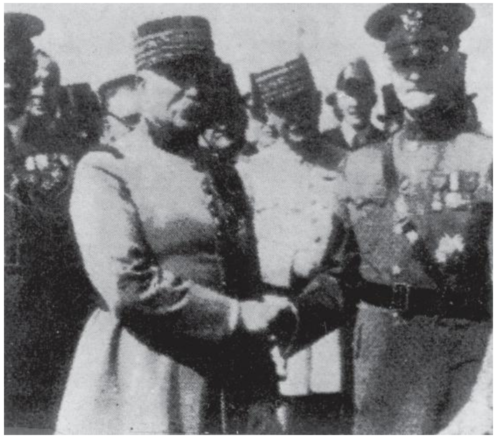
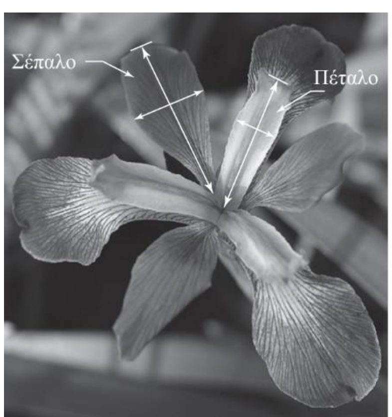


ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

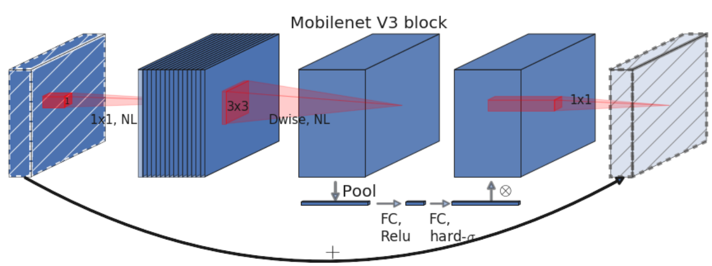
Διαφάνειες διαλέξεων

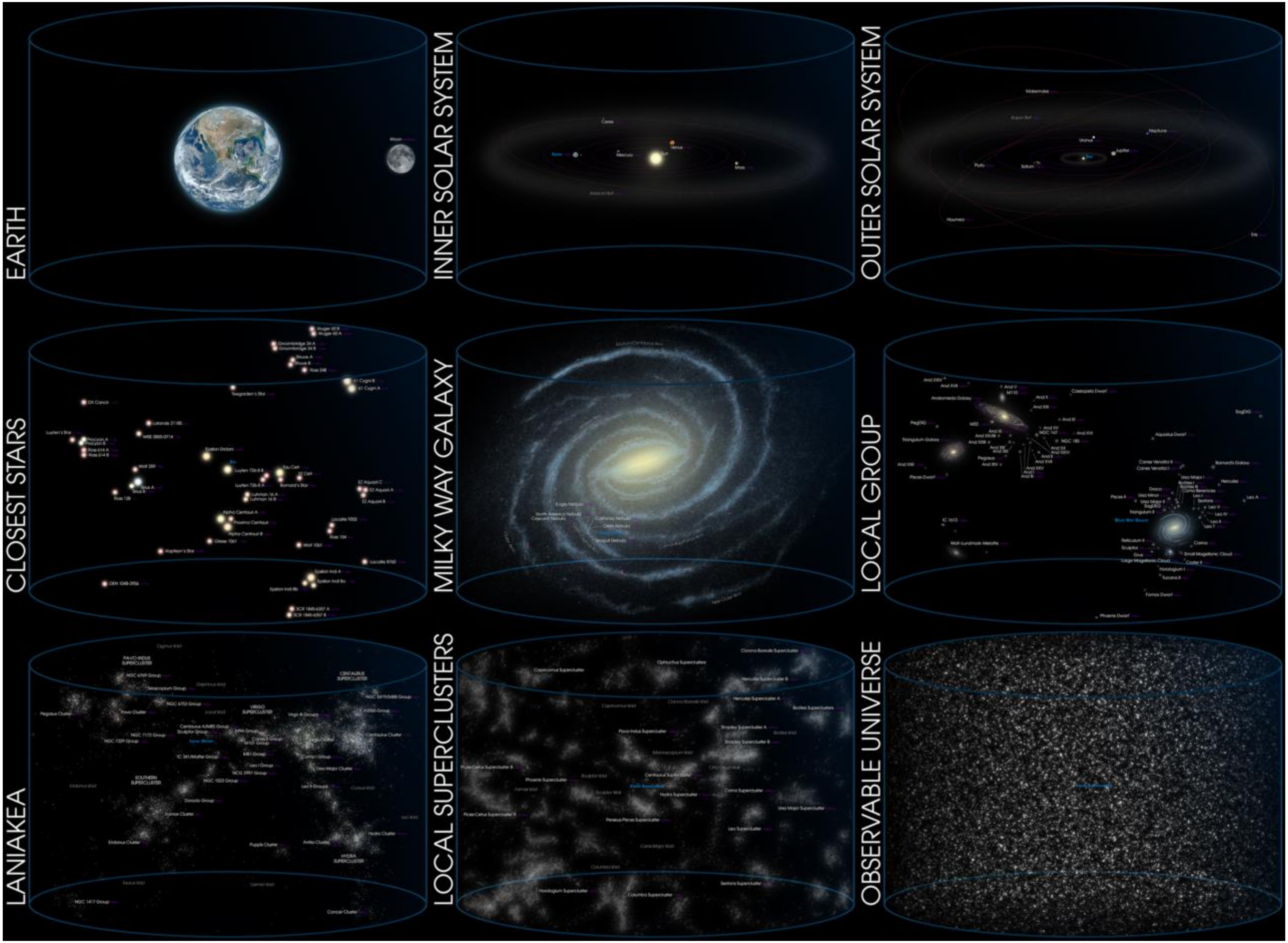
Υπολογιστική Νοημοσύνη
Βαθιά μάθηση



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

x_1 = πλάτος πετάλου
x_2 = μήκος πετάλου
x_3 = πλάτος σεπάλου
x_4 = μήκος σεπάλου




EARTH
INNER SOLAR SYSTEM
OUTER SOLAR SYSTEM
CLOSEST STARS
MILKY WAY GALAXY
LOCAL GROUP
LANIAKEA
LOCAL SUPERCLUSTERS
OBSERVABLE UNIVERSE

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ &
ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ

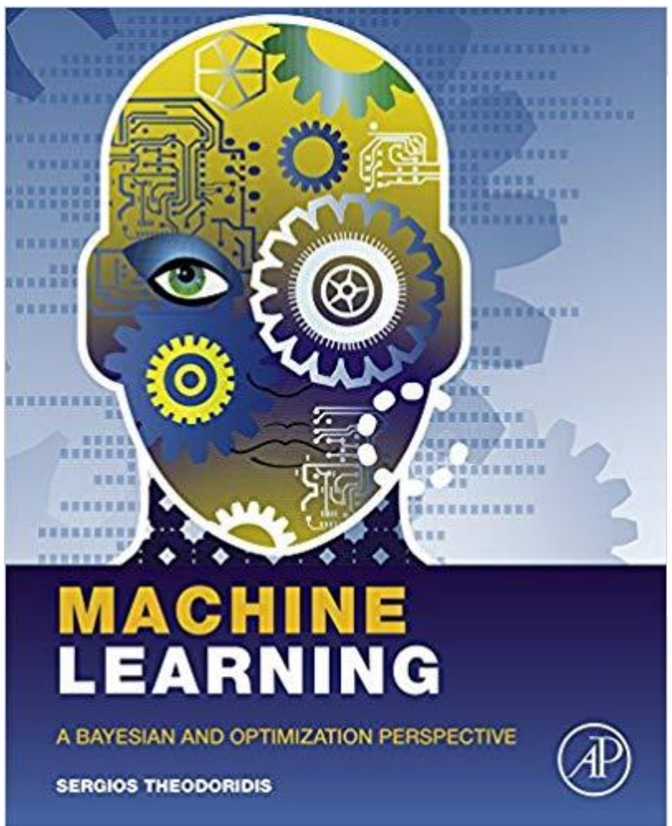
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

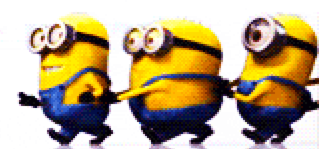
Στοιχεία Μηχανικής Μάθησης

Βασικά στοιχεία πιθανοτήτων

Πηγές

- Sergios Theodoridis: Machine Learning A Bayesian and Optimization Perspective, Academic Press, 2015





Η έννοια της τυχαίας μεταβλητής – ΤΜ (Random Variable-RV)

- Μια τυχαία μεταβλητή (ΤΜ) x, είναι μια μεταβλητή τη οποίας οι μεταβολές οφείλονται στην τύχη/τυχαιότητα. Μια ΤΜ μπορεί να θεωρηθεί ως μια συνάρτηση η οποία αναθέτει μια τιμή στο αποτέλεσμα ενός πειράματος. Για παράδειγμα, σε ένα πείραμα ρίψης νομίσματος η ΤΜ *x*, μπορεί να λάβει τις τιμές $x_1 = 0$ αν το αποτέλεσμα του πειράματος είναι *κεφάλι* και $x_2 = 1$ αν είναι *γράμματα*.
- Μια ΤΜ θα δηλώνεται ως x, ενώ οι τιμές τις οποίες λαμβάνει μετά την εκτέλεση ενός πειράματος θα δηλώνονται ως x (*mathmode italics*).
- Μια ΤΜ περιγράφεται με όρους ενός συνόλου από πιθανότητες αν οι τιμές τις είναι διακριτές ή με συναρτήσεις πυκνότητας πιθανότητας - probability density function (pdf) -αν οι τιμές λαμβάνονται οπουδήποτε σε ένα διάστημα.

Πιθανότητα (Ορισμοί)

- Σχετική συχνότητα (Relative Frequency): Η πιθανότητα P(A), ενός γεγονότος, A, είναι το όριο:

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n},$$

όπου: *n* είναι ο συνολικός αριθμός πειραμάτων και n_A είναι ο αριθμός των φορών που συνέβη το γεγονός A.

- Στην πράξη αν το *n* είναι μεγάλο: $P(A) \approx \frac{n_A}{n},$
  - Προσοχή στο τι σημαίνει (*n μεγάλο*) ειδικά αν η πιθανότητα P(A) είναι μικρή.
- Φυσικό περιεχόμενο: Η πιθανότητα είναι ένα μέτρο της αβεβαιότητας σε σχέση με ένα γεγονός.

