Τυχαία Μεταβλητή (τ.μ.) Random Variable (r.v.)

είναι μία συνάρτηση που αντιστοιχεί το σύνολο των δυνατών αποτελεσμάτων ενός πειράματος (Δειγματικός Χώρος) στο σύνολο των πραγματικών αριθμών.

$$X : \Omega \to \mathbb{R}$$

Ρίψη Νομίσματος

$X : \{Δ, Ε\} \longrightarrow \{0,1\}$

$P(X=0) = P(\text{Εμφ. Διεθνή Όψη})$

$P(X=1) = P(\text{Εμφ. Εθνική Όψη})$

Ρίψη Ζαριού

$X : \{\text{έξι διαφορετικές όψεις}\} \to \{\cancel{0},1,2,\dots,6\}$

$P(X=i) = P(\text{εμφ. η όψη με τον αριθμό } i)$

Διακριτή ή απαριθμητή τ.μ. αν παίρνει πεπερασμένο ή αριθμήσιμο πλήθος τιμών με θετική πιθανότητα

Συνεχής τ.μ.

Χρόνος ζωής λαμπτήρα

Βάρος ψαριών

Ποσό χρημάτων για επιδότηση του λαδιού

Πληθυσμός μιας αποικίας φυκιών

Θερμοκρασία τη νύχτα κατά τη διάρκεια του χειμώνα

(Αθροιστική) Συνάρτηση Κατανομής (σ.κ.) της τ.μ. X

$F_X : \mathbb{R} \to [0,1]$ $\quad F_X(x) = P(\{\omega\in\Omega : X(\omega) \le x\}) = P(X \le x) \quad \forall x \in \mathbb{R}$

Η αθροιστική συνάρτηση κατανομής έχει τις εξής ιδιότητες:

i) είναι αύξουσα συνάρτηση του x:

$$x_1 \le x_2 \Rightarrow F_X(x_1) \le F_X(x_2)$$

ii) τουλάχιστον από δεξιά συνεχής

$$\lim_{h\to 0^+} F_X(x+h) = F_X(x)$$

iii) $\lim_{x\to -\infty} F_X(x) =: F_X(-\infty) = 0$

$$F_X(+\infty) := \lim_{x\to +\infty} F_X(x) = 1$$

• Διακριτή τ.μ. X

όταν το $X(\Omega)$ είναι πεπερασμένο ή αριθμήσιμο τότε η συνάρτηση $f_X(x) := P(X=x)$ λέγεται συνάρτηση πιθανότητας (σ.π.) της τ.μ. X και έχει τις ιδιότητες

i) $f_X(x) \ge 0 \quad \forall x \qquad f_X(x) \le 1$

ii) $\sum_x f_X(x) = 1$

σε αυτή την περίπτωση οι συναρτήσεις $f_X(x)$ και $F_X(x)$ συνδέονται με τις εξής σχέσεις

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} f_X(x_i)$$

$$f_X(x_i) = P(X = x_i) = F_X(x_i) - F_X(x_{i-1}) = \left(P(X \le x_i) - P(X \le x_{i-1})\right)$$

Δείτε επίσης τις σελ. 2,3

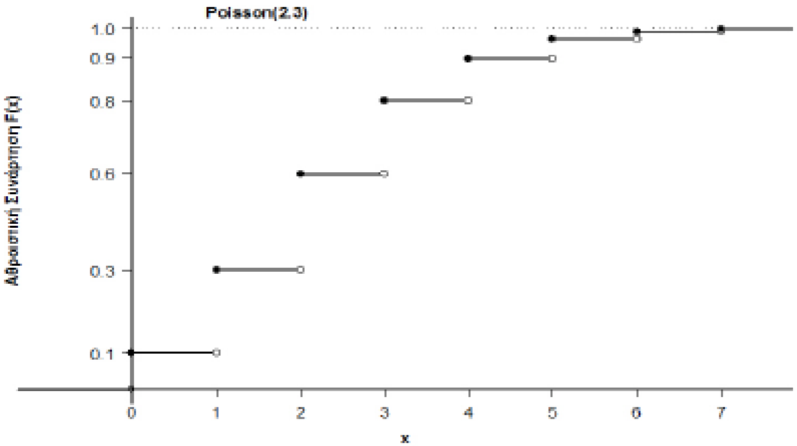
- Τυχαία Μεταβλητή (τμ), X, ονομάζεται η συνάρτηση $X : \Omega \to \mathbb{R}$ για την οποία ισχύει ότι το σύνολο

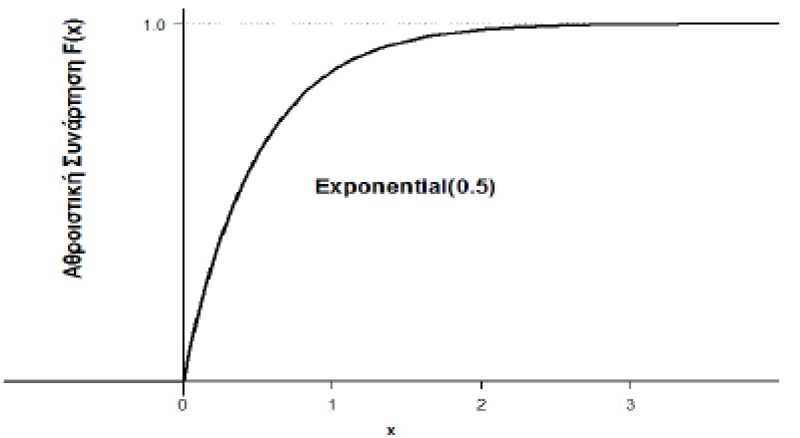
$$(X \leq x) := \{\omega \in \Omega : X(\omega) \leq x\} = A \in \mathcal{A}, \qquad \forall\, x \in \mathbb{R}.$$

