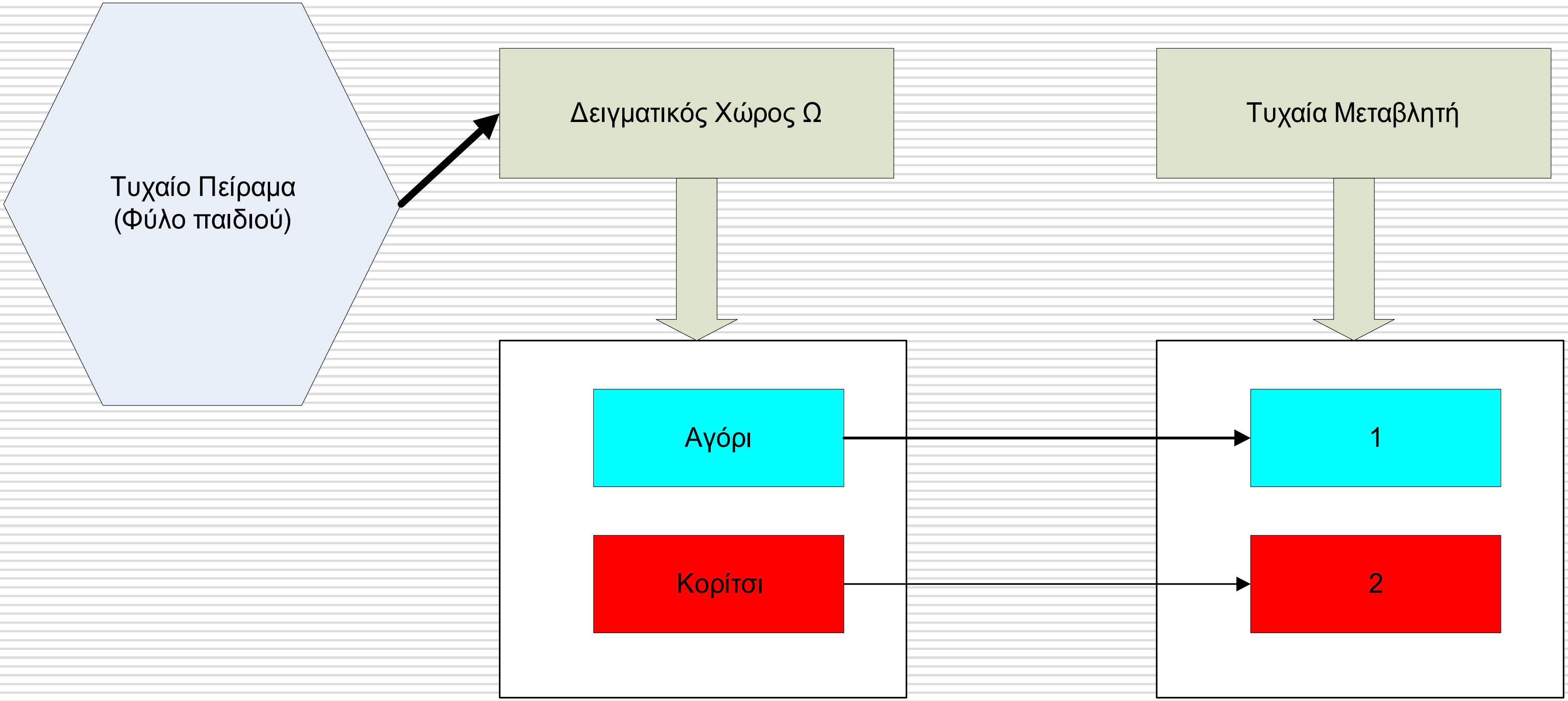
Πανεπιστήμιο Πελοποννήσου

Τυχαίες μεταβλητές
Κατανομές

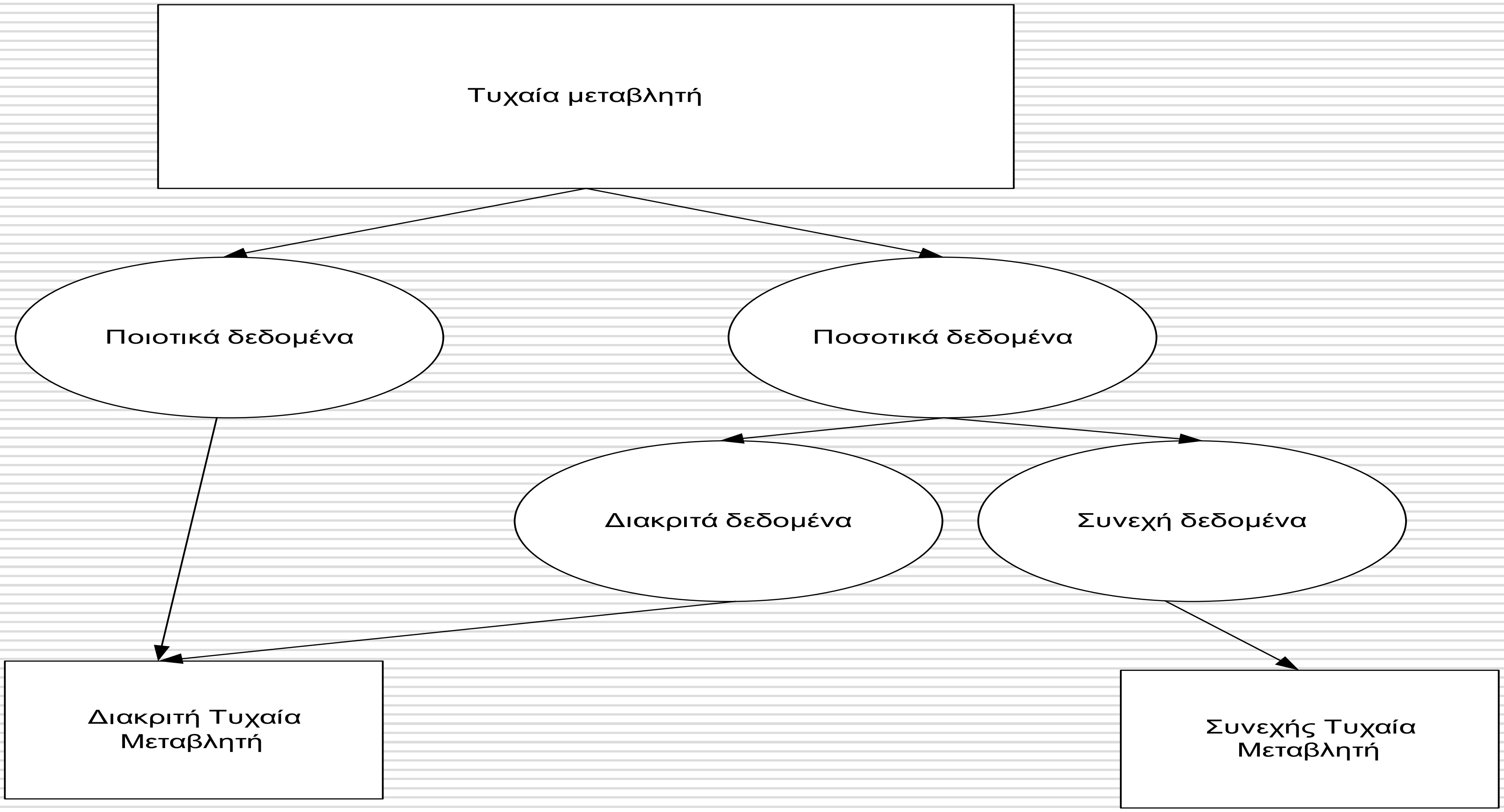
Τυχαία Μεταβλητή (τ.μ.)

- Τυχαία μεταβλητή (τ.μ.) ονομάζεται η συνάρτηση που απεικονίζει το σύνολο των δυνατών αποτελεσμάτων ενός πειράματος στο σύνολο των πραγματικών αριθμών

- Οι τμ συμβολίζονται με κεφαλαία γράμματα πχ Χ,Υ,Ζ


Τυχαίο Πείραμα
(Φύλο παιδιού)
Δειγματικός Χώρος Ω
Τυχαία Μεταβλητή
Αγόρι
Κορίτσι
1
2

Ποσοτικά και Ποιοτικά Δεδομένα



Συνάρτηση Πιθανότητας

- Αν η τμ είναι διακριτή, τότε η συνάρτηση που δίνει την πιθανότητα η τμ X να παίρνει την τιμή x, λέγεται **συνάρτηση πιθανότητας** (σ.π.) της τμ X, Συμβολίζεται με $f(x)=P(X=x)$

Ιδιότητες της Συνάρτησης Πιθανότητας

1. f(x)≥0, για κάθε τιμή x της τμ X
2. Το άθροισμα των πιθανοτήτων για όλες τις τιμές της τμ X, πρέπει να είναι ίσο με 1

$$\sum_x f(x) = \sum_x P(X = x) = 1$$

Άσκηση

- Έστω η διακριτή τμ X. Ποια η πιθανότητα P(X>3)

X	**P(X=x)**
1	4a
2	3a
3	2a
4	1a

Συνάρτηση Αθροιστικής Κατανομής

- Η συνάρτηση F(x) που ορίζεται:
  - $F(x)=P(X\leq x)$, για κάθε x που ανήκει στο R

ονομάζεται **συνάρτηση αθροιστικής κατανομής (σ.α.κ)** της τμ X και δίνει την πιθανότητα η τμ X να πάρει όλες τις τιμές της μέχρι το σημείο x

Συνάρτηση Πυκνότητας Πιθανότητας

- Αν η τμ X είναι συνεχής, τότε υπάρχει συνάρτηση f(x) που ονομάζεται **<u>συνάρτηση πυκνότητας πιθανότητας</u>** (σ.π.π.) της τμ X, τέτοια ώστε:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \ \mathrm{x} \in \Re$$

Ιδιότητες της Συνάρτησης Πυκνότητας Πιθανότητας

1. f(t)≥0
2. Το εμβαδόν κάτω από την καμπύλη f(t) ισούται με 1

$$\int_{-\infty}^{\infty} f(t)dt = 1$$

Μέση τιμή ή Αναμενόμενη τιμή

- **<u>Μέση τιμή</u>** ή **<u>αναμενόμενη τιμή</u>** (ή μαθηματική ελπίδα) της τμ Χ ορίζεται η συνάρτηση

$$E(X) = \begin{cases} \sum_{x} x \cdot f(x), \ \text{αν η X είναι διακριτή} \\ \int_{-\aleph}^{\infty} x \cdot f(x)dx, \ \text{αν η X είναι συνεχής} \end{cases}$$

Πείραμα Bernoulli

- Το τυχαίο πείραμα που έχει μόνο δύο πιθανά αποτελέσματα (επιτυχία ή αποτυχία), ονομάζεται πείραμα Bernoulli.

