ΑΡΙΣΤΟΤΕΛΕΙΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ

ΑΝΟΙΚΤΑ
ΑΚΑΔΗΜΑΪΚΑ
ΜΑΘΗΜΑΤΑ

Μαθηματικά Και Στατιστική Στη Βιολογία

Ενότητα 3 : Κατανομές και Παράμετροι

Ι. Αντωνίου, Χ. Μπράτσας
Τμήμα Μαθηματικών

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

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» 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

Κατανομές και Παράμετροι

Περιεχόμενα Ενότητας (1 από 4)

1. Κατανομή Πιθανότητας Παρατηρήσιμων Μεταβλητών
2. Ισοδύναμες & Ισόνομες Μεταβλητές
3. Κατανομές Εγχειρίδιο Κατάλογος
4. Διακριτές Κατανομές με Φορέα Πεπερασμένο
5. Διακριτές Κατανομές με Φορέα Άπειρο
6. Συνεχείς Κατανομές με Φορέα Φραγμένο Διάστημα
7. Συνεχείς Κατανομές με Φορέα Ημι-ευθεία
8. Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία
9. Στατιστικές Παράμετροι Κατανομών Πιθανότητας
10. Παράμετροι Θέσης
11. Παράδειγμα

Περιεχόμενα Ενότητας (2 από 4)

Περιεχόμενα Ενότητας (3 από 4)

Περιεχόμενα Ενότητας (4 από 4)

Σκοποί Ενότητας

- Στην Ενότητα 3 παρουσιάζονται οι κατανομές πιθανοτήτων καθώς και οι παράμετροι θέσης, διασποράς και σχήματος των κατανομών.

Κατανομή Πιθανότητας Παρατηρήσιμων Μεταβλητών (1 από 2)

- Η Κατανομή Πιθανότητας Παρατηρησιμων Μεταβλητών
- Αριθμητικές ΠΜ: X:Y→ $\mathbb{R} \cup \{-\infty, \infty\}$
- Πεπερασμένες Διακριτές ΠΜ: σ ={ x1 , x2 , …, xN}
- Άπειρες Διακριτές ΠΜ: σ ={x1 , x2 , …}
- Συνεχείς ΠΜ: σ ⊆ $\mathbb{R} \cup \{-\infty, +\infty\}$
- F(x)=P[X(y)≤x]
- **Διακριτές ΠΜ**: σ ={ x1 , x2 , …}
  $p(x_\kappa)=F(x_\kappa) - F(x_\kappa - 1)$ & $F(x_\kappa)= p(x_1) + p(x_2) + \ldots + p(x_\kappa)$, κ=1,2,…
- **Συνεχείς ΠΜ**: σ ⊆ {-∞, +∞}

$$p(\xi) = \frac{dF}{dx}(\xi), \quad F(\xi) = \int_{-\infty}^{\xi} dy p(y)$$

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Κατανομή Πιθανότητας Παρατηρήσιμων Μεταβλητών (2 από 2)

- F(x) Συνάρτηση Κατανομής της ΠΜ Χ(Cumulative Distribution Function of the RV T)
- ρ(x) Συνάρτηση Πιθανότητας της ΠΜ Χ (probability Function of the RV T)
- $\overline{F}(x) = 1 - F(x)$ Συνάρτηση Επιβίωσης ή Αξιοπιστίας της ΠΜ Χ (Survival Function or Reliability Function of the RV X)

- Τα Προβλήματα Αριθμητικών ΠΜ επιλύονται ως

  Προβλήματα Πραγματικών Κατανομών Πιθανότητας

  Με φάσμα (σύνολο τιμών) ένα σύνολο πραγματικών αριθμών

  Συμβολικές Μεταβλητές με φάσμα (σύνολο τιμών) ένα σύνολο σύμβολων: Θεωρία Πληροφορίας.

Παράδειγμα

Μεταβλητή X: Άθροισμα Ενδείξεων 2 Ζαριών	Παρατηρησιμα Γεγονότα Observable Events	Πιθανότητα Probability P(x)	Συνάρτηση Κατανομής F(x)	Συνάρτηση Επιβίωσης 1− *F(x)*
2	Ξ_2 **={ (1,1)}**	1/36	**1/36=2.78%**	**35/36**
3	Ξ_3 **={ (1,2), (2,1)}**	2/36	**3/36=8.4%**	**33/36**
4	Ξ_4 **={ (2,2), (1,3),(3,1)}**	3/36	**6/36=16.7%**	**30/36**
5	Ξ_5 **={ (1,4), (2,3),(3,2), (4,1)}**	4/36	**10/36=27.8%**	**26/36**
6	Ξ_6 **={ (1,5), (2,4),(3,3), (4,2), (5,1)}**	5/36	**15/36=41.7%**	**21/36**
7	Ξ_7 **={ (1,6), (2,5),(3,4), (4,3), (5,2), (6,1)}**	6/36	**21/36=58.3%**	**15/36**
8	Ξ_8 **={ (2,6), (3,5),(4,4), (5,3), (6,2)}**	5/36	**26/36=72.2%**	**10/36**
9	Ξ_9 **={ (3,6), (4,5),(5,4), (6,3)}**	4/36	**30/36=83.3%**	**6/36**
10	Ξ_{10} **={ (4,6), (5,5),(6,4)}**	3/36	**33/36=91.7%**	**3/36**
11	Ξ_{11} **={ (5,6), (6,5)}**	2/36	**35/36=97.2%**	**1/36**
12	Ξ_{12} **={ (6,6)}**	1/36	**36/36=100%**	0

Ισοδύναμες & Ισόνομες Μεταβλητές

- Ορισμός: **Ισοδύναμες Μεταβλητές**

$P[A(y) \neq B(y)]=0$

- Ορισμός: **Ισόνομες Μεταβλητές**

$P[A] = P[B]$, $E[A] = E[B]$

- Θεώρημα

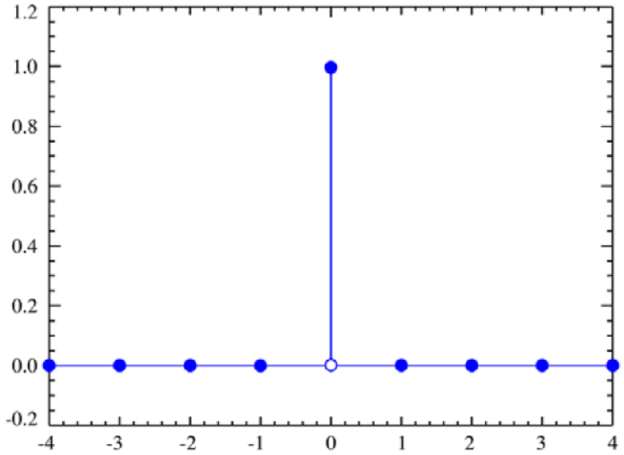
Οι Ισοδύναμες Μεταβλητές είναι Ισονομες.

Περισσότερα για τις Κατανομές

- Κατανομές Εγχειρίδιο Κατάλογος

www.stat.rice.edu/~dobelman/textfiles/DistributionsHandbook.pdf
http://en.wikipedia.org/wiki/List_of_probability_distributions

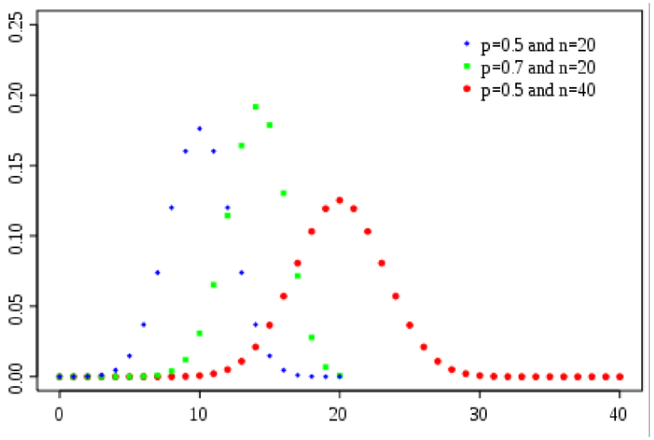
Διακριτές Κατανομές με Φορέα Πεπερασμένο (1 από 5)



- Διακριτές κατανομές με φορέα πεπερασμένο

  deterministic distribution: the support has only one value
- The discrete uniform distribution, where all elements of a finite set are equally likely. This is the theoretical distribution model for a balanced coin, an unbiased die, a casino roulette, or the first card of a well-shuffled deck.

Διακριτές Κατανομές με Φορέα Πεπερασμένο (2 από 5)



- The Bernoulli distribution, which takes value 1 with probability p and value 0 with probability q = 1 − p.
- The Rademacher distribution, which takes value 1 with probability 1/2 and value −1 with probability 1/2.
- The binomial distribution, which describes the number of successes in a series of independent

  Yes/No experiments all with the same probability of success.

Διακριτές Κατανομές με Φορέα Πεπερασμένο (3 από 5)

- Binomial Distribution

$$\rho(x;t,\pi) = \binom{t}{x} \pi^{x} (1-\pi)^{t-x}$$

$\binom{t}{x} = \dfrac{t!}{x!(t-x)!} = C_t^x$ Ο Διωνυμικος Συντελεστης = Binomial Coefficient

n= t = 0,1,2,…,M the number of trials,

x=0,1,2,…,t the number of relevant events among the t trials

π ∈ [0,1] the success probability in each trial

$$\rho[x_1, x_2, \ldots, x_n] = \binom{t}{x_1 x_2 \ldots x_n} \pi_1^{\ x_1} \pi_2^{\ x2} \ldots \pi_n^{\ x_n}$$

Διακριτές Κατανομές με Φορέα Πεπερασμένο (4 από 5)

- **Multinomial Distribution**

$$\binom{t}{x_1 x_2 \ldots x_n} = \frac{t!}{x_1! x_2! \ldots x_n!}$$ Ο Πολυωνυμικός Συντελεστής=Multinomial Coefficient

$$(\pi_1 + \pi_2 + \ldots + \pi_n)^t = \sum_{x_1, x_2, \ldots, x_n} \binom{t}{x_1 x_2 \ldots x_n} \pi_1^{x_1} \pi_2^{x_2} \ldots \pi_n^{x_n}$$

The sum over all n-ads (x1, x2,…, xn): xν = 0,1,2,…,n with x1+ x2+ …+xn= t

x1, x2,…, xn the occupation numbers of n independent states

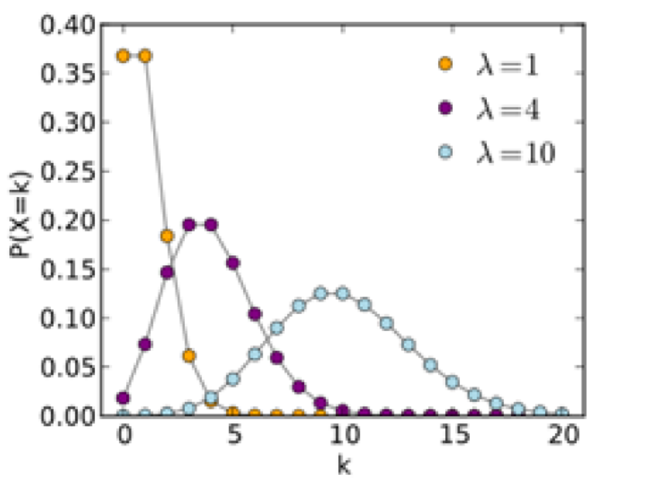
with probabilities p1 , p2 ,…,pn

ρ[x1, x2,…, xn] is the probability of occurrences of the events {1,2,3…n},

x1, x2,…, xn times correspondingly in t Bernoulli trials

The beta-binomial distribution, describes the number of successes in a series of independent Yes/No experiments with heterogeneity in the success probability.