- Αθροιστική συνάρτηση κατανομής (ασκ) είναι η πιθανότητα

$$F(x) := P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \qquad \forall\, x \in \mathbb{R}$$

  - είναι αύξουσα
  - είναι από δεξιά συνεχής
  - $\lim_{x \to -\infty} F(x) = 0$ και $\lim_{x \to \infty} F(x) = 1$





τότε ορίζονται οι εξής πιθανότητες

- $P(X > x) = 1 - F(x)$
- $P(X = x) = F(x) - F(x^-)$ όπου $F(x^-) = \lim\limits_{t \to x^-} F(t)$
- $P(\alpha < X \leq \beta) = F(\beta) - F(\alpha)$

Εάν η ασκ είναι απολύτως συνεχής, η τμ X ονομάζεται συνεχής τμ και

$$\exists\, f : \mathbb{R} \to [0, +\infty) \text{ ώστε } F(x) = \int_{-\infty}^{x} f(t)dt.$$

Η f ονομάζεται συνάρτηση πυκνότητας της τμ X και το στήριγμα της, $S = \{x \in \mathbb{R} : f(x) > 0\}$ είναι άπειρο μη αριθμήσιμο σύνολο.

Εάν η ασκ δεν είναι συνεχής, η τμ X ονομάζεται διακριτή τμ με

$$P(x) := P(X = x) = F(x) - F(x^-) > 0$$

στα σημεία ασυνέχειας. Η συνάρτηση P ονομάζεται συνάρτηση μάζας πιθανότητας της τμ X και το $S = \{x \in \mathbb{R} : P(x) > 0\}$ είναι είτε πεπερασμένο είτε άπειρο αριθμήσιμο.

Κατανομή Bernoulli

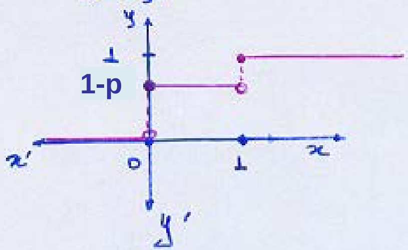
Δοκιμή Bernoulli είναι ένα πείραμα με δύο δυνατά αποτελέσματα (ενδεχόμενα: Ε = επιτυχία, Α = αποτυχία)

$X: \{E, A\} \to \{0,1\}$ $X(\varepsilon)=1$, $X(A)=0$

$P(\varepsilon) = p = 1 - P(A)$ $P(A) = 1-p = q$ $0 \le p \le 1$

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

$$F_X(x) = \begin{cases} 0 & x<0 \\ 1\text{-}p & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$



$X \sim \text{Bernoulli}(p)$

Διωνυμική κατανομή

Έχουμε n επαναλαμβανόμενες δοκιμές Bernoulli με την ίδια πιθανότητα επιτυχίας $P(\varepsilon)=p$

Η Διωνυμική κατανομή μετρά τον συνολικό αριθμό των επιτυχιών στις n ανεξάρτητες δοκιμές

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x=0,1,\ldots,n$$

αλλιώς $X = X_1 + X_2 + \cdots + X_n$ όπου X_i $i=1,\ldots,n$ ανεξάρτητες και ισόνομες τ.μ. $X_i \sim \text{Bernoulli}(p)$

$X \sim \text{Binomial}(n,p) \equiv B(n,p)$ $n \in \mathbb{N}^*$ $p \in [0,1]$

n, p παράμετροι   Bernoulli(p) ≡ Binomial(1,p

Υπεργεωμετρική Κατανομή (Δειγματοληψία χωρίς επανάθεση)

Πληθυσμός N αντικειμένων χωρίζεται σε δύο κατηγορίες με πλήθος κ, και N−κ η κάθε μία.
Από τον πληθυσμό παίρνουμε δείγμα μεγέθους n (χωρίς επανάθεση), η τ.μ. X μετράει το πλήθος των ατόμων x του δείγματος που ανήκουν στην πρώτη κατηγορία, τα υπόλοιπα n−x ανήκουν στη δεύτερη κατηγορία

$$P(X=x) = \frac{\binom{\kappa}{x}\binom{N-\kappa}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n+\kappa-N\}, \ldots, \min\{\kappa, n\}$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \qquad n! = 1\times 2\times \ldots \times n \qquad 0! = 1$$

Σημείωση: **Χρησιμοποιώ τη Διωνυμική κατανομή όταν έχω το ίδιο p σε κάθε δοκιμή και οι δοκιμές είναι ανεξάρτητες. Αυτό συμβαίνει σε δειγματοληψία με επανάθεση.**

Σε δειγματοληψία χωρίς επανάθεση, το αποτέλεσμα σε μια δοκιμή εξαρτάται από τα αποτελέσματα σε όλες τις προηγούμενες δοκιμές. Χρησιμοποιούμε την υπεργεωμετρική κατανομή.

Για μεγάλο n, οι πιθανότητες της Διωνυμικής υπολογίζονται δύσκολα. Την προσεγγίζουμε από την κατανομή Poisson, **βλέπε σελ. 5**

Poisson Κατανομή.