Αναμενόμενη τιμή μιας συνάρτησης της τμ X

- **<u>Μέση τιμή</u>** ή **<u>αναμενόμενη τιμή</u>** μιας συνάρτησης g(X) της τμ X συμβολίζεται σαν **E(X)** ή **μ** και ορίζεται:

$$E\left(g\left(X\right)\right)=\begin{cases}\sum\limits_{x} g\left(x\right)\cdot f\left(x\right), & \text{αν η X είναι διακριτή}\\ \int_{-\aleph}^{\infty} g\left(x\right)\cdot f\left(x\right)dx, & \text{αν η X είναι συνεχής}\end{cases}$$

Ιδιότητες της αναμενόμενης τιμής

1. $E(c)=c$, όπου c είναι μία σταθερά
2. $E(a*X)=a*E(X)$
3. $E(a*g(X))=a*E(g(X))$
4. $E(X \pm Y)=E(X) \pm E(Y)$
5. $E(g(X) \pm h(Y))=E(g(X)) \pm E(h(Y))$
6. $E(aX \pm b)=a*E(X) \pm b$

Διασπορά της τμ Χ

- Η **<u>διασπορά ή διακύμανση</u>** της τμ Χ συμβολίζεται V(X) ή σ^2 και δίνεται από την παρακάτω σχέση:

$$V(X) = E(X-\mu)^2 = \begin{cases} \sum_x (x-\mu)^2 f(x), & \text{Χ διακριτή} \\ \int_x (x-\mu)^2 f(x)dx, & \text{Χ συνεχής} \end{cases}$$

Ιδιότητες της διασποράς

1. $V(c)=0$, όπου c είναι μία σταθερά
2. $V(aX)=a^2V(X)$
3. $V(aX\pm b)=a^2V(X)$
4. $V(X)=E(X^2)-\mu^2$
5. $V(X\pm Y)=V(X)+V(Y)$, αν οι τμ X και Y είναι ανεξάρτητες
6. Η θετική τετραγωνική ρίζα της διασποράς ονομάζεται τυπική απόκλιση και συμβολίζεται με **σ**

Ροπές

- ν-οστή ροπή της τμ X ως προς το α δίνεται από την σχέση $E(X-\alpha)^{\nu}$
- ν-οστή **<u>κεντρική ροπή</u>** της τμ X δίνεται από την σχέση $E(X-\mu)^{2}$
- Η δεύτερη κεντρική ροπή είναι η διασπορά

Κατανομές

Διακριτές Κατανομές

Πείραμα ή δοκιμή Bernoulli

- Το τυχαίο πείραμα που έχει μόνο δύο αποτελέσματα (επιτυχία ή αποτυχία) ονομάζεται πείραμα ή δοκιμή Bernoulli

Κατανομή Bernoulli B(1,p)

- Η τμ X του ενός πειράματος Bernoulli παίρνει τιμή 1 όταν έχουμε επιτυχία και τιμή 0 όταν έχουμε αποτυχία. Όταν p είναι η πιθανότητα επιτυχίας, η κατανομή της τμ X δίνεται:

$$P(X=x)=p^x(1-p)^{1-x}, \quad x=0,1$$

Διωνυμική (Bionomial) B(n,p)

- Η διωνυμική κατανομή είναι μια διαδικασία n ανεξάρτητων δοκιμών Bernoulli με την ίδια πιθανότητα επιτυχίας p σε κάθε δοκιμή. Η τμ X που ακολουθεί διωνυμική κατανομή, εκφράζει το πλήθος των επιτυχιών στις n δοκιμές (επαναλήψεις) Bernoulli και παίρνει τιμές x=0,1,…,n. Η συνάρτηση πιθανότητας δίνεται από τον τύπο:

$$P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}$$

- E(X)=np
- V(X)=np(1-p)

Γεωμετρική (Pascal)

- Η τμ Χ που ακολουθεί Γεωμετρική κατανομή, εκφράζει το πλήθος των επαναλήψεων που πρέπει να γίνουν σε μια διαδικασία Bernoulli μέχρι την πρώτη επιτυχία.
- Το πλήθος των δοκιμών είναι άγνωστο και το πείραμα σταματά στην πρώτη επιτυχία, δηλ x=1,2,…
- Η συνάρτηση πιθανότητας είναι:

$$P(X=x)=p(1-p)^{x-1}$$

- E(X)=1/p
- $V(X)=(1-p)/p^2$

Αρνητική Διωνυμική (Negative Bionomial)