Διακριτές Κατανομές με Φορέα Πεπερασμένο (5 από 5)



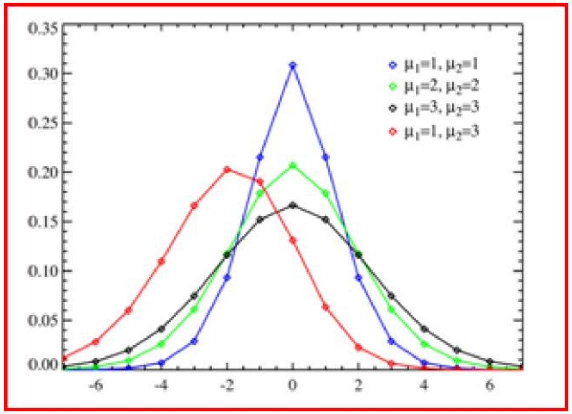
Fisher's noncentral hypergeometric distribution

- The Deterministic distribution at ξ, where X is certain to take the value ξ.
- The hypergeometric distribution, describes the number of successes in the first m of a series of n consecutive Yes/No experiments, if the total number of successes is known. This distribution arises when there is no replacement.
- The Poisson binomial distribution, describes the number of successes in a series of independent Yes/No experiments with different success probabilities.

Διακριτές Κατανομές με Φορέα Άπειρο (1 από 2)

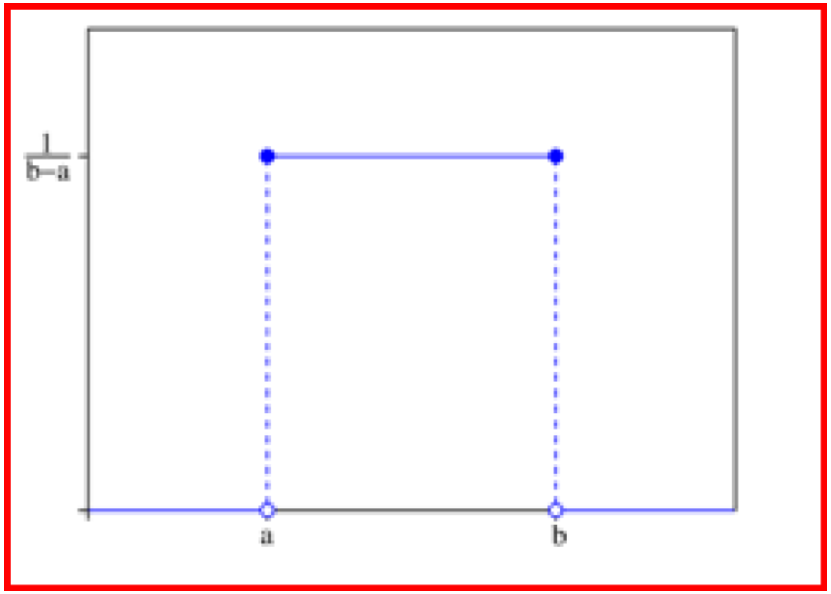
- **Wallenius' noncentral hypergeometric distribution**
- **Διακριτές κατανομές με φορέα άπειρο:**
- **The beta negative binomial distribution.**
- **The Boltzmann distribution**, important in statistical physics, describes the probabilities of the various discrete energy levels of a system in thermal equilibrium. It has a continuous analogue. Special cases include: The Gibbs distribution, The Maxwell–Boltzmann distribution, The Bose–Einstein distribution, The Fermi–Dirac distribution.
- **The extended negative binomial distribution.**
- **The geometric distribution**, describes the number of attempts needed to get the first success in a series of independent Yes/No experiments.

Διακριτές Κατανομές με Φορέα Άπειρο (2 από 2)



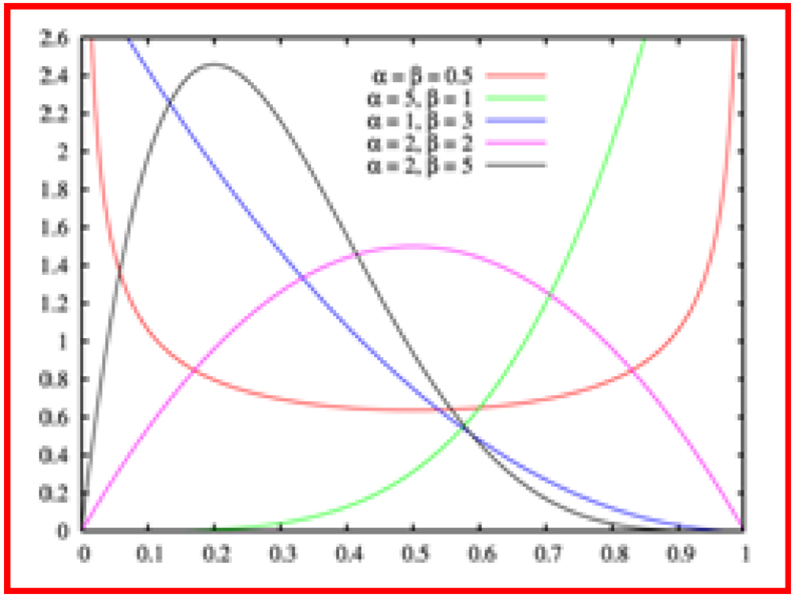
- **The Poisson distribution,** describes a very large number of individually unlikely events that happen in a certain time interval.
- **The Skellam distribution**, the distribution of the difference between two independent Poisson-distributed random variables.
- **The Yule–Simon distribution**
- **Zipf's law or the Zipf distribution** power-law distribution, the most famous example of which is the description of the frequency of words in the English language.

Συνεχείς Κατανομές με Φορέα Φραγμένο Διάστημα (1 από 2)



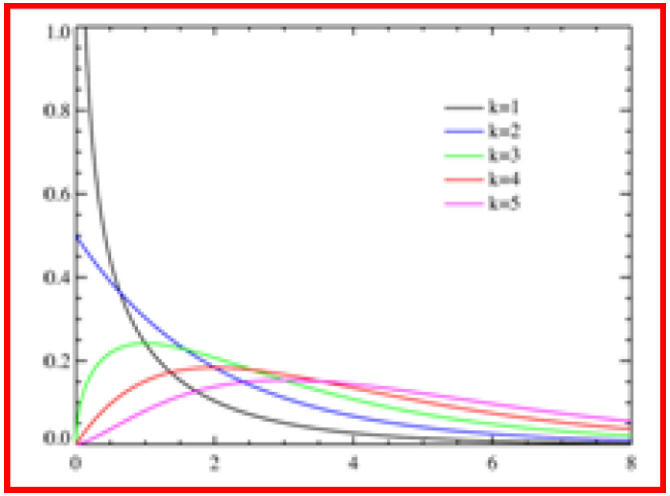
- **The Uniform distribution** on the Interval [a,b], all points are equally likely.

Συνεχείς Κατανομές με Φορέα Φραγμένο Διάστημα(2 από 2)



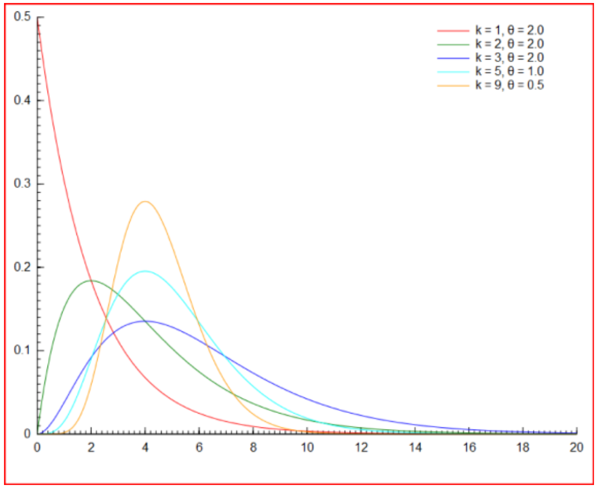
- **The Beta distribution** on [0,1], of which the uniform distribution is a special case, useful in estimating success probabilities.

Συνεχείς Κατανομές με Φορέα Ημι-ευθεία (1 από 3)



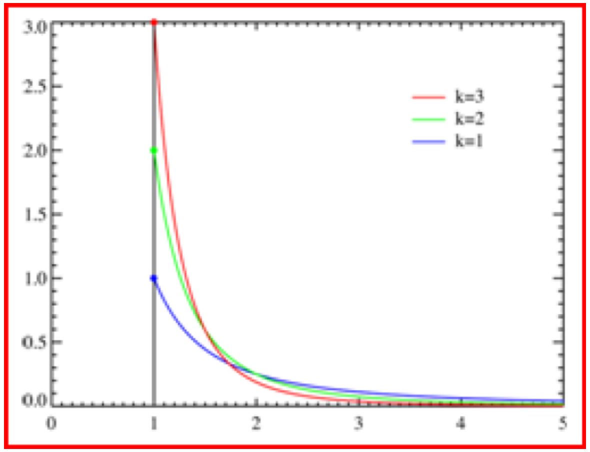
- **The chi distribution.**
- **The chi-squared distribution,** is the sum of the squares of N independent Gaussian Variables.
- It is a special case of the Gamma distribution, and it is used in goodness-of-fit tests in statistics.

Συνεχείς Κατανομές με Φορέα Ημι-ευθεία (2 από 3)



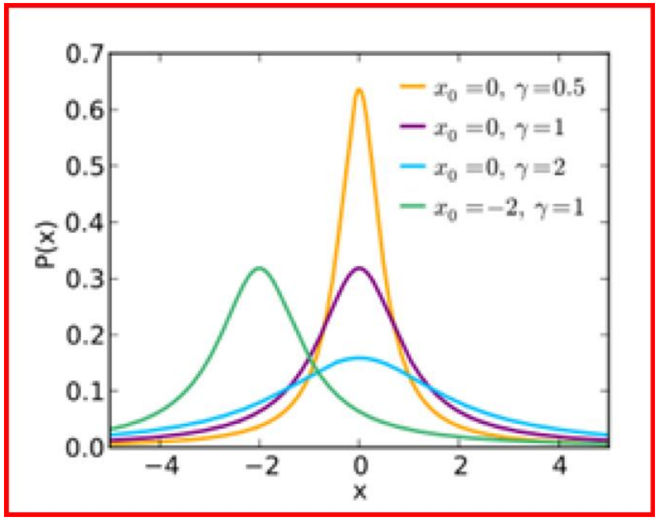
- **The exponential distribution**, describes the time between consecutive rare random events in a process with no memory.
- **The F-distribution**, is the distribution of the ratio of two (normalized) chi-squared distributed random variables, used in the analysis of variance. It is referred to as the beta prime distribution when it is the ratio of two chi-squared variates which are not normalized by dividing them by their numbers of degrees of freedom.
- **The noncentral F-distribution**
- **Fisher's z-distribution**
- **The Gamma distribution**, describes the time until N consecutive rare random events occur in a process with no memory.

Συνεχείς Κατανομές με Φορέα Ημι-ευθεία (3 από 3)



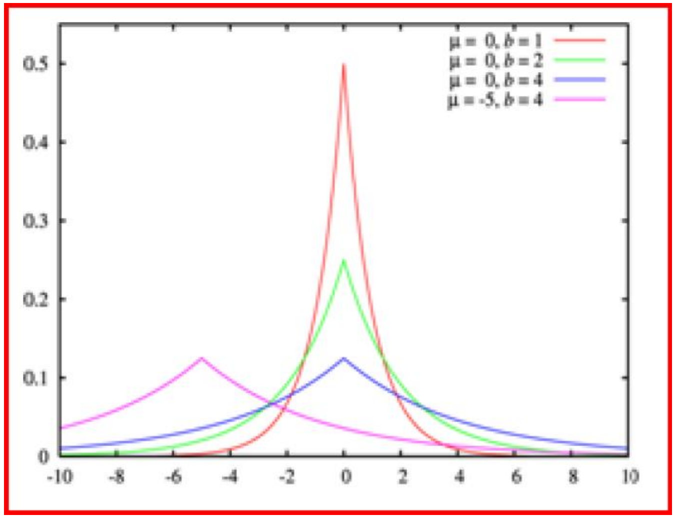
- **The Lévy distribution**
- **The log-logistic distribution**
- **The log-normal distribution**, describing variables which can be modelled as the product of many small independent positive variables.
- **The Mittag–Leffler distribution**
- **The Pareto distribution**, or "power law" distribution, used in the analysis of financial data and critical behavior.

Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (1 από 6)



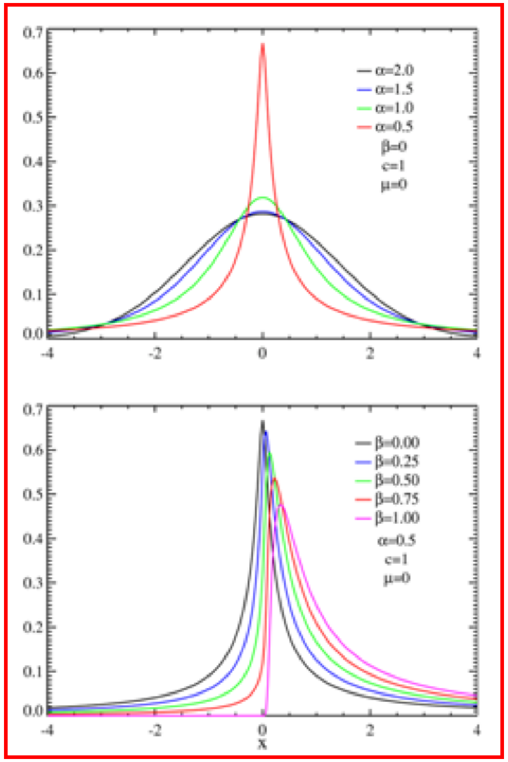
- **The Cauchy distribution,** has no expected value nor variance. In physics it is usually called a Lorentzian profile, and is associated with resonances, impact and natural spectral line broadening and quadratic stark line broadening.
- **Chernoff's distribution**
- **The Fisher-Tippett, extreme value, or log-Weibull distribution**
- **Fisher's z-distribution**

Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (2 από 6)



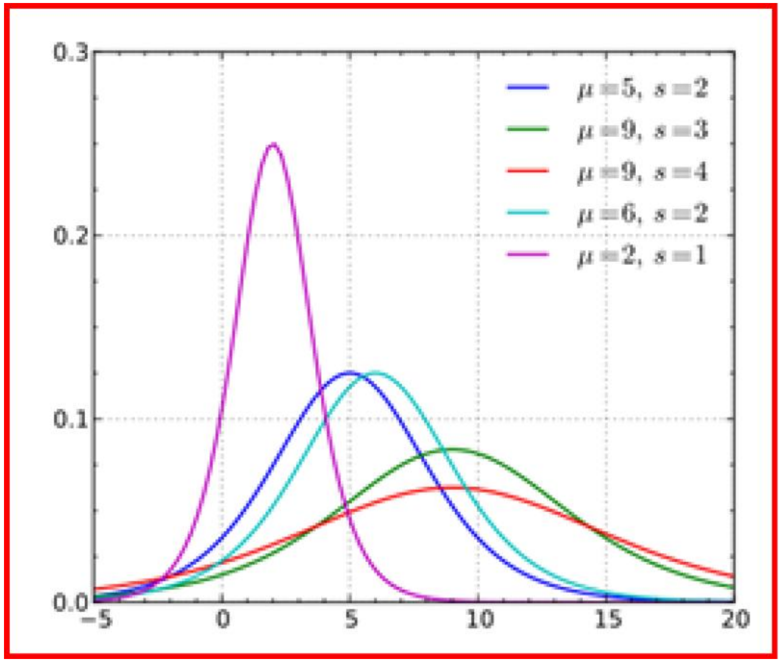
- **The generalized logistic distribution**
- **The generalized normal distribution**
- **The geometric stable distribution**
- **The Holtsmark distribution,** has finite expected value but infinite variance.
- **The hyperbolic distribution**
- **The hyperbolic secant distribution**
- **The Landau distribution**
- **The Laplace distribution**

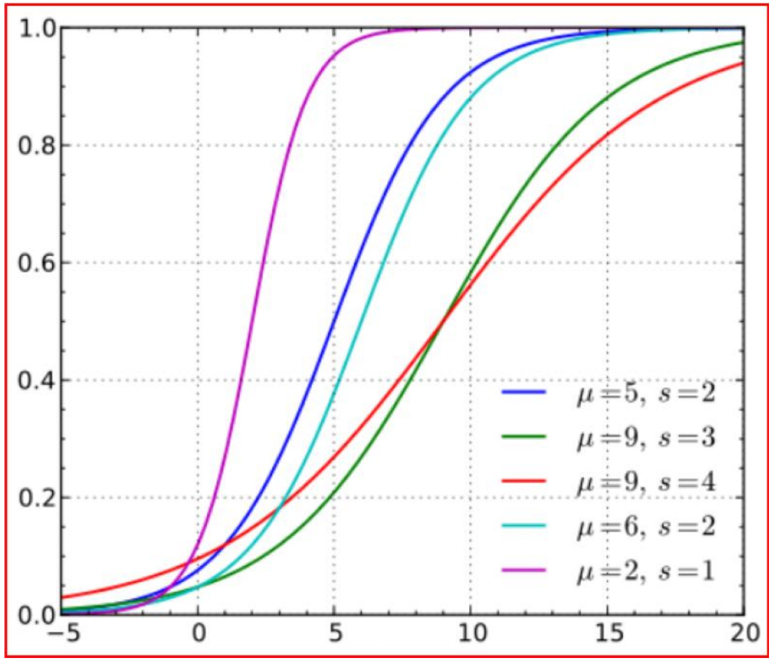
Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (3 από 6)



- **The Lévy skew alpha-stable distribution or stable distribution** is often used to characterize financial data and critical behavior; the Cauchy distribution, Holtsmark distribution, Landau distribution, Lévy distribution and normal distribution are special cases.

Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (4 από 6)

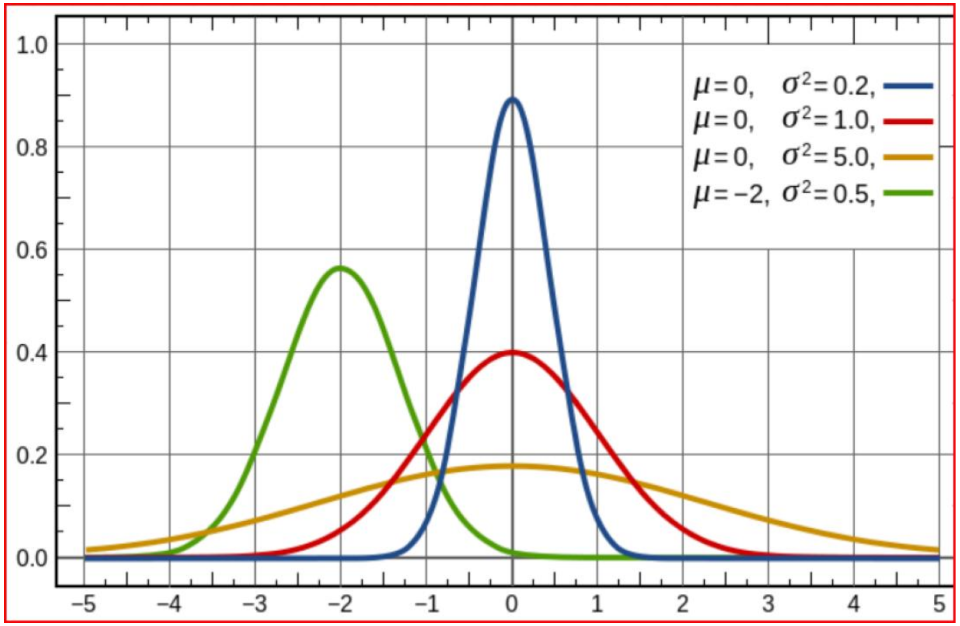




- The logistic distribution: Population Dynamics, Innovations, Neural networks.

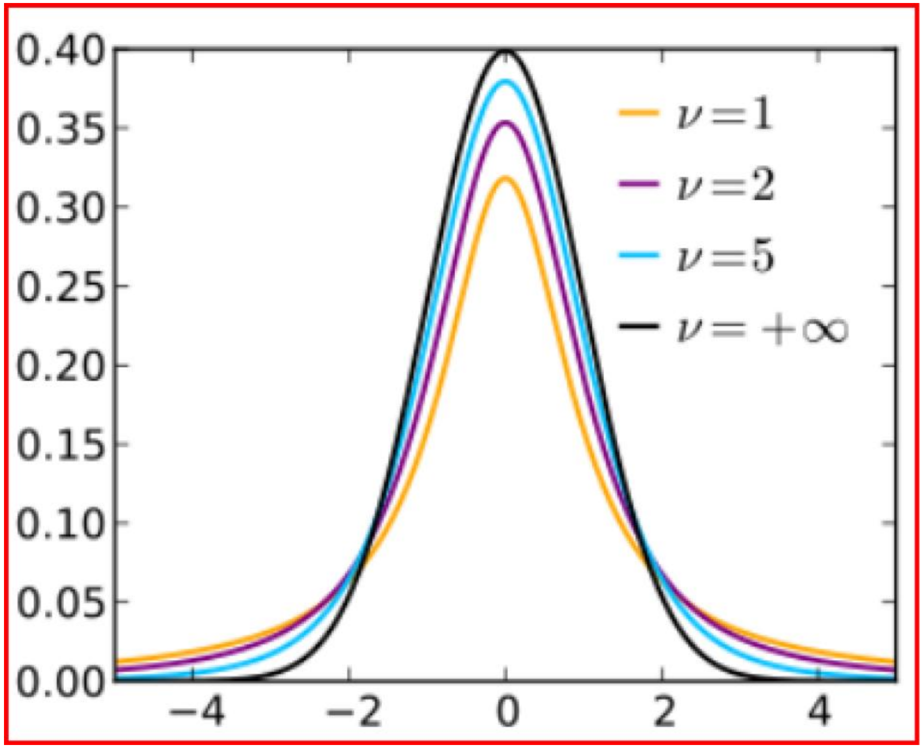
$$F(x)=\frac{1}{1+e^{-\frac{x-\mu}{s}}}$$

Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (5 από 6)



- **The normal distribution, also called the Gaussian or the bell curve.** Ubiquitous in nature and statistics due to the central limit theorem:
- every variable that can be modelled as a sum of many small independent, identically distributed variables with finite mean and variance is approximately normal.

Συνεχείς Κατανομές με Φορέα την Πραγματική Ευθεία (6 από 6)



- **Student's t-distribution,** useful for estimating unknown means of Gaussian populations.

Στατιστικές Παράμετροι Κατανομών Πιθανότητας

- **Παράμετροι Θέσης Κατανομής (Location Parameters):**
  - Μέση Τιμή (mean)
  - Ροπές (Moments)
  - Κορυφες (modes)
  - Διάμεσος (median)
  - Ποσοστημόρια (Quantiles, Percentiles)
- **Παράμετροι Διασποράς Κατανομής (Dispersion Parameters):**
  - Εύρος (range).
  - Μεταβλητότητα (variance)
  - Τυπική Απόκλιση (standard deviation)
  - Σχετικό Σφάλμα (relative error)= Συντελεστής μεταβλητότητας (CV)
  - Αποστάσεις Ποσοστημοριων
  - Εντροπία (και για κατηγορικές μεταβλητές)
- **Παράμετροι Σχήματος Κατανομής (Shape Parameters):**
  - Λοξότητα (skewness).
  - Κύρτωση (kurtosis).

