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

# Συνεχείς διδιάστατες τ.μ.

- Περιθωριακές συναρτήσεις κατανομής:

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

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

- Από κοινού σ.κ. των X και Y:

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

# Πυκνότητα πιθανότητας

- Από κοινού πυκνότητα πιθανότητας:

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\,\partial y}$$

- Ιδιότητες:

f(x, y) συνεχής σχεδόν παντού

f(x, y) $\geq$ 0 για κάθε x, y

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dy\,dx = 1$$

$$F(x_0,y_0) = \int_{-\infty}^{x_0}\int_{-\infty}^{y_0} f(x,y)\,dy\,dx$$

# Συνεχείς διδιάστατες τ.μ.

- Υπολογισμός πιθανοτήτων από την από κοινού π.π.

$$P\left(X \le x_0, Y \le y_0\right) = F(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f\left(x, y\right) dy\, dx$$

$$P\left(a \le X \le b, c \le Y \le d\right) = \int_a^b \int_c^d f\left(x, y\right) dy\, dx$$

$P\left(X, Y \in C\right) = \iint_C f\left(x, y\right) dy\, dx$ όπου C μια περιοχή του επιπέδου

# Παράδειγμα

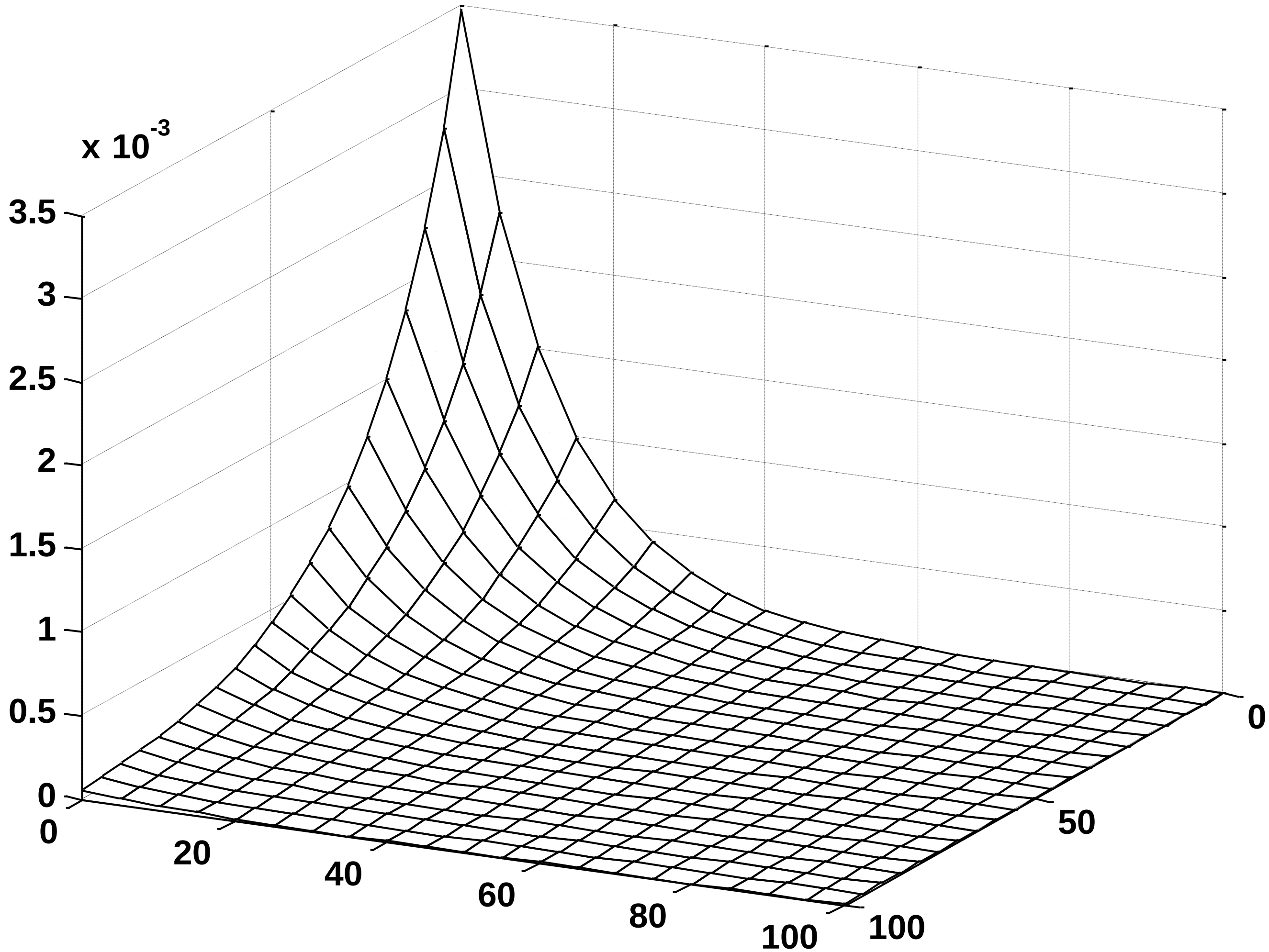
- Από κοινού σ.κ. :