- Η τμ X που ακολουθεί Αρνητική Διωνυμική κατανομή, εκφράζει το πλήθος των επαναλήψεων Bernoulli που πρέπει να γίνουν για να συμπληρωθούν r επιτυχίες.
- Το πλήθος των δοκιμών είναι άγνωστο και το πείραμα σταματά στην r επιτυχία, δηλ x=r,+1,r+2,…
- Η συνάρτηση πιθανότητας είναι:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

- $E(X)=r/p$
- $V(X)=r(1-p)/p^2$
- Όταν r=1, τότε έχουμε τη Γεωμετρική κατανομή

Poisson P(λ)

- Η κατανομή Poisson είναι η κατανομή των σπάνιων γεγονότων και χρησιμοποιείται όταν θέλουμε να μετρήσουμε τον αριθμό των συμβάντων στη μονάδα μέτρησης
- Τα συμβάντα μπορεί να είναι: τυπογραφικά λάθη ανά σελίδα, τηλεφωνικές κλήσεις στη μονάδα του χρόνου, αυτοκίνητα που περνούν από ένα σημείο στη μονάδα του χρόνου, κτλ

Poisson P(λ)

- Η τμ X που ακολουθεί την Poisson κατανομή εκφράζει το πλήθος των γεγονότων στην μονάδα του χρόνου.
- Η παράμετρος **λ** της Poisson είναι η μέση τιμή εμφάνισης των γεγονότων στην μονάδα του χρόνου
- Η συνάρτηση πιθανότητας δίνεται από:

$$P(X = x) = e^{-\lambda} \frac{\lambda^{x}}{x!}$$

- x=0,1,2,… και λ>0
- E(X)=λ
- V(X)=λ

Υπεργεωμετρική Hypergeometrical

- Έστω ένας πληθυσμός Ν ατόμων ή αντικειμένων
- Από τον πληθυσμό αυτό, k μέλη έχουν μια συγκεκριμένη ιδιότητα (πχ από τους 100 φοιτητές, 10 είναι από τον νομό Λακωνίας), και κατά συνέπεια Ν-k μέλη δεν έχουν την συγκεκριμένη ιδιότητα
- Τυχαίο δείγμα **ν** ατόμων επιλέγεται (**χωρίς επανάθεση**)
- Η τμ Χ που εκφράζει τον αριθμό των μελών που έχουν την ιδιότητα και ανήκουν στο τυχαίο δείγμα ακολουθεί την **Υπεργεωμετρική** κατανομή

Υπεργεωμετρική

- Η συνάρτηση πιθανότητας της Υπεργεωμετρικής κατανομής δίνεται από τον παρακάτω τύπο:

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{\nu - x}}{\binom{N}{\nu}}$$

$$x = 0, 1, 2, \ldots, \nu$$

$$E(X) = \frac{\nu k}{N}$$

$$V(X) = \nu \frac{k}{N} \frac{N-k}{N} \frac{N-\nu}{N-1}$$

Συνεχείς Κατανομές

Ομοιόμορφη U(a,b) Uniform

- Η συνεχής τμ Χ ακολουθεί την ομοιόμορφη κατανομή (με παραμέτρους a και b) με σ.π.π και αθροιστική κατανομή:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \mathrm{a} \le \mathrm{x} \le \mathrm{b} \\ 0, & \mathit{otherwise} \end{cases}$$

$$E(X) = \frac{a+b}{2}, \qquad V(X) = \frac{(b-a)^2}{12}$$

$$F(x) = P(X \le x) = \frac{x-a}{b-a}$$

Εκθετική Exp(λ) Exponential

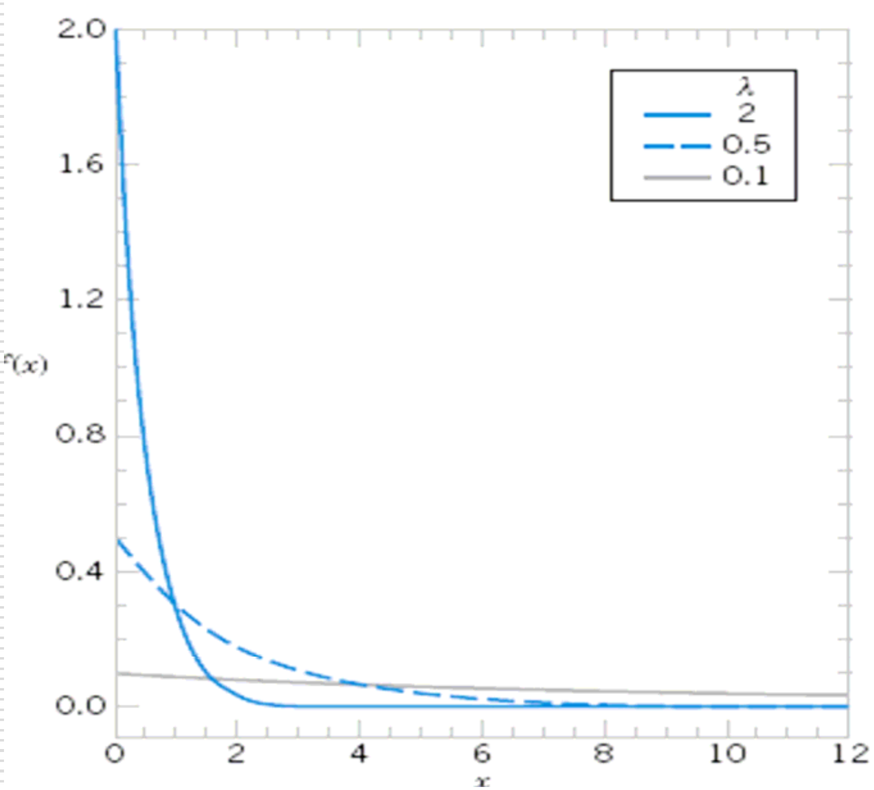
- Η συνεχής τμ Χ ακολουθεί την **εκθετική** κατανομή (με παράμετρο λ) με σ.π.π και αθροιστική κατανομή:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x > 0 \\ 0 & \textit{otherwise} \end{cases}$$