Παράμετροι Θέσης

- **Μέση Τιμή**

$$E[X]=m=X(\mathrm{y}_1)p_1+X(\mathrm{y}_2)p_2+...=x_1p_1+x_2p_2+...=x_1(\mathrm{F}(x_1)-0)+x_2(F(x_2)-F(x_1))+...$$

$$E[X]=m=\int_Y X(\mathrm{y})\,\mathrm{dP}=\int_{-\infty}^{+\infty}xp(x)dx=\int_{-\infty}^{+\infty}xdF$$

- **ΣΧΟΛΙΑ**
  1) Πρέπει τα Αθροίσματα και τα Ολοκληρώματα να είναι πεπερασμένα.
  2) Η Μέση Τιμή είναι η πιο πιθανή τιμή, η βέλτιστη πρόβλεψη για την μέτρηση της μεταβλητής Χ.
  3) Η Μέση Τιμή είναι η βασική παράμετρος δια της όποιας ορίζονται οι περισσότερες άλλες παράμετροι.

Παράδειγμα

Μεταβλητή X: Άθροισμα Ενδείξεων 2 Ζαριών	Παρατηρησιμα Γεγονότα Observable Events	Πιθανότητα Probability P(x)	Συνάρτηση Κατανομής F(x)
2	Ξ_2 **={ (1,1)}**	1/36	**1/36=2.78%**
3	Ξ_3 **={ (1,2), (2,1)}**	2/36	**3/36=8.4%**
4	Ξ_4 **={ (2,2), (1,3),(3,1)}**	3/36	**6/36=16.7%**
5	Ξ_5 **={ (1,4), (2,3),(3,2), (4,1)}**	4/36	**10/36=27.8%**
6	Ξ_6 **={ (1,5), (2,4),(3,3), (4,2), (5,1)}**	5/36	**15/36=41.7%**
7	Ξ_7 **={ (1,6), (2,5),(3,4), (4,3), (5,2), (6,1)}**	6/36	**21/36=58.3%**
8	Ξ_8 **={ (2,6), (3,5),(4,4), (5,3), (6,2)}**	5/36	**26/36=72.2%**
9	Ξ_9 **={ (3,6), (4,5),(5,4), (6,3)}**	4/36	**30/36=83.3%**
10	Ξ_{10} **={ (4,6), (5,5),(6,4)}**	3/36	**33/36=91.7%**
11	Ξ_{11} **={ (5,6), (6,5)}**	2/36	**35/36=97.2%**
12	Ξ_{12} **={ (6,6)}**	1/36	**36/36=100%**

Μέση Τιμή

$$E[X]=m=2\frac{1}{36}+3\frac{2}{36}+4\frac{3}{36}+5\frac{4}{36}+6\frac{5}{36}+7\frac{6}{36}+8\frac{5}{36}+9\frac{4}{36}+10\frac{3}{36}+11\frac{2}{36}+12\frac{1}{36}$$

$$m=\frac{2}{36}+\frac{6}{36}+\frac{12}{36}+\frac{20}{36}+\frac{30}{36}+\frac{42}{36}+\frac{40}{36}+\frac{36}{36}+\frac{30}{36}+\frac{22}{36}+\frac{12}{36}=7$$

Ιδιότητες Μέσης Τιμής

- **Γραμμικότης**

  E[cX]= cE[X], c πραγματικος αριθμος

  E[X1+X2]= E[X1] + E[X2]
- **Θετικότης**

  E[X] ≥0, εάν X ≥0
- **Κανονικοποίηση**

  E[1]=1, 1(y)=1, η Μεταβλητη με σταθερη τιμη 1

  E[O]=0, O(y)= 0, η Μεταβλητη με σταθερη τιμη 0
- **E[X1] ≤ E[X2], αν X1 ≤ X2**
- **|E[X]| ≤ E[|X|]**
- **E[φ(X)] ≠ φ(E[X]), φ:X→$\mathbb{R}$ πραγματική συνάρτηση της μεταβλητής X.**

Οπου: E[φ(X)] = φ(x1) p1+ φ(x2)p2+… για διακριτές μεταβλητές

$= \int_{-\infty}^{+\infty} \varphi(x)\rho(\mathrm{x})\,\mathrm{dx}$ **για συνεχείς μεταβλητές**

Ανισότητα Markov

- **Θεώρημα: Ανισότητα Markov**

  $P[|\mathrm{X}| \geq \alpha] \leq \frac{\mathrm{E}|\mathrm{X}|}{a}, a > 0$ όπου $\frac{\mathrm{E}|\mathrm{X}|}{a}$ είναι το ποσοστό τιμών της Μεταβλητής X με μέγεθος τουλάχιστον α
- **Παράδειγμα**: X = Το Άθροισμα των ενδείξεων 2 Ζαριών

$$P[|\mathrm{X}| \geq 7] \leq \frac{\mathrm{E}|\mathrm{X}|}{8} = \frac{7}{7} = 1$$

$$P[|\mathrm{X}| \geq 7] = \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{21}{36} = 0.58$$

$$P[|\mathrm{X}| \geq 8] \leq \frac{\mathrm{E}|\mathrm{X}|}{8} = \frac{7}{8} = 0.875$$

$$P[|\mathrm{X}| \geq 8] = \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{15}{36} = 0.41$$

$$P[|\mathrm{X}| \geq 9] \leq \frac{\mathrm{E}|\mathrm{X}|}{8} = \frac{7}{9} = 0.778$$

$$P[|\mathrm{X}| \geq 9] = \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = 0.278$$

Ροπή

- **Ορισμός: Ροπη r-τάξεως, r=1,2,3,…**

$$m_r = E[X^r] = (x_1)^r p_1 + (x_2)^r p_2 + \ldots$$

$$= \int_{-\infty}^{+\infty} x^r \rho(x) dx$$

- ΣΧΟΛΙΟ

  1. m1 = m = E[X], η πρώτη ροπή είναι η Μέση Τιμή.

  2. m2 = E[X2] η «Ισχύς» ή «Ροπή Αδρανείας» ή Μέση Τιμή Τετραγώνου (mean square) της Μεταβλητής X.

  3. Αν γνωρίζουμε τις ροπές, γνωρίζουμε την κατανομή, υπό προϋποθέσεις (moment problem). Συνήθως αρκούν οι 4 πρώτες ροπές για προσέγγιση της κατανομής στην πράξη.

Παράδειγμα

- **Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών**

$$E[X^2] = 2^2\frac{1}{36} + 3^2\frac{2}{36} + 4^2\frac{3}{36} + 5^2\frac{4}{36} + 6^2\frac{5}{36} + 7^2\frac{6}{36} + 8^2\frac{5}{36} + 9^2\frac{4}{36} + 10^2\frac{3}{36} + 11^2\frac{2}{36} + 12^2\frac{1}{36}$$

$$E[X^2] = 4\frac{1}{36} + 9\frac{2}{36} + 16\frac{3}{36} + 25\frac{4}{36} + 36\frac{5}{36} + 49\frac{6}{36} + 64\frac{5}{36} + 81\frac{4}{36} + 100\frac{3}{36} + 121\frac{2}{36} + 144\frac{1}{36}$$

$$E[X^2] = 0.11 + 0.5 + 1.33 + 2.78 + 8 + 8.17 + 8.89 + 9 + 8.33 + 6.72 + 4$$

$E[X^2] = 54.83$ **Η ισχύς της X**

Κορυφές

- **Ορισμός: Κορυφές η Επικρατούσες Τιμές (Μονό)**

  Οι τιμές x=ξmode για τις όποιες η Κατανομή ρ(x) έχει (τοπικά) μέγιστα.
- **Μονοκόρυφες κατανομές (Unimodal)** έχουν 1 μέγιστο.
- **Δικόρυφες κατανομές (Bimodal)** έχουν 2 μέγιστα.
- Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών.

Διάμεσος

- **Ορισμός: Διάμεσοι**

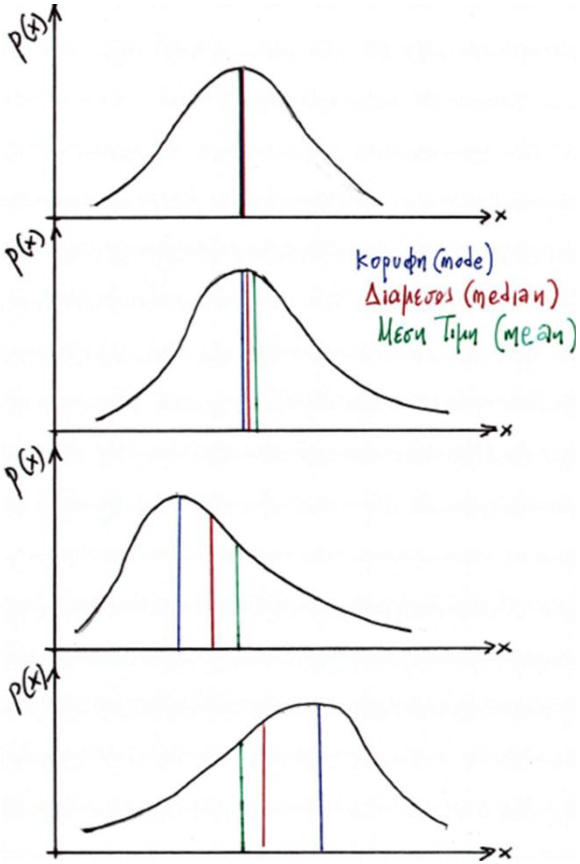
  Η τιμή $x=\xi_{1/2} : P[x < x_{1/2}] \le \frac{1}{2} \le P[x \le x_{1/2}]$

- **Παράδειγμα X:**

  το άθροισμα των ενδείξεων 2 ζαριών

  Διάμεσος $\xi_{1/2}$ **= 7 = m =** $\xi_{\mathbf{mode}}$

Σχέση Μέσης Τιμής, Διάμεσου, Κορυφής



- Θεώρημα
- **Για συμμετρικές κατανομές:**
  Μέση Τιμή = Διάμεσος = Κορυφή
- **Για λίγο ασύμμετρες κατανομές:**
  Μέση Τιμή − Κορυφή ≈ 3(Μέση Τιμή Διάμεσος)
- **Για ασύμμετρες κατανομές προς τα αριστερά: α3 > 0**
  Μέση Τιμή > Διάμεσος > Κορυφή
- **Για ασύμμετρες κατανομές προς τα δεξιά: α3 < 0**
  Μέση Τιμή < Διάμεσος < Κορυφή