Μετράμε τον αριθμό των "συμβάντων" στη μονάδα μέτρησης του χρόνου, όταν σε ένα πολύ μικρό χρονικό διάστημα $(t, t+\Delta t)$ μπορεί να συμβεί το πολύ ένα συμβάν με πιθανότητα $P(\text{συμβάν } (t, t+\Delta t)) = \lambda \Delta t$

$$P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x=0,1,\dots, \qquad \lambda>0$$

$[0,t]$

$$P(X=x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!} \qquad x=0,1,\dots,$$

$X \sim$ Poisson(λ)

λ: συχνότητα εμφάνισης των συμβάντων στη μονάδα του χρόνου

Διασύνδεση Διωνυμικής Κατανομής με Poisson

Έστω $X \sim$ Binomial(n,p) όπου το n αρκετά μεγάλο $(n>20)$ και το np αρκετά μικρό $(np<10)$ τότε

$$P(X=k) \xrightarrow[np\to\lambda]{n\to+\infty} e^{-\lambda}\frac{\lambda^k}{k!}$$

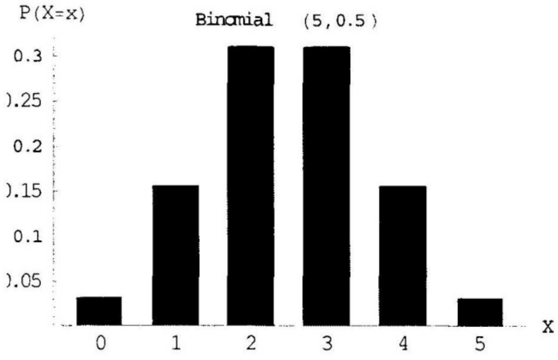
$X \sim$ Poisson(np)

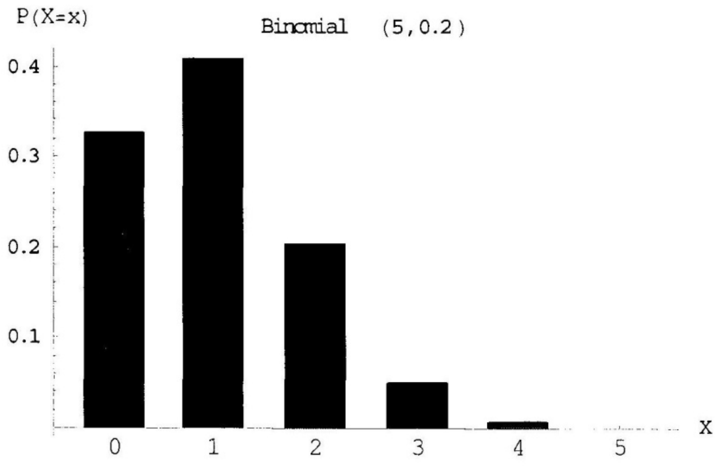
Σημείωση: **Χρησιμοποιώ την κατανομή Poisson όταν μελετώ**

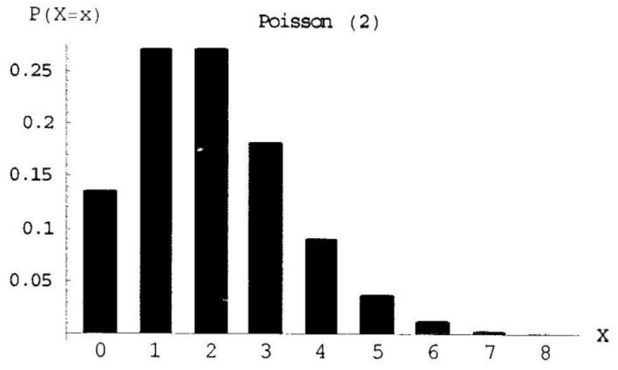
- **τον αριθμό των αφίξεων**
- **των γεννήσεων ή των θανάτων**

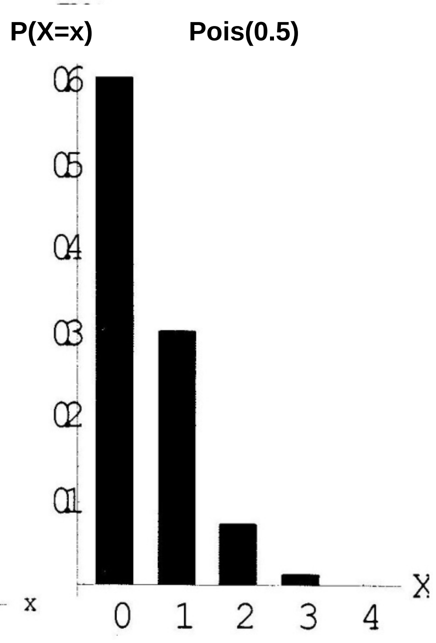
Γενικά όταν έχω συμβάντα στο χρόνο ή «άτομα» που κατανέμονται στο χώρο

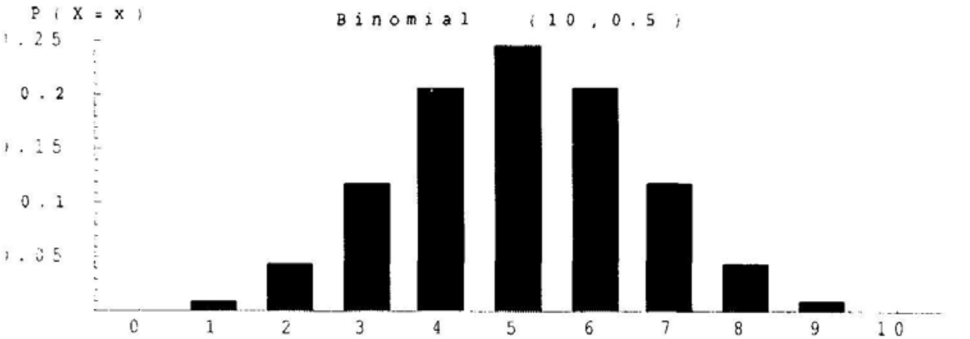
Εάν αλλάξει η μονάδα του χρόνου ή του χώρου αλλάζει αντίστοιχα και η παράμετρος λ της Poisson