$$F(x) = 1 - e^{-\lambda x}$$

$$E(X) = 1/\lambda, \quad V(X) = 1/\lambda^2$$

Γραφική Παράσταση



Κανονική Κατανομή $N(\mu,\sigma^2)$ Normal

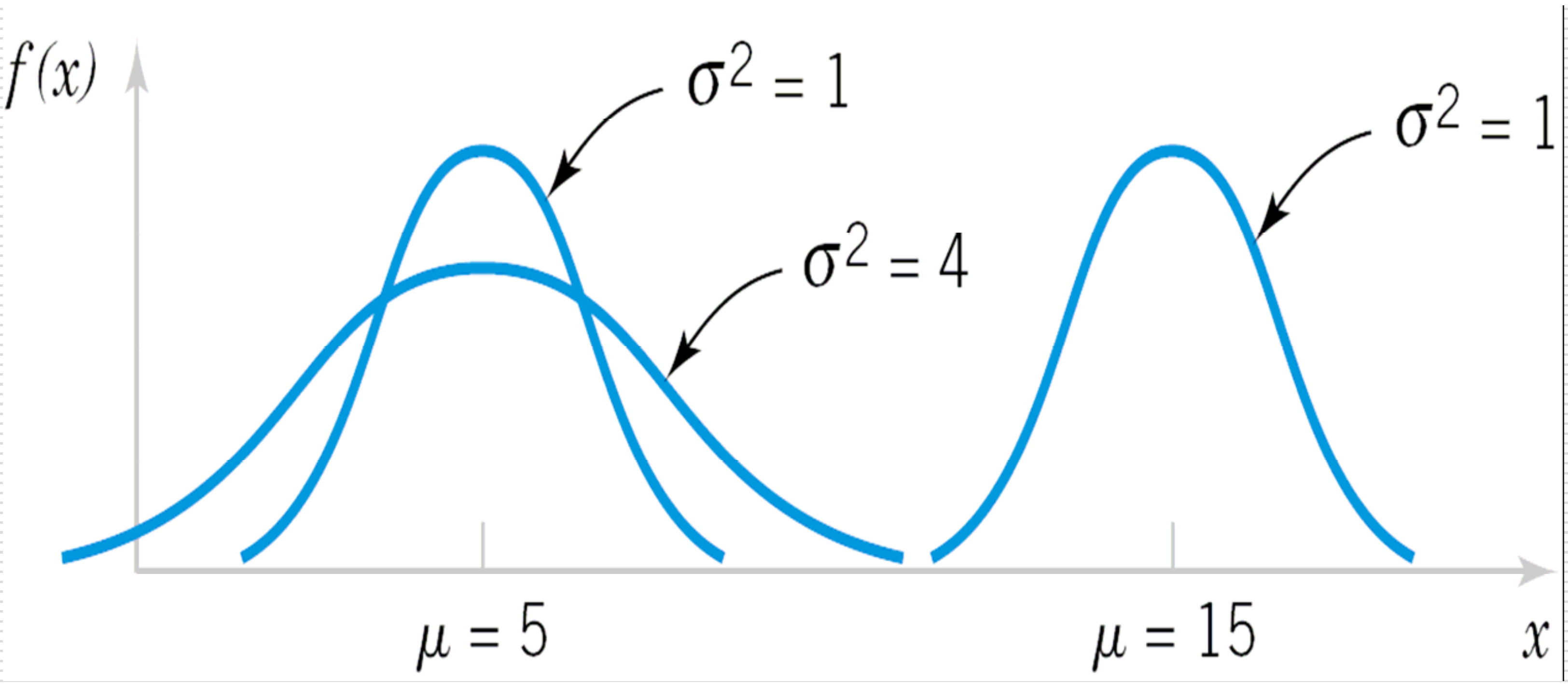
- Η συνεχής τμ Χ ακολουθεί την **κανονική** κατανομή (με παραμέτρους μ και σ^2) με σ.π.π :

