Μέρος V

Οριακά Θεωρήματα

Άθροισμα η τυχαίων μεταβλητών

• Έστω \mathbf{n} τυχαίες μεταβλητές X_1 , X_2 ,..., X_n . Το άθροισμά τους είναι μία τυχαία μεταβλητή:

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

 Μέση τιμή αθροίσματος = άθροισμα μέσων τιμών των η μεταβλητών

$$E(S_n) = E(X_1) + E(X_2) + \dots + E(X_n) = \sum_{i=1}^n E(X_i)$$

Απόδειξη

Για n=2 τυχαίες μεταβλητές με από-κοινού σ.π.π. f_{1,2}(x₁, x₂) έχουμε:

$$\begin{split} E\big(X_1 + X_2\big) &= \\ &= \int\!\!\int (x_1 + x_2) f_{1,2}(x_1, x_2) dx_1 dx_2 = \\ &= \int\!\!\int x_1 f_{1,2}(x_1, x_2) dx_1 dx_2 + \int\!\!\int x_2 f_{1,2}(x_1, x_2) dx_1 dx_2 = \\ &= \int\!\!x_1 \Big(\!\!\int f_{1,2}(x_1, x_2) dx_2\Big) dx_1 + \int\!\!x_2 \Big(\!\!\int f_{1,2}(x_1, x_2) dx_1\Big) dx_2 = \\ &= \int\!\!x_1 f_1(x_1) dx_1 + \int\!\!x_2 f_2(x_2) dx_2 = \\ &= E(X_1) + E(X_2) \end{split}$$

- Για n=3 έχουμε πάλι 2 τυχαίες μεταβλητές την $X_{12} = X_1 + X_2$ και την X_3 .
- Έτσι ισχύει ότι

$$E(X_1 + X_2 + X_3) = E(X_{12} + X_3) =$$

$$= E(X_{12}) + E(X_3) = E(X_1 + X_2) + E(X_3) =$$

$$= E(X_1) + E(X_2) + E(X_3)$$

• Αναδρομικά για τις *n* τυχαίες μεταβλητές έχουμε

$$E(X_{1} + X_{2} + \dots + X_{n-1} + X_{n}) = E(X_{1,n-1} + X_{n}) =$$

$$= E(X_{1,n-1}) + E(X_{n}) = E(X_{1,n-2} + X_{n-1}) + E(X_{n}) =$$

$$= E(X_{1,n-2}) + E(X_{n-1}) + E(X_{n}) =$$

$$\vdots$$

 $= E(X_1) + E(X_2) + \cdots + E(X_{n-1}) + E(X_n)$

• Διακύμανση του αθροίσματος *η* τυχαίων μεταβλητών

$$S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

$$V(S_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i=1}^{n} \sum_{j=1}^{n} COV(X_i, X_j)$$

$$j \neq i$$

Απόδειξη

• Έστω n=2 τυχαίες μεταβλητές

$$S_2 = X_1 + X_2$$

$$V(S_{2}) = E[(S_{2} - E(S_{2}))^{2}] = E[(X_{1} + X_{2} - (E(X_{1}) + E(X_{2})))^{2}] =$$

$$= E[((X_{1} - E(X_{1})) + (X_{2} - E(X_{2})))^{2}] =$$

$$= E[(X_{1} - E(X_{1}))^{2} + (X_{2} - E(X_{2}))^{2} + 2(X_{1} - E(X_{1}))(X_{2} - E(X_{2}))] =$$

$$= E[(X_{1} - E(X_{1}))^{2}] + E[(X_{2} - E(X_{2}))^{2}] + 2E[(X_{1} - E(X_{1}))(X_{2} - E(X_{2}))] =$$

$$= V(X_{1}) + V(X_{2}) + 2COV(X_{1}, X_{2})$$

(1) Άθροισμα ανεξάρτητων μεταβλητών

> Η μέση τιμή του αθροίσματος παραμένει το άθροισμα, δηλ.