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

$$\frac{\partial}{\partial x}: \quad \frac{1}{24}e^{-\frac{x}{24}} - \frac{1}{24}e^{-\frac{x}{24}-\frac{y}{12}}$$

$$\frac{\partial}{\partial y}: \quad \frac{1}{288}e^{-\frac{x}{24}-\frac{y}{12}} = f(x,y) \qquad (\text{για } x \geq 0,\ y \geq 0)$$

# Παράδειγμα: γραφική παράσταση της π.π.



# Παράδειγμα

- Περιεκτικότητα ενός δείγματος μεταλλεύματος σε Zn, Fe:

$$f(x,y) = 0.0975 - 0.34(x-1)^2 - 10^{-4}(y-25)^2$$

για $0.5 \leq x \leq 1.5$, $20 \leq y \leq 35$ (και 0 αλλού)

---

$$\int_{x=0.5}^{1.5} \int_{y=20}^{35} f(x,y)\,dy\,dx = \int_{y=20}^{35} \left[0.692 - 10^{-4}(y-25)^2\right] dy = 1$$

$$P(0.8 \leq Zn \leq 1.0 \;\&\; 25 \leq Fe \leq 30) = \int_{x=0.8}^{1.0} \int_{y=25}^{30} f(x,y)\,dy\,dx = 0.092$$

# Περιθωριακές (marginal) κατανομές

- Περιθωριακή σ.κ.

$$F_X(x) = F(x,+\infty) = \int_{s=-\infty}^{x} \int_{t=-\infty}^{+\infty} f(s,t)\,dt\,ds$$

- Περιθωριακή π.π.

$$f_X(x) = \frac{dF_X(x)}{dx} = \int_{t=-\infty}^{+\infty} f(x,t)\,dt$$

- X, Y ανεξάρτητες: $F(x, y) = F_X(x)F_Y(y) \Rightarrow$

$$f(x,y) = \frac{\partial^2}{\partial x\,\partial y}\left[F_X(x)F_Y(y)\right] = \frac{dF_X}{dx}\frac{dF_Y}{dy} = f_X(x)f_Y(y)$$

# Παραδείγματα

- Από κοινού π.π.:

$$f(x,y) = \frac{1}{288} e^{-\frac{x}{24}-\frac{y}{12}} \quad \text{για } x \geq 0,\ y \geq 0$$

$$f_X(x) = \int_0^{+\infty} f(x,y)dy = \frac{1}{288} e^{-\frac{x}{24}} \int_0^{+\infty} e^{-\frac{y}{12}} dy = \frac{1}{24} e^{-\frac{x}{24}} \quad \text{για } x \geq 0$$

$$f_Y(y) = \int_0^{+\infty} f(x,y)dx = \frac{1}{12} e^{-\frac{y}{12}} \quad \text{για } y \geq 0$$

- Οι X, Y είναι ανεξάρτητες.