α-Ποσοστημόριο

- **Ορισμός: α-Ποσοστημόριο, 0<α<1 α-Quantile**

  Η τιμή της μεταβλητής x = xα με πιθανότητα το πολύ α

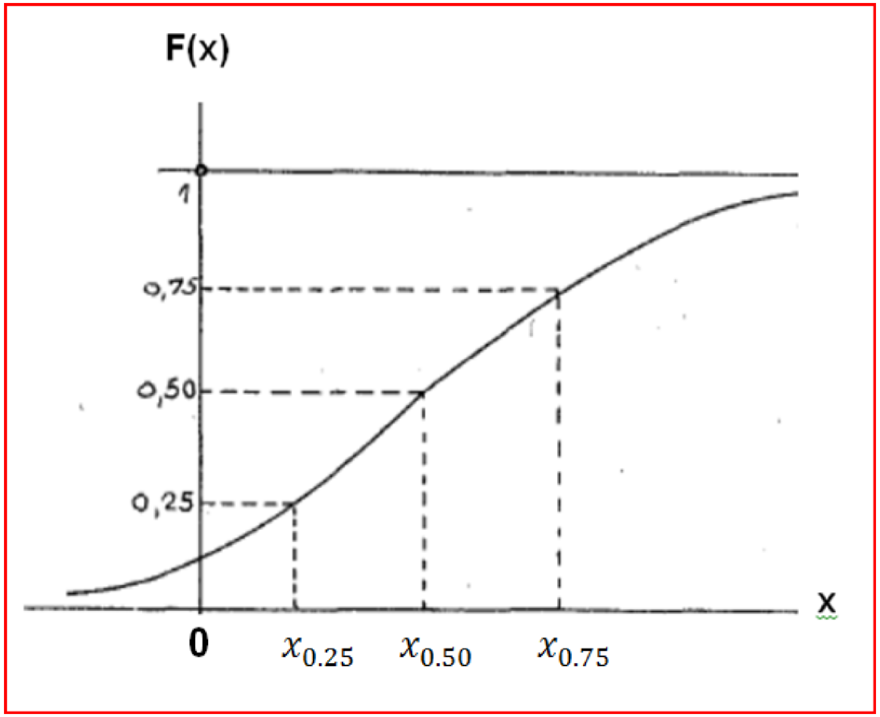
  $\mathbf{P[X< x_{\alpha}] \leq \alpha \leq P[X \leq x_{\alpha}]}$

  Αν F συνεχής και γνησίως αύξουσα, τότε το α Ποσοστημόριο είναι η λύση της Εξίσωσης

  [x] = α

  $\mathbf{x_{\alpha} = F^{-1} [\alpha]}$

Παράδειγμα
(1 από 2)



- Τεταρτημόρια:

  $x_{1/4} = x_{0.25,}$ το πρώτο τεταρτημόριο

  $x_{1/2} = x_{0.50}$ το δεύτερο τεταρτημόριο (η διάμεσος)

  $x_{3/4} = x_{0.75,}$ το τρίτο τεταρτημόριο
- **Παράδειγμα: Χ = Το Άθροισμα των ενδείξεων 2 Ζαριών**

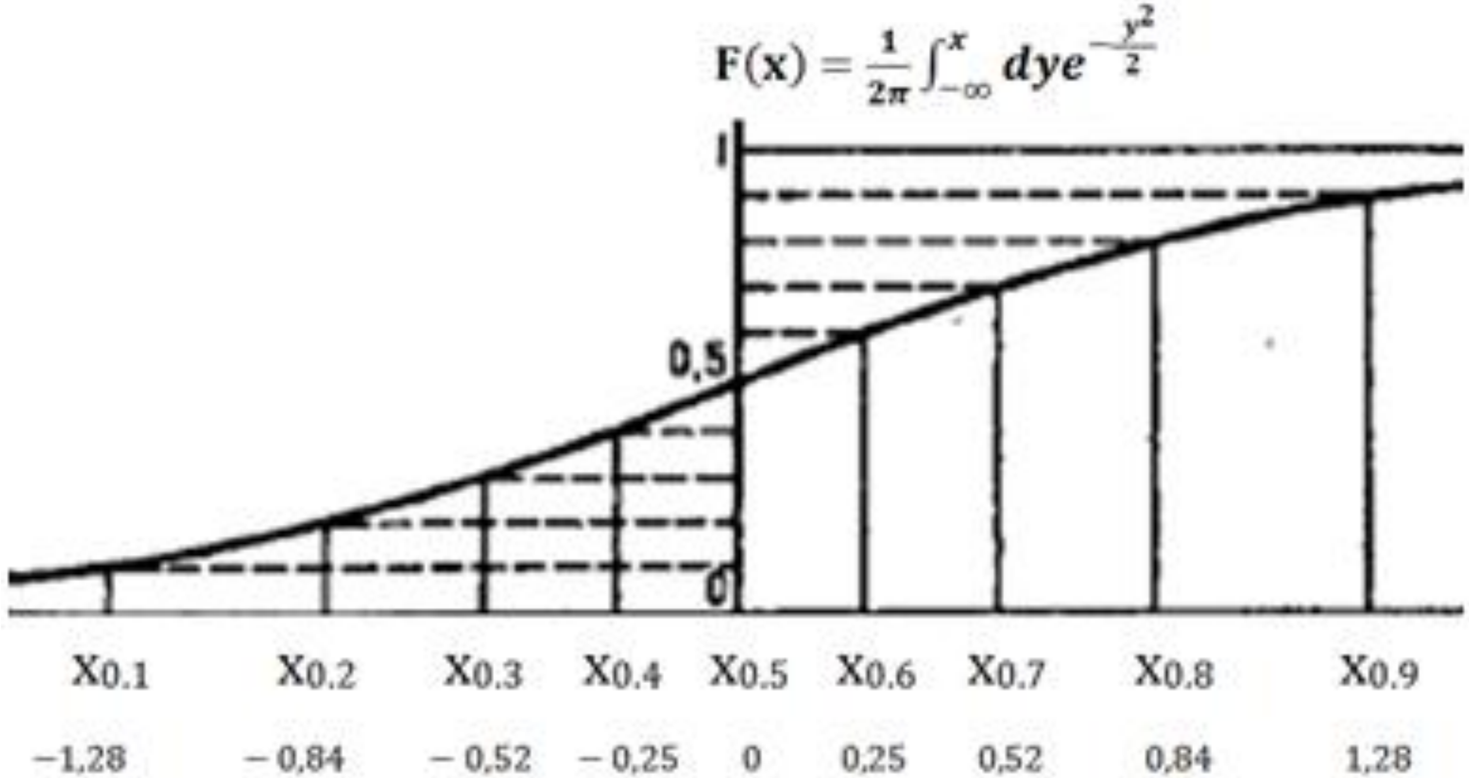
  $x_{1/4} = x_{0.25} = 4$

  $x_{1/2} = x_{0.50} = 7$

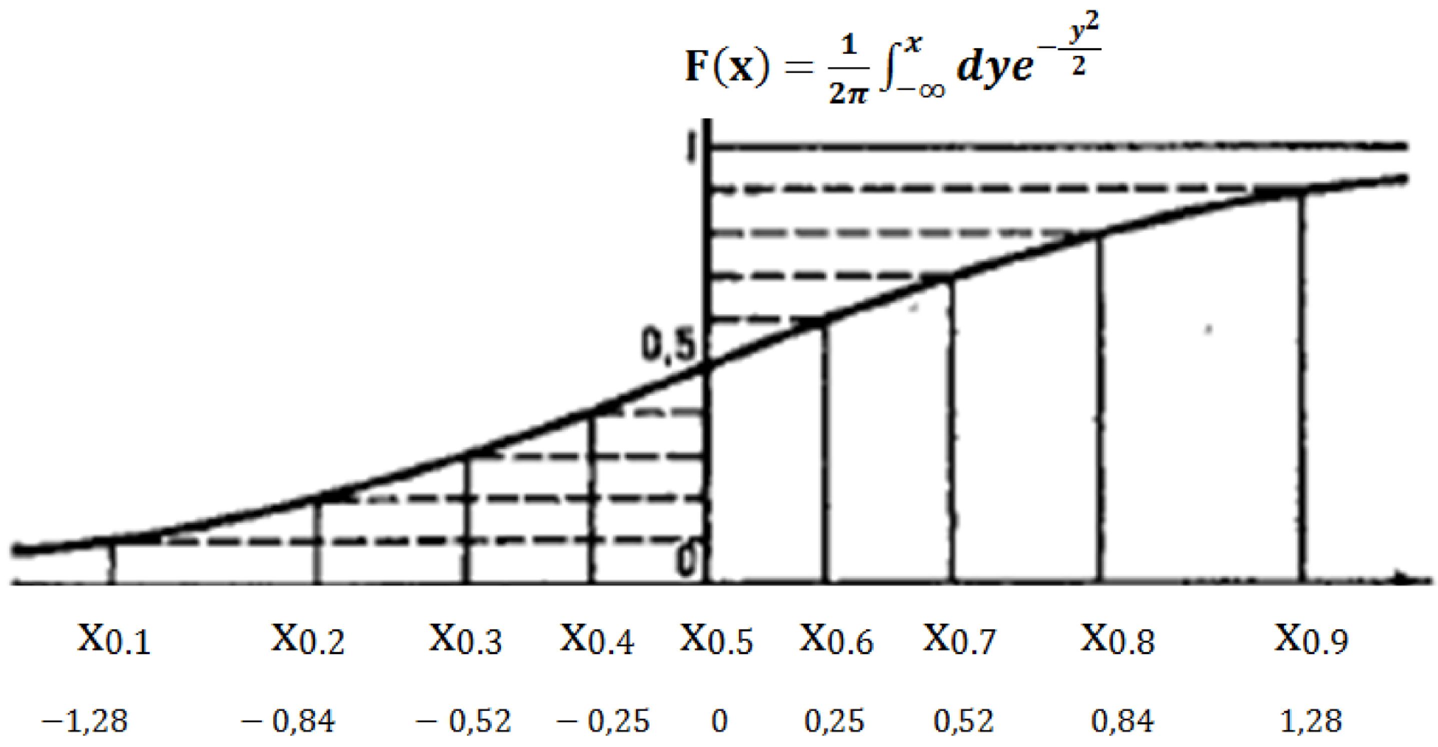
  $x_{3/4} = x_{0.75} = 9$

Παράδειγμα (2 από 2)

- $X_{0.1}$, $x_{0.2}$,…$x_{0.9}$ τα 9 δεκατημόρια (deciles)
- Παράδειγμα: τα 9 δεκατημόρια της Κανονικής Κατανομής:



Παράμετροι Μεταβλητότητας Κατανομής



Παράμετροι Μεταβλητότητας Κατανομής (Dispersion Parameters)

Εύρος

- **Εύρος**:

  Η έκταση του φάσματος: $|x_{max} - x_{min}|$
- **Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών**

  Εύρος $|x_{max} - x_{min}| = 12 – 2 = 10$

Διακύμανση

- **Παράμετροι Διαποράς Κατανομής (Dispersion Parameters)**
- **Διακύμανση (Fluctuation):** Η τιμη (X – m) = (X – E[X]).
- **ΣΧΟΛΙΟ**

  Η Διακύμανση δείχνει ποσό απέχει η μέτρηση από την Μέση Τιμή Συνεπώς η Μέση Διακύμανση είναι εκτίμηση της Μεταβλητότητας.
- **Θεώρημα**

  Η μέση Διακύμανση μηδενίζεται: E[(X – E[X])]=0

  Απόδειξη

  E[(X – E[X])]= E[X] – E[E[X] = E[X] – E[X] = 0
- **ΣΧΟΛΙΟ**

  Είμαστε υποχρεωμένοι να ορίσουμε άλλες παραμέτρους για την Μεταβλητότητα

Κεντρική Ροπή

- **Κεντρική Ροπή r-τάξεως, r = 1,2,3,…**

$$c_r = E[(X-m)^r] = (x_1-m)^r p_1 + (x_2-m)^r p_2 + \ldots$$

$$= \int_{-\infty}^{+\infty} (x-m)^r \rho(x)dx$$

- ΣΧΟΛΙΑ

  1. c_1 = E[(X – E[X])] = 0
  2. c_2 = E[(X – E[X])2] (η ροπή 2ας τάξεως της (X – m))

  Μέση Τιμή του Τετραγώνου (mean square) της Διακύμανσης

  Είναι η απλούστερη εκτίμηση της Μεταβλητότητας
- **Θεώρημα**

  Οι Κεντρικές Ροπές άρτιας Τάξεως Συμμετρικών ως προς τον Μέσο Κατανομών, μηδενίζονται.

