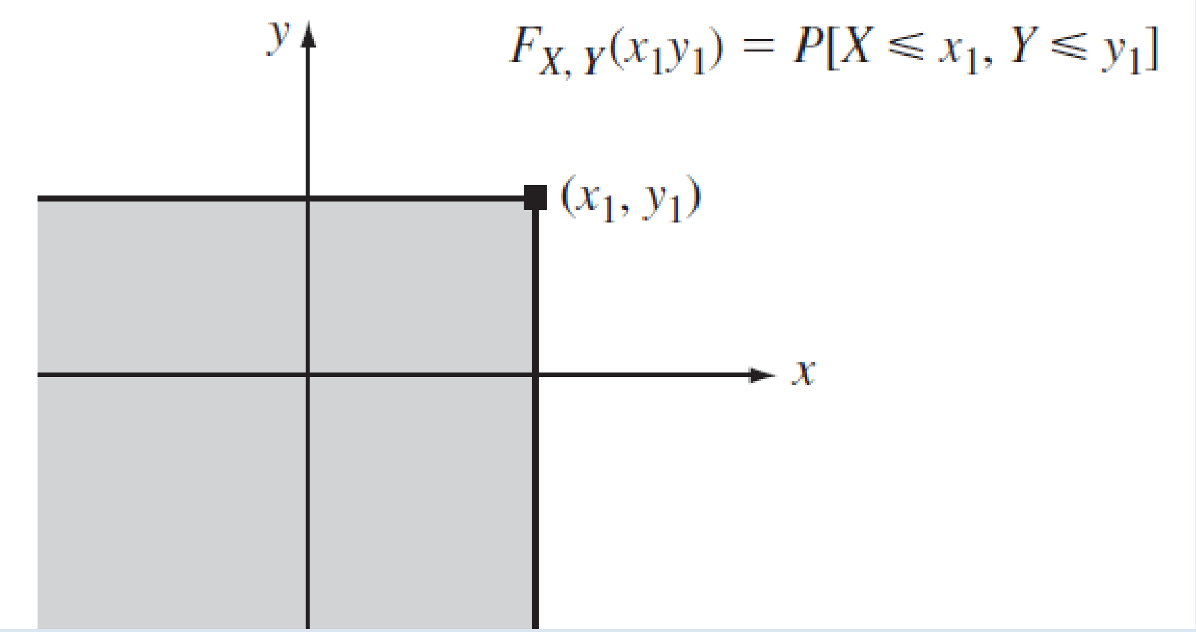
- Έστω 2 τ.μ. X, Y με από-κοινού σ.π.π. $f_{XY}(x, y)$
- Ορίζεται η **δεσμευμένη συνάρτηση πυκνότητας πιθανότητας** μιας εκ των 2 μεταβλητών, θεωρώντας ότι η άλλη έχει **σταθερή τιμή**

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \qquad \kappa\alpha\iota \qquad f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

(*Δεσμευμένη = πηλίκο της από-κοινού προς την περιθώρια*)

Η από-κοινού αθροιστική συνάρτηση κατανομής πιθανότητας $F_{X,Y}$ για δυσδιάστατες τυχαίες μεταβλητές

- Η **αθροιστική συνάρτηση κατανομής πιθανότητας** $F_{X,Y}(x,y)$ εκφράζει την **ποσότητα μάζας πιθανότητας** που περιέχεται στην **ορθογώνια περιοχή** που ορίζεται από το σημείο (x,y)



Η από-κοινού αθροιστική συνάρτηση κατανομής πιθανότητας $F_{X,Y}$ για δυσδιάστατες τυχαίες μεταβλητές

Ορισμός:

Άθροισμα τιμών της συνάρτησης πυκνότητας (πιθανοτήτων) για τις τιμές της ορθογώνιας περιοχής

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = \begin{cases} \displaystyle\sum_{x' \le x} \sum_{y' \le y} f_{X,Y}(x', y') & \text{X, Y διακριτές} \\ \displaystyle\int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y') dx' dy' & \text{X, Y συνεχείς} \end{cases}$$