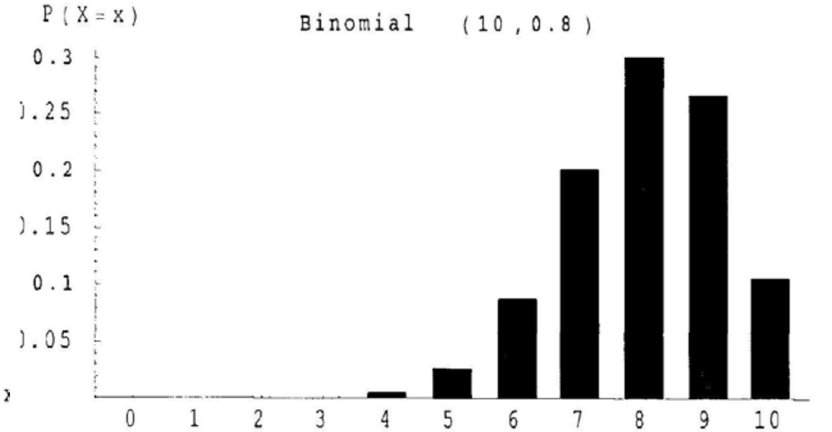

P(X=x)
Binomial (5,0.5)
0.3
0.2
0.1
x
0
1
2
3
4
5

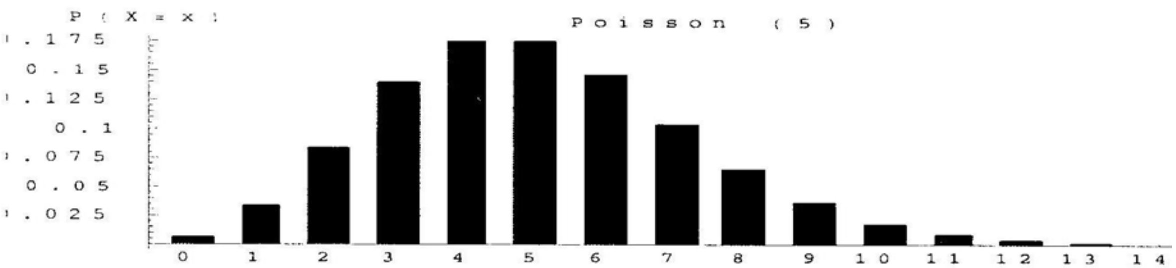

P(X=x)
Binomial (5,0.2)
0.4
0.3
0.2
0.1
X
0
1
2
3
4
5

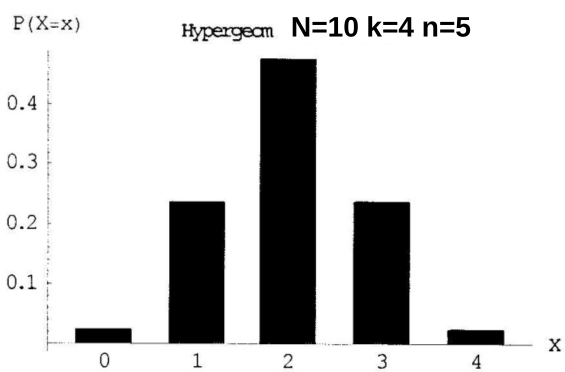

P(X=x)
Poisson (2)
0.25
0.2
0.15
0.1
0.05
x
0
1
2
3
4
5
6
7
8

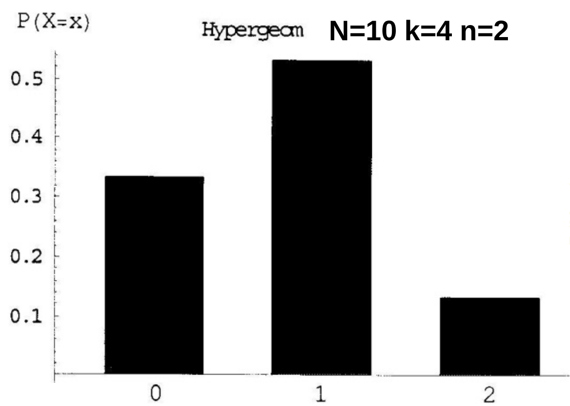

P(X=x)
Pois(0.5)
X
0
1
2
3
4

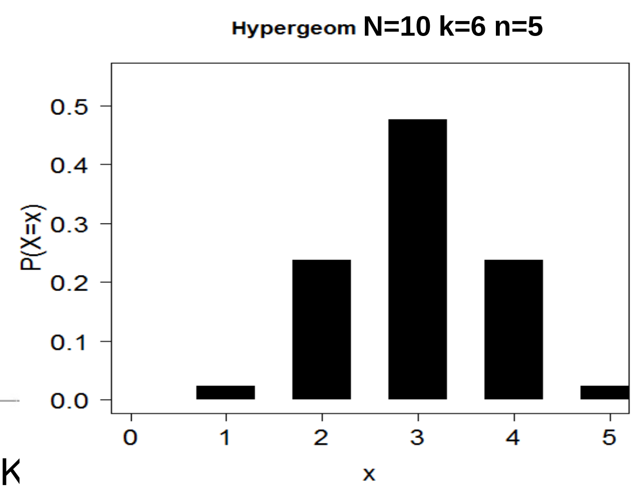

P(X=x)
Binomial (10, 0.5)
0.2
0.1
0 1 2 3 4 5 6 7 8 9 10


P(X=x)
Binomial (10, 0.8)
0.3
0.2
0.1
X
0 1 2 3 4 5 6 7 8 9 10


P(X=x)
Poisson (5)
0.15
0.1
0.05
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14


P(X=x)
Hypergeom N=10 k=4 n=5
0.4
0.3
0.2
0.1
0 1 2 3 4
x


P(X=x)
Hypergeom N=10 k=4 n=2
0.5
0.4
0.3
0.2
0.1
0 1 2


Hypergeom N=10 k=6 n=5
P(X=x)
0.5
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5
x

• Συνεχής τ.μ. X

όταν το $X(\Omega)$ είναι μη αριθμήσιμο τότε η συνάρτηση

$$f_X(x) := \lim_{\Delta x \to 0} \frac{P(x \le X \le x+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x}$$

ονομάζεται συνάρτηση πυκνότητας πιθανότητας

Δηλαδή $F_X'(x) = f_X(x)$ και ισχύουν τα εξής

i) $f_X(x) \ge 0$

ii) $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

Από τον ορισμό προκύπτει

- $f_X(x)\,\Delta x \simeq P(x < X \le x+\Delta x)$
- $P(\alpha \le X \le b) = P(\alpha < X \le b) = P(\alpha \le X < b) = \int_{\alpha}^{b} f_X(x)\,dx = F_X(b) - F_X(\alpha)$
- $F_X(+\infty) = 1 = \int_{-\infty}^{\infty} f_X(t)\,dt$
- $P(X = x) = 0$! ! !