Διασπορά
(1 από 3)

- **Μεταβλητότητα (Variance)**

$$\operatorname{var}(X) = E[(X-m)^2] = (x_1 - m)^2 p_1 + (x_2 - m)^2 p_2 + \dots$$

$$= \int_{-\infty}^{+\infty} (x-m)^2 \rho(x) dx$$

- Θεώρημα
- Ιδιότητες της Μεταβλητότητας
- **var[X] ≥0**
- **var [X+c] = var[X]**
- **var [cX]= c^2 var [X] , c πραγματικός αριθμός**
- **var [X_1+X_2] = var [X_1] + var [X_2]**
- **var [X] = E[X^2]-(E[X])^2 = E[X^2]- m^2**
- Απόδειξη: Από τον ορισμό με Άλγεβρα

Διασπορά
(2 από 3)

- **Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών**

$$\mathrm{E}[(\mathrm{X}-7)^2] = (2-7)^2\frac{1}{36} + (3-7)^2\frac{2}{36} + (4-7)^2\frac{3}{36} + (5-7)^2\frac{4}{36} + (6-7)^2\frac{5}{36} + (7-7)^2\frac{6}{36} +$$
$$(8-7)^2\frac{5}{36} + (9-7)^2\frac{4}{36} + (10-7)^2\frac{3}{36} + (11-7)^2\frac{2}{36} + (12-7)^2\frac{1}{36}$$

$$\mathrm{E}[(\mathrm{X}-7)^2] = 5^2\frac{1}{36} + 4^2\frac{2}{36} + 3\frac{3}{36} + 2^2\frac{4}{36} + 1^2\frac{5}{36} + 0^2\frac{6}{36} + 1^2\frac{5}{36} + 2^2\frac{4}{36} + 3^2\frac{3}{36} + 4^2\frac{2}{36} + 5\frac{1}{36}$$

$$\mathrm{E}[(\mathrm{X}-7)^2] = 25\frac{1}{36} + 16\frac{2}{36} + 9\frac{3}{36} + 4\frac{4}{36} + 1\frac{5}{36} + 0 + 1\frac{5}{36} + 4\frac{4}{36} + 9\frac{3}{36} + 16\frac{2}{36} + 25\frac{1}{36}$$

$$\mathrm{E}[(\mathrm{X}-7)^2] = 4.945$$

Διασπορά
(3 από 3)

- **Θεώρημα**
- Ανισότητα Chebychev

$$P[|X-m|\geq \alpha] = \frac{\mathrm{var}[X]}{\alpha^2}, \alpha > 0$$

- όπου $\frac{\mathrm{var}[X]}{\alpha^2}, \alpha > 0$ είναι το ποσοστό τιμών της Μεταβλητής X με απόσταση από τη μέση τιμή τουλάχιστον α.
- Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών $P[|X-7|\geq 5] = \frac{4.995}{5^2} = 0.1998$

Τυπική Απόκλιση

- **Τυπική Απόκλιση**
- (Standard Deviation , the root mean square fluctuation, the rms fluctuation)

$$\sigma = \sqrt{\operatorname{var}[X]}$$

- Η Τυπική Απόκλιση είναι η συνήθης εκτίμηση των σφαλμάτων (θεωρούνται ως αποκλίσεις από την μέση τιμή).
- Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών

$$E[(\mathrm{X}-7)^2] = 4.945$$

$$\sigma = \sqrt{\operatorname{var}[X]} = \sqrt{4.945} = 2.23$$

- ΣΧΟΛΙΟ

  Για να συγκρίνουμε τα σφάλματα διαφορετικών Μεταβλητών χρησιμοποιούμε το σχετικό σφάλμα.

Σχετικό ή Ποσοστιαίο Σφάλμα ή Απόκλιση

- **Σχετικό η Ποσοστιαίο Σφάλμα ή Απόκλιση (από την Μέση Τιμή) relative deviation**

$$\frac{\sigma}{m} = \frac{\sqrt{\mathrm{var}[X]}}{E[X]}$$

- ΣΧΟΛΙΟ
- $\frac{\sigma}{m}$ μικρό ⇔ οι τιμές της X είναι πλησίον της μέσης τιμής m με μεγάλη πιθανότητα
- $\frac{\sigma}{m}$ μεγάλο ⇔ οι τιμές της X είναι μικράν της μέσης τιμής m με μεγάλη πιθανότητα
- Παράδειγμα: X = Το Άθροισμα των ενδείξεων 2 Ζαριών

$$\frac{\sigma}{m} = \frac{\sqrt{\mathrm{var}[X]}}{E[X]} = \frac{2.23}{7} = 0.319$$

Άσκηση

- **Υπολογίστε το σχετικό σφάλμα για το Άθροισμα των ενδείξεων 2 Ζαριών**

1.Για 2 όμοια ζαριά με $p(1)=\frac{1}{3}, p(2)=\mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{2}{15}$

2. Για 2 όμοια ζαριά με $p(1)=\frac{1}{3}, p(2)=\frac{1}{6}, \mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{1}{8}$

3. Ένα ζάρι με $p(1)=p(2)=\mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{1}{6}$

Ένα ζάρι με $p(1)=\frac{1}{3}, p(2)=\mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{2}{15}$

4. Ένα ζάρι με $p(1)=p(2)=\mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{1}{6}$

Ένα ζάρι με $p(1)=\frac{1}{3}, p(2)=\frac{1}{6}, \mathrm{p}(3)=\mathrm{p}(4)=p(5)=p(6)=\frac{1}{8}$

Συγκρίνατε τα Αποτελέσματα των 5 περιπτώσεων.

Αποστάσεις Ποσοστημορίων

- Η απόσταση των συμπληρωματικών α-Ποσοστημορίων $x_α$ και $x_{1-α}$, $|x_α - x_{1-α}|$ είναι μια Εκτιμηση της συγκέντρωσης των τιμών της μεταβλητής Χ.
- Η διαφορά του 1ου από το 3ο τεταρτημόριο, είναι ανάλογη της διασποράς.
- $|x_{0.75} – x_{0.25}|$ το ενδοτεταρτημοριακό πλάτος (interquartile range).

Εντροπία Κατανομών (1 από 2)

- **Εντροπία Διακριτής Κατανομής**

$$J = -\sum_{\mathrm{N}} p_\nu \log_2 p_\nu$$

ν = οι τιμές μιας Μεταβλητής (Αριθμητικής ή Κατηγορικής) τις οποίες Παρατηρώ στο Πείραμα ή Υποθέτω στο πλαίσιο μιας Θεωρίας

$$\log_2 x = \frac{\ln x}{\ln 2}, \;\; \ln 2 = 0.69314718055994 5$$

Εντροπία Κατανομών
(2 από 2)

- **Εντροπία Συνεχούς Κατανομής**

$$J = -\int_{-\infty}^{\infty} dx p(x) \ln p(x)$$

- Η Εντροπία είναι εκτιμήτρια:
  - **της Αταξίας-Τυχαιότητος της κατανομής,**
  - **της Ποικιλότητος- Πολυπλοκότητος της κατανομής,**
  - **της Αβεβαιότητας- Ρίσκου Πρόβλεψης με βάση την κατανομή,**
  - **του πλήθους των bits που χρειάζομαι,**
  - **για να περιγράψω-κωδικοποιήσω-καταγράψω το μήνυμα-γεγονός.**

Εντροπία
(1 από 3)

- **(Δυαδική) Πληροφορία Μονάδες Μέτρησης**

1Byte=1B=2^3 bits=8bits

1KB=2^{10} B=1024B=8142 bits

1MB=2^{10} KB=1024KB=1048576B=8337408 bits

1GB=2^{10} MB=1024MB=1048576KB=1073741824B $\cong$ 1.1x10^9B $\cong$ 8.8 x10^9bits

1TB=2^{10} GB=1024GB=1048576MB=1073741824KB $\cong$ 1.1x10^{12}B $\cong$ 8.8

x10^{12}bits

Εντροπία
(2 από 3)

1 Text Character	1 Byte = 8 bits
TV Image	Log2=10414720bits =1.4 x 10^6 bits (576 lines , 720 columns) = 414720 px and 10 luminosity scales
1 chromosome DNA as 4 Symbol Message	ld4100000bits = 2 x 10^5 bits
Information in Bacteria Memory Cells, E. Coli (2011)	900000 GB
Cells in the Human Body	> 10^{14}
Brain Neurons	$\sim 10^{11}$
Brain Synaptic Links	$\sim 10^{15}$

Εντροπία
(3 από 3)

Brain Memory	2.5 PetaBytes = 1048576 GB ≈ 8.8×10^{18} bits ≈ 300 years of TV and Audio recording!
Cyberspace 2007:	281 billion GB=281×10^{9}GB≅2.5×10^{21}bits
Cyberspace 2012:	3.6×10^{22} bits
Cyberspace Indexed Google 0.004%	10^{18} bits 2007 1.4×10^{18} bits 2012
Atoms in 12gr C	$6{,}022 \times 10^{23}$
Universe	10^{100} bits
Chess	10^{43} bits
GO	10^{200} bits ?
Eternity II	10^{550} bits
Borges Babel Library	$2.0 \times 10^{183410}4$ bits

Genetic Alphabet

- **Genetic Alphabet**

  Eors Szathmary 1992. What is the Optimum Size for the Genetic Alphabet? Proc. Natl. Acad. Sci. USA 89, 2614-2618.

DNA Digital Storage
(1 από 2)

- **DNA Digital Storage**
- Church G. Gao Y., Kosuri S. 2012, Next-Generation Digital Information Storage in DNA. Science DOI: 10.1126/science.1226355.

  "DNA is among the most dense and stable information media known. The development of new technologies in both DNA synthesis and sequencing make DNA an increasingly feasible digital storage medium. We develop a strategy to encode arbitrary digital information in DNA, write a 5.27-megabit book (HTML draft) using DNA microchips, and read the book using next-generation DNA sequencing.
- A,C → 0, G,T → 1

DNA Digital Storage (2 από 2)

- DNA Advantages over traditional digital storage media.

  1. DNA can be easily copied, and is often still readable after thousands of years in non-ideal conditions.

  2. the Techniques required to read and write DNA information are as old as life on Earth, unlike ever-changing electronic storage formats such as magnetic tape and DVDs.

References

- Shannon C. , Weaver W. 1949, The Mathematical Theory of Communication, University of Illinois, Urbana, Illinois.
- Kempton R. and Wedderburn R. 1978, A Comparison of Three Measures of Species Diversity, Biometrics 34, 25-37.
- Kuppers B.-O. 1990, Information and the Origin of Life, MIT Press, Cambridge, Massachusetts
- Traub J. , Werschulz A. 1998, Complexity and Information, Cambridge University Press, Cambridge.
- McDonald G. 2003, Biogeography: Space, Time and Life, Wiley, New York.
- Yockey H. 2005, Information theory, Evolution and the origin of Life, Cambridge University Press, Cambridge.