# Παραδείγματα

- Στο παράδειγμα του μεταλλεύματος:

$$f_X(x) = \int_{20}^{35} \left[0.0975 - 0.34(x-1)^2 - 10^{-4}(y-25)^2\right] dy$$

$$= 1.425 - 5.1(x-1)^2 \quad , \qquad 0.5 \le x \le 1.5$$

E(X) = 1, Var(X) = 0.055, σ = 0.23

$$f_Y(y) = \int_{0.5}^{1.5} f(x,y)\,dx = 0.0692 - 10^{-4}(y-25)^2, \quad 20 \le y \le 35$$

E(Y) = 27.36, Var(Y) = 18.23, σ = 4.27

# Παραδείγματα

$$f(x,y) = xe^{-x(y+1)} \text{ για } x \geq 0,\ y \geq 0$$

$$f_X(x) = \int_0^{+\infty} xe^{-x(y+1)}\,dy = e^{-x}\int_0^{+\infty} xe^{-xy}\,dy = e^{-x}\left[-e^{-xy}\right]_{y=0}^{+\infty} = e^{-x}$$

για $x \geq 0$

$$f_Y(y) = \int_0^{+\infty} xe^{-x(y+1)}\,dx = \left[-\frac{x}{y+1}e^{-x(y+1)}\right]_{x=0}^{+\infty} + \frac{1}{y+1}\int_0^{+\infty} e^{-x(y+1)}\,dx$$

$$= \frac{1}{y+1}\left[-\frac{1}{y+1}e^{-x(y+1)}\right]_{x=0}^{+\infty} = \frac{1}{(y+1)^2} \quad \text{για } y \geq 0$$

# Παραδείγματα

$$f(x,y) = 2e^{-x-y} \text{ για } x \geq 0,\ y \geq 0,\ y \geq x$$

$$f_X(x) = \int_x^{+\infty} 2e^{-x-y}\,dy = 2e^{-x}\int_x^{+\infty} e^{-y}\,dy = 2e^{-x}\left[-e^{-y}\right]_{y=x}^{+\infty}$$

$$= 2e^{-2x}, \quad x \geq 0$$

$$f_Y(y) = \int_0^{y} 2e^{-x-y}\,dx = 2e^{-y}\int_0^{y} e^{-x}\,dx = 2e^{-y}\left(1-e^{-y}\right), \qquad y \geq 0$$

# Υπό συνθήκη κατανομές

- Αν οι X, Y είναι συνεχείς τ.μ. με από κοινού π.π. f(x, y) και περιθωριακές π.π. $f_X(x)$ και $f_Y(y)$, η υπό συνθήκη π.π. της X δεδομένου ότι Y = y είναι:

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

- Για δεδομένο y η η υπό συνθήκη π.π. $f_{X|Y=y}(x)$ είναι έγκυρη π.π. (ολοκλήρωμα = 1) και μπορεί να χρησιμοποιηθεί για υπολογισμό πιθανοτήτων:

$$P(a \leq X \leq b \mid Y = y) = \int_a^b f_{X|Y=y}(x)\,dx = \frac{\int_a^b f(x,y)\,dx}{f_Y(y)}$$

# Υπό συνθήκη μέσες τιμές

- Αντίστοιχα με τις διακριτές τ.μ.:

$$E\{X \mid Y = y\} = \int_{-\infty}^{+\infty} x\, f_{X|Y=y}(y)\,dy = \frac{\int_{-\infty}^{+\infty} x\, f(x,y)\,dy}{f_X(x)}$$

$$E\{g(X) \mid Y = y\} = \int_{-\infty}^{+\infty} g(x) f_{X|Y=y}(y)\,dy$$

# Παραδείγματα

- Στο παράδειγμα του μεταλλεύματος, αν ξέρουμε ότι η περιεκτικότητα σε Zn είναι X = 0.55, πόσο αλλάζει η γνώση μας για την περιεκτικότητα σε Fe;

$$f_{Y|X=0.55}(y) = \frac{f(0.55, y)}{f_X(0.55)} = 0.073 - 2.55 \times 10^{-4}(y-25)^2$$

- Η a priori π.π. ήταν:

$$f_Y(y) = 0.0692 - 10^{-4}(y-25)^2$$

- A priori: E{Y} = 27.36, $\sigma_Y$ = 4.27
- A posteriori: E{Y|X=0.55} = 27.14, $\sigma_{Y|X=0.55}$ = 4.14