$$E(S_n) = E(X_1) + E(X_2) + \dots + E(X_n) = \sum_{i=1}^n E(X_i)$$

> Η διακύμανση είναι ίση με το **άθροισμα** των διακυμάνσεων

$$V(S_n) = \sum_{i=1}^n V(X_i) + 2\sum_{i=1}^n \sum_{\substack{j=1 \ j \neq i}}^n COV(X_i, X_j) = \sum_{i=1}^n V(X_i)$$

(2) Άθροισμα ανεξάρτητων και ισόνομων μεταβλητών

- Οι μεταβλητές είναι ανεξάρτητες και ισόνομες, (i.i.d. independent identically distributed), δηλ.
- \succ έχουν κοινό μέσο και διακύμανση $E(X_i)=\mu$, $V(X_i)=\sigma^2$ $\forall i$:
- ≥ То́тє

$$E(S_n) = \sum_{i=1}^n E(X_i) = n\mu$$

$$V(S_n) = \sum_{i=1}^n V(X_i) = n\sigma^2$$

(3) Ο μέσος όρος *η* ανεξάρτητων και ισόνομων τυχαίων μεταβλητών

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} S_n$$

• Μέση τιμή:

$$E(M_n) = \frac{1}{n}E(S_n) = \frac{1}{n}n\mu = \mu$$

(3) Ο μέσος όρος *η* ανεξάρτητων και ισόνομων τυχαίων μεταβλητών

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} S_n$$

• Μέση τιμή:

$$E(M_n) = \frac{1}{n}E(S_n) = \frac{1}{n}n\mu = \mu$$

• Διακύμανση:

$$V(M_n) = \frac{1}{n^2}V(S_n) = \frac{1}{n^2}[n\sigma^2] = \frac{\sigma^2}{n}$$

$$V(aX) = a^2V(X)$$

Ανισοτικές σχέσεις

Ανισότητα του *Markov*

Έστω τ.μ. X με θετικές τιμές, δηλ. $\Omega_X = \{x: x>0\}$. Τότε

$$P(X \ge a) \le \frac{E(X)}{a}$$

Απόδειξη

$$E(X) = \int_{0}^{\infty} xf(x)dx = \int_{0}^{a} xf(x)dx + \int_{a}^{\infty} xf(x)dx \ge \int_{a}^{\infty} xf(x)dx \ge$$

$$\ge \int_{a}^{\infty} af(x)dx = a\int_{a}^{\infty} f(x)dx = aP(X \ge a)$$

Ανισότητα του Chebyshev

- Έστω τ.μ. X με μέσο $E(X)=\mu$ και διακύμανση $V(X)=\sigma^2$.
- Εφαρμόζοντας την ανίσωση του Markov $P(X \ge a) \le \frac{E(X)}{a}$ για $D^2 = (X \mu)^2$ και $a = \varepsilon^2$ και $E(D^2) = E((X \mu)^2) = V(X)$ παίρνουμε:

$$P(D^2 \ge \varepsilon^2) \le \frac{V(X)}{\varepsilon^2} \Rightarrow P(|X - \mu| \ge \varepsilon) \le \frac{V(X)}{\varepsilon^2}$$

Εφαρμογή της ανισότητας του Chebyshev για τον δειγματικό μέσο *M*_n

• Γνωρίζουμε για τον **μέσο όρο** : $E(M_n) = \mu$ και $V(M_n) = \sigma^2/n$

$$\boldsymbol{M_n} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} S_n$$

$$E(M_n) = \frac{1}{n}E(S_n) = \frac{1}{n}n\mu = \mu$$

$$V(M_n) = \frac{1}{n^2}V(S_n) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$P(|X - \mu| \ge \varepsilon) \le \frac{V(X)}{\varepsilon^2}$$