Παράδειγμα: Ρίψη 2 Ζαριών (1 από 4)

- **Δειγματοχώρος**

Y={y|y= (1,1), (1,2), (1,3), (1,4), (1,5), (1,6),
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }

Τυχαία Μεταβλητή: Z(y) = y, το Αποτέλεσμα της ρίψης των 2 ζαριών.

$$J[Z] = -\sum_{\nu=1}^{36}\left(\frac{1}{36}\log_2\frac{1}{36}\right) = \log_2 36 \cong 5.17$$

Παράδειγμα: Ρίψη 2 Ζαριών
(2 από 4)

- **Τυχαία Μεταβλητή: Χ = το Άθροισμα των Ενδείξεων των 2 Ζαριών**

$$J[X] = -(2\frac{1}{36}\log_2\frac{1}{36} + 2\frac{2}{36}\log_2\frac{2}{36} + 2\frac{3}{36}\log_2\frac{3}{36} + 2\frac{4}{36}\log_2\frac{4}{36} + 2\frac{5}{36}\log_2\frac{5}{36} + 2\frac{6}{36}\log_2\frac{6}{36})$$

$$J[X] = (\frac{1}{18}5.17 + \frac{1}{9}4.17 + \frac{1}{6}3.58 + \frac{1}{8}3.17 + \frac{5}{18}2.85 + \frac{1}{6}2.58)$$

$$J[X] = 3.031$$

Μεταβλητή Χ: Άθροισμα Ενδείξεων 2 Ζαριών	Παρατηρησιμα Γεγονότα Observable Events	Πιθανότητα Probability P(x)	Συνάρτηση Κατανομής F(x)
2	Ξ_2 **={ (1,1)}**	1/36	**1/36=2.78%**
3	Ξ_3 **={ (1,2), (2,1)}**	2/36	**3/36=8.4%**
4	Ξ_4 **={ (2,2), (1,3),(3,1)}**	3/36	**6/36=16.7%**
5	Ξ_5 **={ (1,4), (2,3),(3,2), (4,1)}**	4/36	**10/36=27.8%**
6	Ξ_6 **={ (1,5), (2,4),(3,3), (4,2), (5,1)}**	5/36	**15/36=41.7%**
7	Ξ_7 **={ (1,6), (2,5),(3,4), (4,3), (5,2), (6,1)}**	6/36	**21/36=58.3%**
8	Ξ_8 **={ (2,6), (3,5),(4,4), (5,3), (6,2)}**	5/36	**26/36=72.2%**
9	Ξ_9 **={ (3,6), (4,5),(5,4), (6,3)}**	4/36	**30/36=83.3%**
10	Ξ_{10} **={ (4,6), (5,5),(6,4)}**	3/36	**33/36=91.7%**
11	Ξ_{11} **={ (5,6), (6,5)}**	2/36	**35/36=97.2%**
12	Ξ_{12} **={ (6,6)}**	1/36	**36/36=100%**

Παράδειγμα: Ρίψη 2 Ζαριών (3 από 4)

- Τυχαία Μεταβλητή

  Y = η απόλυτη τιμή της Διαφοράς των ενδείξεων των 2 Ζαριών.

  Η Διαμέριση της Y: η = { H_0, H_1, H_2, H_3, H_4, H_5 }

Cell	Probability
H0 **={ (1,1), (2,2),(3,3), (4,4), (5,5),(6,6)}**	6/36
H1 **={ (1,2), (2,3), (3,4), (4,5), (5,6), (6,5), (5,4), (4,3),(3,2), (2,1)}**	10/36
H2 **={ (1,3), (2,4), (3,5), (4,6), (6,4), (5,3),(4,2), (3,1)}**	8/36
H3 **={ (1,4), (2,5), (3,6), (6,3), (5,2),(4,1)}**	6/36
H4 **={ (1,5), (2,6), (6,2), (5,1)}**	4/36
H5 **={ (1,6), (6,1)}**	2/36

$$J[\mathrm{Y}] = -(2\frac{6}{36}\log_2\frac{6}{36} + 2\frac{10}{36}\log_2\frac{10}{36} + 2\frac{8}{36}\log_2\frac{8}{36} + 2\frac{4}{36}\log_2\frac{4}{36} + 2\frac{2}{36}\log_2\frac{2}{36})$$

$$J[\mathrm{Y}] = 2.43$$

Παράδειγμα: Ρίψη 2 Ζαριών (4 από 4)

- **J[Ένδειξη Ζαριών]>J[Άθροισμα Ζαριών]>J[Διαφορά Ζαριών]**

$$J[Z]=5.17$$

$$J[X]=3.031$$

$$J[\mathrm{Y}]=2.43$$

Ποια η Αβεβαιότητα «Πειραγμένου» Ζαριού;

- Εντροπία Ισοπίθανου Ζαριού $J_{\max} = 6(-\frac{1}{6}\log_2\frac{1}{6}) = \log_2 6 = 2.586$
- Εντροπία «Πειραγμένου» Ζαριού

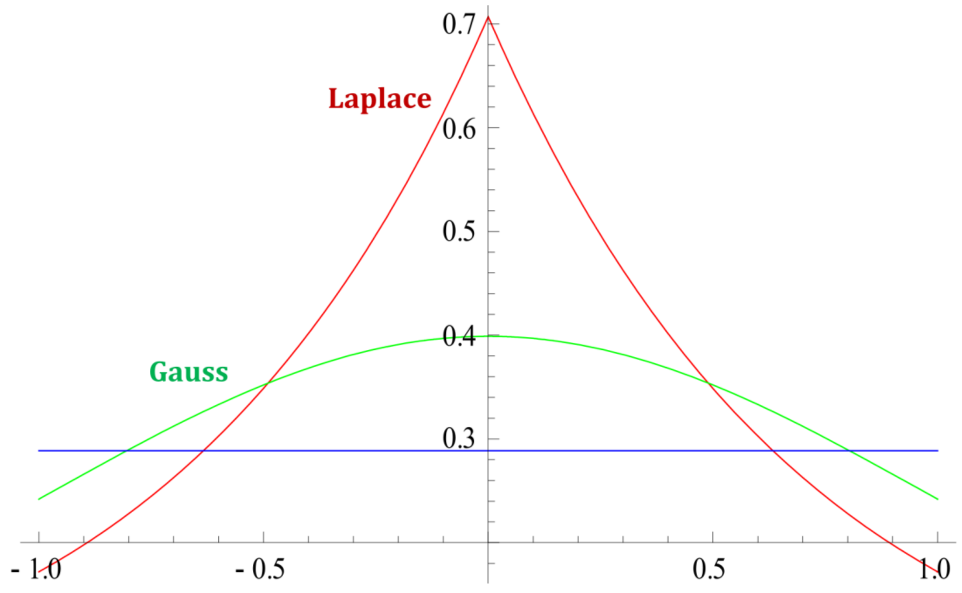
	Ψηφίο	1	2	3	4	5	6
Ζάρι Α	Συχνότητα	1/10	1/10	1/10	1/10	1/10	1/2
Ζάρι Β	Συχνότητα	1/20	1/20	1/20	1/20	1/20	3/4

$$\mathcal{J}_A = 5\left(-\frac{1}{10}\log_2\frac{1}{10}\right) + \left(-\frac{1}{2}\log_2\frac{1}{2}\right) = \frac{1}{2}\log_2 10 + \frac{1}{2}\log_2 2 = 1.661 + \frac{1}{2} = \mathbf{2.161}$$

$$\mathcal{J}_B = 5\left(-\frac{1}{20}\log_2\frac{1}{20}\right) + \left(-\frac{3}{4}\log_2\frac{3}{4}\right) = \frac{1}{4}\log_2 20 - \frac{3}{4}\log_2\frac{3}{4} = 1.080 + 0.331 = \mathbf{1.411}$$

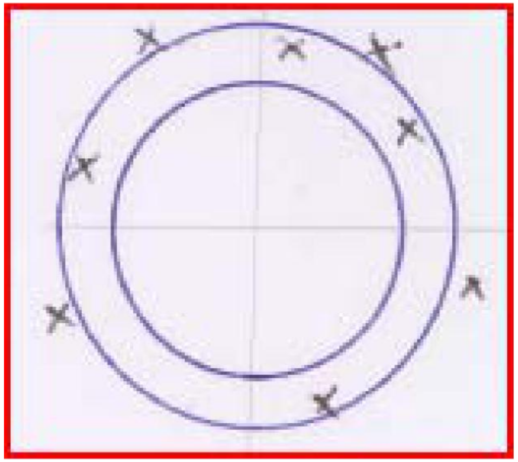
$$\mathcal{J}_B < \mathcal{J}_A < \mathcal{J}_{MAX}$$

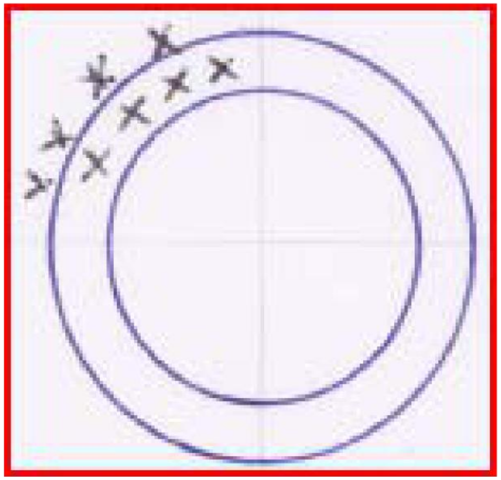
Εντροπία Κατανομών Gauss, Laplace

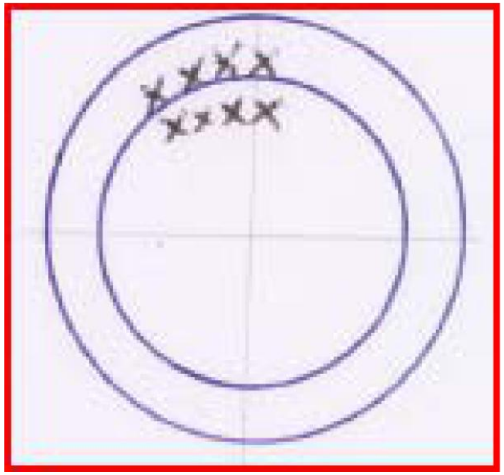


Κατανομή	Τύπος	Εντροπία
Gauss	$\rho(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}, m=0, \sigma=1$	1.42
Laplace	$\rho(x)=\frac{\sqrt{2}}{2}e^{-\lvert x\rvert\sqrt{2}}, m=0, \sigma=1$	1.35

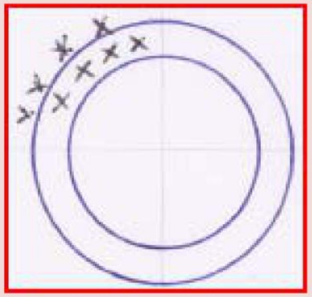
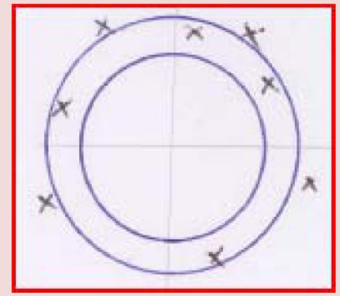
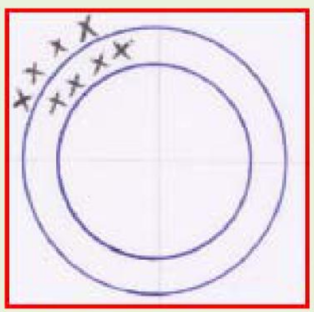
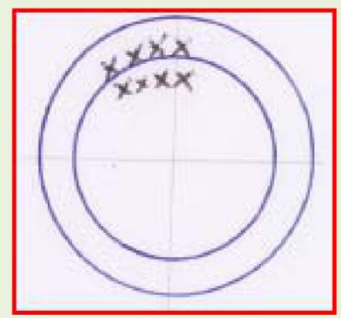
Entropy and Variance
(1 από 2)







Entropy and Variance
(2 από 2)

Same Variance (Accuracy)	 **Low Entropy** **High Precision**	 **High Entropy** **Low Precision**
Same Entropy (Precision)	 **High Variance** **Low Accuracy**	 **Low Variance** **High Accuracy**

Error
(1 από 2)

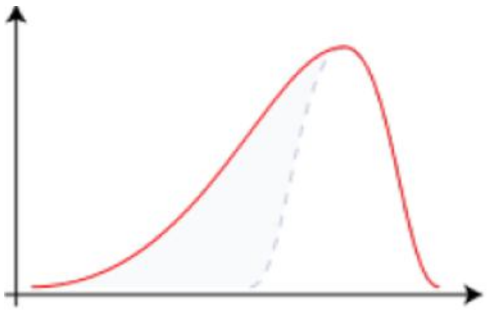
Error is the discrepancy-difference-deviation of
some Observation-Approximation-Estimation-Opinion about Reality and Reality

