

## Στατιστικές μετρήσεις & Κατανομές

### Περιεχόμενα:

#### 1. Στατιστικές μετρήσεις

- (α') Μέσος
- (β') Διάμεσος
- (γ') Επικρατούσα τιμή
- (δ') Έυρος
- (ε') Τυπική απόκλιση
- (ϛ') Διασπορά

#### 2. Κατανομές

- (α') Διωνυμική
- (β') Poisson
- (γ') Κανονική

#### 3. Παραδείγματα

- Οι μετρήσεις Μέσος, Διάμεσος και Επικρατούσα τιμή ονομάζονται μέσοι όροι ή μέτρα θέσης.
- Οι μετρήσεις Τυπική απόκλιση και Διακύμανση ονομάζονται μέτρα διασποράς.

Ο Μέσος

Ο αριθμητικός μέσος (ή απλά μέσος) ενός συνόλου αριθμών  $\{x_1, x_2, \dots, x_n\}$  συμβολίζεται με  $\bar{x}$  και ορίζεται ως:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Θεωρήστε τη συχνότητα μιας διακριτής κατανομής που παίρνει τιμές  $\{x_1, x_2, \dots, x_n\}$  με αντίστοιχες συχνότητες  $\{f_1, f_2, \dots, f_n\}$ . Ο μέσος  $\bar{x}$  δίνεται από:

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}.$$

### Παράδειγμα

- Βρείτε το μέσο της σειράς  $\{-3, -1, 0, 2, 3, 4\}$ .

$$\bar{x} = (-3 - 1 + 0 + 2 + 3 + 4)/6 = 0.83.$$

- Βρείτε το μέσο της πιο κάτω συχνότητας κατανομής:

$x_i$	-3	-2	-1	0	1	2	3
$f_i$	6	5	4	3	2	1	1
$f_i x_i$	-18	-10	-4	0	2	2	3

$$\sum f_i = 22, \quad \sum f_i x_i = -25, \quad \bar{x} = -25/22 = -1.14.$$

Θεωρούμε ότι η σειρά  $\{x_1, x_2, \dots, x_n\}$  μετασχηματίζεται σε  $\{X_1, X_2, \dots, X_n\}$ , όπου  $X_i = (x_i - a)/b$ . Το  $\bar{x}$  πηγάζει από

$$\boxed{\bar{x} = a + b\bar{X}}$$

### Απόδειξη

Από το  $X_i = (x_i - a)/b$  συμπεραίνουμε ότι  $x_i = bX_i + a$ . Τώρα,

$$\begin{aligned}\bar{x} &= \frac{1}{n} \sum x_i = \frac{1}{n} \sum (a + bX_i) \\ &= \frac{1}{n} \sum a + \frac{b}{n} \sum X_i \\ &= a + b\bar{X}.\end{aligned}$$

□

### Παράδειγμα

Βρέστε το μέσο της σειράς  $\{2678, 4678, 8678, 5678, 6678\}$ .

Θεωρούμε  $X_i = (x_i - a)/b$ , όπου  $a = 2678$  και  $b = 1000$ . Αυτό δίνει μια νέα σειρά  $X = \{0, 2, 6, 3, 4\}$  με μέσο  $\bar{X} = 3$ . Έτσι,

$$\bar{x} = 2678 + 1000 \times 3 = 5678$$

## Διάμεσος

Ο μέσος ως μέθοδος μέτρησης έχει το μειονέκτημα ότι λαμβάνει υπόψη ακραίες τιμές, ειδικά για ένα μικρό σύνολο αριθμών.

Η διάμεσος ενός συνόλου αριθμών  $\{x_1, x_2, \dots, x_n\}$  ορίζεται ως η μεσαία τιμή του συνόλου αριθμών όταν αυτοί παραταχθούν σε αύξουσα σειρά. Αν η σειρά έχει άρτιο αριθμό στοιχείων, τότε η διάμεσος υπολογίζεται ως η μέση τιμή των δύο μεσαίων τιμών.

### Παράδειγμα

1. Οι μισθοί που κατατάσσονται σε αύξουσα σειρά είναι:  $\{28, 29, 32, 35, 36, 38, 41, 103\}$ .

Ο μέσος είναι  $\bar{x} = 41.89$  και η διάμεσος  $x^* = 35.5$ .

2. Βρείτε τη διάμεσο της σειράς:  
 $\{65, 68, 68, 66, 64, 65, 65, 67\}$ .

Ταξινομώντας τους σε αύξουσα σειρά:  
 $\{64, 65, 65, 65, 66, 67, 68, 68\}$ .

Η διάμεσος δίνεται από:  $(65 + 66)/2 = 65.5$ .

Θεωρείστε τη διακριτή κατανομή συχνοτήτων που παίρνει τιμές  $\{x_1, x_2, \dots, x_n\}$  με αντίστοιχες συχνότητες  $\{f_1, f_2, \dots, f_n\}$ . Η διάμεσος δίνεται από τον αριθμό

$$\left(\frac{1 + \sum f}{2}\right)^{\text{th}}$$

όταν καταταχθούν οι τιμές.

### Παράδειγμα

Βρέστε τη διάμεσο της ακόλουθης διακριτής κατανομής:

$x_i$	0	1	2	3	4	5	6
$f_i$	5	5	10	20	30	20	10
Cum $f$	5	10	20	40	70	90	100

$$\sum f_i = 100, \quad (1 + \sum f_i)/2 = 50.5.$$

Ο 50.5ος όρος εμπίπτει  $x = 4$  χρησιμοποιώντας αθροιστική συχνότητα (Cum  $f$ ). Έτσι η διάμεσος είναι 4.

### Η επικρατούσα τιμή

Η επικρατούσα τιμή (ή κορυφή) ενός συνόλου τιμών ορίζεται ως η τιμή που παρουσιάζεται με τη μεγαλύτερη συχνότητα.

Σημιώστε ότι για ένα σύνολο που δεν περιλαμβάνει επαναλαμβανόμενες τιμές η επικρατούσα τιμή δεν υπάρχει.

#### Παράδειγμα

Η επικρατούσα τιμή του συνόλου  $\{2, 3, 3, 1, 3, 2, 4, 5, 8, 3, 2, 4, 4, 3\}$  είναι το 3.

Να σημειωθεί, ότι θα πρέπει να υπάρχουν περισσότερες από μία επικρατούσες τιμές σε ένα σύνολο αριθμών.