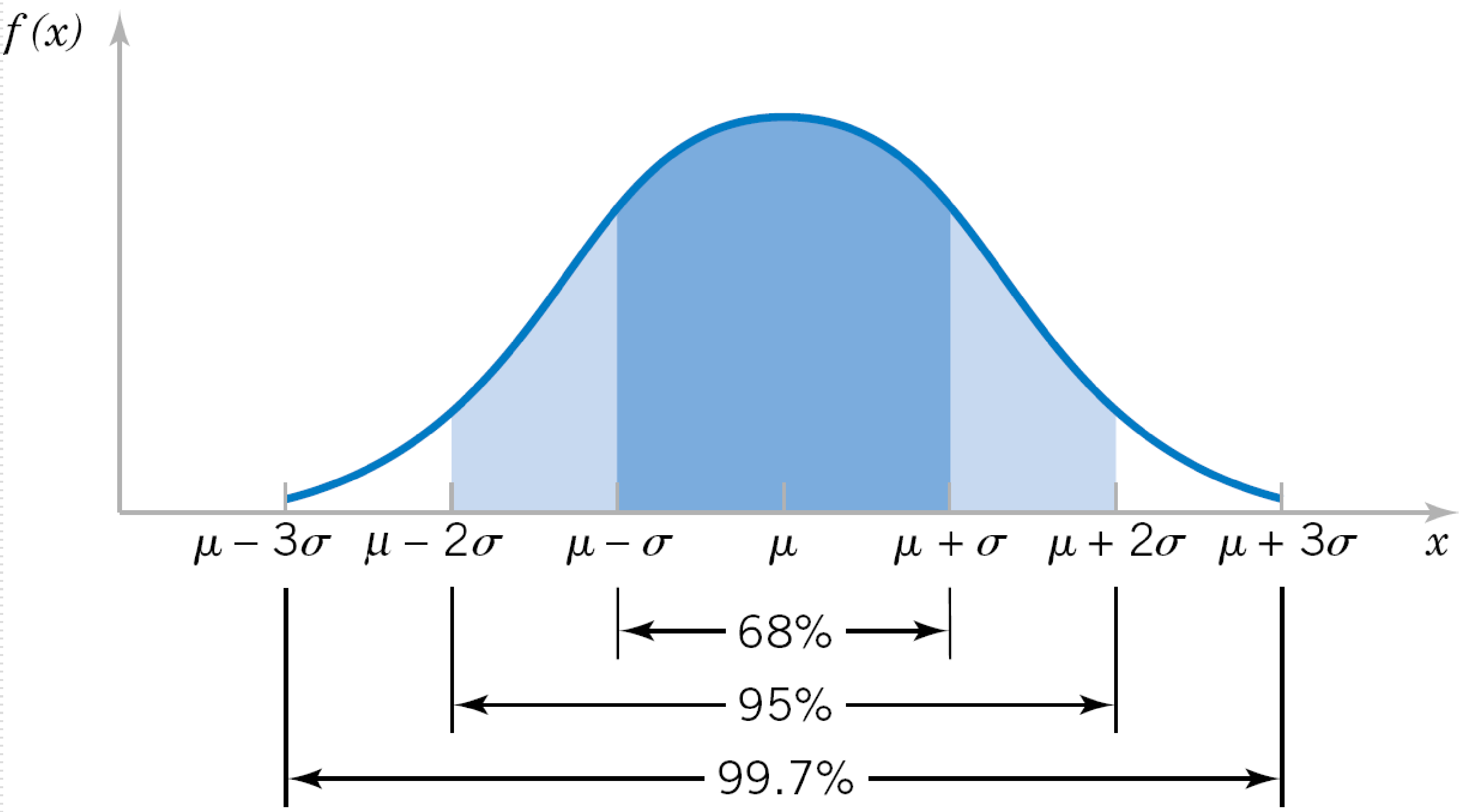
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$-\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0,$$

$$E(X) = \mu, \quad V(X) = \sigma^2$$

Γραφική Παράσταση





$f(x)$
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$
x
68%
95%
99.7%

Τυπική Κανονική Κατανομή N(0,1) Standard Normal

- Η συνεχής τμ Z ακολουθεί την **τυπική κανονική** κατανομή (με παραμέτρους μ=0 και $\sigma^{2=1}$) με σ.π.π :