Process	Errors		Error Estimation
Syntactic Processing	Observation Errors	Accuracy (Σφάλμα Ακρίβειας Τυχαίο)	Τυπική Απόκλιση $\tilde{\sigma}$ Σχετικό Σφάλμα Μετρήσεων $\frac{\tilde{\sigma}}{\tilde{m}}$
		Precision (Σφάλμα Στόχευσης Συστηματικό)	Εντροπία Μετρήσεων
	Approximation Erros	Analytic Approximations	
		Asymptotic Approximations	
		Statistical Estimations	
	Computation Errors due to finite representation of numbers in the computers	Rounding of numbers, Στρογγυλοποίηση	
		Truncation of numbers, Αποκοπή	
Semantic Processing	Cognitive Errors	Perception	
		Cognitive Bias	
		Misapprehension	
	Categorical Errors	Ontological Relations	
	Logical Errors	Inference, Contradictions, Paradoxes	
Significance Assessment	Validation Errors	Testing Hypothesis Errors	

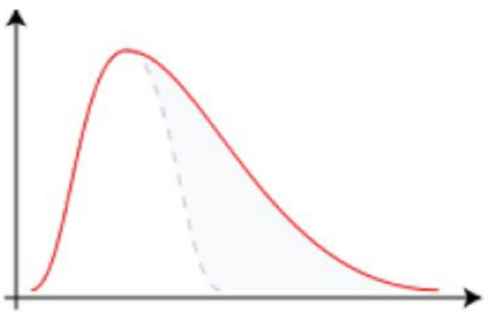
Error
(2 από 2)

Testing Hypothesis Errors		
Observation	Reality	
Test Result (Positive or Negative for H1) Decision	H1 is True Είναι Ένοχος Pregnancy	H1 is False Είναι Αθώος No Pregnancy
H0 is rejected by the Test ⇔The test is Positive for H1 ⇔ There is Evidence for H1 ⇔ Διαπιστώθηκε Ενοχή ⇔ Pregnancy Indicated	True Positive outcome Convicting the Guilty Pregnancy Indicated and The Lady is Pregnant Η Υπόθεση Επαληθεύεται	False Positive outcome Convicting the Innocent Pregnancy Indicated but The Lady is not Pregnant Error of 1st kind
H0 is not rejected by the Test ⇔The test is Negative for H1 ⇔ There is No Evidence for H1 ⇔ Δεν διαπιστωθηκε Ενοχη ⇔ Pregnancy not Indicated	False Negative outcome Releasing the Guilty Pregnancy not Indicated but The Lady is Pregnant Error of 2nd kind	True Negative outcome Releasing the Innocent Pregnancy not Indicated and The Lady is not Pregnant Η Υπόθεση Διαψεύδεται

Παράμετροι Σχήματος Κατανομής (1 από 2)



$\alpha_3 < 0$



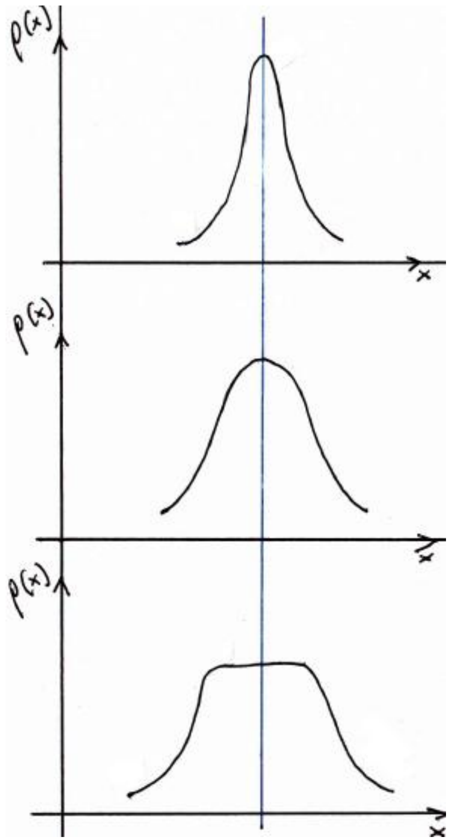
$\alpha_3 > 0$

- Παράμετροι Σχήματος Κατανομής (Shape Parameters)

Λοξότητα (skewness)

$$a_3 = \frac{c_3}{\sigma_3}$$

Παράμετροι Σχήματος Κατανομής (2 από 2)



- Κύρτωση
- $\alpha_4 = \frac{c_4}{\sigma^4}$ Λεπτόκυρτη κατανομή
- $\frac{c_4}{\sigma^4} > 3$ Μεσόκυρτη κατανομή που προσεγγίζεται από την κανονική κατανομή
- $\frac{c_4}{\sigma^4} = 3$ Πλατύκυρτη κατανομή
- $\frac{c_4}{\sigma^4} - 3$ Excess Kurtosis

Άσκηση

- Επιλέξατε μια συμμετρική και μια ασύμμετρη Διακριτή κατανομή πιθανότητας. Υπολογίστε τις 8 Παραμέτρους:
    - Μέση Τιμή (Mean),
    - Κορυφές (Modes),
    - Διάμεσος (Median),
    - Τυπική Απόκλιση (Standard Deviation),
    - Σχετικό Σφάλμα (Relative Error),
    - Λοξότητα (Skewness),
    - Κύρτωση (Kurtosis) και
    - Εντροπία (Entropy)

  Εξετάστε αν ισχύουν οι Σχέσεις Μέσης Τιμής, Διάμεσου, Κορυφής.

  Απαραίτητο να είναι μια συμμετρική και μια ασύμμετρη κατανομή.
- Άσκηση Συνεχών Κατανομών. Όπως η Άσκηση Διακριτών Κατανομών

Συμμεταβλητότητα των Μεταβλητών X,Y

- Συμμεταβλητότητα των Μεταβλητών X,Y (Covariance)

$$\sigma_{XY} = \mathrm{cov}(X,Y) = E[(\mathrm{X}-\mathrm{E}[\mathrm{X}])-(\mathrm{Y}-\mathrm{E}[\mathrm{Y}]) = \mathrm{E}[(\mathrm{X}-\mathrm{m}_X)(Y-m_Y)]$$

$\sigma_{XX} = \sigma^2 = \mathrm{var}(X)$ the Variance of X

- Θεώρημα

$$\sigma_{XY} = E[XY] - m_X m_Y$$

Συσχέτιση των Μεταβλητών X,Y

- Συσχέτιση των Μεταβλητών X,Y (Correlation)

$$Cor(XY) = E[XY] = < X, Y >$$

$$Cor(XX) = E[X^2] = < X, X > = \| X^2 \|$$

Συντελεστής Pearson των Μεταβλητών X,Y

- Συντελεστής Συμμεταβλητότητας Pearson των Μεταβλητών X,Y (Covariance)

$$r_{XY} = r(\mathrm{X},\mathrm{Y}) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(\mathrm{Y})}} = \frac{\sigma_{XY}}{\sigma_{\mathrm{X}}\sigma_{Y}}$$

Συντελεστής Pearson
(1 από 2)

- Θεώρημα

  1. Ο Συντελεστής Pearson λαμβάνει τιμές στο διάστημα [-1,1]: $-1 \leq r_{XY} \leq 1$.

  2. $r_{XY} = \quad 1$ Οι Μεταβλητές X,Y συνδέονται με γραμμική σχέση: Y = α+βX, α,β πραγματικοί αριθμοί.

$$Y = m_Y - \frac{\sigma_Y}{\sigma_X} m_X + \frac{\sigma_Y}{\sigma_X} X,\ \ r_{XY} = 1$$

$$Y = m_Y - \frac{\sigma_Y}{\sigma_X} m_X - \frac{\sigma_Y}{\sigma_X} X,\ \ r_{XY} = -1$$

Απόδειξη Από την Ανισότητα Cauchy–Schwarz.

Συντελεστής Pearson
(2 από 2)

- **Συντελεστής Pearson**
  - Galton 1888 (Κληρονομικότητα Ύψους), ξάδελφος Darwin
  - Pearson 1895 συνεργάτης και συνεχιστής του έργου του Galton.

- Pearson K. 1920, Notes on the History of Correlation, Biometrika 13, 25-45.
- Rodgers J. L. , Nicewander W. A. 1988, Thirteen ways to look at the correlation coefficient, The American Statistician, 42(1), 59–66.
- Stigler S. M. 1989, Francis Galton's Account of the Invention of Correlation, Statistical Science 4 (2): 73–79.

Συντελεστής Αλληλοεξάρτησης Αμοιβαίας Πληροφορίας των Μεταβλητών X,Y (1 από 2)

- **Συντελεστής Αλληλοεξάρτησης Αμοιβαίας Πληροφορίας των Μεταβλητών X,Y**

$$r_I = \frac{J[X;Y]}{\min(J[X], J[Y])}$$

- Όπου $J[X;Y] = J[X] + J[Y] - J[X,Y] = \sum_{x,y} \rho(x,y)\log_2(\frac{\rho(x|y)}{\rho(x)}) = \sum_{x,y} \rho(x,y)\log_2(\frac{\rho(x,y)}{\rho(x)\rho(\mathrm{y})})$
- Η Αμοιβαία Πληροφορία (Mutual Information) των Μεταβλητών X,Y
- Η εντροπία της Μεταβλητής X

$$J[X] = -\sum_x \rho(x)\log_2\rho(x)$$

- Η εντροπία της Μεταβλητής Y

$$J[Y] = -\sum_y \rho(y)\log_2\rho(y)$$

- Η κοινή εντροπία των Μεταβλητών X,Y

$$J[\mathrm{X},Y] = -\sum_{x,y} \rho(\mathrm{x},y)\log_2\rho(\mathrm{x},y)$$

Συντελεστής Αλληλοεξάρτησης Αμοιβαίας Πληροφορίας των Μεταβλητών X,Y (2 από 2)

- **Θεώρημα**
  - $0 \leq r_I \geq 1$
  - $r_I = 0 \Leftrightarrow$ X, Y Ανεξάρτητες Μεταβλητές
  - $r_I = 1 \Leftrightarrow$ η μεταβλητή μικρότερης εντροπίας είναι συνάρτηση της άλλης μεταβλητής μεγαλύτερης εντροπίας (αιτιώδης – καθορισμένη εξάρτηση).

Παράδειγμα

- **Παράδειγμα**
  Αν οι μεταβλητές X, Y ακολουθούν κοινή κανονική κατανομή:

$$\boldsymbol{\rho}(\mathbf{x},\mathbf{y}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\, exp\left(-\frac{1}{2(1-r^2)}\left[\left(\frac{\mathbf{x}-m_X}{\sigma_X}\right)^2 + \left(\frac{\mathbf{y}-m_Y}{\sigma_Y}\right)^2 - 2r\left(\frac{\mathbf{x}-m_X}{\sigma_X}\right)\left(\frac{\mathbf{y}-m_Y}{\sigma_Y}\right)\right]\right)$$

  τότε ο συντελεστής αλληλοεξάρτησης είναι:

$$r_I = -\frac{ln\left(1 - r^2\right)}{ln\left(2\pi e\,\sigma^2\right)}$$

  σ=min(σX,σY)

  $|r|=0 \Leftrightarrow r_I = 0$

Σημείωμα Αδειοδότησης

Τέλος ενότητας

Επεξεργασία: Βασιλική Αλμπανίδου
Θεσσαλονίκη, Χειμερινό Εξάμηνο 2014-2015

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

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.