#### Παράδειγμα

Το σύνολο  $\{8, 6, 8, 5, 5, 7, 6, 8, 6, 9\}$  έχει δύο επικρατούσες τιμές το 6 και το 8.

### Το Έυρος

Το έυρος ενός συνόλου  $S = \{x_1, x_2, \dots, x_n\}$  δίνεται από:

$$\mathbf{Range} = \max(S) - \min(S).$$

Το έυρος είναι η πιο απλή μέθοδος μέτρησης της διασποράς και μπορεί να υπολογιστεί γρήγορα και εύκολα.

Δεν θεωρείται όμως αξιόπιστο μέτρο υπολογισμού της αφού χρησιμοποιεί μόνο ακραίες τιμές.

### Παραδείγματα

- Η σειρά  $\{6, \boxed{5}, 7, \boxed{10}, 8, 9\}$  έχει έυρος  $= 10 - 5 = 5$ .
- Η σειρά  $\{600, 610, 620, \boxed{600}, 610, \boxed{650}, 640, 650, 650\}$  έχει έυρος  $= 650 - 600 = 50$ .
- Η σειρά  $\{600, 610, 620, \boxed{200}, 610, \boxed{1000}, 640, 650, 650\}$  έχει έυρος  $= 800$ .

Τυπική απόκλιση και Διασπορά

Η τυπική απόκλιση είναι το μέτρο διασποράς που χρησιμοποιείται περισσότερο στη στατιστική. Βασίζεται στον αριθμητικό μέσο.

Η τυπική απόκλιση ενός συνόλου αριθμών  $\{x_1, x_2, \dots, x_n\}$  με μέσο  $\bar{x}$  συμβολίζεται  $S$  και ορίζεται ως

$$\begin{aligned} S &= \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n}} \\ &= \sqrt{\frac{1}{n} \sum_i^n x_i^2 - \bar{x}^2} \end{aligned}$$

### Περάδειγμα

Η σειρά  $\{3, 4, 6, 2\}$  έχει  $\bar{x} = 15/4 = 3.75$ ,  $\bar{x}^2 = 14.063$  και  $\sum_i^n x_i^2 = 65$ .

Έτσι,

$$\begin{aligned} S &= \left( \frac{65}{4} - (3.75)^2 \right)^{\frac{1}{2}} = (16.25 - 14.063)^{\frac{1}{2}} \\ &= 1.48 \end{aligned}$$

Για μια διακριτή κατανομή συχνοτήτων, η τυπική απόκλιση ορίζεται ως:

$$S = \sqrt{\frac{\sum_i f_i (x_i - \bar{x})^2}{\sum_i f_i}}$$

$$= \sqrt{\sum_i \frac{f_i x_i^2}{\sum_i f_i} - \bar{x}^2}$$

όπου  $\bar{x} = \sum_i f_i x_i / \sum_i f_i$  είναι ο μέσος της κατανομής συχνότητας.

### Παράδειγμα

Να υπολογιστεί η διάμεσος της ακόλουθης διακριτής κατανομής:

							$\Sigma$
$x_i$	0	1	2	3	4	5	
$f_i$	1	3	11	9	5	2	31
$f_i x_i$	0	3	22	27	20	10	82
$x_i^2$	0	1	4	9	16	25	
$f_i x_i^2$	0	3	44	81	80	50	258

$$\begin{aligned}
 S &= \sqrt{\sum_i \frac{f_i x_i^2}{\sum_i f_i} - \left( \frac{\sum_i f_i x_i}{\sum_i f_i} \right)^2} = \sqrt{\frac{258}{31} - \left( \frac{82}{31} \right)^2} \\
 &= 1.15
 \end{aligned}$$

Η Διασπορά ή Διακύμανση ενός συνόλου αριθμών ή μιας κατανομής ορίζεται ως το τετράγωνο της τυπικής απόκλισης και συμβολίζεται  $S^2$ .

Για ένα σύνολο τιμών:

$$\begin{aligned}
 S^2 &= \frac{\sum_i^n (x_i - \bar{x})^2}{n} \\
 &= \frac{1}{n} \sum_i^n x_i^2 - \bar{x}^2.
 \end{aligned}$$

Για μια κατανομή συχνότητας:

$$\begin{aligned}
 S^2 &= \frac{\sum_i f_i (x_i - \bar{x})^2}{\sum_i f_i} \\
 &= \sum_i \frac{f_i x_i^2}{\sum_i f_i} - \bar{x}^2.
 \end{aligned}$$

Θεωρούμε τις τιμές  $\{x_1, x_2, \dots, x_n\}$  που μετασχηματίζονται σε  $\{X_1, X_2, \dots, X_n\}$ , όπου  $X_i = (x_i - a)/b$ . Εάν η τυπική απόκλιση του  $X$  και  $x$  συμβολίζονται αντίστοιχα  $S_X$  και  $S_x$ , τότε

$$S_x = bS_X.$$

### Απόδειξη

Για  $X_i = (x_i - a)/b$  συνεπάγεται ότι  $x_i = bX_i + a$ .

Τώρα,

$$\begin{aligned} S_x &= \sqrt{\frac{(\sum x_i - \bar{x})^2}{n}} = \sqrt{\frac{(\sum bX_i + a - (a + b\bar{X}))^2}{n}} \\ &= b \sqrt{\frac{(\sum X_i - \bar{X})^2}{n}} = bS_X. \end{aligned}$$

### Παράδειγμα

Βρείτε την τυπική απόκλιση των παρατηρήσεων  $\{2678, 4678, 8678, 5678, 6678\}$ .

Θεωρούμε ότι  $X_i = (x_i - a)/b$ , όπου  $a = 2678$  και  $b = 1000$ . Αυτό δίνει ένα νέο σύνολο τιμών

$X = \{0, 2, 6, 3, 4\}$  με μέσο  $\bar{X} = 3$  και τυπική απόκλιση  $S_X = \sqrt{20/5} = 2$ . Έτσι,

$$S_x = bS_X = 1000 \times 2 = 2000.$$

## Σταστικές μετρήσεις & Κατανομές

### Διάλεξη 3

#### Κατανομές Πιθανοτήτων

1. Διωνυμική
2. Poisson
3. Κανονική

## Διωνυμική Κατανομή

### Συνδιασμοί

Οι Συνδιασμοί είναι ένα υποσύνολο στοιχείων που δε βρίσκονται σε σειρά. Η διάταξη των στοιχείων δεν έχει σημασία όταν προσδιορίζουμε τους συνδιασμούς.