Εφαρμογή της ανισότητας του Chebyshev για τον δειγματικό μέσο M_n

- Γνωρίζουμε για τον μέσο όρο : E(M_n)=μ και V(M_n)=σ²/n
- Χρησιμοποιώντας την ανισότητας Chebyshev για το M_n :

$$P(|X - \mu| \ge \varepsilon) \le \frac{V(X)}{\varepsilon^2}$$



$$|P(|M_n - \mu| \ge \varepsilon) \le \frac{\sigma^2}{n\varepsilon^2} \Rightarrow P(|M_n - \mu| < \varepsilon) \ge 1 - \frac{\sigma^2}{n\varepsilon^2}$$

$$P(|M_n - \mu| \ge \varepsilon) \le \frac{\sigma^2}{n\varepsilon^2} \Rightarrow P(|M_n - \mu| < \varepsilon) \ge 1 - \frac{\sigma^2}{n\varepsilon^2}$$

- Αν επιλέξουμε ένα άνω φράγμα ε για το σφάλμα (ανοχή σφάλματος), τότε
- ο δειγματικός μέσος (M_n) δηλ. ο μέσος όρος των μετρήσεων, θα είναι κατά ε προσέγγιση ίσος με τον πραγματικό μέσο (μ) με πιθανότητα τουλάχιστον 1-δ,

όπου
$$\delta = \sigma^2 / (n \epsilon^2)$$
.

Νόμοι των μεγάλων αριθμών

$$ho$$
 Ασθενής νόμος $\lim_{n\to\infty} P(|M_n - \mu| < \varepsilon) = 1$

$$P(|M_n - \mu| < \varepsilon) \ge 1 - \frac{\sigma^2}{n\varepsilon^2}$$

Μετά από πολλές μετρήσεις, ο δειγματικός μέσος θα προσεγγίσει τον πραγματικό μέσο με υψηλή πιθανότητα

$$M_n \xrightarrow{P} \mu$$

 $M_{\perp} \xrightarrow{P} \mu$ σύγκλιση κατά πιθανότητα

Ισχυρός νόμος

$$P(\lim_{n\to\infty}M_n=\mu)=1$$

Με πιθανότητα 1 (πλήρη βεβαιότητα) ο δειγματικός μέσος θα προσεγγίσει τον πραγματικό μετά από πολλές μετρήσεις

$$M_n \xrightarrow{\sigma.\beta.} \mu$$

 $M_{\mu} \xrightarrow{\sigma.\beta.} \mu$ σύγκλιση με βεβαιότητα

Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)

- ✓ Εξετάζει την ασυμπτωτική συμπεριφορά του αθροίσματος πολλών τυχαίων μεταβλητών.
- ✓ Πολλές τυχαίες μεταβλητές που περιγράφουν διάφορα χαρακτηριστικά είναι αποτέλεσμα συνάθροισης πολλών τυχαίων μεταβλητών π.χ. μηνιαίο ή ετήσιο εισόδημα εργαζομένου, ύψος αποζημίωσης, θέση και ταχύτητα των μορίων ενός αερίου, βάρος/ύψος ατόμων, κ.ο.κ.
- ✓ Εξηγεί την κανονικότητα τυχαίων μεταβλητών σε εφαρμογές και τη σημαντικότητα της κανονικής κατανομής.

Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)

- ✓ Στη θεωρία της μέτρησης σφαλμάτων, το παρατηρούμενο σφάλμα μπορεί να εκφραστεί ως το άθροισμα μεγάλου πλήθους ανεξάρτητων επιμέρους σφαλμάτων (ανεξάρτητων τυχαίων μεταβλητών).
- ✓ Στη Στατιστική ο μέσος πολλών ανεξάρτητων τυχαίων μεταβλητών είναι κομβικής σημασίας για την εκτίμηση και τον στατιστικό έλεγχο παραμέτρων.

Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)

- Έστω n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, X_2, ..., X_n$, δηλ. κοινό μέσο ($\mu_i = \mu$) και διακύμανση ($\sigma_i^2 = \sigma^2$)
- Τότε το άθροισμα τους $S_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^n X_i$ ακολουθεί <u>προσεγγιστικά</u> την Κανονική κατανομή (ανεξάρτητα της κατανομής των X_i) με μέσο $n\mu$ και διακύμανση $n\sigma^2$, δηλ.

$$S_n \sim N(n\mu, n\sigma^2)$$
 $\dot{\eta}$ $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$