Δείτε επίσης σελ. 3

Διασύνδεση της Εκθετικής κατανομής με Poisson

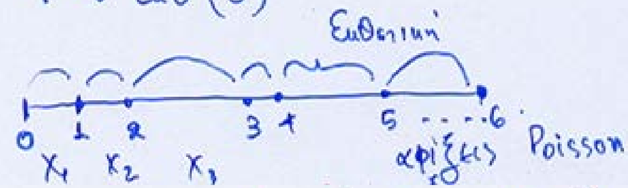
Έστω $X \sim Poisson(\theta)$. Ποιά είναι η πιθανότητα στο χρονικό διάστημα $(0, t]$ να μη συμβεί κανένα γεγονός

$$P(X=0) = e^{-\theta t}\frac{(\theta t)^0}{0!} = e^{-\theta t}$$

Αυτό είναι ισοδύναμο με τον χρόνο αναμονής μέχρι την πρώτη άφιξη (T) να είναι τουλάχιστον t

$$P(T > t) = e^{-\theta t} \Rightarrow P(T \le t) = 1 - e^{-\theta t}$$

Άρα $T \sim ΕκΘ(\theta)$



Εκθετική κατανομή (χρόνος ζωής)

Έχει σ.π.π. της μορφής

$$f(x) = \begin{cases} \theta e^{-\theta x} & x > 0 \\ 0 & \text{αλλού} \end{cases} \qquad (\theta > 0 \text{ παράμετρος})$$

Η αθροιστική σ.κ. δίδεται από τον τύπο

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} f(t)\,dt = \int_{0}^{x} \theta e^{-\theta t}\,dt$$

$$= -e^{-\theta t}\Big|_{0}^{x} = 1 - e^{-\theta x} \quad x > 0 \qquad X \sim ΕκΘ(\theta)$$

• Στα 150 ποντίκια ενός εργαστηρίου 20 είναι μαύρα. Παίρνουμε 4 ποντίκια. Ποιά είναι η πιθανότητα να πάρουμε 1 μαύρο;

(i) Με επανάθεση

(ii) Χωρίς επανάθεση

• Ένας εντομολόγος μελετά τον αριθμό των ζωυφίων στα φύλλα ενός δένδρου. Παρατηρεί δε ότι κατά μέσο όρο εμφανίζονται 10 ζωύφια σε κάθε φύλλο.

i) Ποιά η πιθανότητα να πάρει ένα φύλλο με τουλάχιστον 5 ζωύφια

ii) Ποιά η πιθανότητα να πάρει τρία φύλλα από τα οποία τα 2 να έχουν τουλάχιστον 5 ζωύφια

iii) Ποιά η πιθανότητα στα τρία φύλλα να υπάρχουν συνολικά 12 τουλάχιστον ζωύφια

Αεροπορική εταιρεία έχει διαπιστώσει ότι κατά μέσο όρο 5% των προσώπων με κρατημένες θέσεις δεν εμφανίζονται κατά την αναχώρηση. Πουλάει 75 εισιτήρια για αεροπλάνο 70 θέσεων. Ποιά η πιθανότητα κανένα πρόσωπο να μείνει χωρίς θέση.

Έχει παρατηρηθεί ότι 3 άτομα το μήνα κατά μέσο όρο πεθαίνουν στην Πάτρα από μια σπάνια ασθένεια. Να υπολογισθούν οι πιθανότητες

α) να υπάρξουν το πολύ 2 θάνατοι από την ασθένεια αυτή σε ένα μήνα

β) να υπάρξουν το πολύ 4 θάνατοι σε χρονικό διάστημα 2 μηνών

γ) να υπάρξουν 2 τουλάχιστον μήνες με 2 το πολύ θανάτους, στο επόμενο τρίμηνο.

Σε κατάστημα καταφθάνουν 20 πελάτες την ώρα. Ποια η πιθανότητα ο χρόνος μεταξύ δύο διαδοχικών αφίξεων

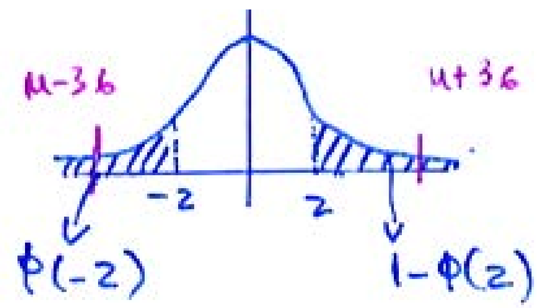
α) να είναι μικρότερος από 3 λεπτά

β) —"— μεγαλύτερος από 4 λεπτά

Κανονική κατανομή $\mathrm{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{για} \quad x \in \mathbb{R}$$
$$\text{με} \quad \mu \in \mathbb{R}, \quad \sigma > 0$$

- Εάν η τ.μ. $X \sim \mathrm{N}(\mu, \sigma^2)$ τότε για την τ.μ. $Z = \frac{X-\mu}{\sigma}$ ισχύει $Z \sim \mathrm{N}(0,1)$ δηλαδή είναι μια τυπική κανονική κατανομή
- για την τυπική κανονική η α.σ.κ. συμβολίζεται με $\Phi(x) = P(Z \leq x)$ δεν υπολογίζεται αναλυτικά αλλά χρησιμοποιούμε πίνακες.
- ισχύει $\Phi(-x) = 1 - \Phi(x)$