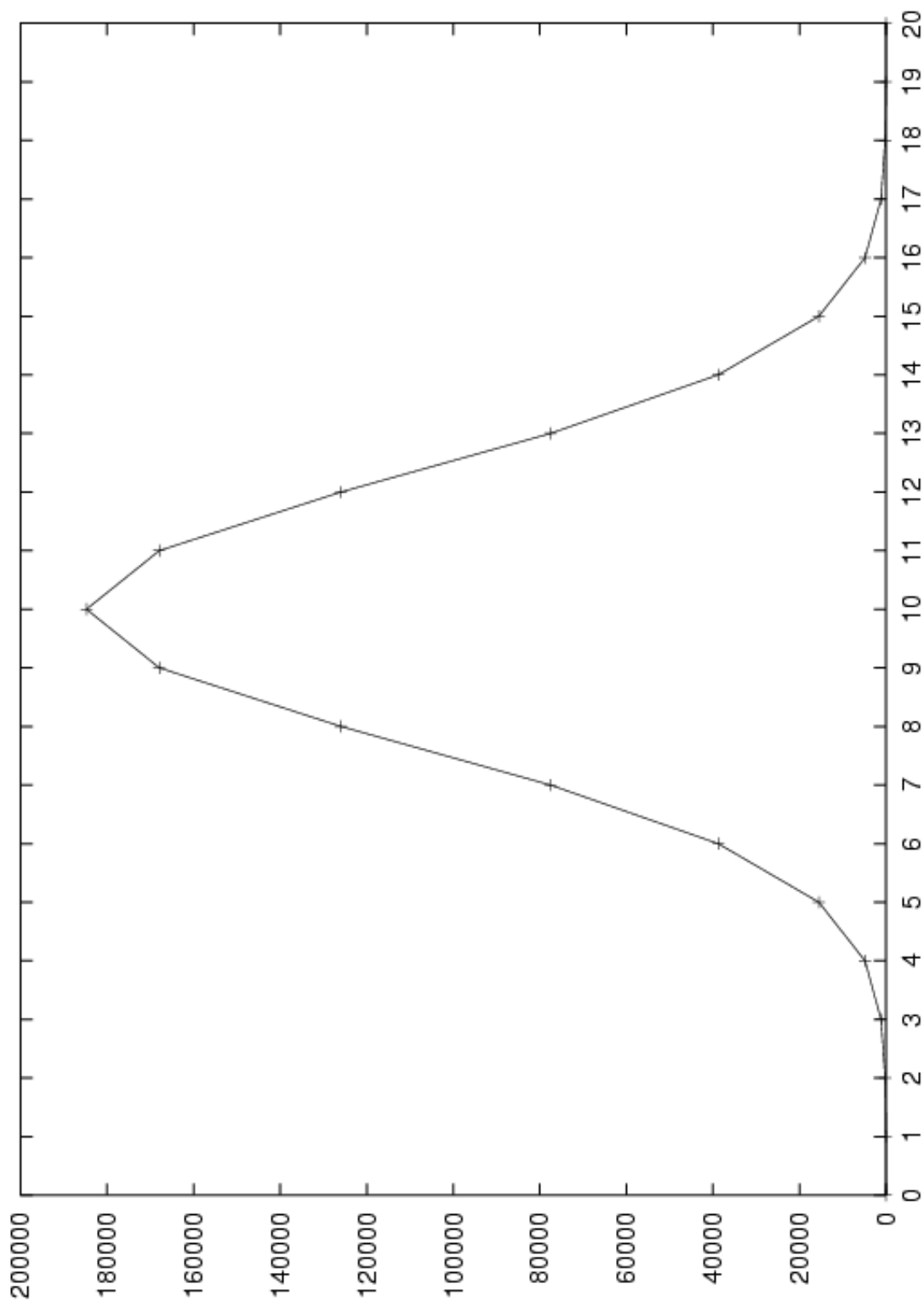
Ο αριθμός των συνδιασμών των  $r$  στοιχείων που παίρνουμε από ένα σύνολο  $n$  στοιχείων δίνεται από:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \quad \text{όπου} \quad r \leq n.$$

Το  $n!$  ( $n$  παραγοντικό) δίνεται από  
 $n! = 1 \times 2 \times \cdots \times n$  και  $0! = 1$ .

Η έκφραση  $\binom{n}{r}$  μπορεί να γραφτεί ως  $C_r^n$ , ή  ${}_n C_r$ , ή  $C(n, r)$ .

$\binom{20}{x}$



## Παράδειγμα

Τέσσερα αγαθά που συμβολίζονται  $A, B, C$  και  $D$  εξετάζονται από ένα διαχειστή κεφαλαίου. Υπάρχει όμως ο περιορισμός ότι μπορεί να επενδύσει μόνο σε τρία από αυτά. Πόσα διαφορετικά χαρτοφυλάκια μπορούμε να δημιουργήσουμε;

## Λύση

Ο διαχειριστής χαρτοφυλακίου κάνει επιλογή 3 αγαθών από  $A, B, C$  και  $D$  χωρίς η σειρά να παίζει ρόλο. Έχει 4 πιθανούς συνδιασμούς:

$$ABC, \quad ABD, \quad ACD, \quad BCD.$$

Διαφορετικά,  $n = 4$  και  $r = 3$  έτσι ώστε

$$\binom{n}{r} = \binom{4}{3} = \frac{4!}{3!(1)!} = 4.$$

Η Διωνυμική κατανομή έχει εφαρμογή σε περιπτώσεις όπου καταμετρούμε γεγονότα.

Για τη διωνυμική κατανομή προκύπτει από τις ακόλουθες υποθέσεις:

1. Έχει σταθερό αριθμό  $n$  δοκιμών.
2. Κάθε δοκιμή έχει μόνο δύο δυνατά αποτελέσματα (επιτυχία ή αποτυχία.)
3. Οι δοκιμές είναι ανεξάρτητες.
4. Η πιθανότητα επιτυχίας  $p$ , είναι σταθερή για όλες τις δοκιμές.