Πιθανότητα (Ορισμοί)

- Αξιωματικός ορισμός:
  - 1933 Andrey Kolmogorov.
  - Θεωρία συνόλων και θεωρία μετρήσεων.
  - Με τα ακόλουθα 3 αξιώματα όλα τα άλλα είναι εύκολα…
- Η πιθανότητα ενός γεγονότων A, *P(A)* είναι ένας μη αρνητικός αριθμός ($P(A) \geq 0$).
- Η πιθανότατα ενός γεγονότος C, που είναι βέβαιο ότι θα συμβεί είναι ίση με 1: $P(C) = 1$.
- Αν δύο γεγονότα A, B είναι αμοιβαία αποκλειόμενα τότε η πιθανότητα του να συμβεί ή το ένα ή το άλλο ($A \cup B$) δίνεται από:

$$P(A \cup B) = P(A) + P(B)$$

Διακριτές τυχαίες μεταβλητές Discrete Random Variables

- Μια ΔΤΜ x, μπορεί να πάρει οποιαδήποτε τιμή από ένα πεπερασμένο σύνολο τιμών Χ.
  - Η πιθανότητα του γεγονότος "x = x « συμβολίζεται ως: P(x = x) ή P(x).
- Υποθέτοντας ότι δεν μπορούν να συμβούν 2 γεγονότα ταυτόχρονα στο Χ και ότι το πείραμα πάντα επιστρέφει μια τιμή: $\sum_{x \in \mathcal{X}} P(x) = 1,$
  - Ο χώρος Χ ονομάζεται «ο δειγματικός χώρος- sample space»

Διακριτές τυχαίες μεταβλητές
Discrete Random Variables

- Από κοινού πιθανότητα (Joint propability): Η κοινή πιθανότητα δύο γεγονότων A και B να συμβούν από κοινού, $P(A,B)$.
- Έστω δύο ΤΜ $x \in X,\ y \in Y$.

Τότε ο ακόλουθος κανόνας αθροίσματος:

$$P(x) = \sum_{x \in \mathcal{Y}} P(x, y).$$

Διακριτές τυχαίες μεταβλητές
Discrete Random Variables

- Υπό συνθήκη πιθανότητα - Conditional probability: Η υπό συνθήκη πιθανότητα ενός γεγονότος A, δεδομένου ότι ένα άλλο γεγονός B, έχει συμβεί, συμβολίζεται ως P(A|B) και ορίζεται ως:

$$P(A|B) := \frac{P(A,B)}{P(B)}.$$

- Κανόνας γινομένου: $P(A,B) = P(A|B)P(B).$

Διακριτές τυχαίες μεταβλητές
Discrete Random Variables

- Ο κανόνας του γινομένου για δύο τυχαίες μεταβλητές x,y: $P(x, y) = P(x|y)P(y).$
- Οι P(x), P(y) ονομάζονται περιθωριακές (marginal) πιθανότητες.
- Στατιστική ανεξαρτησία: Δύο ΤΜ x and y, χαρακτηρίζονται ως στατιστικά ανεξάρτητες αν και μόνο αν:

$$P(x, y) = P(x)P(y), \ \forall x \in \mathcal{X}, \ y \in \mathcal{Y}.$$

Διακριτές τυχαίες μεταβλητές Discrete Random Variables

- Εφόσον: P(x,y) = P(y,x) συνάγεται:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)},$$
$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}.$$

- **Θεώρημα Bayes**:
  - Πολύ σημαντικό θεώρημα στην Μηχανική Μάθηση.

- Το θ. Bayes λέει ότι: Η αβεβαιότητα όπως εκφράζεται από την υπό συνθήκη πιθανότητα P(y|x) μιας μεταβλητής εξόδου y, δεδομένης της τιμής μιας εισόδου x, μπορεί να εκφραστεί με τον ανάποδο τρόπο:
  - Δλδ, με την υπό συνθήκη αβεβαιότητα P(x|y) και τις δύο περιθωριακές πιθανότητες P(x), P(y).

Συνεχείς ΤΜ (Continuous Random Variables)

- Μια ΣΤΜ x, μπορεί να λάβει οποιαδήποτε τιμή σε ένα πραγματικό διάστημα του R.

Ορισμοί:

- Η συσσωρευτική συνάρτηση κατανομής – (cumulative distribution function-cdf):

$$F_{\mathrm{x}}(x) := P(\mathrm{x} \leq x).$$

- Η cdf είναι η πιθανότητα του διακριτού γεγονότος: “η μεταβλητή x λαμβάνει τιμές μικρότερες ή ίσες από x”
- Άρα:

$$P(x_1 < \mathrm{x} \leq x_2) = F_{\mathrm{x}}(x_2) - F_{\mathrm{x}}(x_1).$$

Συνεχείς ΤΜ (Continuous Random Variables)

- Θεωρώντας την $\mathrm{F}_{\mathrm{x}}(x)$ ως διαφορίσιμη η συνάρτηση πυκνότητας πιθανότητας (probability density function -pdf) συμβολίζεται με μικρό *p* και ορίζεται ως:

$$p_{\mathrm{x}}(x) := \frac{dF_{\mathrm{x}}(x)}{dx}.$$

Τότε:

$$P(x_1 < \mathrm{x} \leq x_2) = \int_{x_1}^{x_2} p_{\mathrm{x}}(x)dx,$$

και

$$F_{\mathrm{x}}(x) = \int_{-\infty}^{x} p_{\mathrm{x}}(z)dz.$$

Συνεχείς ΤΜ (Continuous Random Variables)

- Καθώς ένα γεγονός είναι βέβαιο ότι θα συμβεί στο διάστημα $-\infty < x < \infty$ έχουμε ότι:

$$\int_{-\infty}^{+\infty} p_{\mathrm{X}}(x)dx = 1.$$

- Προφανώς ισχύουν και τα ακόλουθα:

$$p(x|y) = \frac{p(x,y)}{p(y)}, \quad p_{\mathrm{X}}(x) = \int_{-\infty}^{+\infty} p(x,y)dy.$$

Μέση τιμή, Διασπορά και συν-διακύμανση (Mean, Variance and Covariance)

- Δύο διάσημες ποσότητες που σχετίζονται με μια ΤΜ x είναι:
  - Η μέση τιμή (mean):
    $$\mathbb{E}[\mathrm{x}] := \int_{-\infty}^{+\infty} x p(x) dx.$$
  - Η διασπορά (variance):
    $$\sigma_x^2 := \int_{-\infty}^{+\infty} (x - \mathbb{E}[\mathrm{x}])^2 p(x) dx,$$
- (Αν είχαμε διακριτές ΤΜ π.χ):
  $$\mathbb{E}[\mathrm{x}] := \sum_{x \in \mathcal{X}} x P(x).$$
- Γενικότερα για μια f(x) :
  $$\mathbb{E}[f(\mathrm{x})] := \int_{-\infty}^{+\infty} f(x) p(x) dx.$$

Μέση τιμή, Διασπορά και συν-διακύμανση (Mean, Variance and Covariance)

- Η μέση τιμή σε σχέση με 2 ΤΜ:

$$\mathbb{E}[\mathrm{x}, \mathrm{y}] := \mathbb{E}_{\mathrm{x}}\left[\mathbb{E}_{\mathrm{y}|\mathrm{x}}[f(\mathrm{x}, \mathrm{y})]\right].$$

- Ομοίως η συν-διακύμανση(covariance):

$$\mathrm{cov}(\mathrm{x}, \mathrm{y}) := \mathbb{E}\big[(\mathrm{x} - \mathbb{E}[\mathrm{x}])(\mathrm{y} - \mathbb{E}[\mathrm{y}])\big].$$

- Η συσχέτιση (correlation):

$$r_{x,y} := \mathbb{E}[\mathrm{xy}] = \mathrm{cov}(\mathrm{x}, \mathrm{y}) + \mathbb{E}[\mathrm{x}]\mathbb{E}[\mathrm{y}].$$

- Ένα **τυχαίο διάνυσμα (random vector)** είναι μια συλλογή τυχαίων μεταβλητών $x := [x_1, x_2, \dots, xl]^T$ και η από κοινού κατανομή τους συμβολίζεται ως:

$$p(\boldsymbol{x}) = p(x_1, \dots, x_l), \;\; \boldsymbol{x} = [x_1, \dots, x_l]^T.$$

- Ο πίνακας συν-διακύμανσης ενός τυχαίου διανύσματος $\boldsymbol{x} \in Rl$ είναι:

$$\mathrm{Cov}(\mathbf{x}) := \mathbb{E}\big[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^T\big],$$

$$\mathrm{Cov}(\mathbf{x}) = \begin{bmatrix} \mathrm{cov}(\mathrm{x}_1, \mathrm{x}_1) & \dots & \mathrm{cov}(\mathrm{x}_1, \mathrm{x}_l) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(\mathrm{x}_l, \mathrm{x}_1) & \dots & \mathrm{cov}(\mathrm{x}_l, \mathrm{x}_l) \end{bmatrix}.$$

- Ο πίνακας συσχέτισης ενός τυχαίου διανύσματος $\boldsymbol{x} \in Rl$ είναι: $R_x := \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right],$

$$\begin{aligned} R_x &= \begin{bmatrix} \mathbb{E}[\mathrm{x}_1, \mathrm{x}_1] & \dots & \mathbb{E}[\mathrm{x}_1, \mathrm{x}_l] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[\mathrm{x}_l, \mathrm{x}_1] & \dots & \mathbb{E}[\mathrm{x}_l, \mathrm{x}_l] \end{bmatrix} \\ &= \mathrm{Cov}(\mathbf{x}) + \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}^T]. \end{aligned}$$

Παράδειγμα: Υπολογισμός covariance

- Έστω $x_i \in \mathbb{R}^{p\times 1}$ ένα διάνυσμα στήλης

  Το μέγεθος $x_i x_i^T \in \mathbb{R}^{p\times p}$

- **Αριθμητική μέση τιμή $\overline{x}$ (sample mean vector)**:

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}, \quad j = 1, \ldots, K.$$

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_j \\ \vdots \\ \bar{x}_K \end{bmatrix}$$

Παράδειγμα: Υπολογισμός covariance

Αριθμητικός πίνακας συν-διακύμανσης (sample covariance):