Καθώς ισχύει ότι: Αν X $^{\sim}$ N(μ,σ 2) \Rightarrow Z=(X-μ)/σ $^{\sim}$ N(0,1)

- Έστω n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές $X_1, X_2, ..., X_n$, δηλ. κοινό μέσο ($\mu_i = \mu$) και διακύμανση ($\sigma_i^2 = \sigma^2$)
- Τότε το άθροισμα τους είναι κατά προσέγγιση «κανονικό»

$$S_n = X_1 + X_2 + ... + X_n = \sum_{i=1}^n X_i$$
 $S_n \sim N(n\mu, n\sigma^2)$

• Το ίδιο ισχύει και για τον μέσο όρο ως τυχαία μεταβλητή

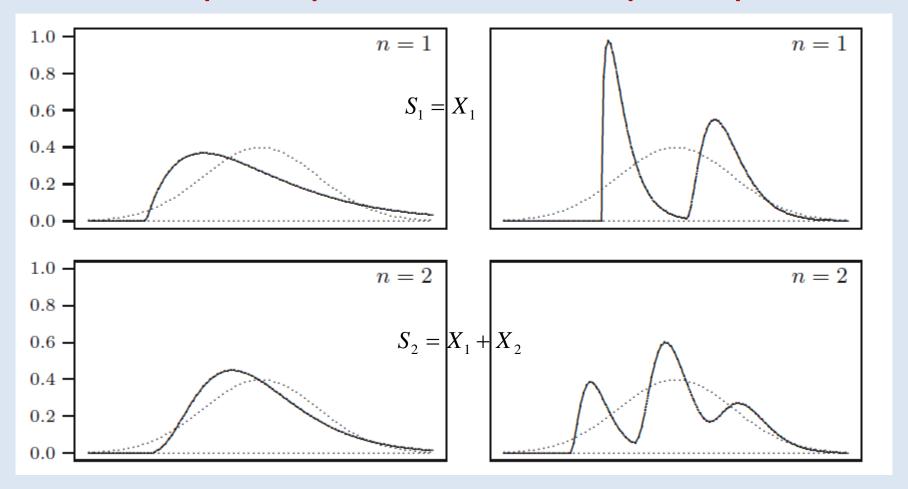
$$\boldsymbol{M_n} = \frac{\boldsymbol{S_n}}{n} \sim N\left(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}\right)$$