### Παράδειγμα

Θεωρείστε ένα επενδυτή που αγοράζει 3 χρεόγραφα. Αν η τιμή των χρεογράφων αυξηθεί (μειωθεί), τότε ο επενδυτής κερδίζει (ζημιώνει). Υποθέστε ότι οι τιμές των 3 χρεογράφων του χαρτοφυλακίου είναι ανεξάρτητες. Θεωρούμε  $p$  ( $q = 1 - p$ ) την πιθανότητα η τιμή να πάει πάνω (κάτω). Έτσι,  $P(U) = p$  και  $P(D) = q = 1 - p$ . Η πιθανότητα οι τιμές των δύο πρώτων χρεογράφων να κινηθούν ανοδικά και ένα άλλο να κινηθεί καθοδικά δίνεται από:

$$P(UUD) = P(U)P(U)P(D) = p p q = p^2 q.$$

Σημειώστε ότι  $P(UUD) = P(UDU) = P(DUU)$ .

Ας θεωρήσουμε ότι  $B(x; 3; p)$  υποδηλώνει την πιθανότητα οι τιμές  $x$  χρεογράφων να ανέβουν. Τότε,

Συνδ.	# αύξηση	Πιθαν.
DDD	0	$q^3$
DDU	1	$pq^2$
DUD	1	$pq^2$
UDD	1	$pq^2$
DUU	2	$p^2q$
UDU	2	$p^2q$
UUD	2	$p^2q$
UUU	3	$p^3$

ή

$x$	$B(x; 3; p)$
0	$q^3$
1	$3pq^2$
2	$3p^2q$
3	$p^3$

ή

$$B(x; 3; p) = \binom{3}{x} p^x (1-p)^{3-x} \quad \text{όπου} \quad x = 0, 1, 2, 3.$$

Σημειώστε ότι

$$\sum_1^3 B(x; 3; p) = q^3 + 3pq^2 + 3p^2q + p^3 = (p+q)^3 = 1.$$

Η Διωνυμική κατανομή δίνεται από τον τύπο:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{για } x = 0, 1, \dots, n.$$

Οι παράμετροι είναι ο ακέραιος  $n \geq 0$  και η πιθανότητα  $p$  ( $0 \leq p \leq 1$ ).

Ο μέσος και η διακύμανση σε μια Διωνυμική κατανομή δίνονται από:

$$\mathbf{E}(x) = n p$$

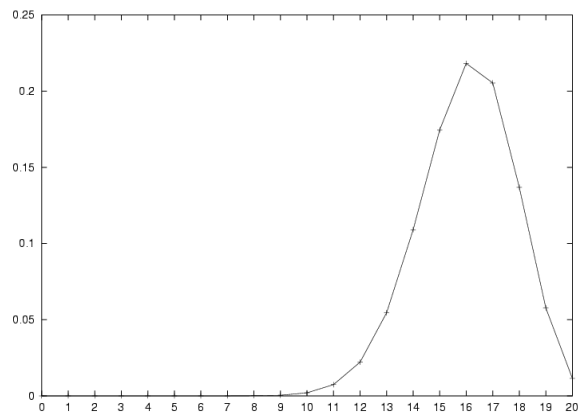
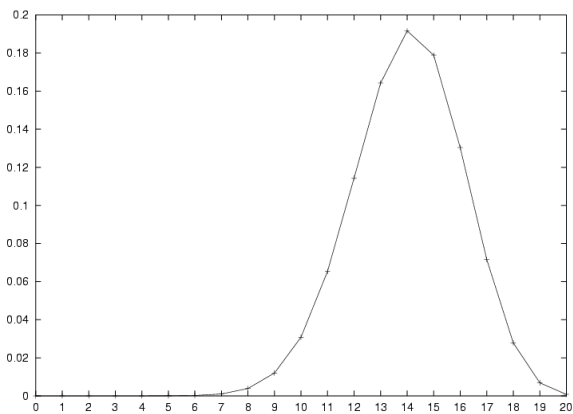
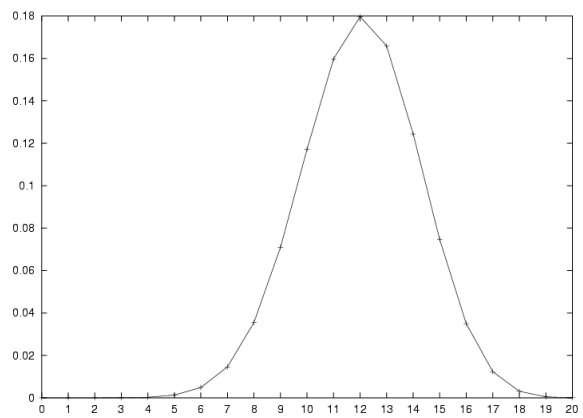
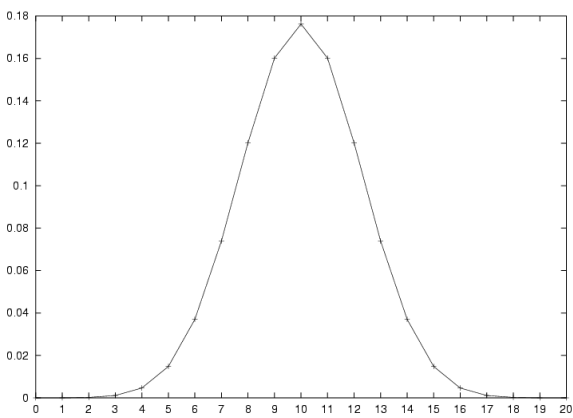
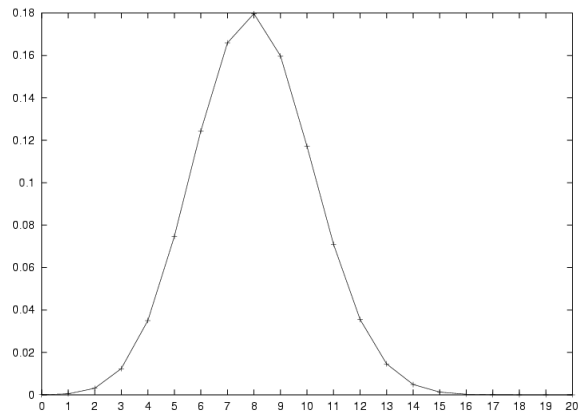
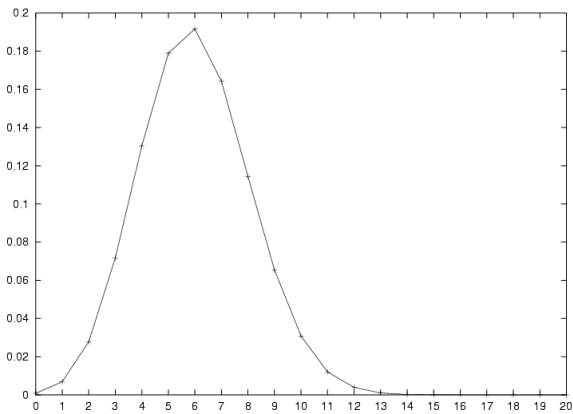
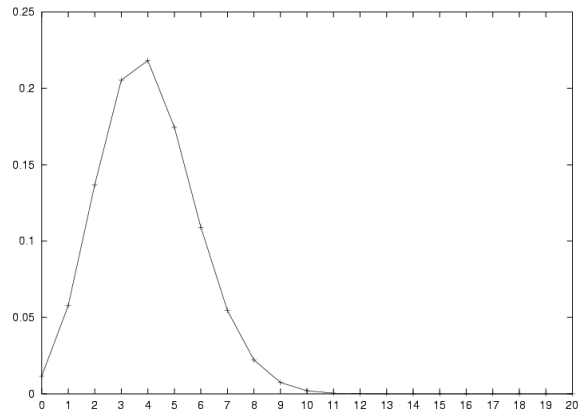
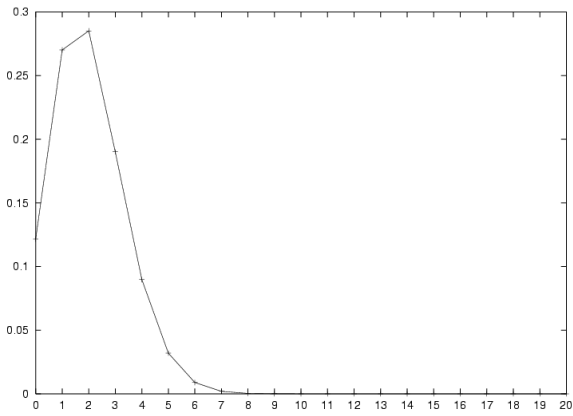
$$\mathbf{Var}(x) = n p (1 - p).$$