- εάν οι τ.μ. $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ και $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ είναι ανεξάρτητες τότε για την τ.μ. $Y = X_1 + X_2$ ισχύει $Y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

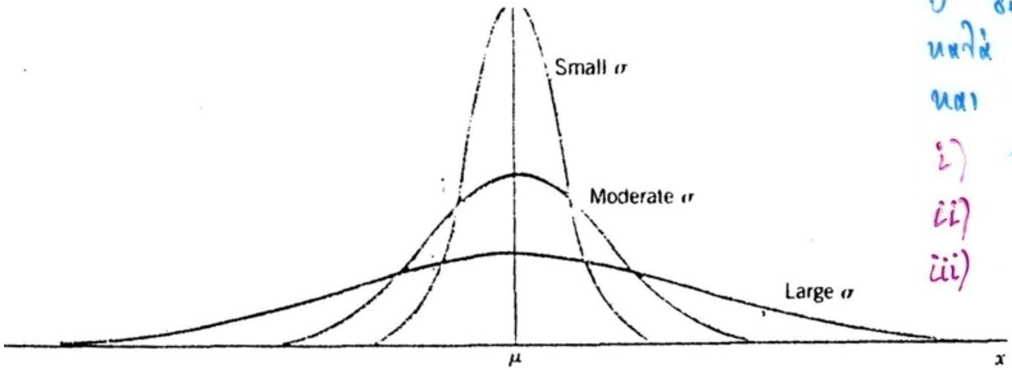


Figure 7 Decreasing σ increases the maximum height and the concentration of probability about μ.

Ο δείκτης IQ σε μια ομάδα ανθρώπων προσεγγίζεται καλά από την κανονική κατανομή με $\mu = 105$ και $\sigma = 20$. Ποιό ποσοστό ανθρώπων έχει IQ

i) τουλάχιστον 50;

ii) το πολύ 80;

iii) ανάμεσα σε 95 και 125;

όταν $X: N(\mu, \sigma^2)$

$$P(\alpha < X \le \beta) = P\left(\frac{\alpha-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{\beta-\mu}{\sigma}\right)$$

$$= P\left(\frac{\alpha-\mu}{\sigma} < Z \le \frac{\beta-\mu}{\sigma}\right) = \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right)$$

όπου $Z: N(0,1)$ **και** $\Phi(.)$ **η ασκ της Z**

έτσι

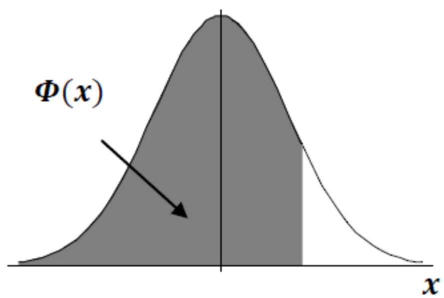
$$P(\mu-\sigma < X \le \mu+\sigma) = \Phi(1) - \Phi(-1) = \Phi(1) - (1-\Phi(1)) = 2\Phi(1) - 1 = 0.682$$

$$P(\mu-2\sigma < X \le \mu+2\sigma) = 2\Phi(2) - 1 = 0.954$$

$$P(\mu-3\sigma < X \le \mu+3\sigma) = 2\Phi(3) - 1 = 0.997$$

Πίνακας τιμών της σ.κ. Φ της τυπικής κανονικής κατανομής ($x = 0, .01, .02, ..., 3.49$)

$$\boldsymbol{\Phi(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt}$$



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5674	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.6315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Έστω τμ X και συνάρτηση $g : \mathbb{R} \to \mathbb{R}$. Τότε ορίζεται η $E[g(X)]$ ως

- $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$ όταν η τμ X είναι συνεχής
- $E[g(X)] = \sum_{x=0}^{\infty} g(x)P(x)$ όταν η τμ X είναι διακριτή

Μέτρα Θέσης

Η Μέση τιμή (προσδοκώμενη, μαθηματική ελπίς) μιας τ.μ. X, συμβολίζεται με $E(X)$, και ορίζεται ως εξής:

$$E(x) = \sum_i x_i P(X = x_i) \qquad E(x) = \int_{-\infty}^{+\infty} x f(x)\, dx$$

για διακριτή ή για συνεχή τ.μ.

Ιδιότητες Μέσης τιμής

- $E(X+c) = E(X) + c$ όπου c = σταθερά
- $E(aX) = a E(X)$ όπου a = σταθερά
- Έστω X, Y, τ.μ. και υπάρχουν $E(X)$, $E(Y)$ τότε
  $E(X+Y) = E(X) + E(Y)$
  γενικά
  $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$ α, β = σταθερές

Μέτρα Διασποράς

Η Διασπορά (διακύμανση, κεντρική ροπή δεύτερης τάξεως) μιας τ.μ. X, συμβολίζεται με $\Delta(X)$ ή $Var(X)$, και ορίζεται ως

$$\Delta(X) = E\left((X-\mu)^2\right) \qquad \text{όπου } \mu = E(X)$$

κατά συνέπεια $\Delta(X) = E(X^2) - [E(X)]^2$

τυπική απόκλιση σ ονομάζουμε την ποσότητα

$$\sigma^2 = \Delta(X) \Rightarrow \sigma = \sqrt{\Delta(X)}$$

Ιδιότητες Διασποράς

- $\Delta(X+c) = \Delta(X)$ όπου c: σταθερά
- $\Delta(\alpha X) = \alpha^2 \Delta(X)$ όπου α: σταθερά
- Έστω οι τ.μ. X, Y, είναι ανεξάρτητες μεταξύ τους
  $\Delta(X+Y) = \Delta(X) + \Delta(Y)$

Μέση Τιμή και Διασπορά για Διακριτές κατανομές.