ολοκληρώνουμε την από-κοινού συνάρτηση πυκνότητας στην ορθογώνια περιοχή

Ιδιότητες της από-κοινού αθροιστικής συνάρτησης κατανομής $F_{X,Y}$

(α) $0 \leq F_{X,Y}(x, y) \leq 1$

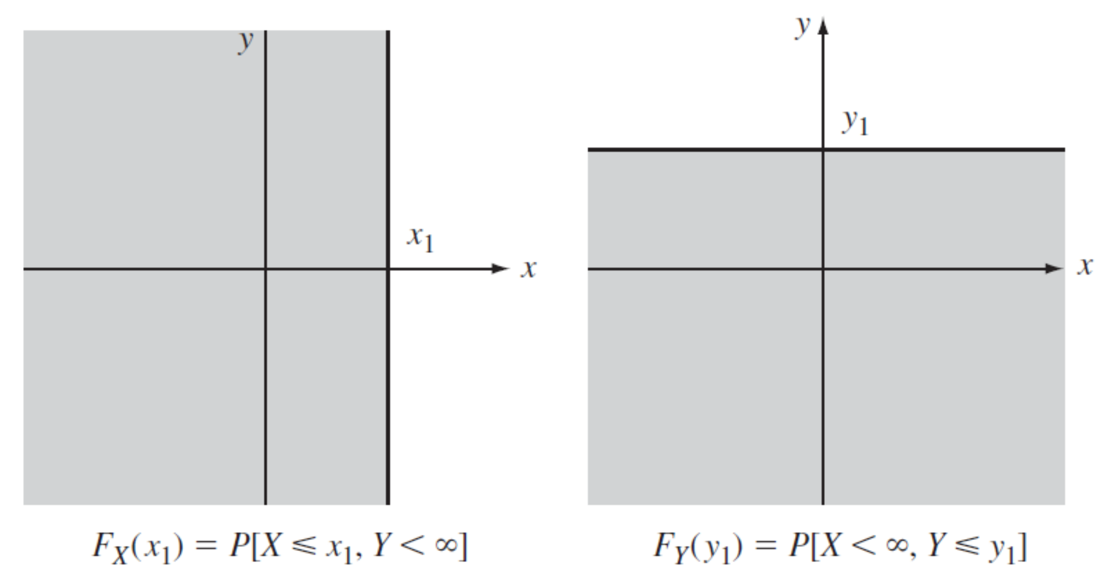
(β) $F_{X,Y}(x, -\infty) = F_{X,Y}(-\infty, y) = 0$

$$F_{X,Y}(\infty, \infty) = 1$$

(γ) $x_1 \leq x_2 \ \wedge \ y_1 \leq x_2 \Rightarrow F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$

δηλ. είναι **μη-φθίνουσα** συνάρτηση (δ) **Περιθώριες** συναρτήσεις κατανομής πιθανότητας

$$F_X(x) = F_{X,Y}(x,\infty) \qquad F_Y(y) = F_{X,Y}(\infty, y)$$



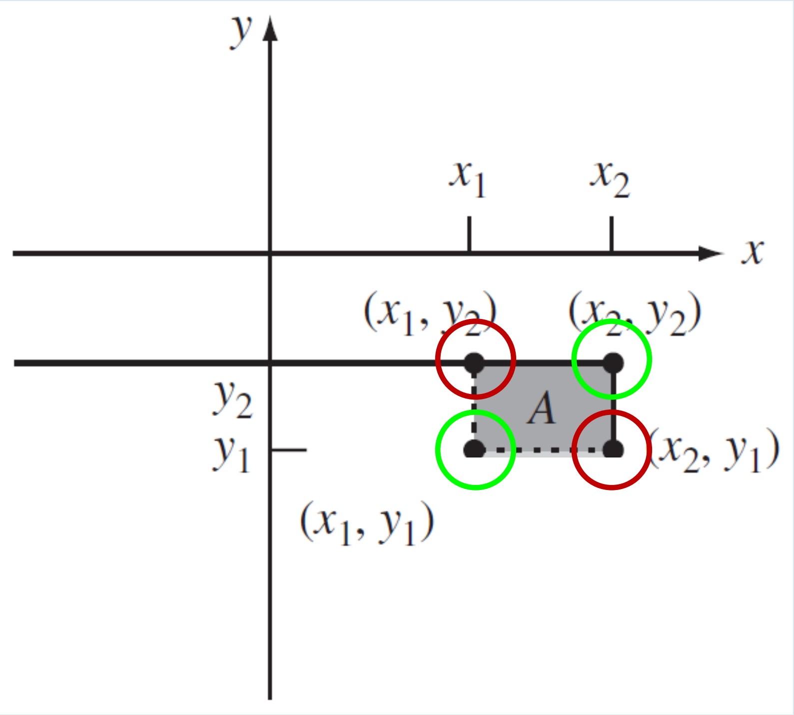
(ε) Συνεχής από δεξιά:

$$F_{X,Y}(x,y) = \lim_{h\to 0} F_{X,Y}(x+h, y) = F_{X,Y}(x^+, y)$$

$$F_{X,Y}(x,y) = \lim_{h\to 0} F_{X,Y}(x, y+h) = F_{X,Y}(x, y^+)$$

(στ) Υπολογισμός της πιθανότητας ενός **ορθογώνιου ενδεχομένου**

$$P(x_1 < X < x_2, y_1 < Y < y_2) = P(A) =$$
$$= F_{X,Y}(x_1, y_1) + F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$$

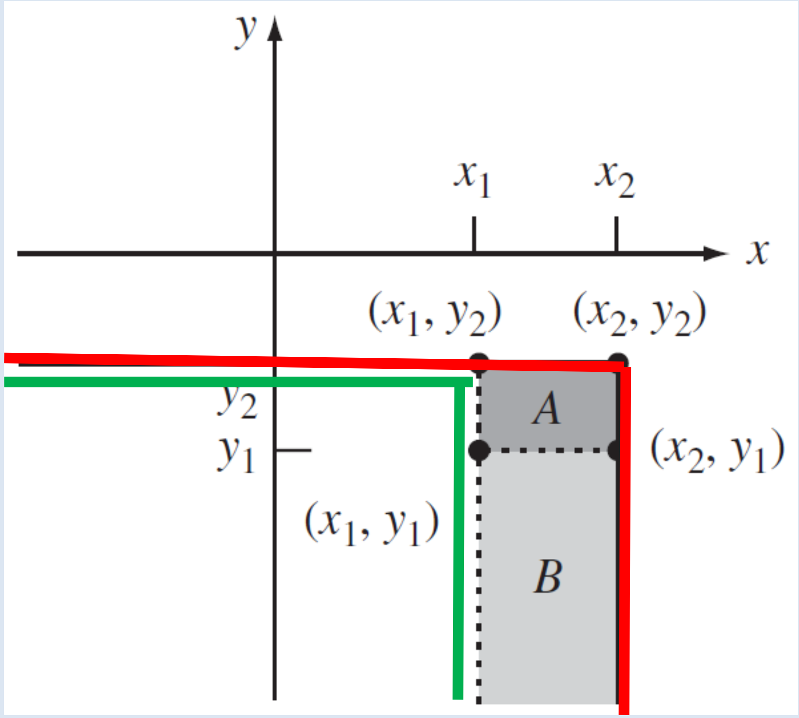


(στ) Υπολογισμός της πιθανότητας ενός **ορθογώνιου ενδεχομένου**

$$P(x_1 < X < x_2, y_1 < Y < y_2) = P(A) =$$
$$= F_{X,Y}(x_1, y_1) + F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$$

Απόδειξη

$$\{X < x_2, Y < y_2\} = A + B + \{X < x_1, Y < y_2\} \Rightarrow$$
$$P(X < x_2, Y < y_2) = P(A) + P(B) + P(X < x_1, Y < y_2) \Rightarrow$$
$$F_{X,Y}(x_2, y_2) = P(A) + P(B) + F_{X,Y}(x_1, y_2) \Rightarrow$$
$$P(A) = \boxed{F_{X,Y}(x_2, y_2)} - P(B) - \boxed{F_{X,Y}(x_1, y_2)} \qquad \textbf{(1)}$$