Η σχέση

$$P(X = x + 1) = \frac{p}{1 - p} \frac{n - x}{x + 1} P(X = x)$$

είναι γνωστή ως Διωνυμικός Αναδρομικός Τύπος .

$B(x, 20, p)$ , όπου  $p = 0.1, 0.2, \dots, 0.8, 0.9$ .



## Παράδειγμα

Στο πείραμα ρήψης νομίσματος η πιθανότητα να φέρουμε κορώνα είναι 0.5. Αν ρίξουμε το νόμισμα 5 φορές, τότε ποιά είναι η πιθανότητα (α) να φέρουμε ακριβώς 2 φορές κορώνα και (β) περισσότερες από μία φορές κορώνα;

Θεωρούμε  $x$  τον αριθμό των κορώνων,  $n = 5$  και  $p = q = 1 - p = 1/2$ .

1.  $P(X = 2) = B(2; 5; p) = \binom{5}{2} p^2 (1 - p)^3 = 5/16.$

2.

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) \\ &= 1 - \left( P(x = 0) + P(x = 1) \right) \\ &= 1 - B(0; 5; 0.5) - B(1; 5; 0.5) \\ &= 13/16. \end{aligned}$$

Δεδομένου της κατανομής συχνότητας η διωνυμική κατανομή μπορεί να υπολογιστεί ως ακολούθως:

1. Υπολογίζοντας τις τιμές των παραμέτρων  $n$  και  $p$  είτε  
είτε  
(α') Από μια γνωστή διωνυμική κατανομή  
(β') Υπολογίζοντας το μέσο  $\bar{x}$  της κατανομής συχνότητας και χρησιμοποιώντας τη σχέση  $\bar{x} = np$ .
2. Παράγουμε τη διωνυμική κατανομή πιθανότητας χρησιμοποιώντας  $n$  και  $p$ .
3. Παράγουμε τις Αναμενόμενες συχνότητες πολλαπλασιάζοντας τις συνολικές συχνότητες  $\sum_i f_i$  με την πιθανότητα.

### Παράδειγμα

Μια μεροληπτική ρήψη ζαριού 5 φορές αποτελεί ένα πείραμα. Το πείραμα επαναλαμβάνεται 250 φορές.

Συνολικά η συχνότητα εμφάνισης των άρτιων (ζυγών) αριθμών στο πείραμα καταγράφεται δίνοντας τα ακόλουθα αποτελέσματα:

# ζυγοί	0	1	2	3	4	5	Σύνολο
Παρατηρ. συζν.	11	41	83	73	36	6	250

Προσαρμογή Διωνυμικής κατανομής στα δεδομένα.

### Στάδιο 1

Ο αριθμός των ρήψεων  $n = 5$ . Η σχέση  $\bar{x} = np$  μπορεί να χρησιμοποιηθεί έτσι ώστε να υπολογίσουμε το  $p$ .

Τώρα,

$x_i$	0	1	2	3	4	5	Σύνολο
$f_i$	11	41	83	73	36	6	250
$f_i x_i$	0	41	166	219	144	30	600

Συνεπάγεται ότι

$$\bar{x} = \sum_i f_i x_i / \sum_i f_i = 600/250 = 2.4.$$

Έτσι, από  $\bar{x} = np$  συνεπάγεται  $2.4 = 5p$ , δηλαδή  $p = 0.48$ .

### Στάδιο 2

Θεωρούμε την τυχαία μεταβλητή  $X$  που αντιπροσωπεύει τον αριθμό των άρτιων στο πείραμα για να έχουμε τη διωνυμική κατανομή:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad \text{για } x = 0, 1, 2, 3, 4, 5,$$

όπου  $n = 5$  και  $p = 0.48$ .