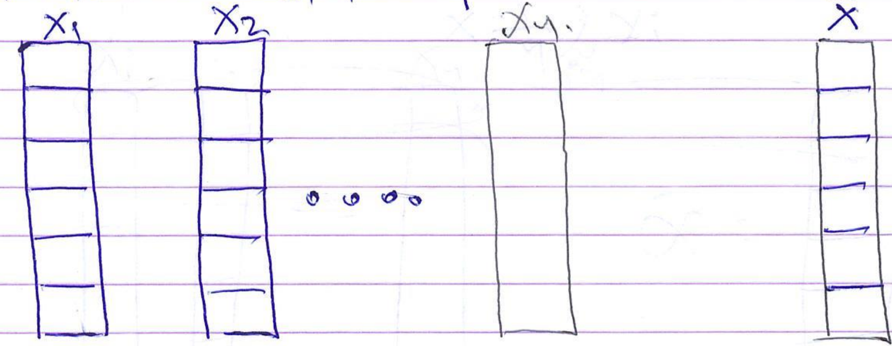
$$\Sigma \in \mathbb{R}^{p \times p},$$

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})\,(x_i - \bar{x})^T$$

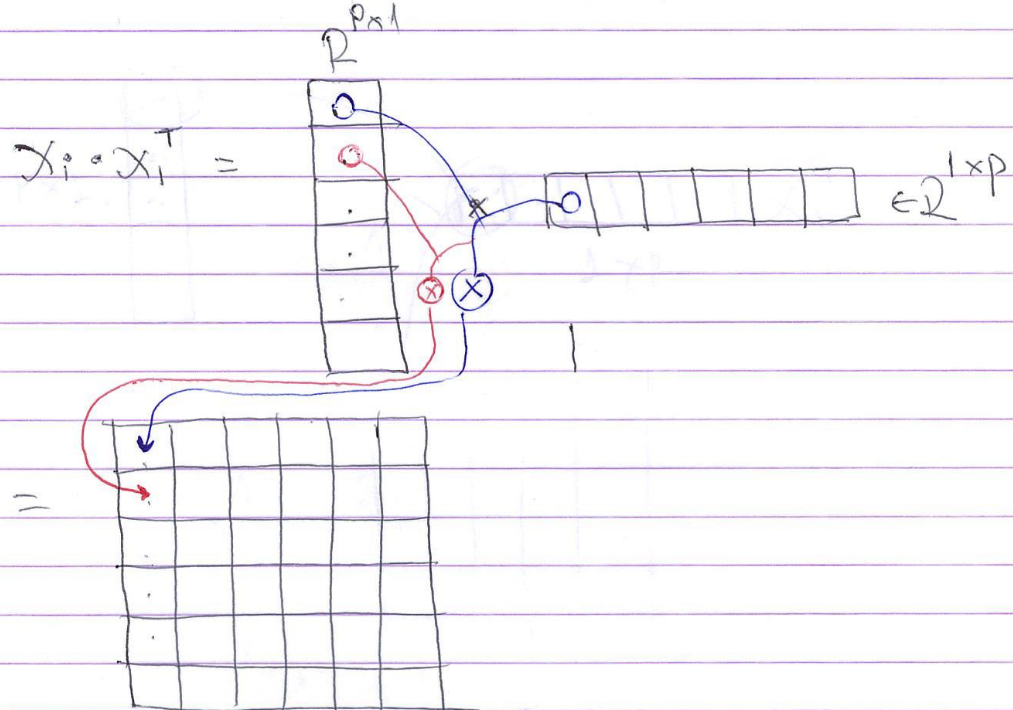
- Ο παραπάνω τύπος δίνει μια μη πολωμένη εκτίμηση.
- Αν έχω δείγματα από κανονική κατανομή…

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,(x_i - \bar{x})^T$$

$x_i \in \mathbb{R}^p \qquad i = 1, 2, \dots, n$



$x_i \cdot x_i^T =$



$$C = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (x_i - \bar{x})^T$$

Fore (1)

- Αν: $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{ll} \end{bmatrix}$ $\boldsymbol{x} = [x_1, x_2, \ldots, x_l]^T$

- Τότε: $\boldsymbol{x}^T A \boldsymbol{x} = \sum_{i=1}^{l} \sum_{j=1}^{l} a_{ij} x_i x_j$
- Αν ο Α είναι συμμετρικός:

$$\boldsymbol{x}^T A \boldsymbol{x} = \sum_{i=1}^{l} a_{ii} x_i^2 + 2 \sum_{i=1}^{l} \sum_{j>i}^{l} a_{ij} x_i x_j$$

- Αν ο Α διαγώνιος: $\boldsymbol{x}^T A \boldsymbol{x} = \sum_{i=1}^{l} a_{ii} x_i^2$

play.. (2)

- Αν $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{ll} \end{bmatrix}$ $\boldsymbol{x} = [x_1, x_2, \ldots, x_l]^T$

- Ο Α οριστικά θετικός (μη αρνητικός) (***positive definite***) αν:

$$\boldsymbol{x}^T A \boldsymbol{x} > (\geq) 0$$

Μέση τιμή, Διασπορά και συν-διασπορά (Mean, Variance and Covariance)

- Ιδιότητες: Ο πίνακας Α (ή **Σ**) της συν-διακύμανσης καθώς και ο πίνακας της συσχέτισης είναι positive semidefinite.
- Ένας πίνακας είναι positive semidefinite αν:

$$\boldsymbol{y}^T A \boldsymbol{y} \geq 0, \ \forall \boldsymbol{y} \in \mathbb{R}^l,$$

- Ένας πίνακας είναι positive definite αν:

$$\boldsymbol{y}^T A \boldsymbol{y} > 0, \ \forall \boldsymbol{y} \in \mathbb{R}^l,$$

Μετασχηματισμοί των ΤΜ (Transformation of Random Variables)

- Έστω **x**, **y** δυο τυχαία διανύσματα που σχετίζονται με την ακόλουθη σχέση: **y**=f(**x**);
- Θεωρούμε ότι η συνάρτηση f είναι αντιστρέψιμη, δλδ πάντα μπορούμε να: **x**=f^{-1}(**y**).
- Δεδομένης της pdf p_x(**x**) μπορεί να δειχτεί ότι:

$$p_{\mathbf{y}}(\boldsymbol{y}) = \left.\frac{p_{\mathbf{x}}(\boldsymbol{x})}{\left|\det(J(\mathbf{y};\mathbf{x}))\right|}\right|_{\boldsymbol{x}=\boldsymbol{f}^{-1}(\boldsymbol{y})}$$

- Όπου J είναι ο Ιακωβιανός πίνακας του μετασχηματισμού:

Ο Ιακωβιανός πίνακας του μετ/μού

$$p_{\mathbf{y}}(\boldsymbol{y}) = \left.\frac{p_{\mathbf{x}}(\boldsymbol{x})}{\left|\det(J(\mathbf{y};\mathbf{x}))\right|}\right|_{\boldsymbol{x}=\boldsymbol{f}^{-1}(\boldsymbol{y}}$$

$$J(\mathbf{y};\mathbf{x}) := \frac{\partial(y_1, y_2, \ldots, y_l)}{\partial(x_1, x_2, \ldots, x_l)} := \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_l} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_l}{\partial x_1} & \cdots & \frac{\partial y_l}{\partial x_l} \end{bmatrix}$$

det(.) είναι η ορίζουσα ενός πίνακα και |.| η απόλυτη τιμή.

Για την περίπτωση 2 μεταβλητών:

$$p_{\mathrm{y}}(y) = \left.\frac{p_{\mathrm{x}}(x)}{\left|\frac{dy}{dx}\right|}\right|_{x=f^{-1}(y)}.$$

Παράδειγμα

- Έστω ότι τα δύο τυχαία διανύσματα **x**, **y** συνδέονται με ένα γραμμικό μετασχηματισμό μέσω ενός αντιστρέψιμου πίνακα A: **y**=A**x**.

- Η Jacobian: $\mathrm{J}(\mathbf{y};\mathbf{x}) = \mathrm{A}$.

- Τότε: $$p_{\mathbf{y}}(\boldsymbol{y}) = \frac{p_{\mathbf{x}}(A^{-1}\boldsymbol{x})}{|\det A|}.$$

Παραδείγματα διακριτών κατανομών

Bernoulli κατανομή

- Περιγράφει **<u>ένα</u>** τυχαίο πείραμα με δυο πιθανά αποτελέσματα (επιτυχία =1, αποτυχία=0) και πιθανότητα επιτυχίας p.

Παραδείγματα διακριτών κατανομών

- **Bernoulli κατανομή** : Μια ΤΜ έχει κατανομή Bernoulli αν είναι δυαδική στο X={0,1} με:

  $$P(x = 1) = p;\ P(x = 0) = 1 - p.$$

- λέμε ότι x~*Bern*(x|p) με:

  $$P(x)=Bern(x\ ;\ p) := p^{x}\,(1-p)^{1-x}$$

- Μέση τιμή: $\mathbb{E}[x] = 1p + 0(1-p) = p.$

- Διασπορά: $\sigma_x^2 = (1-p)^2 p + p^2(1-p) = p(1-p).$

Διωνυμική κατανομή

- Περιγράφει **ένα** τυχαίο πείραμα με δυο πιθανά αποτελέσματα (επιτυχία - αποτυχία) και πιθανότητα επιτυχίας p **που επαναλαμβάνεται n φορές**.

- Δειγματοληψία με επαναφορά.

- Αν η δειγματοληψία είναι χωρίς επαναφορά, η τυχαία μεταβλητή που δηλώνει τον αριθμό των λευκών μπαλών ακολουθεί την υπεργεωμετρική κατανομή.

Παραδείγματα διακριτών κατανομών

- **Η διωνυμική κατανομή** (**binomial**): Μια ΤΜ ακολουθεί την διωνυμική κατανομή με παραμέτρους: *n, p:* $x \sim Bin(x|n, p)$ αν X={0,1,…,*n*} και:

$$P(\mathrm{x}=k) := \mathrm{Bin}(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}, \; k = 0, 1, \ldots, n.$$

- **Παράδειγμα**: Με την ΔΚ μοντελοποιούμε ένα πείραμα πολλαπλών ρίψεων π.χ. ενός νομίσματος με 2 ενδεχόμενα.
  - Έτσι, μπορούμε να μοντελοποιήσουμε τον αριθμό εμφανίσεων του γεγονότος: «κορώνα» σε επαναλαμβανόμενα πειράματα ρίψης ενός νομίσματος όπου P(«κορώνα») = p.

Διωνυμική Κατανομή

- Θεωρούμε μια κάλπη με Κ λευκές μπάλες και Ν-Κ μαύρες.
  - Η πιθανότητα να τραβήξουμε μια λευκή μπάλα είναι p=K/N.
- Τραβάμε **μια-μια** τις μπάλες από την κάλπη επανατοποθετώντας τις κάθε φορά πίσω στην κάλπη (δειγματοληψία με επαναφορά) μέχρι να τραβήξουμε n μπάλες.
- Ζητάμε την πιθανότητα οι k από αυτές να είναι λευκές.