ακολουθεί *προσεγγιστικά* την Κανονική κατανομή με μέσο μ και διακύμανση σ²/n

Παράδειγμα εφαρμογής του Κ.Ο.Θ. σε δύο περιπτώσεις κατανομής του *X*_i

Περίπτωση 1

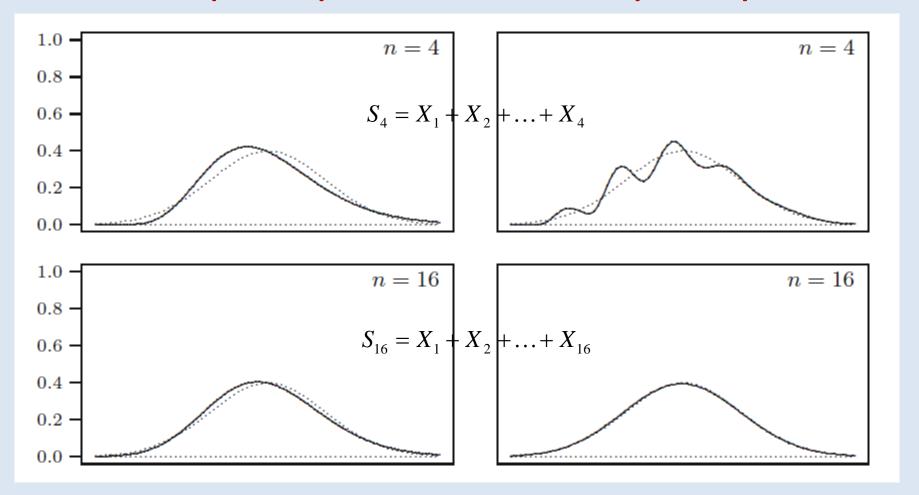
Περίπτωση 2



Παράδειγμα εφαρμογής του Κ.Ο.Θ. σε δύο περιπτώσεις κατανομής του *X*_i

Περίπτωση 1

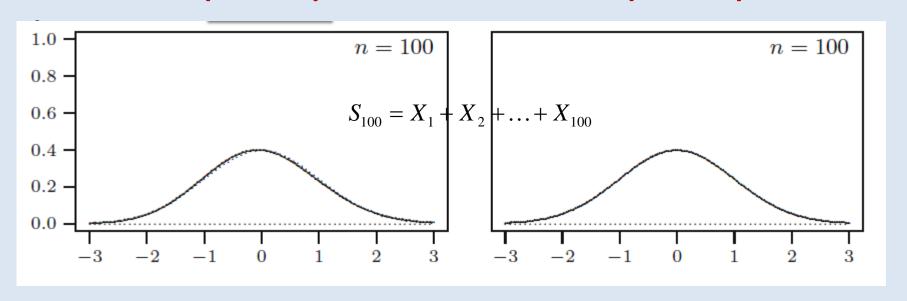
Περίπτωση 2



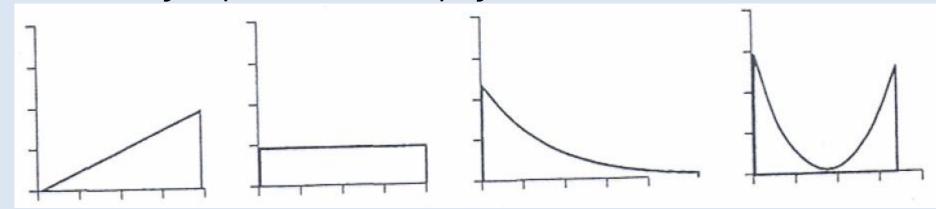
Παράδειγμα εφαρμογής του Κ.Ο.Θ. σε δύο περιπτώσεις κατανομής του *X*_i

Περίπτωση 1

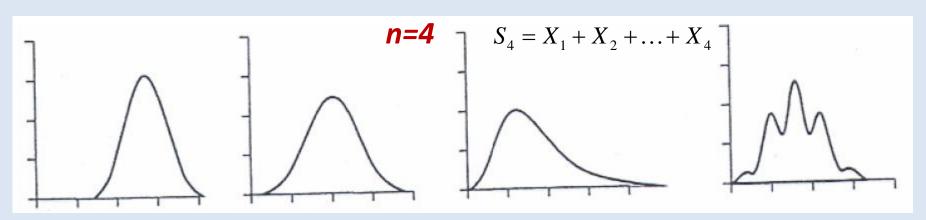
Περίπτωση 2



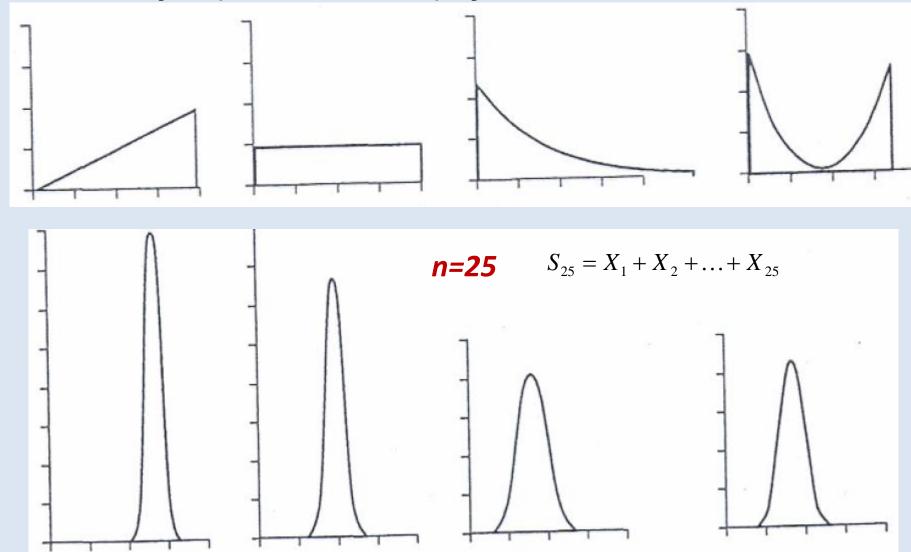
Αν για παράδειγμα θεωρήσουμε πληθυσμούς μεγέθους n
 από τις παρακάτω κατανομές