# Παραδείγματα

- Από κοινού π.π.:

$$f(x,y) = \frac{1}{288} e^{-\frac{x}{24}-\frac{y}{12}} \quad \text{για } x \geq 0,\ y \geq 0$$

- Αν γνωρίζουμε ακριβώς την τιμή που παίρνει η τ.μ. Y, τί μπορούμε να πούμε για την X;

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{1/288\, e^{-x/24-y/12}}{1/12\, e^{-y/12}} = \frac{1}{24} e^{-x/24} = f_X(x)$$

- Επομένως η γνώση της τιμής της Y δεν μας δίνει καμμία πληροφορία για την X (αναμενόμενο, αφού είναι ανεξάρτητες).

# Ανεξάρτητες τ.μ.

Οι παρακάτω προτάσεις είναι ισοδύναμες:

- Οι τ.μ. X, Y είναι ανεξάρτητες
- $F(x, y) = F_X(x)\, F_Y(y)$ για κάθε x, y
- $f(x, y) = f_X(x)\, f_Y(y)$ για κάθε x, y
- $f_{X \mid Y = y}(x, y) = f_X(x)$ για κάθε x, y
- $f_{Y \mid X = x}(x, y) = f_Y(y)$ για κάθε x, y

και αντίστοιχα αν οι X, Y είναι διακριτές.

# Συναρτήσεις δύο τ.μ.

- X, Y συνεχείς τ.μ. με από κοινού π.π. $f_{XY}(x, y)$
- Z, W τ.μ. που προκύπτουν από τις X, Y μέσω του μετασχηματισμού $Z = g_1(X, Y)$, $W = g_2(X, Y)$.
- Αν ο μετασχηματισμός είναι 1 προς 1, υπάρχει ο αντίστροφος $X = h_1(W, Z)$, $Y = h_2(W, Z)$.
- Η από κοινού π.π. των W, Z είναι:

$$f_{ZW}(z, w) = f_{XY}\left(h_1(z, w), h_2(z, w)\right) \times |J|$$

όπου $J = \begin{vmatrix} \dfrac{\partial h_1}{\partial z} & \dfrac{\partial h_1}{\partial w} \\ \dfrac{\partial h_2}{\partial z} & \dfrac{\partial h_2}{\partial w} \end{vmatrix}$ (Ιακωβιανή ορίζουσα)

# Ειδική περίπτωση: άθροισμα τ.μ.

- Αν ξέρουμε την από κοινού π.π. $f_{XY}(x, y)$ των X, Y, ποιά είναι η π.π. της X+Y;

  Θέτουμε Z = X+Y, W = X. Ο αντίστροφος μετασχηματισμός είναι $X = h_1(W, Z) = W$, $Y = h_2(W, Z) = Z - W$. Η Ιακωβιανή είναι $J = -1$ ($|J| = 1$). Η από κοινού π.π. των Z, W είναι:

$$f_{ZW}(z,w) = f_{XY}(w, z-w)$$

  Η περιθωριακή π.π. της Z είναι: $f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(w, z-w)\,dw$

  Αν X, Y ανεξάρτητες: $f_Z(z) = \int_{-\infty}^{+\infty} f_X(w) f_Y(z-w)\,dw$

# Παράδειγμα

- X, Y ανεξάρτητες, ομοιόμορφες στο [0,1]:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(w) f_X(z-w)\,dw$$

όπου $f_X(x) = 1$ για $0 \le x \le 1$. Επομένως τα όρια ολοκλήρωσης είναι:

$$\left.\begin{array}{c} 0 \le w \le 1 \\ 0 \le z - w \le 1 \end{array}\right\} \Rightarrow \left.\begin{array}{c} 0 \le w \le 1 \\ z - 1 \le w \le z \end{array}\right\} \Rightarrow \max(0, z-1) \le w \le \min(1, z)$$

# Παράδειγμα

Η π.π. της Ζ είναι:

$$f_Z(z) = \int_{\max(0,z-1)}^{\min(1,z)} dw = \begin{cases} 0 & z < 0 \\ \int_0^z dw = z & 0 \le z \le 1 \\ \int_{z-1}^1 dw = 2 - z & 1 \le z \le 2 \\ 0 & z > 2 \end{cases}$$

δηλ. η Ζ = Χ + Υ είναι τριγωνική στο [0, 2].