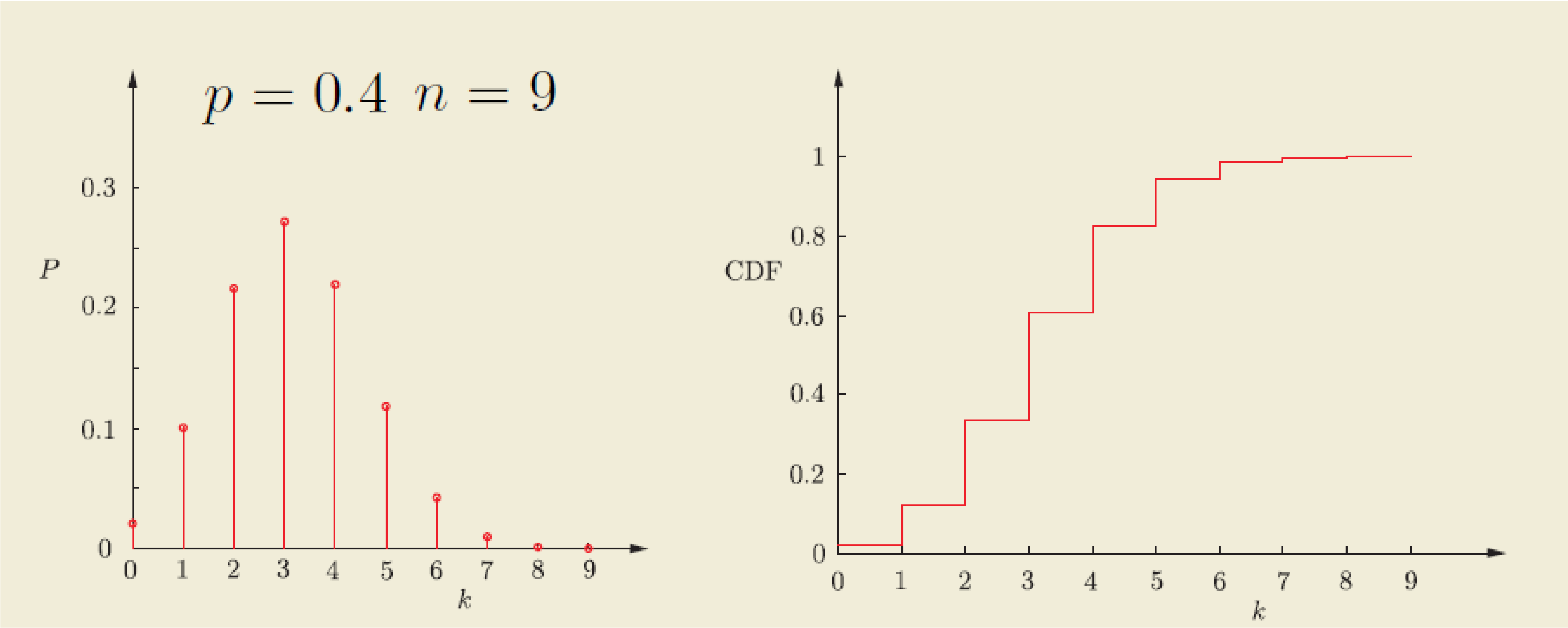
$$\mathrm{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

Παραδείγματα διακριτών κατανομών

- Η διωνυμική κατανομή είναι μια γενίκευση της κατανομής Bernoulli αν θέσουμε n = 1.
- Η μέση τιμή και η διασπορά της διωνυμικής κατανομής είναι :

$$\mathbb{E}[\mathrm{x}] = np, \text{ and } \sigma_x^2 = np(1-p).$$

Διωνυμική Κατανομή



Πολυωνυμική κατανομή (Multinomial Distribution)

- Είναι η γενίκευση της διωνυμικής κατανομής.
- Το αποτέλεσμα (ενδεχόμενο) κάθε πειράματος δεν είναι δυαδικό αλλά μπορεί να πάρει μια από K δυνατές τιμές:
  - Π.χ: Αντί να ρίξουμε ένα νόμισμα, ρίχνουμε ένα ζάρι με K πλευρές (K ενδεχόμενα).
- Κάθε ένα από τα K (E_1=«1», E_2=«2», …, E_k=«K») ενδεχόμενα έχει πιθανότητα P_1, P_2,…, P_K να συμβεί. Αυτό συμβολίζεται ως:

$$\boldsymbol{P} = [P_1, P_2, \ldots, P_K]^T.$$

Πολυωνυμική κατανομή (Multinomial Distribution)

- Έστω ότι κάνουμε n πειράματα:
- Επιθυμούμε να υπολογίζουμε την πιθανότητα του ακριβή πλήθους $\{x_1, x_2, x_K\}$ να εμφανιστεί το κάθε ένα από τα K ενδεχόμενα E_1=«1», E_2=«2», …, E_k=«K».
- Το τυχαίο διάνυσμα $\mathbf{x} = [x_1, x_2, \ldots, x_K]^T$ ακολουθεί την πολυωνυμική κατανομή $x \sim \mathrm{Mult}(\mathbf{x}|n; \mathbf{P})$:

$$P(\boldsymbol{x}) = \mathrm{Mult}(\boldsymbol{x}|n, \boldsymbol{P}) := \binom{n}{x_1, x_2, \ldots, x_K} \prod_{i=1}^{K} P_k^{x_k}$$

Πολυωνυμική κατανομή

$$P(\boldsymbol{x}) = \mathsf{Mult}(\boldsymbol{x}|n, \boldsymbol{P}) := \left(\begin{array}{c} n \\ x_1, x_2, \ldots, x_K \end{array} \right) \prod_{i=1}^{K} P_k^{x_k}$$

Όπου:

$$\left(\begin{array}{c} n \\ x_1, x_2, \ldots, x_K \end{array} \right) = \frac{n!}{x_1! x_2! \ldots x_K!}.$$

Και με τους ακόλουθους περιορισμούς:

$$\sum_{k=1}^{K} \mathrm{x}_k = n, \ \sum_{k=1}^{K} P_K = 1.$$

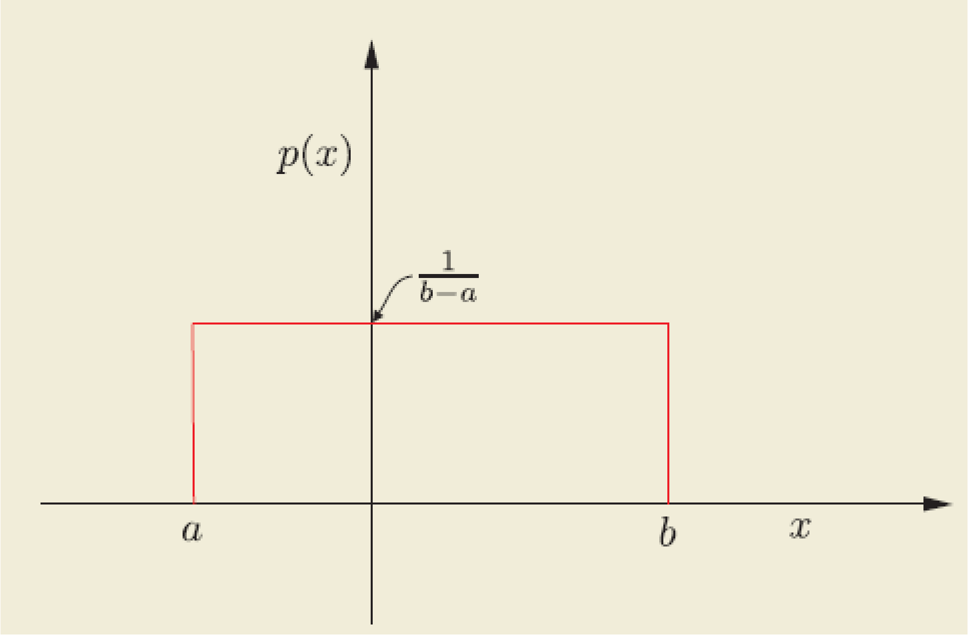
Πολυωνυμική κατανομή

- Μέσες τιμές: $\mathbb{E}[\mathbf{x}] = n\boldsymbol{P},$
- Διασπορές: $\sigma_k^2 = nP_k(1 - P_k),\ k = 1, 2, \ldots, K,$
- Συνδιασπορές: $\mathsf{cov}(\mathrm{x}_i, \mathrm{x}_j) = -nP_iP_j,\ i \neq j.$

Ομοιόμορφη Κατανομή (Uniform Distribution)

- Μια ΤΜ x, ακολουθεί την ομοιόμορφη κατανομή σε ένα διάστημα [a, b]: x~U(a,b) με a > -∞ and b < + ∞, αν:

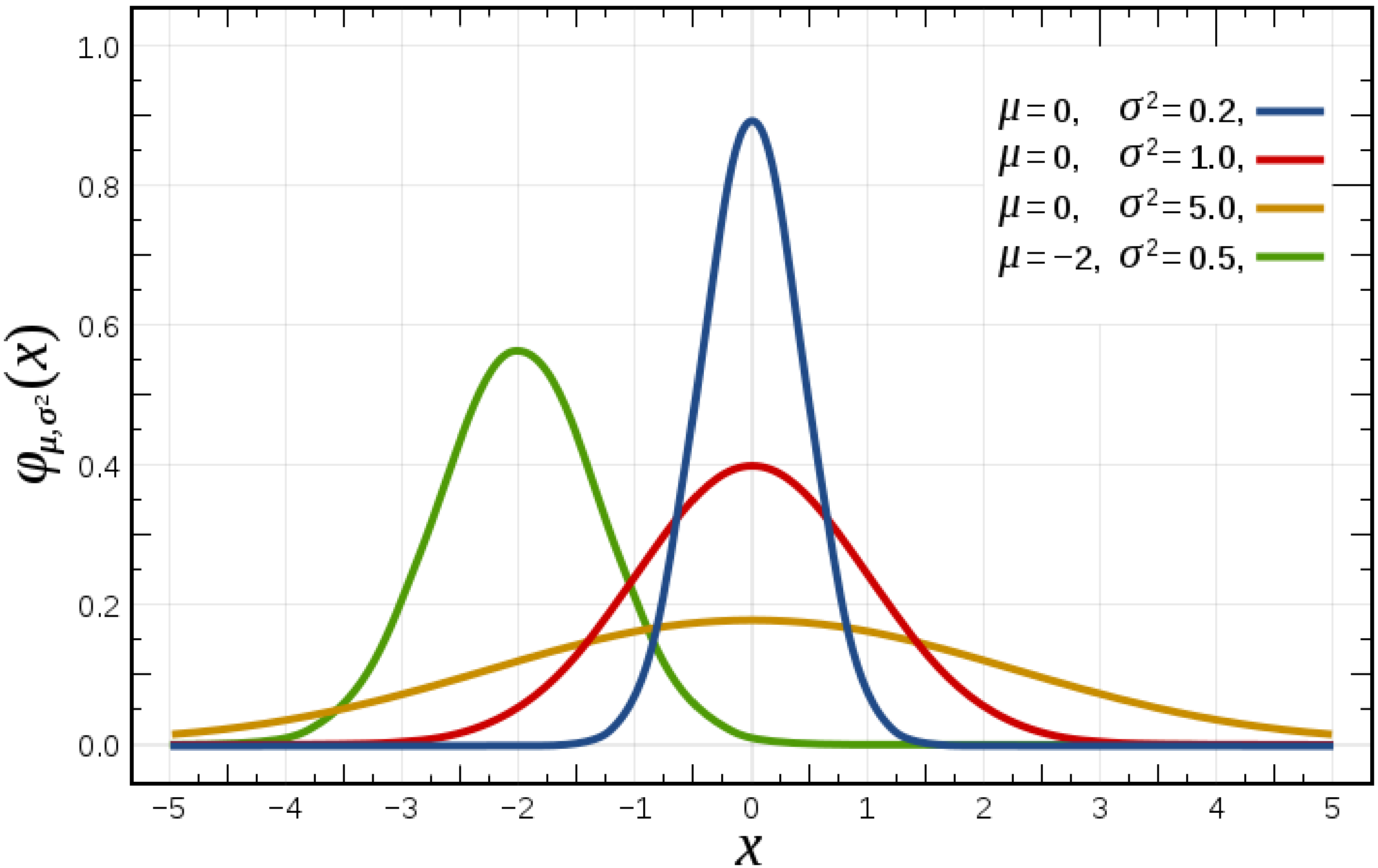
$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$



$$\mathbb{E}[\mathrm{x}] = \frac{a+b}{2}, \ \sigma_x^2 = \frac{1}{12}(b-a)^2.$$

Τυπικά παραδείγματα κατανομών με συνεχείς κατανομές

Η Gaussian Κατανομή
(The Gaussian Distribution)