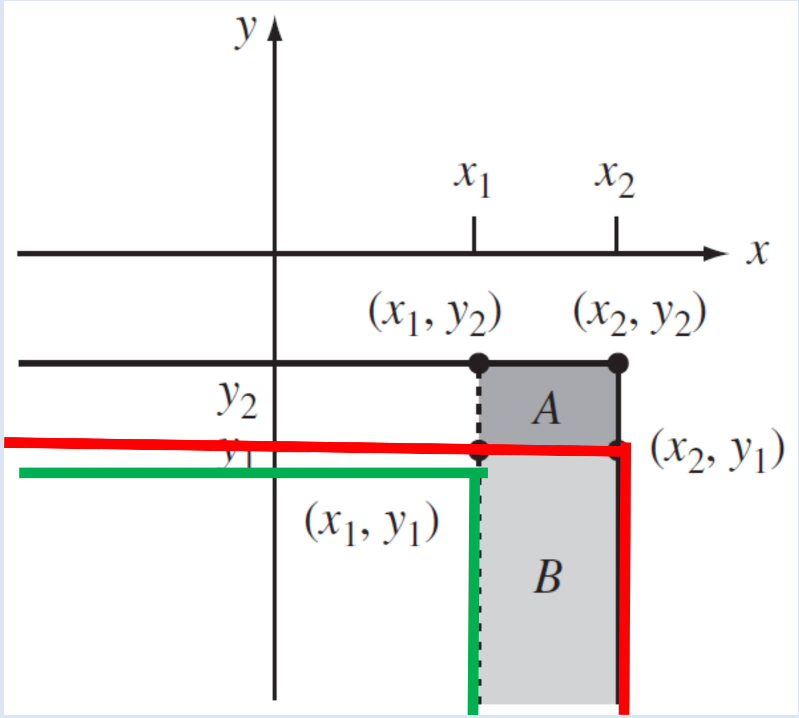


(στ) Υπολογισμός της πιθανότητας ενός **ορθογώνιου ενδεχομένου**

$$P(x_1 < X < x_2, y_1 < Y < y_2) = P(A) =$$
$$= F_{X,Y}(x_1, y_1) + F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$$

Απόδειξη

$$\{X < x_2, Y < y_2\} = A + B + \{X < x_1, Y < y_2\} \Rightarrow$$
$$P(X < x_2, Y < y_2) = P(A) + P(B) + P(X < x_1, Y < y_2) \Rightarrow$$
$$F_{X,Y}(x_2, y_2) = P(A) + P(B) + F_{X,Y}(x_1, y_2) \Rightarrow$$
$$P(A) = F_{X,Y}(x_2, y_2) - P(B) - F_{X,Y}(x_1, y_2) \qquad \textbf{(1)}$$



παρόμοια βρίσκουμε ότι:

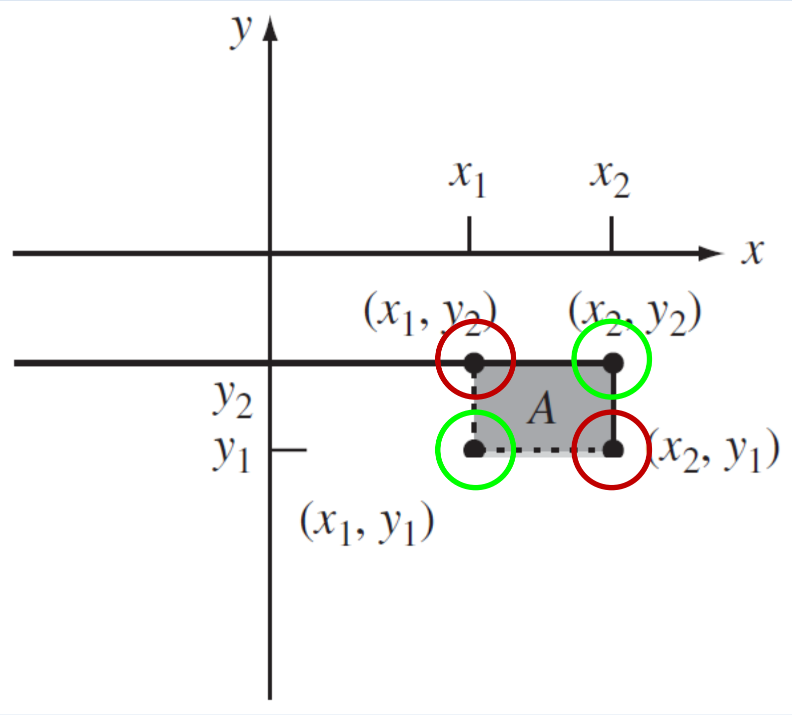
$$\{X < x_2, Y < y_1\} = B + \{X < x_1, Y < y_1\} \Rightarrow$$
$$P(X < x_2, Y < y_1) = P(B) + P(X < x_1, Y < y_1) \Rightarrow$$
$$F_{X,Y}(x_2, y_1) = P(B) + F_{X,Y}(x_1, y_1) \Rightarrow$$
$$P(B) = \boxed{F_{X,Y}(x_2, y_1)} - \boxed{F_{X,Y}(x_1, y_1)} \qquad \textbf{(2)}$$