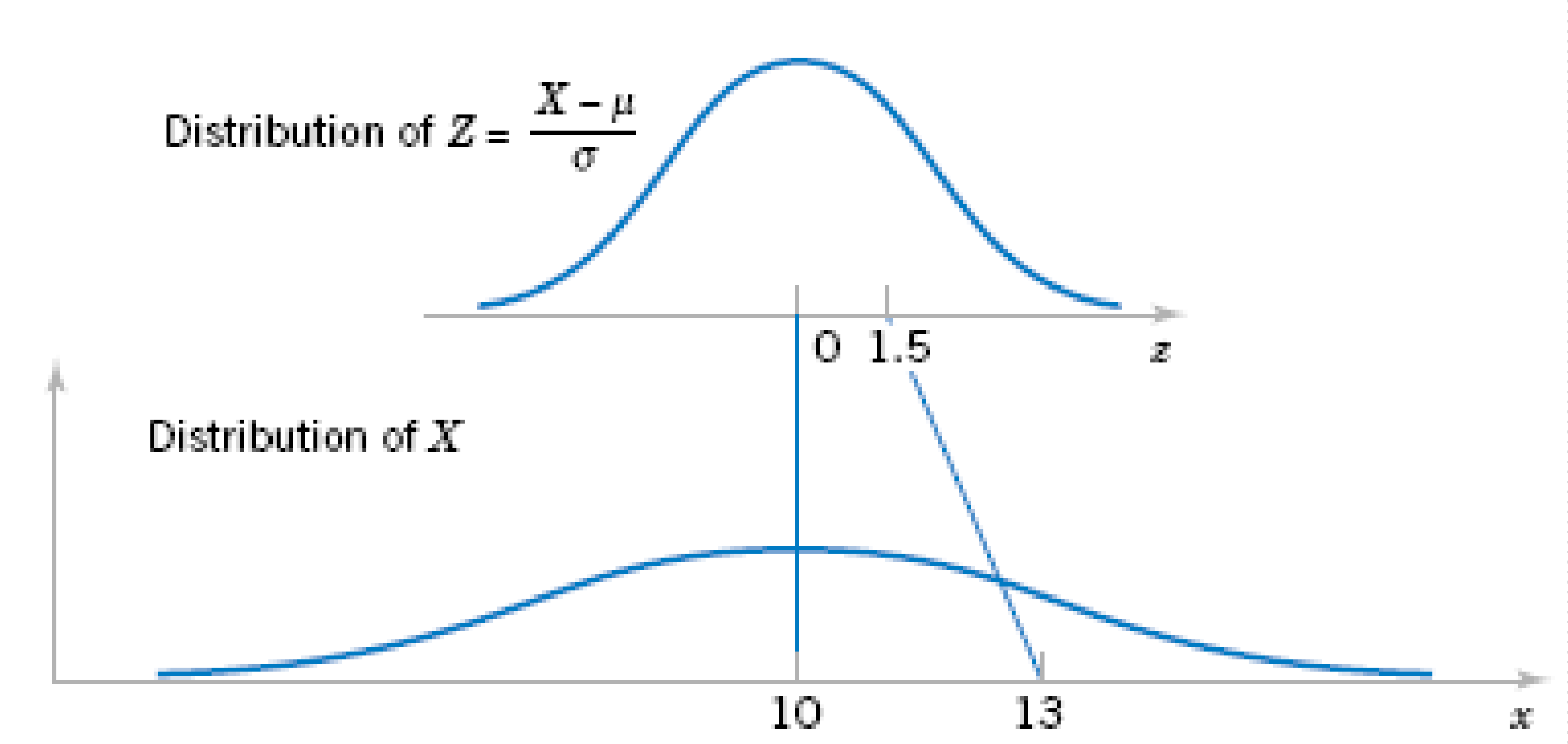
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$$E(X) = 0, \quad V(X) = 1$$

Ιδιότητες Κανονικής Κατανομής

- Αν η τμ X ακολουθεί κανονική κατανομή $N(\mu,\sigma^2)$, τότε η τμ $Z=(X-\mu)/\sigma$ ακολουθεί την τυπική κατανομή $N(0,1)$
- Η κανονική κατανομή είναι συμμετρική κατανομή σε σχέση με την μέση τιμή **μ**
- Η αθροιστική κατανομή της κανονικής κατανομής συμβολίζεται **$\Phi(z)$**
- **$\Phi(-a)=1-\Phi(a)$**

Γραφική Παράσταση



Λογαριθμο-Κανονική Κατανομή LogNormal LN(μ,σ²)

- Έστω ότι η τμ W ακολουθεί την κανονική κατανομή με παραμέτρους μ και σ². Τότε η τμ X=exp(W) ακολουθεί την **λογαριθμό-κανονική** κατανομή (με παραμέτρους μ και σ²) με σ.π.π :

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{ln(x)-\mu}{\sigma}\right)^2}$$

$$x > 0, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0,$$

$$E(X) = \exp\left(\mu + \sigma^2/2\right),$$

$$V(X) = \exp\left(2\mu + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right)$$

Γάμμα κατανομή G(a,b)

- Η συνεχής τμ Χ ακολουθεί την **γάμμα** κατανομή (με παραμέτρους a, b) με σ.π.π:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad x > 0$$

- Η εκθετική κατανομή είναι ειδική περίπτωση της Γάμμα.

$$X \sim \exp(\lambda) \Rightarrow X \sim G(1,\lambda)$$

Άλλες συνεχείς Κατανομές

- Student ή t με παράμετρο **ν** που ονομάζεται βαθμοί ελευθερίας (t_{ν})
- **X^2** με παράμετρο **ν** που ονομάζεται βαθμοί ελευθερίας
- $F_{n,m}$ με n και m βαθμούς ελευθερίας

Σχέσεις μεταξύ κατανομών

- Έστω η τμ X με σπп f(x), και θεωρούμε την τμ Y=g(X), όπου g() είναι μία συνάρτηση της X. Τότε
- Αν η y=g(x) λύνεται μονοσύμαντα, δηλαδή $x=g^{-1}(y)$, τότε η σπп της τμ Y είναι

$$f_Y(y)=f_X\left(g^{-1}(y)\right)\left|\frac{dx}{dy}\right|$$

- Αν η y=g(x) έχει παραπάνω από μία λύση, τότε η σπп της Y είναι

$$f_Y(y)=\sum_i f_X\left(g_i^{-1}(y)\right)\left|\frac{dx_i}{dy}\right|$$

Σχέσεις μεταξύ κατανομών (1)