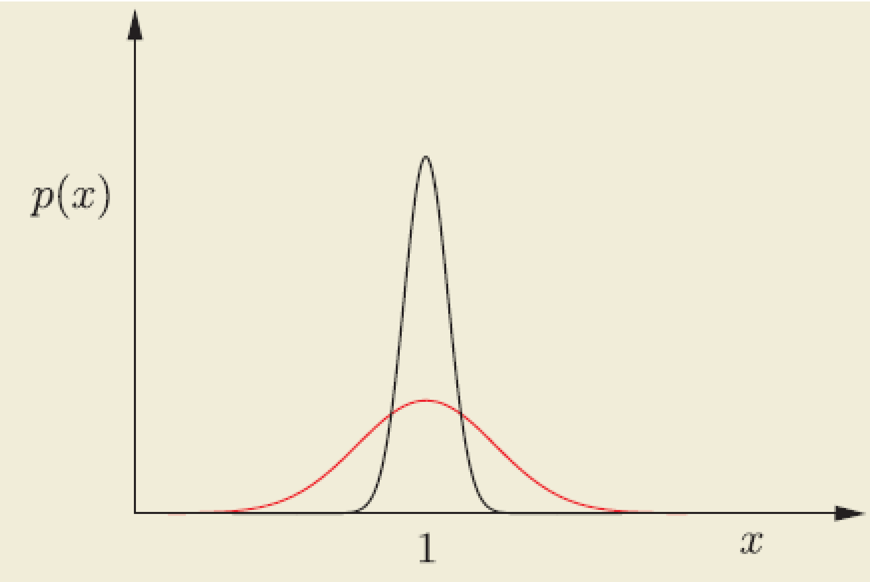
Η Gaussian Κατανομή
(The Gaussian Distribution)

- Η κανονική κατανομή αποτελεί την πιο σημαντική κατανομή της στατιστικής μεθοδολογίας για τους εξής λόγους:
  - Την κανονική κατανομή ακολουθούν είτε με ακρίβεια είτε με μεγάλη προσέγγιση τα περισσότερα συνεχή φαινόμενα.
  - Πολλές ασυνεχείς κατανομές πιθανοτήτων μπορούν να προσεγγιστούν μέσω της κανονικής κατανομής.
    - Για παράδειγμα πολλά πληθυσμιακά χαρακτηριστικά, όπως το ύψος, το βάρος **η βαθμολογία σε διαγώνισμα**, κ.λπ.
  - Η κανονική κατανομή αποτελεί σύμφωνα με το κεντρικό οριακό θεώρημα τη βάση της στατιστικής συμπερασματολογίας ή επαγωγικής στατιστικής.
  - Τυχαία σφάλματα που εμφανίζονται σε διάφορες μετρήσεις έχουν κανονική κατανομή.
    - Γι' αυτό το λόγο η Κανονική κατανομή αναφέρεται πολλές φορές και ως κατανομή σφαλμάτων.

Η Gaussian Κατανομή

- Μια ΤΜ ακολουθεί την **Gaussian ή κανονική κατανομή** με παραμέτρους μ, σ^2: $x \sim N(\mu, \sigma^2)$ ή $N(x|\mu, \sigma^2)$ αν:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right).$$



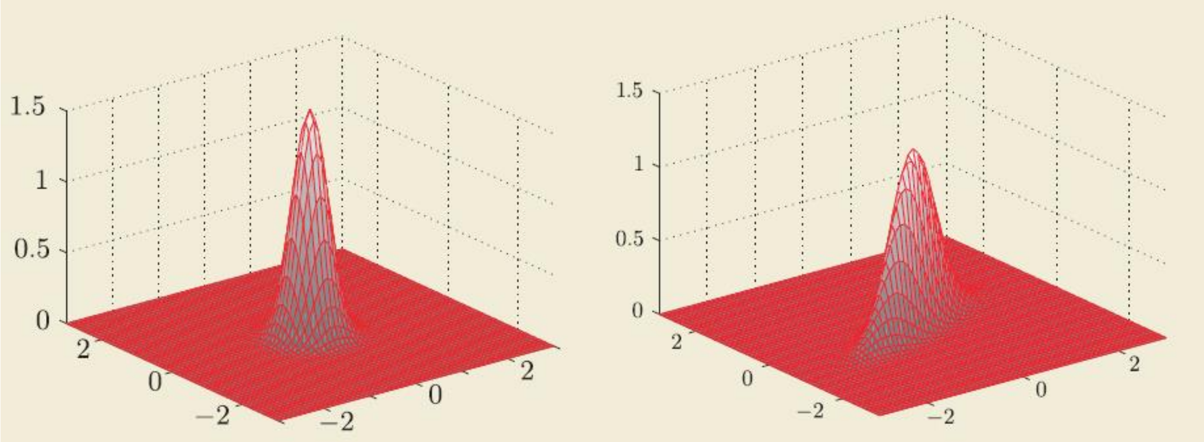
$$\mathbb{E}[\mathrm{x}] = \mu,$$

$$\sigma_x^2 = \sigma^2.$$

Gaussian πολλών μεταβλητών (Multivariate – Joint normal distribution)

- Γενίκευση της Gaussian ή κανονικής κατανομής σε πολλές διαστάσεις: $\mathbf{x} \in \Re^l$: x~N(**x**|**μ**,Σ):

$$p(x) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$



$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu}$$
$$\mathrm{Cov}(\mathbf{x}) = \Sigma.$$

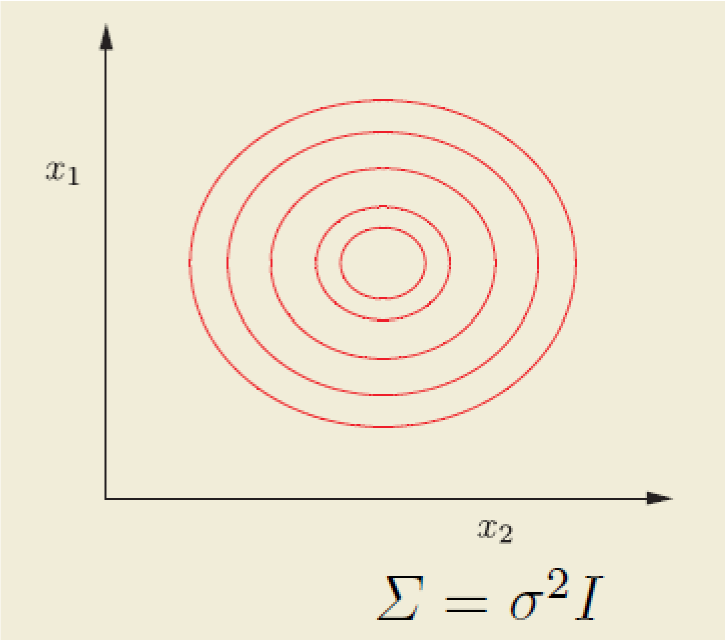
Gaussian πολλών μεταβλητών

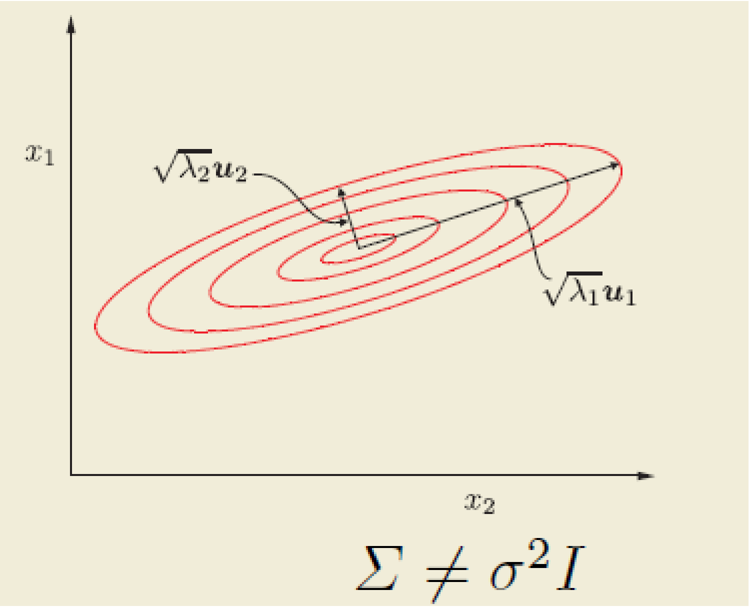
- Ισοδυναμικές επιφάνειες (isovalue curves) των multivariate Gaussians:
  - Οι ισοδυναμικές επιφάνειες αποτελούνται από τα σημεία εκείνα τα οποία αντιστοιχούν στην ίδια τιμή της pdf, δλδ., p(**x**) = c.

$$(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = c.$$

- Οι ισοδυναμικές επιφάνειες είναι τετραγωνικής φύσης: κύκλοι (υπερ-σφαίρες) ή ελλείψεις (υπερ-ελλειψοειδή) με κέντρο την μέση τιμή.
- Οι άξονες των ισοδυναμικών επιφανειών καθορίζοναι από τα ιδιοδιανύσματα και τις ιδιοτιμές του Σ.