Τώρα  $P(X = 0) = \binom{5}{0} p^0 (1 - p)^5 = 0.038$  και

$$\begin{aligned} P(X = x + 1) &= \frac{p}{1 - p} \frac{n - x}{x + 1} P(X = x) \\ &= 0.9231 \frac{5 - x}{x + 1} P(X = x). \end{aligned}$$

Από τα πιο πάνω συνεπάγεται ότι

$$P(X = 1) = 0.1754, P(X = 2) = 0.324,$$

$$P(X = 3) = 0.2990, P(X = 4) = 0.1380 \text{ και}$$

$$P(X = 5) = 0.0255.$$

### Σύνοψη

$x_i$ : Ο αριθμός των άρτιων.

$f_i$ : Πραγματική συχνότητα.

$P(X = x_i)$ : Πιθανότητα να πάρουμε  $x_i$  άρτιους.

$P(X = x_i) \sum_i f_i$ : Αναμενόμενη συχνότητα.

$x_i$	0	1	2	3	4	5	Σύνολο
$f_i$	11	41	83	73	36	6	250
$P(X = x_i)$	0.04	0.18	0.32	0.3	0.14	0.03	0.9999
$P(X = x_i) \sum_i f_i$	10	44	81	75	34	6	250

**Κατανομή Poisson**

Μια διακριτή τυχαία μεταβλητή  $X$  λέμε ότι ακολουθεί την κατανομή Poisson αν έχει pdf της μορφής

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{για } x = 0, 1, \dots, n,$$

όπου  $\lambda > 0$  είναι η παράμετρος.

Ο μέσος και η διακύμανση της κατανομής Poisson δίνονται από:

$$\mathbf{E}(x) = \mathbf{Var}(x) = \lambda.$$

Η κατανομή Poisson μπορεί να οριστεί ως η περιοριστική περίπτωση διωνυμικής κατανομής για  $n \rightarrow \infty$  με σταθερό  $np = \lambda$ . Συνεπώς, περιγράφει τη συμπεριφορά μεγάλου αριθμού  $n$  ανεξάρτητων πειραμάτων από τα οποία μόνο ένα μικρό τμήμα  $np$  αναμένεται να παράγει ενδεχόμενα παρόμοιου τύπου.

## Παράδειγμα

Η διεκδίκηση 5000 ασφαλιστήριων συμβολαίων, που το καθένα εκτίθεται σε κίνδυνο για ένα χρόνο, δίνονται στον πιο κάτω πίνακα:

$x_i$	0	1	2	3	Σύνολο
$f_i$	3695	1120	160	25	5000

Όπου  $x_i$  είναι ο αριθμός των αιτημάτων και  $f_i$  είναι ο αριθμός των συμβολαίων.

## Ερωτήματα

1. Υπολογίστε το μέσο αριθμό απαιτήσεων.
2. Υποθέστε κατανομή Poisson. Συγκρίνετε τις προβλέψεις της θεωρητικής κατανομής με τον αριθμό των συμβολαίων που έχει παρατηρηθεί.

## Απάντηση

1. Ο μέσος αριθμός των απαιτήσεων υπολογίζεται από:

$$\lambda = \frac{\sum_i f_i x_i}{\sum_i f_i} = \frac{1515}{5000} = 0.303$$

2. Πρώτα υπολογίζουμε  $P(X = x)$  για  $x = 0, 1, 2, 3$ .

$$\bullet P(X = 0) = \frac{e^{-\lambda} \lambda^0}{0!} = e^{-\lambda} = 0.7386.$$

$$\bullet P(X = 1) = \frac{e^{-\lambda} \lambda^1}{1!} = P(X = 0) \times \frac{\lambda}{1} = 0.2238.$$

$$\bullet P(X = 2) = \frac{e^{-\lambda} \lambda^2}{2!} = P(X = 1) \times \frac{\lambda^2}{2} = 0.0339.$$

$$\bullet P(X = 3) = P(X = 2) \times \frac{\lambda^3}{3} = 0.0034.$$

Ο εκτιμώμενος αριθμός συμβολαίων για  $x$  απαιτήσεις δίνεται από  $\hat{f}_i = P(X = x_i) \sum_i f_i$ .

Έτσι,

$$\hat{f}_1 = P(X = x_i) \sum_i f_i = 0.7386 \times 5000 = 3693.$$

Αρ.	Πιθ. απαιτ. ανά συμβ.	Αρ. απαιτ. Αναμενόμενη	Πραγματική
$x_i$	$p_i$	$\hat{f}_i$	$f_i$
0	0.7386	3693.0	3695
1	0.2238	1119.0	1120
2	0.0339	169.5	160
3	0.0034	17.0	25
Σύνολο	0.9997	4998.5	5000

Συγκρίνοντας τη θεωρητική με την εμπειρική τιμή παρατηρούμε ότι μια κατανομή Poisson με παράμετρο  $\lambda = 0.3$  περιγράφει αυτή τη συγκεκριμένη τυχαία μεταβλητή πολύ καλά.

### Παρατηρήσεις

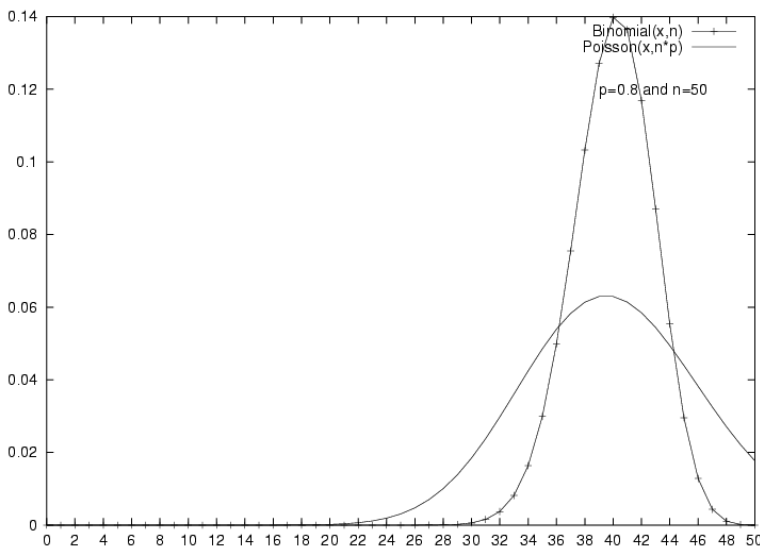
- $\sum_i p_i = 0.9997 \approx 1.$
- $E(x) = \sum_i x_i p_i = 0.3018 \equiv \lambda.$
- $E(x^2) = \sum_i x_i^2 p_i = 0.39.$
- $\text{Var}(x) = E(x^2) - (E(x))^2 = 0.299 \approx \lambda.$