(στ) Υπολογισμός της πιθανότητας ενός **ορθογώνιου ενδεχομένου**

$$P(x_1 < X < x_2, y_1 < Y < y_2) = P(A) =$$
$$= F_{X,Y}(x_1, y_1) + F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$$

Απόδειξη

$$\{X < x_2, Y < y_2\} = A + B + \{X < x_1, Y < y_2\} \Rightarrow$$
$$P(X < x_2, Y < y_2) = P(A) + P(B) + P(X < x_1, Y < y_2) \Rightarrow$$
$$F_{X,Y}(x_2, y_2) = P(A) + P(B) + F_{X,Y}(x_1, y_2) \Rightarrow$$
$$P(A) = F_{X,Y}(x_2, y_2) - P(B) - F_{X,Y}(x_1, y_2) \qquad \textbf{(1)}$$



παρόμοια βρίσκουμε ότι:

$$\{X < x_2, Y < y_1\} = B + \{X < x_1, Y < y_1\} \Rightarrow$$
$$P(X < x_2, Y < y_1) = P(B) + P(X < x_1, Y < y_1) \Rightarrow$$
$$F_{X,Y}(x_2, y_1) = P(B) + F_{X,Y}(x_1, y_1) \Rightarrow$$
$$P(B) = F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_1) \qquad \textbf{(2)}$$

Έτσι, **συνδυάζοντας τις (1), (2) σχέσεις** προκύπτει η παραπάνω σχέση.

- **Η από-κοινού** συνάρτηση κατανομής πιθανότητας στο **συνεχή χώρο** είναι παντού συνεχής

Ισχύει:

$$F_{X,Y}(a,b)=P(X\le a,Y\le b)=\int_{-\infty}^{a}\int_{-\infty}^{b}f_{X,Y}(x,y)dxdy$$

$$f_{X,Y}(x,y)=\frac{dF_{X,Y}(x,y)}{dxdy}$$

Ανεξαρτησία τυχαίων μεταβλητών

- Η γνωστή **έννοια της ανεξαρτησίας** των ενδεχομένων **επεκτείνεται** στην περίπτωση πολλών μεταβλητών.
- Υπενθύμιση:

Δύο ενδεχόμενα A, B είναι **ανεξάρτητα** αν:

$$P(A|B) = P(A)$$

ή

$$P(A \cap B) = P(A,B) = P(A)P(B)$$

Ανεξαρτησία τυχαίων μεταβλητών

- Παρόμοια, 2 τυχαίες μεταβλητές *X* και *Y* είναι ανεξάρτητες εάν ισχύει:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

A ∩ B ***A*** ***B***

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

Ανεξαρτησία τυχαίων μεταβλητών

Ισοδύναμες σχέσεις

$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

- **ή ισοδύναμα** $F_{XY}(x, y) = F_X(x)F_Y(y)$
- **ή ισοδύναμα** (***παραγωγίζοντας***): $f_{XY}(x, y) = f_X(x)f_Y(y)$
- **ή ισοδύναμα** η **δεσμευμένη ισούται με την περιθώρια**

$$f_{X|Y}(x \mid y) = f_X(x) \qquad f_{Y|X}(y \mid x) = f_Y(y)$$

Ανεξαρτησία τυχαίων μεταβλητών

2 μεταβλητές είναι ανεξάρτητες όταν ισχύει

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

Η από-κοινού γράφεται ως γινόμενο των περιθωρίων

Ένα παράδειγμα

- Εστω η από-κοινού συνάρτηση πυκνότητας

$$f_{X,Y}(x,y) = abe^{-(ax+by)} \quad , \quad x,y>0 \quad (a,b>0)$$

Να εξεταστούν οι 2 μεταβλητές ως προς την ανεξαρτησία

Λύση

1. Βρίσκουμε τις περιθώριες συναρτήσεις πυκνότητας

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy=\int_{0}^{+\infty} abe^{-(ax+by)}dy=ae^{-ax}\int_{0}^{+\infty} be^{-by}dy=ae^{-ax}$$

$$f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)dx=\int_{0}^{+\infty} abe^{-(ax+by)}dx=be^{-by}\int_{0}^{+\infty} ae^{-ax}dx=be^{-by}$$