# Συνδιασπορά δύο τ.μ.

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

- 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

# Ασυσχέτιστες και ανεξάρτητες τ.μ.

- Αν οι X, Y είναι ανεξάρτητες, τότε Cov{X, Y} = 0 και $\rho_{XY} = 0$
- Αν $\rho_{XY} = 0$ (X, Y ασυσχέτιστες), οι X, Y δεν είναι απαραίτητα ανεξάρτητες (μπορεί να υπάρχει κάποια μη γραμμική εξάρτηση)
- Ειδικότερα αν $\rho_{XY} = 0$ και οι X, Y έχουν από κοινού κανονική π.π., είναι και ανεξάρτητες.

# Παράδειγμα

- Συνδιασπορά των X, Y με $f(x, y) = 6xy^2$ για $x, y \in [0,1]$

$$f_X(x) = \int_0^1 6xy^2 dy = 2x, \quad 0 \le x \le 1$$

$$f_Y(y) = \int_0^1 6xy^2 dx = 3y^2, \ 0 \le y \le 1$$

$$E\{X\} = \int_0^1 x f_X(x) dx = \int_0^1 2x^2 dx = 2/3$$

$$E\{Y\} = \int_0^1 y f_Y(y) dy = \int_0^1 3y^3 dy = 3/4$$

$$E\{XY\} = \int_0^1 \int_0^1 xyf(x,y) dy\, dx = \int_0^1 \int_0^1 6x^2y^3 dy\, dx = \int_0^1 2y^3 dy = 1/2$$

$$\mathrm{Cov}\{X, Y\} = E\{XY\} - E\{X\}\, E\{Y\} = 0$$

# Μέση τιμή και διασπορά: ανακεφαλαίωση

- $E\{aX + b\} = aE\{X\} + b$
- $Var\{aX + b\} = a^2 Var\{X\}$
- $E\{aX + bY\} = aE\{X\} + bE\{Y\}$
- $Var\{aX + bY\} = a^2 Var\{X\} + b^2 Var\{Y\} + 2ab Cov\{X, Y\}$
- X, Y ανεξάρτητες: $E\{XY\} = E\{X\} E\{Y\}$
- X, Y ανεξάρτητες: $Var\{aX + bY\} = a^2 Var\{X\} + b^2 Var\{Y\}$

# Μέση τιμή και διασπορά αθροίσματος τ.μ.

- Αν οι $X_1, ..., X_n$ είναι ανεξάρτητες τ.μ. με την ίδια μέση τιμή μ και διασπορά $\sigma^2$ και $Y = X_1 + ... + X_n$ :

  $E\{Y\} = n\mu$

  $Var\{Y\} = n\sigma^2$

- Αν $Z = (X_1 + ... + X_n) / n$ :

  $E\{Z\} = \mu$

  $Var\{Z\} = \sigma^2 / n$

- Αθροισμα (ή διαφορά) ανεξάρτητων τ.μ. αυξάνει τη διασπορά, ενώ μέσος όρος τη μειώνει.

# Παράδειγμα

- Αν οι $X_i$ είναι ανεξάρτητες Bernoulli με παράμετρο p (π.χ. ρίψη νομίσματος), $E\{X_i\} = p$, $Var\{X_i\} = pq$.
- $Y = X_1 + ... + X_n$ (η Y είναι διωνυμική):

  $E\{Y\} = np$, $Var\{Y\} = npq$
- $Z = (X_1 + ... + X_n) / n$ :

  $E\{Z\} = p$, $Var\{Z\} = pq / n$
- Η $E\{Z\}$ μπορεί να χρησιμοποιηθεί για να εκτιμήσουμε την παράμετρο p, αν δεν την γνωρίζουμε. Το σφάλμα της εκτίμησης $Var\{Z\} \to 0$ όταν $n \to \infty$.