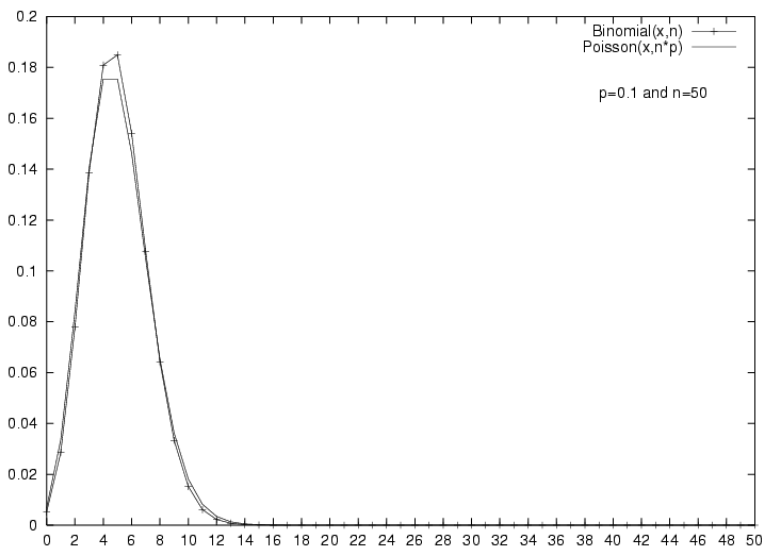
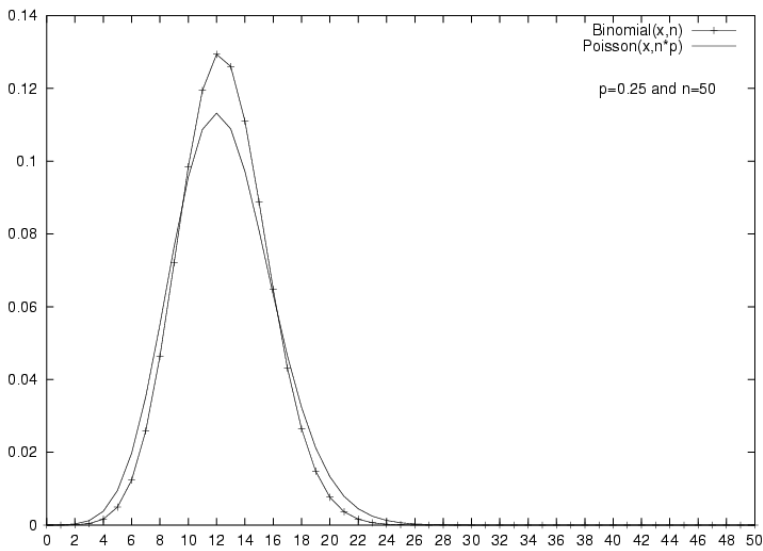
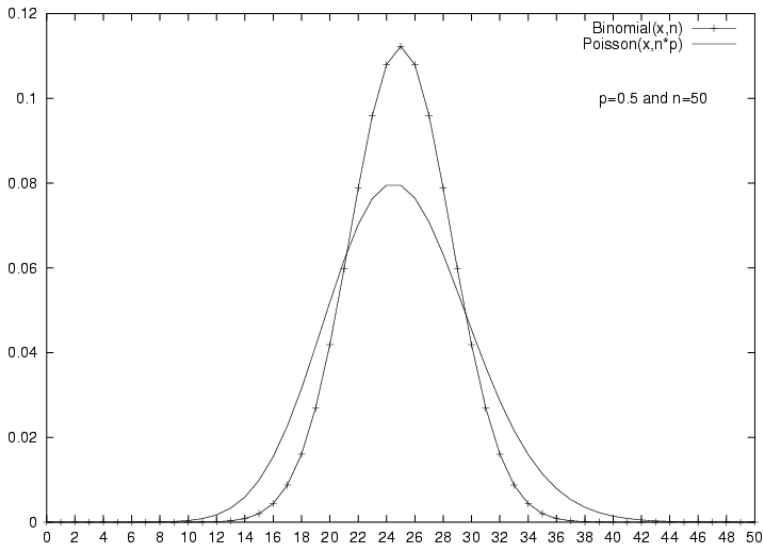
## Προσέγγιση Διωνυμ. κατανομ. από **Poisson**

Μια Διωνυμική κατανομή με παραμέτρους  $n$  και  $p$  προσεγγίζεται από την Poisson με παραμέτρους  $\lambda = np$  αν  $n$  είναι πολύ μεγάλο και  $p$  είναι μικρο. Έτσι,  $n \rightarrow \infty$  και  $p \rightarrow 0$ .

Σημειώστε ότι αν  $p \rightarrow 0$ , τότε  $q = 1 - p \approx 1$ . Έτσι, η διασπορά της διωνυμικής κατανομής δίνεται από:

$$\mathbf{Var}(x) = np(1 - p) \approx np1 = np = \mathbf{E}(x).$$





## Παράδειγμα

Θεωρείστε ότι κάποιος έχει ασφαλίσει το αυτοκίνητό του έναντι κλοπής. Η πιθανότητα κλοπής σε ένα 24-ωρο, οδηγεί σε ασφαλιστική απαίτηση 0.005. Οι πιθανότητες μιας απαίτησης σε μια μέρα χωρίς κλοπή είναι ανεξάρτητες. Επιπλέον θεωρούμε ότι δεν είναι πιθανό να υπάρξουν περισσότερες από μια κλοπές (που να οδηγούν σε απαίτηση) την ίδια μέρα.

Υπολογίστε την πιθανότητα ο ασφαλιζόμενος να προβεί τουλάχιστον σε 3 απαιτήσεις σε ένα χρόνο.

Η πιθανότητα κλοπής είναι  $p = 0.005$ . Η πιθανότητα το αυτοκίνητο να μην κλαπεί (δηλ. καμιά κλοπή) είναι  $q = 1 - p = 0.995$ . Ο συνολικός αριθμός δοκιμών σε ένα χρόνο είναι ο αριθμός των ημερών  $n = 365$ .