2. Ελέγχουμε για ανεξαρτησία

$$f_{X,Y}(x,y)=abe^{-(ax+by)} = f_X(x)f_Y(y)$$

3. **Συμπεραίνουμε** ότι οι δύο μεταβλητές είναι **ανεξάρτητες**

- **<u>Θεώρημα:</u>** Αν η από-κοινού συνάρτηση πυκνότητας γράφεται ως γινόμενο δύο συναρτήσεων μία ως προς x και η άλλη ως προς y, δηλ. $f_{X,Y}(x,y) = g(x)h(y)$

  τότε **αυτομάτως** οι δύο μεταβλητές *X*, *Y* είναι **ανεξάρτητες**

<u>Απόδειξη</u>

Βρίσκουμε τις περιθώριες συναρτήσεις πυκνότητας:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)\int_{-\infty}^{\infty} h(y)dy = g(x)C_y$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)\int_{-\infty}^{\infty} g(x)dx = h(y)C_x$$

$$f_X(x)f_Y(y) = g(x)h(y)\boxed{C_xC_y}$$

αλλά $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)dxdy = \int_{-\infty}^{\infty} g(x)dx\int_{-\infty}^{\infty} h(y)dy = \boxed{C_yC_x = 1}$$

άρα ανεξάρτητες μεταβλητές καθώς: $f_{X,Y}(x,y) = g(x)h(y) = f_X(x)f_Y(y)$

Παραδείγματα

- Δίνεται η από-κοινού συνάρτηση πυκνότητας 2 διακριτών τυχαίων μεταβλητών

$$f_{X,Y}(x,y) = \begin{array}{ccc} x \backslash y & 1 & 2 \\ 1 & 1/8 & 1/4 \\ 2 & 1/8 & 1/2 \end{array}$$

α) Να βρεθούν οι περιθώριες κατανομές των δύο μεταβλητών, *X* και *Y*,

β) Να υπολογίσετε τις μέσες τιμές και τις διακυμάνσεις τους.

- Δύο τυχαίες μεταβλητές *X*, *Y* έχουν από-κοινού συνάρτηση πυκνότητας:

$$f_{X,Y}(x,y) = \begin{cases} xye^{-\frac{x^2+y^2}{2}} & x, y > 0 \\ 0 & \alpha\lambda\lambda o\upsilon \end{cases}$$

Να βρεθεί η από-κοινού συνάρτηση κατανομής $F_{X,Y}(x,y)$ και οι περιθώριες κατανομές των *X* και *Y*.

- Δύο τυχαίες μεταβλητές *X*, *Y* έχουν από-κοινού συνάρτηση πυκνότητας:

$$f_{X,Y}(x,y) = \begin{cases} 15y & x^2 \le y \le x \\ 0 & \alpha\lambda\lambda o\upsilon \end{cases}$$

Να βρεθούν οι περιθώριες κατανομές των *X* και *Y*, και οι μέσες τιμές τους.

- Έστω 2 μεταβλητές X,Y με από-κοινού συνάρτηση πυκνότητας

$$f_{X,Y}(x,y) = c\,x\,y\,(1-x) \quad , \quad 0 < x,y < 1$$

α) Να βρείτε την **σταθερά c**

β) Να βρείτε τις **περιθώριες** κατανομές των *X*, *Y*

γ) Να βρείτε τις **δεσμευμένες** κατανομές *X*|*Y* και *Y*|*X*

δ) Υπολογίστε τις παρακάτω **πιθανότητες** :

δ1) $P(Y<1/2 \cap X>1/2)$

δ2) $P(Y<1/2 \mid X>1/2)$

- Να βρεθούν οι περιθώριες κατανομές μιας δυσδιάστατης τυχαίας μεταβλητής με από-κοινού συνάρτηση πυκνότητας πιθανότητας:

$$f_{X,Y}(x,y) = \begin{cases} ce^{-x}e^{-y} & 0 \leq y \leq x < \infty \\ 0 & \alpha\lambda\lambda o\upsilon \end{cases}$$

- Εστω 2 μεταβλητές X,Y με από-κοινού συνάρτηση πυκνότητας

$$f_{XY}(x,y) = \begin{cases} kx^2y^{-3} & 1 \leq x, y \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

(α) Να εξεταστεί εάν οι 2 μεταβλητές είναι ανεξάρτητες.

(β) Να βρεθούν οι περιθώριες κατανομές.

(γ) Να υπολογιστεί η πιθανότητα P(X>Y).