Gaussian πολλών μεταβλητών





Η απόδειξη…

Το σχήμα των Gaussian

- Τα σημεία μιας ισοδυναμικής επιφάνειας ικανοποιούν την σχέση:

$$(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = c.$$

- Ο πίνακας συνδιακύμανσης Σ είναι positive definite και symmetric: $\Sigma = \Sigma^T$.
  - Έτσι οι ιδιοτιμές του είναι πραγματικές (και θετικές) και τα αντίστοιχα ιδιοδιανύσματα μπορούν να επιλεγούν έτσι ώστε να φτιάξουν ένα ορθοκανονικό σύστημα συντεταγμένων.
  - Ως αποτέλεσμα έχουμε την διαγωνοποίηση (diagonalization), του πίνακα Σ:

$$\Sigma = U \Lambda U^T$$

Διαγωνοποίηση (diagonalization), πίνακα Σ

$$\Sigma = U \Lambda U^T$$

- $\mathrm{U} := [\mathbf{u}_1, \ldots, \mathbf{u}_l]$ όπου:
  - $\mathbf{u}_i$, $\mathrm{i} = 1,2,\ldots,l$ τα ορθοκανονικά ιδιοδιανύσματα.
  - Ο πίνακας $\mathbf{U}$ είναι unitary (δλδ: $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$).
  - $\Lambda := \mathrm{diag}(\lambda_1, \lambda_2, .., \lambda_l,)$ να είναι ένας διαγώνιος πίνακας με στοιχεία τις αντίστοιχες ιδιοτιμές.
- Αν $y := \mathbf{U}^T(\mathbf{x}\text{-}\boldsymbol{\mu})$ τότε:

$$(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) = c.$$

$$\Sigma = U \Lambda U^T$$

και $$\boldsymbol{y}^T \Lambda^{-1} \boldsymbol{y} = c,$$

$$\mathbf{y} := \mathbf{U}^T(\mathbf{x}-\boldsymbol{\mu})$$

- Αντιστοιχεί σε μια περιστροφή των αξόνων κατά **U** και μια μετατόπισης της αρχής στο **μ**.

$$\boldsymbol{y}^T \Lambda^{-1} \boldsymbol{y} = c,$$

$$\frac{y_1^2}{\lambda_1} + \ldots + \frac{y_l^2}{\lambda_l} = c.$$

- Υπερ-ελλειψοειδές στο R^l.
- Κεντραρισμένο στο **μ**
- Οι κύριοι άξονες παράλληλοι με τα $\mathbf{u}_1, \ldots, \mathbf{u}_l$
- Το μέγεθος των αξόνων καθορίζεται από τις αντίστοιχες ιδιοτιμές.

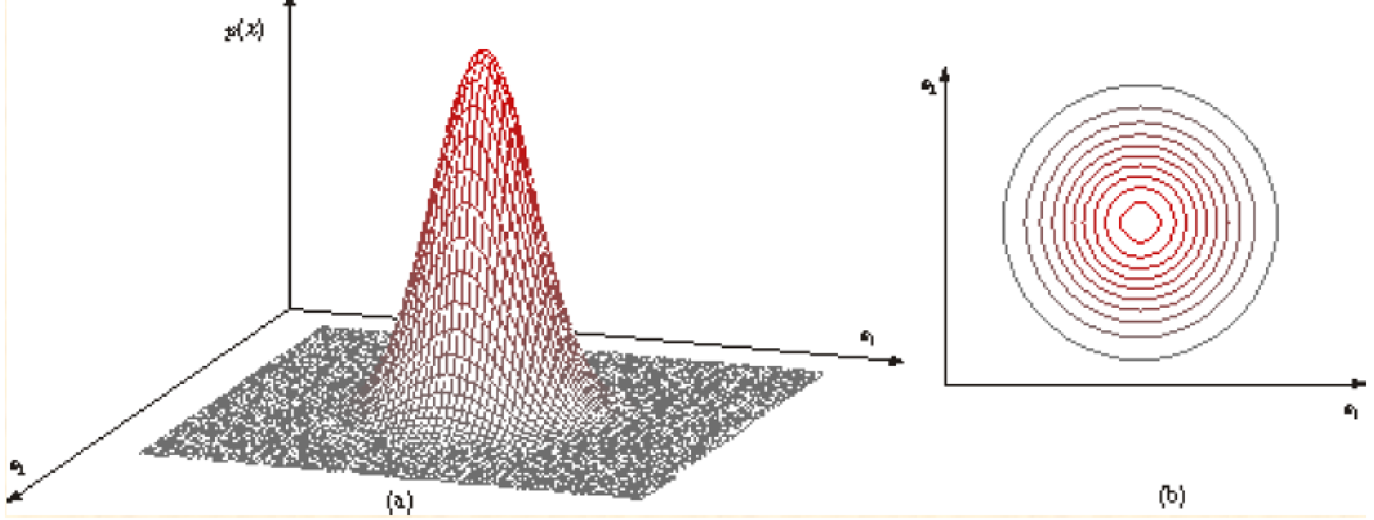
Ιδιότητες Gaussian

- Αν ο Σ είναι διαγώνιος: $\Sigma = \text{diag}\{\sigma_1^2, \ldots, \sigma_l^2\},$
  - Οι τυχαίες μεταβλητές που αποτελούν το διάνυσμα **x** είναι **στατιστικά ανεξάρτητες**.
  - Το παραπάνω δεν είναι πάντα αληθές.
  - Γενικά, οι ασυσχέτιστες μεταβλητές δεν είναι αναγκαία και ανεξάρτητες.
  - Η ανεξαρτησία είναι πολύ πιο ισχυρή συνθήκη.
- Αν ο Σ είναι διαγώνιος:

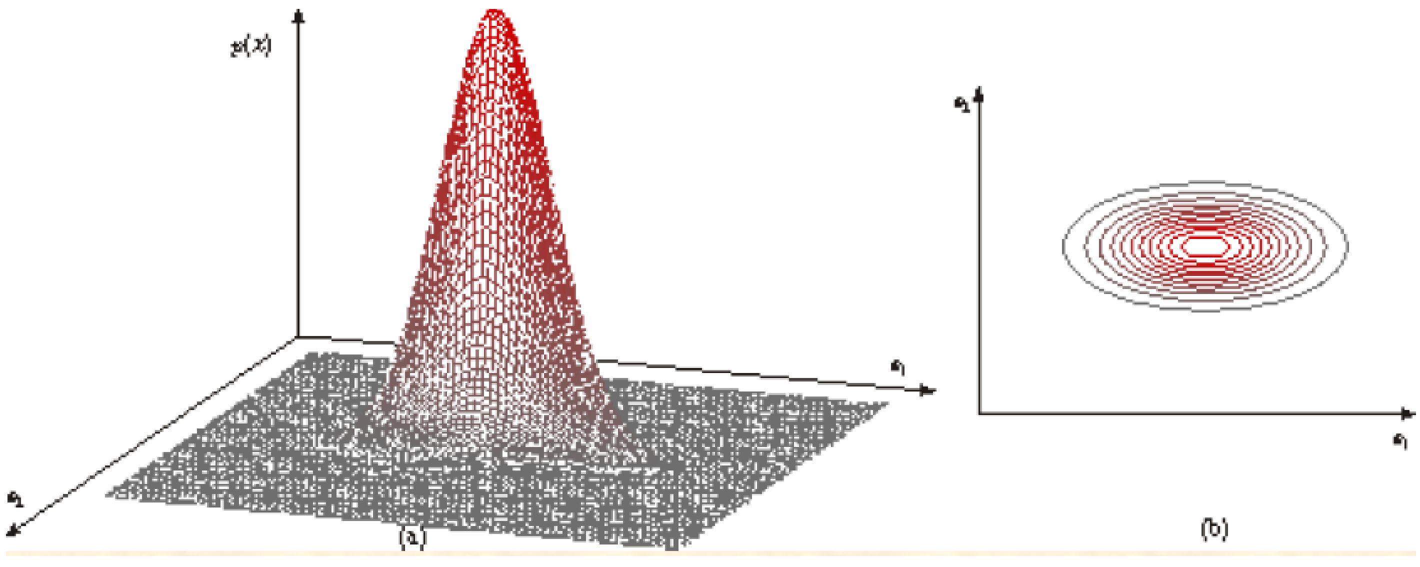
$$p(\boldsymbol{x}) = \prod_{i=1}^{l} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right).$$

Συνθήκη για statistical independence.

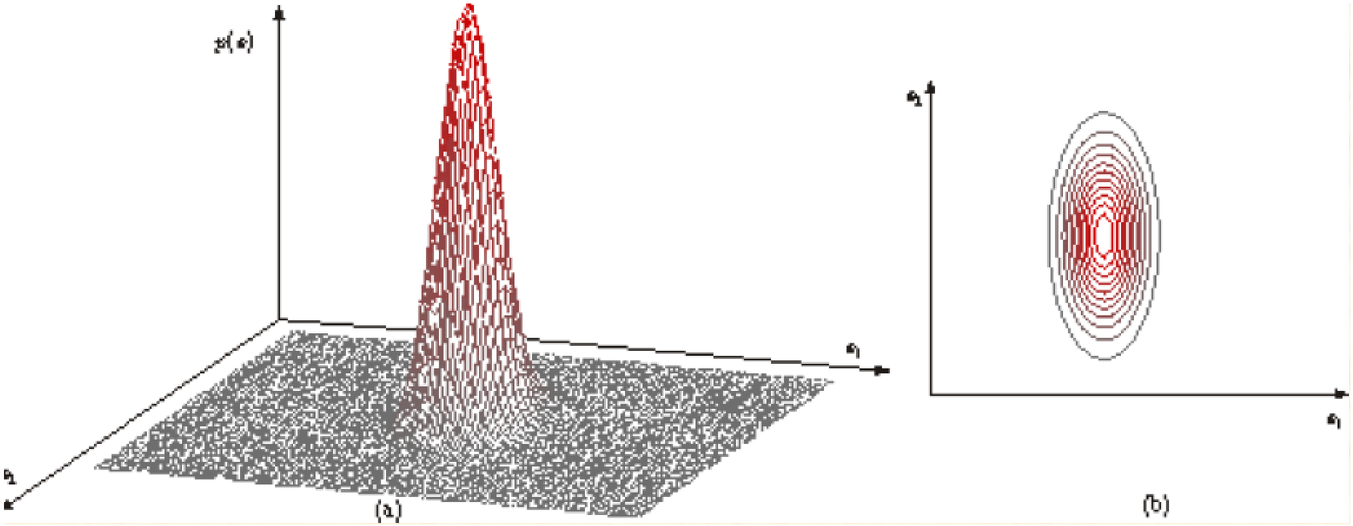
$$p(\boldsymbol{x}) = \prod_{i=1}^{l} p(x_i),$$