• Τότε οι κατανομές των δειγματικών μέσων ή του αθροίσματος θα είναι κανονικές



• Αν για παράδειγμα θεωρήσουμε **πληθυσμούς** μεγέθους *n* από τις παρακάτω κατανομές



Απόδειξη του Κεντρικού Οριακού Θεωρήματος

• Η απόδειξη θα γίνει βασιζόμενοι στη συνάρτηση ροπογεννήτριας.

• Η ροπογεννήτρια της τυπικής Κανονικής κατανομής *Ν(0,1)* είναι (έχει δειχθεί):

$$M(t) = E(e^{tX}) = e^{\frac{t^2}{2}}$$

Απόδειξη του Κ.Ο.Θ.

Ορίζουμε την μεταβλητή

$$Z_{n} = \frac{S_{n} - n\mu}{\sqrt{n\sigma}} = \sum_{i=1}^{n} \frac{(X_{i} - \mu)}{\sqrt{n\sigma}}$$

Υπολογίζουμε την ροπογεννήτρια:

$$M_{Z_n}(t)$$
 = $E(e^{Z_n t})$ = $E(e^{\sum_{i=1}^n \frac{(X_i - \mu)}{\sqrt{n}\sigma}t})$ = Οι μεταβλητές είναι ανεξάρτητες και ισόνομες

$$= \prod_{i=1}^n E\left(e^{\left(\frac{X_i - \mu}{\sqrt{n}\sigma}\right)t}\right) = \left(E\left(e^{\left(\frac{X - \mu}{\sqrt{n}\sigma}\right)t}\right)\right)^n$$
όπου X η ΚΟΙνή μεταβλητή των n μεταβλητών

Απόδειξη του Κ.Ο.Θ. (συν.)

$$M_{Z_n}(t) = \left(E\left(e^{\left(\frac{X-\mu}{\sqrt{n\sigma}}\right)t}\right)\right)^n$$

• Αλλά καθώς ισχύει ότι:

$$e^{x} = \sum_{k=0}^{\infty} \frac{x^{k}}{k!} = 1 + x + \frac{x^{2}}{2!} + \frac{x^{3}}{3!} + \dots$$

• Παίρνουμε ότι

$$e^{\left(\frac{X-\mu}{\sqrt{n}\sigma}\right)t} = 1 + \frac{(X-\mu)^2}{\sqrt{n}\sigma}t + \frac{(X-\mu)^2}{2!(\sqrt{n}\sigma)^2}t^2 + \frac{(X-\mu)^3}{3!(\sqrt{n}\sigma)^3}t^3 + \dots$$

Απόδειξη του Κ.Ο.Θ. (συν.)

• Έτσι η ροπογεννήτρια της Z_n γίνεται: $M_{Z_n}(t) = \left(E^{\left(\frac{X-\mu}{\sqrt{n}\sigma}\right)^t}\right)$

$$M_{Z_{n}}(t) = \left(E\left(1 + \frac{X - \mu}{\sqrt{n\sigma}}t + \frac{(X - \mu)^{2}}{2!(\sqrt{n\sigma})^{2}}t^{2} + \frac{(X - \mu)^{3}}{3!(\sqrt{n\sigma})^{3}}t^{3} + \dots\right)\right)^{n} = \left(1 + \frac{t}{\sqrt{n\sigma}}E(X \neq \mu) + \frac{t^{2}}{2!(\sqrt{n\sigma})^{2}}E[(X \neq \mu)^{2}] + \phi(n)\right)^{n} \approx \left(1 + \frac{t^{2}}{2n}\right)^{n} \approx \left(1 + \frac{t^{2}}{2n}\right)^{n} \approx e^{\frac{t^{2}}{2}}$$

$$\lim_{n \to \infty} \left(1 + \frac{a}{n}\right)^{n} = e^{a}$$

Απόδειξη του Κ.Ο.Θ. (συν.)