- Έστω ανεξάρτητες τμ X_1 και X_2 όπου:

$$X_1 \sim N\left(\mu_1, \sigma_1^2\right)$$

$$X_2 \sim N\left(\mu_2, \sigma_2^2\right)$$

- Τότε

$$Z = X_1 \pm X_2 \sim N\left(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Σχέσεις μεταξύ κατανομών (2)

- Έστω ανεξάρτητες τμ X_1 και X_2 όπου:

$$X_1 \sim N(0,1)$$

$$X_2 \sim \mathrm{X}_n^2$$

- Τότε

$$Y = {X_1}\Big/{\sqrt{{X_2}/{n}}} \sim t_n$$

Σχέσεις μεταξύ κατανομών (3)

- Έστω ανεξάρτητες τμ X_1, X_2,…,X_n που ακολουθούν την τυπική κανονική κατανομή N(0,1) η κάθε μία. Τότε:

$$Y = \sum_{i=1}^{n} X_i^2 \sim \mathrm{X}_n^2$$

Σχέσεις μεταξύ κατανομών (4)

- Έστω ανεξάρτητες τμ $X_1, X_2, \ldots, X_n$ που ακολουθούν την **X^2** με $k_1, k_2, \ldots, k_n$ βαθμούς ελευθερίας αντίστοιχα. Τότε:

$$Y = \sum_{i=1}^{n} X_i \sim \mathrm{X}_r^2$$

$$r = k_1 + k_2 + \ldots + k_n$$

Σχέσεις μεταξύ κατανομών (5)

- Έστω ανεξάρτητες τμ X_1 και X_2 όπου:

$$X_1 \sim \mathrm{X}_k^2$$

$$X_1 + X_2 \sim \mathrm{X}_n^2$$

- Τότε

$$X_2 \sim \mathrm{X}_{n-k}^2$$

Σχέσεις μεταξύ κατανομών (6)

- Έστω ανεξάρτητες τμ X_1 και X_2 όπου:

$$X_1 \sim \mathrm{X}_k^2$$

$$X_2 \sim \mathrm{X}_n^2$$

- Τότε

$$Y = \frac{X_1 / k}{X_2 / n} \sim F_{k,n}$$

Σχέσεις μεταξύ κατανομών (7)

- Έστω τμ Χ $\sim F_{k,n}$. Τότε $Y=1/X \sim F_{n,k}$

Τυχαίο Δείγμα

- Τυχαίο δείγμα τδ είναι μία ακολουθία n ανεξάρτητων τμ (X_1, X_2,...,X_n) που έχουν όλες την ίδια κατανομή, αυτή του πληθυσμού

Στατιστικό

- Κάθε συνάρτηση του τδ που δεν περιέχει άγνωστες παραμέτρους, ονομάζεται ***στατιστική συνάρτηση*** ή ***στατιστικό***.
- Τα στατιστικά είναι και αυτά τυχαίες μεταβλητές αφού προέρχονται από τυχαίο δείγμα που είναι μία συλλογή τυχαίων μεταβλητών

Τυχαίο δείγμα από την Κανονική κατανομή

- Έστω $X_1, X_2, \ldots, X_n$ τδ από την κανονική κατανομή $N(\mu, \sigma^2)$. Τότε η μέση τιμή και η διασπορά του δείγματος είναι ανεξάρτητες τμ και:

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

- Τότε:

$$t = \frac{\left(\overline{X} - \mu\right)}{S \Big/ \sqrt{n}} \sim t_{n-1}$$

Τυχαία δείγματα από την Κανονική κατανομή

- Έστω $X_1, X_2, \ldots, X_n$ και $Y_1, Y_2, \ldots, Y_m$ ανεξάρτητα τδ όπου:

$$X_i \sim N\left(\mu_1, \sigma_1^2\right)$$
$$Y_i \sim N\left(\mu_2, \sigma_2^2\right)$$

- Τότε

$$t = \frac{\bar{X} - \bar{Y} - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_1^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Τυχαία δείγματα από την Κανονική κατανομή

- Έστω $X_1, X_2, \ldots, X_n$ και $Y_1, Y_2, \ldots, Y_m$ ανεξάρτητα τδ όπου:

$$X_i \sim N\left(\mu_1, \sigma_1^2\right)$$

$$Y_i \sim N\left(\mu_2, \sigma_2^2\right)$$

- Τότε

$$F = \frac{\left.\frac{(n-1)S_1^2}{\sigma_1^2}\right/ (n-1)}{\left.\frac{(m-1)S_2^2}{\sigma_2^2}\right/ (m-1)} = \frac{S_1^2}{S_2^2}\frac{\sigma_2^2}{\sigma_1^2} \sim F_{n-1,m-1}$$

Κεντρικό Οριακό Θεώρημα

- Αν οι τμ $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες και ισόνομες (ακολουθούν την ίδια κατανομή) με $E(X_i)=\mu$ και $V(X_i)=\sigma^2$, $i=1,\ldots,n$. Τότε η δειγματική μέση τιμή ακολουθεί ασυμπτωτικά την κανονική κατανομή:

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\left(\overline{X} - \mu\right)}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

$$\text{when} \quad n \to \infty \quad (n \geq 30)$$