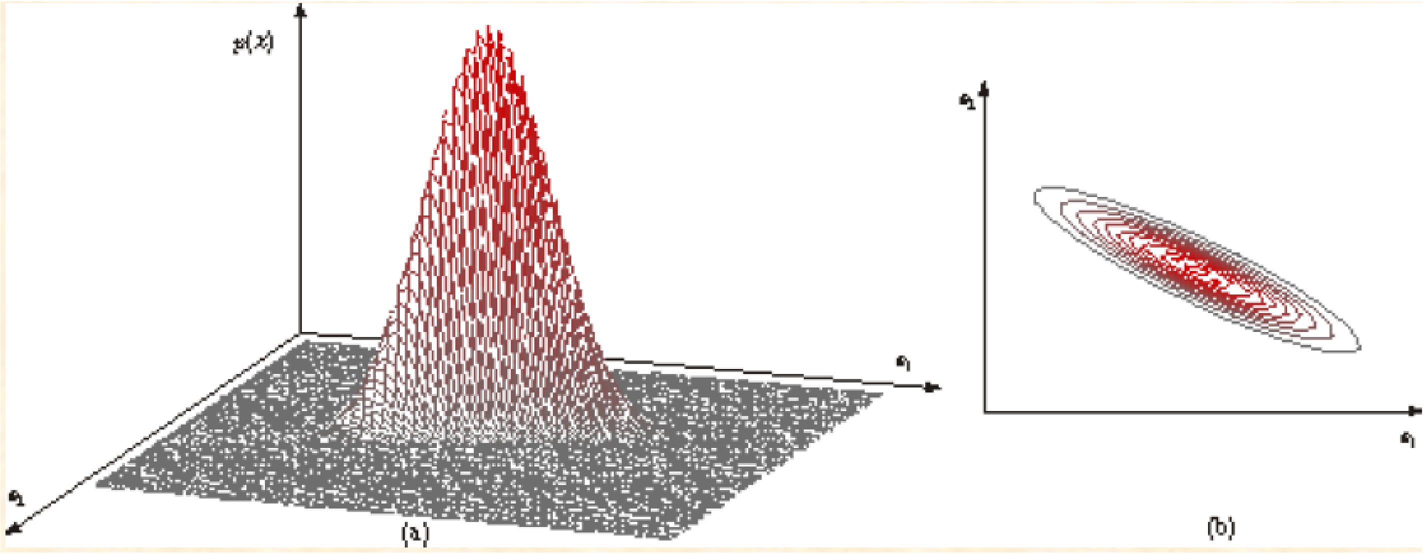
Σ διαγώνιος με ίσα διαγώνια στοιχεία



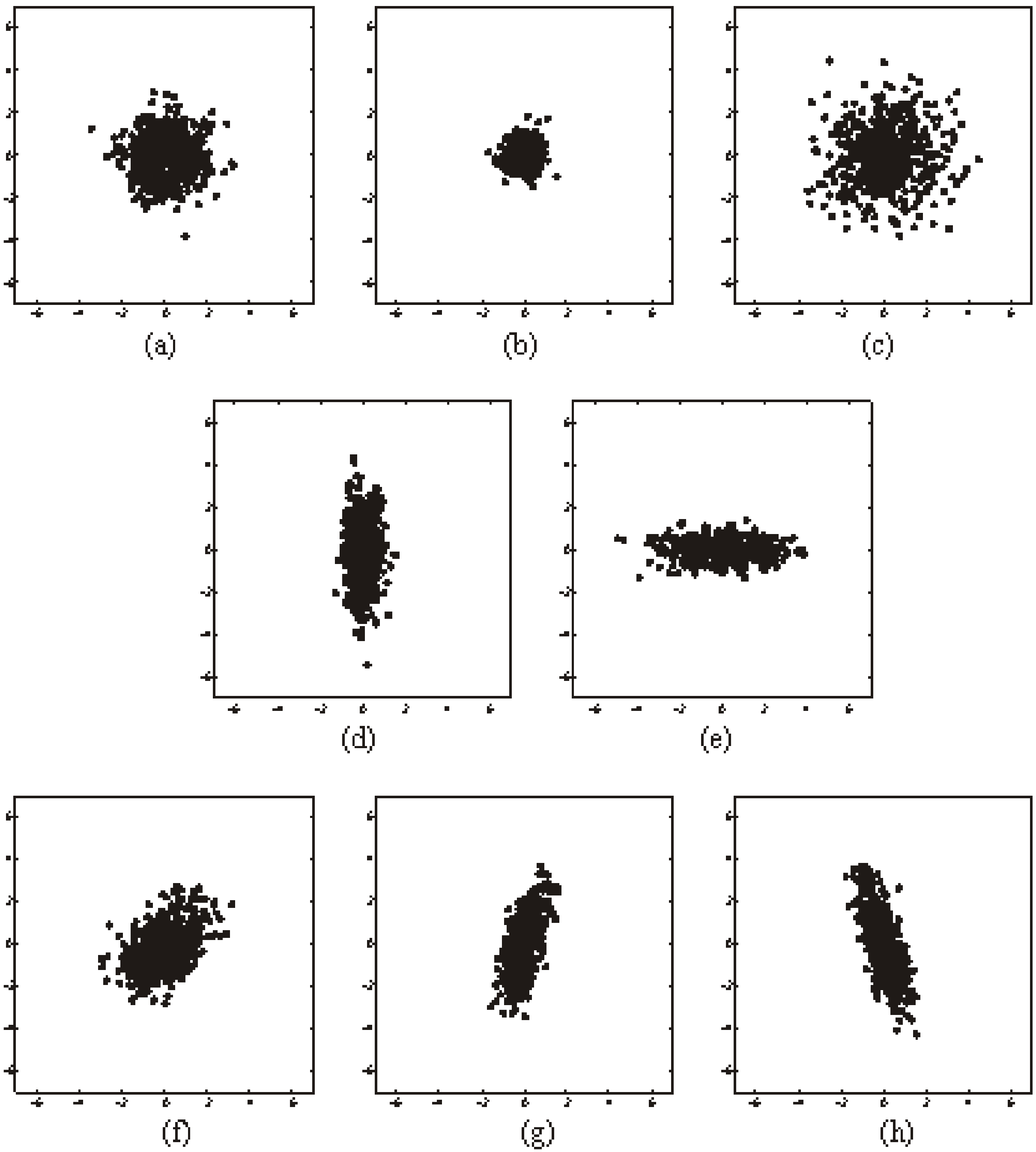
Σ διαγώνιος με $\sigma_1{}^2 \gg \sigma_2{}^2$



Σ διαγώνιος με $\sigma_1{}^2 \ll \sigma_2{}^2$



Σ μη διαγώνιος



$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

(α) $\sigma_1^2=\sigma_2^2=1$, $\sigma_{12}=0$

(β) $\sigma_1^2=\sigma_2^2=0.2$, $\sigma_{12}=0$

(γ) $\sigma_1^2=\sigma_2^2=2$, $\sigma_{12}=0$

(δ) $\sigma_1^2=0.2$, $\sigma_2^2=2$, $\sigma_{12}=0$

(ε) $\sigma_1^2=2$, $\sigma_2^2=0.2$, $\sigma_{12}=0$

(στ) $\sigma_1^2=\sigma_2^2=1$, $\sigma_{12}=0.5$

(ζ) $\sigma_1^2=0.3$, $\sigma_2^2=2$, $\sigma_{12}=0.5$

(η) $\sigma_1^2=0.3$, $\sigma_2^2=2$, $\sigma_{12}=-0.5$

Το Κεντρικό Οριακό Θεώρημα (The Central Limit Theorem)

- Θεωρήσατε Ν ανεξάρτητες ΤΜ $x_i, i = 1, \ldots, N$
  - Κάθε μια ακολουθεί την δικιά της κατανομή με μέση τιμή μ_i και διασπορά ${\sigma_i}^2, i = 1, \ldots, N$
  - Ορίζουμε μια νέα μεταβλητή να είναι το άθροισμα τους: $x = \sum_{i=1}^{N} x_i$
- Τότε η μέση τιμή και η διασπορά της νέας μεταβλητής δίνονται από:

$$\mu = \sum_{i=1}^{N} \mu_i, \quad \text{and} \quad \sigma_x^2 = \sum_{i=1}^{N} \sigma_i^2 .$$

Το Κεντρικό Οριακό Θεώρημα (The Central Limit Theorem)

- Μπορεί να αποδειχτεί ότι:
  - Αν N $\rightarrow \infty$ τότε η κατανομή της κανονικοποιημένης μεταβλητής

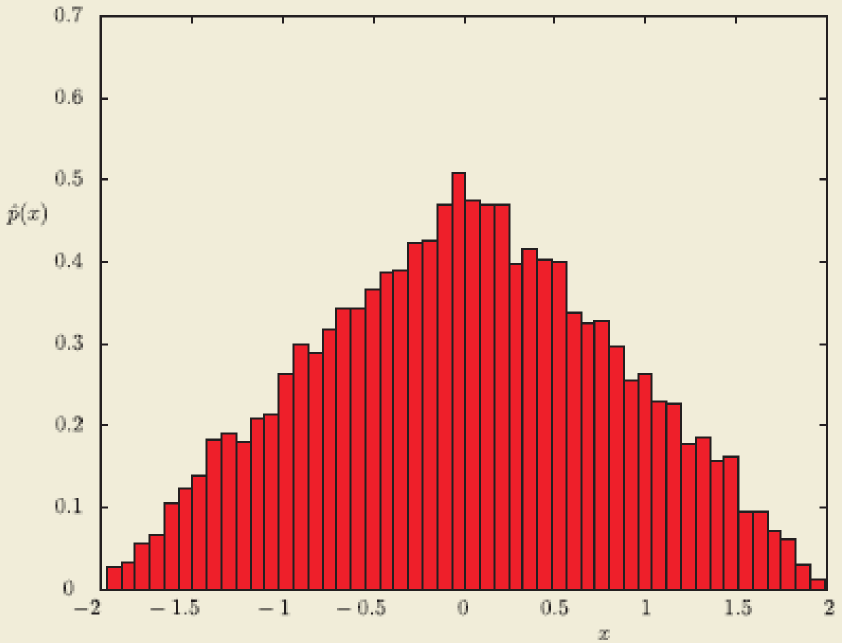
$$z = \frac{x - \mu}{\sigma},$$

  - τείνει στην κανονική κατανομή N(z|0, 1)
- Το Θεώρημα Κεντρικού Ορίου είναι ένα από τα πιο σημαντικά θεωρήματα στην επιστήμη της πιθανότητα και της στατιστικής και εξηγεί εν μέρει τη δημοτικότητα της Gaussian κατανομής.

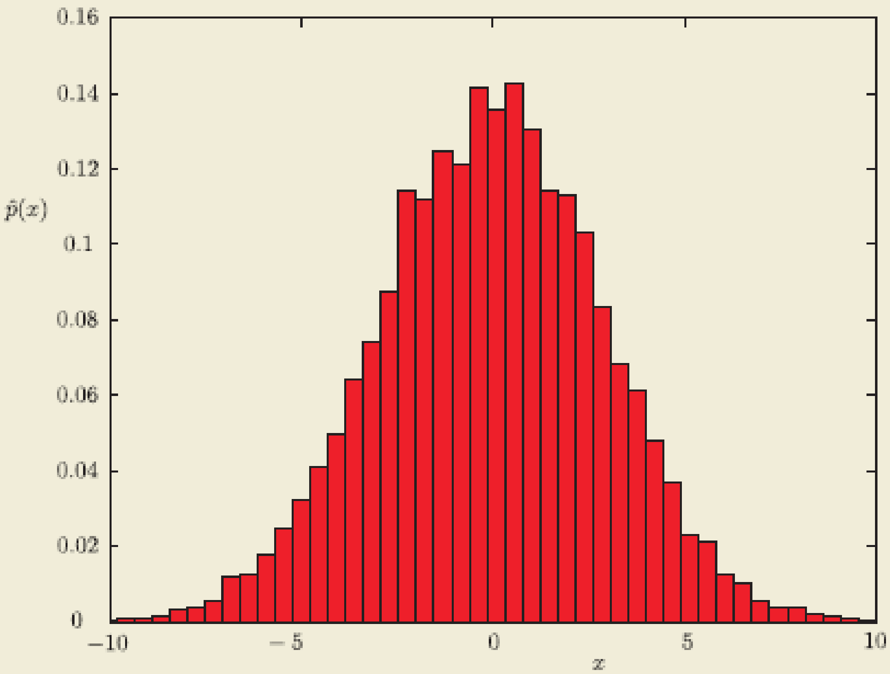
Το Κεντρικό Οριακό Θεώρημα (The Central Limit Theorem)

- Στην πράξη, ακόμη και με το άθροισμα ενός σχετικά μικρού πλήθους τυχαίων μεταβλητών, μπορεί κανείς να αποκτήσει μια καλή προσέγγιση σε Gaussian.
- Για για παράδειγμα, αν οι μεμονωμένες pdf είναι αρκετά ομαλές και οι τυχαίες μεταβλητές είναι (iid), τότε ένας αριθμός μεταξύ 5 και 10 μεταβλητών μπορεί να είναι επαρκής.