• Έτσι η ροπογεννήτρια της $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \sum_{i=1}^n \frac{(X_i - \mu)}{\sqrt{n}\sigma}$

$$M_{Z_n}(t) \approx e^{\frac{t^2}{2}}$$

- Που είναι (προσεγγιστικά) η ροπογεννήτρια της τυπικής Κανονικής.
- Έτσι αποδείξαμε ότι η Z_n ακολουθεί την τυπική Κανονική N(0,1).

Πόρισμα του Κ.Ο.Θ.: Προσέγγιση Διωνυμικής

- Έστω n ανεξάρτητες και ισόνομες Bernoulli τυχαίες μεταβλητές με μέσο μ=ρ και διακύμανση σ²= ρ(1-ρ).
- Τότε το άθροισμα **S**_n είναι μια Διωνυμική μεταβλητή

$$S_n = \sum_{i=1}^n X_i \sim Bin(n, \rho)$$

• Από το **Κ.Ο.Θ.** έχουμε ότι το άθροισμα προσεγγιστικά ακολουθεί την **Κανονική κατανομή** *Ν(nμ, nσ*²), δηλ.

$$S_n \sim Bin(n, \rho) \approx N(n\rho, n\rho(1-\rho))$$

Πόρισμα του Κ.Ο.Θ.: Προσέγγιση Διωνυμικής

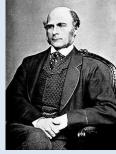
• Εφαρμογή του πορίσματος για **ρ = 1 / 2**:

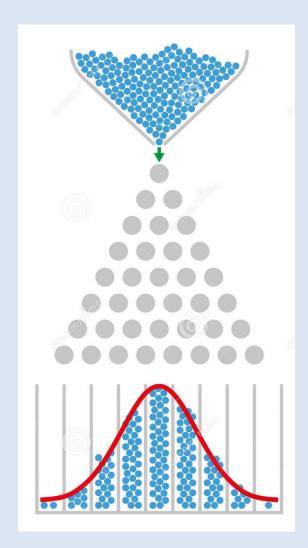
$$S_n \sim Bin\left(n, \frac{1}{2}\right) \approx N\left(n\rho, n\rho(1-\rho)\right) = N\left(\frac{n}{2}, \frac{n}{4}\right)$$

• Έτσι η διωνυμική πιθανότητα προσεγγίζεται ικανοποιητικά από την κανονική:

$$p(x) = {n \choose x} 2^{-n} \approx \sqrt{\frac{2}{\pi n}} e^{-\frac{(2x-n)^2}{2n}}$$

"Μηχανή" ή πίνακας του Galton (1822-1911)

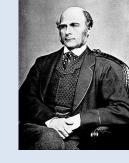


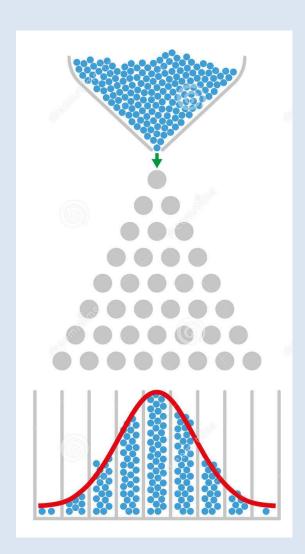


(1873)