Bernoulli (p)

$E(x) = p \qquad \Delta(x) = p(1-p) = pq$

Διωνυμική $B(n,p)$

$E(x) = np \qquad \Delta(x) = npq$

Υπεργεωμετρική n: δείγμα. κ: 1ης κατ. $N-\kappa$: 2ης κατ.

$E(x) = \dfrac{n\kappa}{N} \qquad \Delta(x) = n \dfrac{\kappa}{N} \dfrac{N-\kappa}{N} \dfrac{N-n}{N-1}$

Poisson Poisson (λ)

$E(x) = \lambda \qquad \Delta(x) = \lambda$

Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τ.μ. με $E(X_i) = \mu$ και $\Delta(X_i) = \sigma^2 \quad i = 1, \ldots, n$.

Τότε ορίζω την τ.μ. $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

δηλ. τον μέσο όρο τους

$$E(\overline{X}) = \frac{1}{n} E(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\left[E(X_1) + \cdots + E(X_n)\right]$$
$$= \frac{1}{n}\, n \cdot \mu = \mu$$

$$\Delta(\overline{X}) = \frac{1}{n^2} \Delta(X_1 + X_2 + \cdots + X_n) \overset{\text{ανεξ}}{=} \frac{1}{n^2}\left(\Delta(X_1) + \cdots + \Delta(X_n)\right)$$
$$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Μέση τιμή και Διασπορά για Συνεχείς τ.μ.

Εκθετική $\text{Εκθ}(\theta)$

$E(x) = \dfrac{1}{\theta} \qquad \Delta(x) = \dfrac{1}{\theta^2}$

Κανονική $N(\mu, \sigma^2)$

$E(x) = \mu \qquad \Delta(x) = \sigma^2$

Εάν είναι η τυποποιημένη $E(x) = 0 \quad \Delta(x) = 1$

Ασθενής Νόμος των Μεγάλων Αριθμών

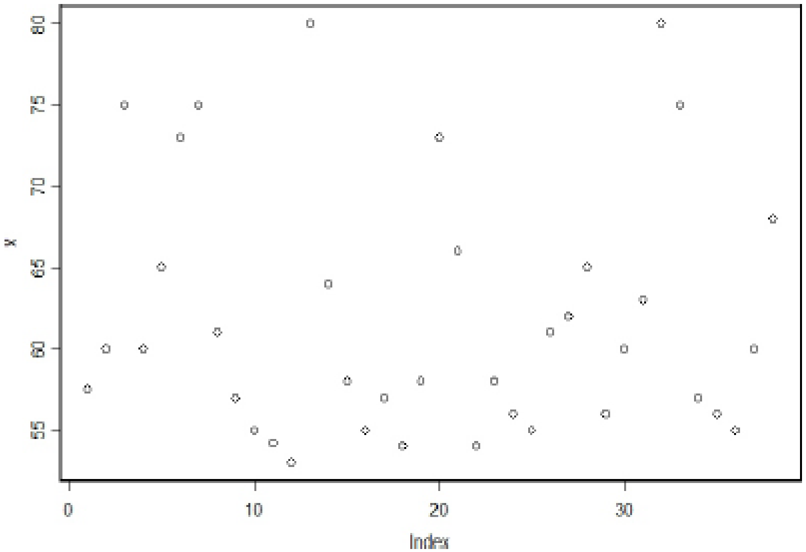
- Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τ.μ. με $E(X_1) = \mu < \infty$ τότε για τον δειγματικό μέσο ισχύει

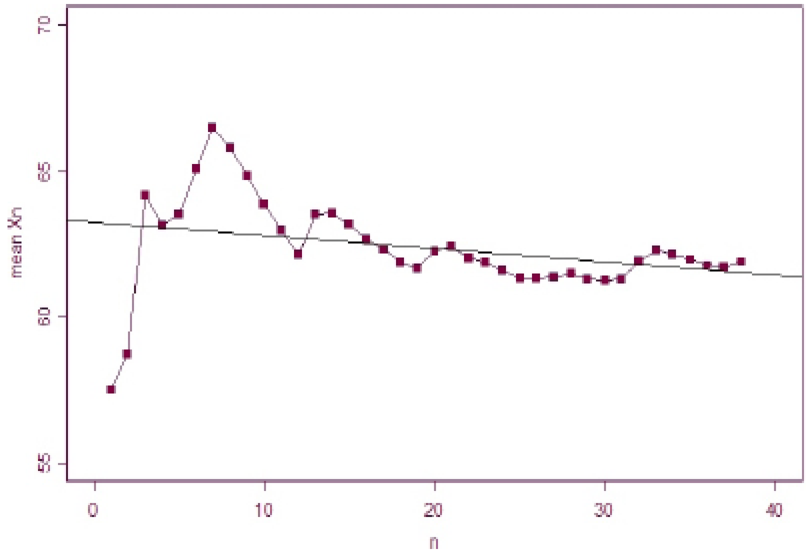
$$\overline{X_n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{\text{κατά πιθανότητα}} \mu$$

δηλαδή

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \leq \varepsilon) = 1 \; \forall \varepsilon > 0$$

- τυχαίο δείγμα μεγέθους $n = 38$ των βαρών των φοιτητών που παρακολούθησαν το μάθημα της Στατιστικής στις 7/10/2009