Το Κεντρικό Οριακό Θεώρημα (The Central Limit Theorem)



Sum of two i.i.d variables from a uniform in [-1,1]



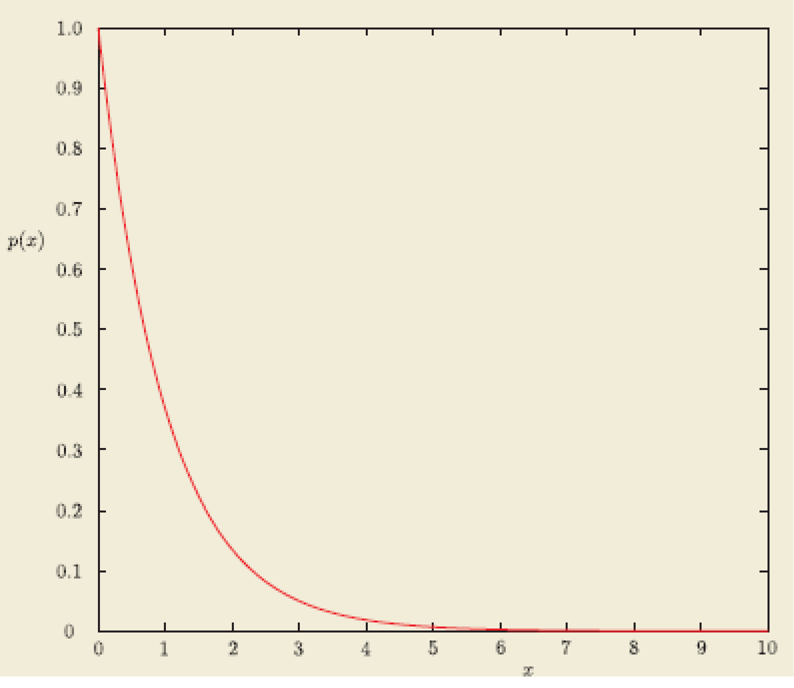
Sum of twenty five i.i.d r.v from a uniform in [-1,1]

Η Εκθετική Κατανομή (The Exponential Distribution)

- Για λ>0:

$$p(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- Η εκθετική κατανομή μπορεί να μοντελοποιήσει τον χρόνο μεταξύ τηλεφωνικών κλήσεων ή τον χρόνο μεταξύ των αφίξεων ενός λεωφορείου σε μία στάση.



$$\mathbb{E}[\mathrm{x}] = \frac{1}{\lambda}, \quad \sigma_x^2 = \frac{1}{\lambda^2}.$$

Η κατανομή Βήτα (The Beta Distribution)

- x~Beta($x|a,b$): Μια ΤΜ $x \in [0,1]$ έχει κατανομή βήτα με θετικές παραμέτρους *a,b* αν:

$$p(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

- Όπου η B(*a,b*):

$$B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}dx, \ \text{and } B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

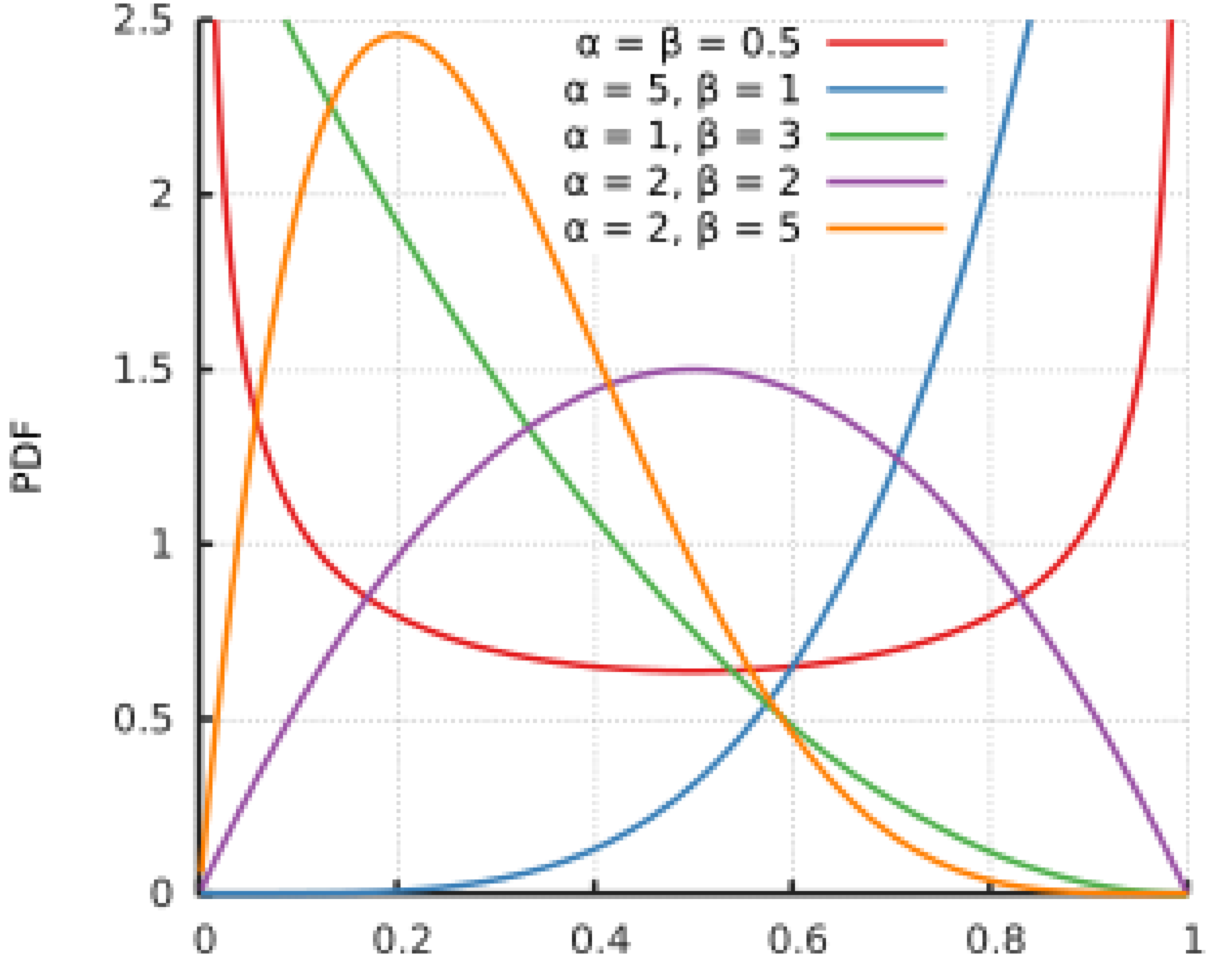
- Και Γ:

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx.$$

Η κατανομή Βήτα

$$\mathbb{E}[\mathrm{x}] = \frac{a}{a+b},$$

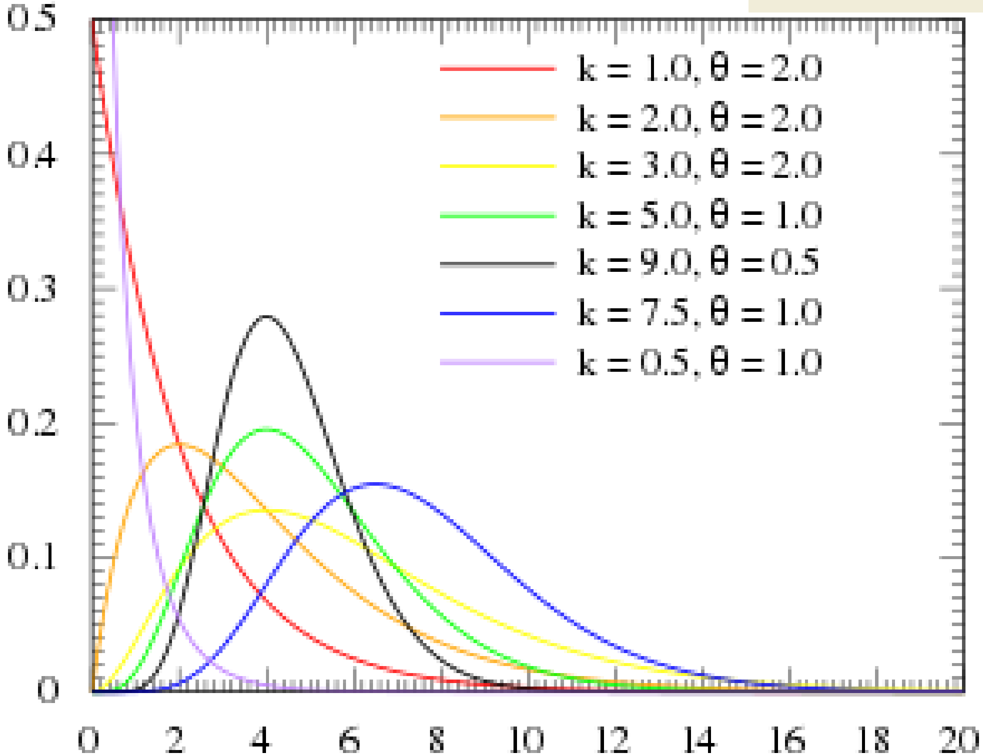
$$\sigma_x^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$



Η κατανομή Γάμα
(The Gamma Distribution)

- x~Gamma($x|a,b$)

$$p(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$



$$\mathbb{E}[\mathrm{x}] = \frac{a}{b}, \quad \sigma_x^2 = \frac{a}{b^2}.$$

The gamma distribution also takes various shapes by varying the parameters. For $a < 1$, it is strictly decreasing and $p(x) \longrightarrow \infty$ as $x \longrightarrow 0$ and $p(x) \longrightarrow 0$ as $x \longrightarrow \infty$.