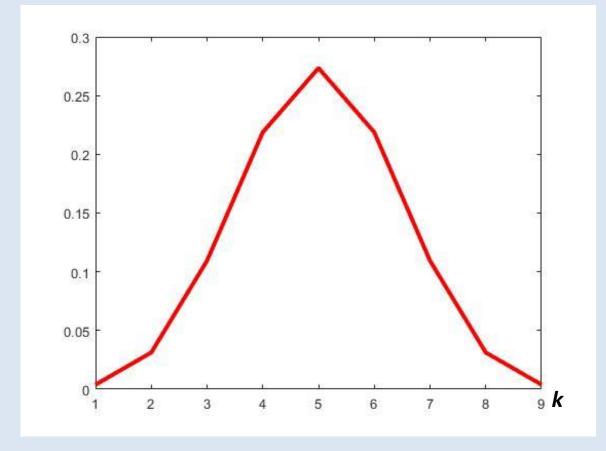


"Μηχανή" ή πίνακας του Galton (1822-1911)





$$p(k) = {n \choose k} 2^{-n} \approx \sqrt{\frac{2}{\pi n}} e^{-\frac{(2k-n)^2}{2n}}$$

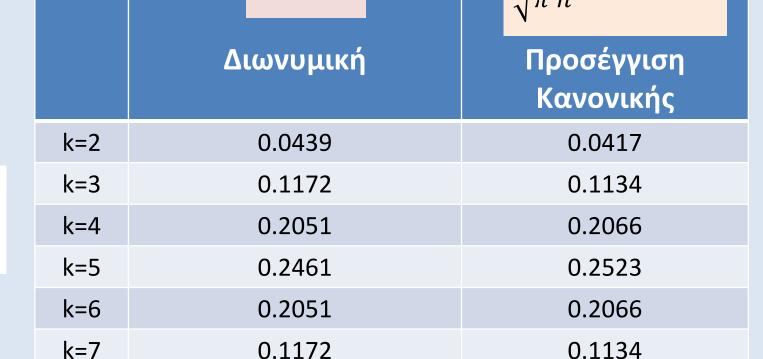


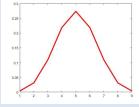


Παράδειγμα για n=10

$$p(k) = {n \choose k} 2^{-n} \approx \sqrt{\frac{2}{\pi n}} e^{-\frac{(2k-n)^2}{2n}}$$

0.0417





k=8

0.0439

Παραδείγματα

Το πάχος ενός βιβλίου είναι μια τυχαία μεταβλητή με μέση τιμή 3cm και τυπική απόκλιση 1cm. Να υπολογιστεί η πιθανότητα σε ένα ράφι μήκους 1.30 cm να χωρέσουν 40 βιβλία.

- Οι παραγγελίες σ' ένα εστιατόριο είναι μια τυχαία μεταβλητή με μέση τιμή 8 ευρώ και διακύμανση 4.
 - (α) Ποια είναι η πιθανότητα οι 100 πρώτοι πελάτες να ξοδέψουν ποσό μεγαλύτερο του 840 ευρώ.
 - (β) Ποια η πιθανότητα το συνολικό ποσό των 100 πρώτων πελατών να είναι μεταξύ 780 και 820 ευρώ.
 - (γ) Μετά από πόσες παραγγελίες θα είμαστε 90% βέβαιοι ότι το συνολικό ποσό των παραγγελιών θα ξεπεράσει τα 1000 ευρώ.

Ο χρόνος μεταξύ διαδοχικών δρομέων ενός μαραθωνίου είναι εκθετική μεταβλητή με παράμετρο λ=1/m. Ποια η πιθανότητα ο 1000°ς δρομέας να εμφανιστεί την χρονική στιγμή (1000 ± 50)m sec.

- Ο κυκλικός τροχός της ρουλέτας ο οποίος είναι χωρισμένος σε 37 (ίσα) τόξα που φέρουν αριθμούς από [0, 36], μπαίνει σε περιστροφική κίνηση και ένα σφαιρίδιο ρίχνεται σε αυτόν. Αν ο αριθμός του τόξου είναι περιττός τότε ο παίκτης κερδίζει 100 ευρώ, ενώ σε αντίθετη περίπτωση (ζυγός ή μηδέν) χάνει 100 ευρώ. Να υπολογιστούν:
 - (α) η πιθανότητα σε 100 παιχνίδια ο παίκτης να μην χάσει χρήματα
 - (β) ο αριθμός των παιχνιδιών που μπορεί να παίζει ο παίκτης, ώστε με πιθανότητα τουλάχιστον 40% να χάσει το πολύ 10 ευρώ.