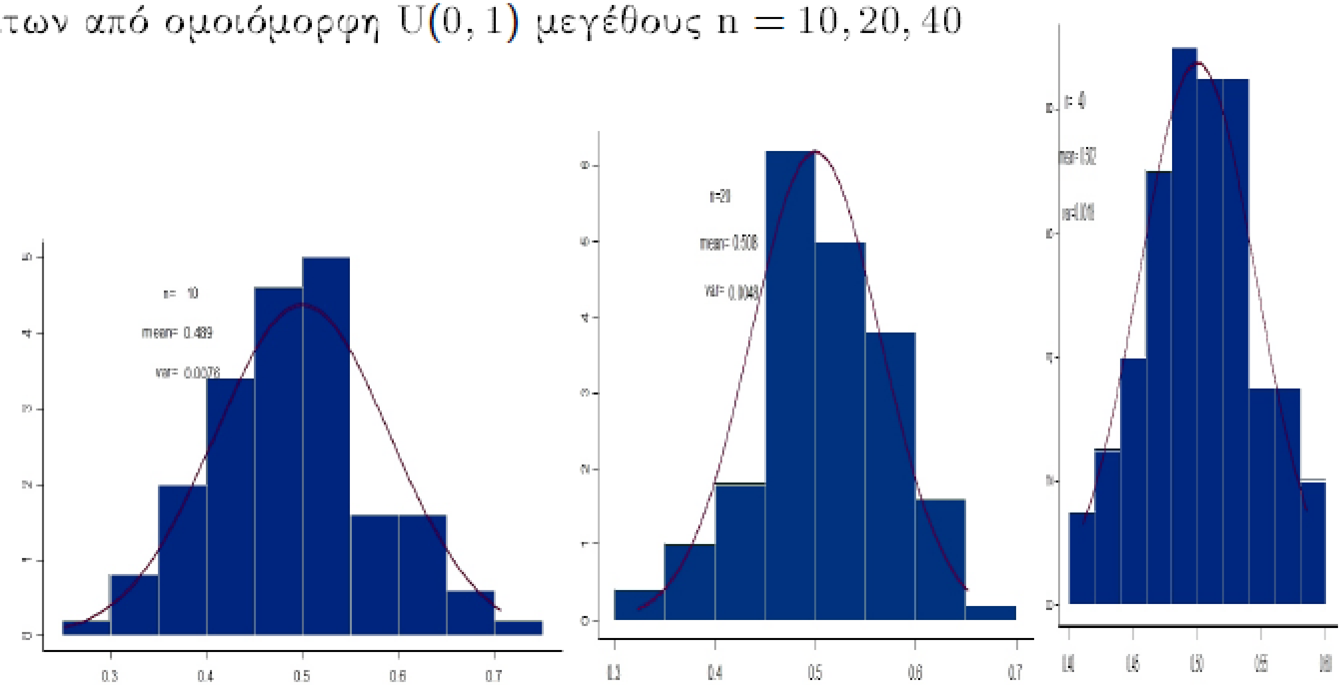
Κεντρικό Οριακό Θεώρημα

- Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τ.μ. με $E(X_1) = \mu < \infty$ και $0 < Var(X_1) = \sigma^2 < \infty$ τότε για τον δειγματικό μέσο ισχύει

$$\overline{X_n} \xrightarrow[n\to\infty]{\text{κατά κατανομή}} N(\mu, \frac{\sigma^2}{n}) \qquad \text{ή} \qquad \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{\text{κ. κ.}} N(0,1)$$

δηλαδή $\lim_{n\to\infty} P(\frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \leq x) = \Phi(x)\ \forall x \in \mathbb{R}$

- ιστογράμματα σχετικών συχνοτήτων (100 παρατηρήσεων) των μέσων τιμών τυχαίων δειγμάτων από ομοιόμορφη $U(0,1)$ μεγέθους $n = 10, 20, 40$



Εναλλακτική μορφή για το Κ.Ο.Θ.

με τις προηγούμενες προϋποθέσεις η ακολουθία

$$\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\text{κατά Νόμο}} N(n\mu, n\sigma^2)$$

• Σε περίπτωση που ζητείται η πιθανότητα

$$P\left(\alpha < \sum_{i=1}^{n} X_i < \beta\right) = P\left(\frac{\alpha - n\mu}{\sqrt{n}\,\sigma} < \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} < \frac{\beta - n\mu}{\sqrt{n}\,\sigma}\right)$$

$$\xrightarrow[n\to+\infty]{} P\left(\frac{\alpha - n\mu}{\sqrt{n}\,\sigma} < Z < \frac{\beta - n\mu}{\sqrt{n}\,\sigma}\right) = \Phi\left(\frac{\beta - n\mu}{\sqrt{n}\,\sigma}\right) - \Phi\left(\frac{\alpha - n\mu}{\sqrt{n}\,\sigma}\right)$$

όπου η τ.μ. $Z \sim N(0,1)$

και $\Phi(\cdot)$ η α.κ.κ. της τυποποιημένης Κανονικής την οποία βρίσκουμε από σχετικούς πίνακες.

Διασύνδεση Διωνυμικής με Κανονική

Αν $X \sim B(n,p)$ και $n \geq 30$ $np \geq 5$

χρησιμοποιούμε το κεντρικό οριακό θεώρημα

$X \sim N(np, np(1-p))$

$$P(X=\kappa) = P\left(\kappa - \tfrac{1}{2} < X < \kappa + \tfrac{1}{2}\right)$$

$$P(\alpha \leq X \leq \beta) = P\left(\frac{\alpha - \frac{1}{2} - np}{\sqrt{np(1-p)}} < Z < \frac{\beta + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

6:5

Ο χρόνος αναμονής ενός φοιτητή κάθε πρωί στη στάση του λεωφορείου, ακολουθεί Εκθετική κατανομή με μέσο χρόνο αναμονής 10 λεπτά.
Εάν ο φοιτητής κατά τη διάρκεια του έτους πηγαίνει 120 ημέρες στο Πανεπιστήμιο με λεωφορείο ποιά η πιθανότητα ο συνολικός χρόνος αναμονής να υπερβαίνει τις 20 ώρες;

• Σε μια εκλογή ο υποψήφιος Α έχει την υποστήριξη του 40% του πληθυσμού. Αν σε 100 ψηφοφόρους που ρωτάμε τυχαία οι Χ μας πουν ότι υποστηρίζουν τον Α

i) ποια είναι η μέση τιμή και η διασπορά της τ.μ. Χ

ii) Να βρεθεί η $P(30 < X \leq 46)$.