Η πιθανότητα να έχουμε τουλάχιστον 3 αιτήσεις δίνεται από:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - \left( P(X = 0) + P(X = 1) + P(X = 2) \right). \end{aligned}$$

Χρησιμοποιώντας τη Διωνυμική κατανομή και τον αναδρομικό τύπο:

$$P(X = 0) = \binom{365}{0} p^0 q^{365} = 0.1605,$$

$$P(X = 1) = 0.2944 \quad \text{και}$$

$$P(X = 2) = 0.2692.$$

$$\text{Άρα,} \quad P(X \geq 3) = 0.276.$$

Χρησιμοποιώντας την προσέγγιση στην κατανομή Poisson. Θεωρούμε  $\lambda = np = 365 \times 0.005 = 1.825$ . Σημειώστε ότι  $npq = 1.815$ .

$$P(X = 0) = \frac{e^{-\lambda} \lambda^0}{0!} = 0.1612,$$

$$P(X = 1) = 0.2942 \quad \text{και}$$

$$P(X = 2) = 0.2685.$$

$$\text{Έτσι,} \quad P(X \geq 3) = 0.2761.$$

Η διαφορά από την εφαρμογή της κατανομής Poisson έχει σχετικό σφάλμα 0.04% το οποίο είναι αμελητέο.

## Κανονική Κατανομή

Η Διωνυμική κατανομή και η κατανομή Poisson αποτελούν σημαντικά παραδείγματα των διακριτών κατανομών. Η Κανονική κατανομή μπορεί να περιγραφεί ως η πιο σημαντική συνεχής κατανομή στη στατιστική.

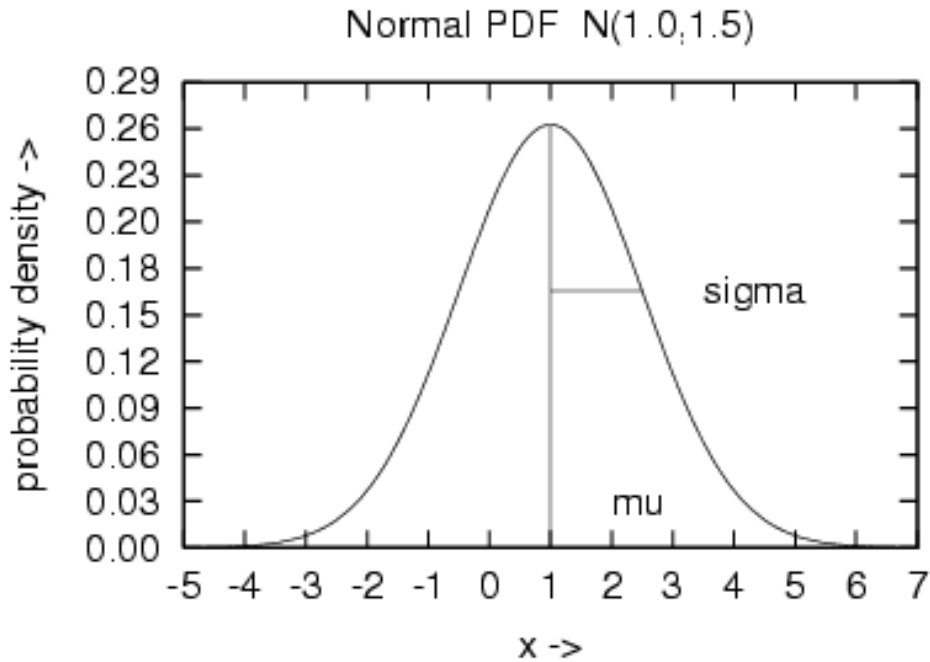
Η Κανονική κατανομή έχει δύο παραμέτρους: το μέσο  $\mu$  και την τυπική απόκλιση  $\sigma$ .

Σε συντομογραφία δίνεται από  $N(\mu, \sigma)$ . Η συνάρτηση πυκνότητας (pdf)  $N(\mu, \sigma)$  έχει σχήμα κωδωνοειδές (bell shaped) και είναι συμμετρική γύρω από το μέσο.

Ο τύπος είναι:

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

όπου  $E(x) = \mu$  και  $Var(x) = \sigma^2$ .



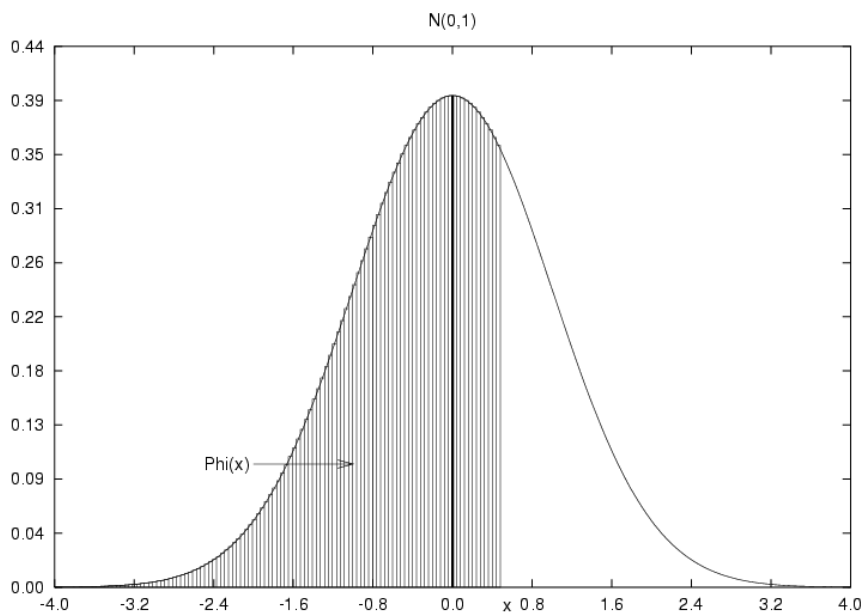
Μία κανονική κατανομή που έχει  $\mu = 0$  και  $\sigma^2 = 1$ , δηλαδή  $N(0, 1)$ , ονομάζεται τυπική κανονική κατανομή. Η τυχαία μεταβλητή που σχετίζεται με αυτήν την κατανομή συμβολίζεται συνήθως με  $Z$ . Δηλαδή,  $Z \sim N(0, 1)$ . Η pdf της  $Z$  δίνεται από:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{for } -\infty < x < \infty.$$

Η συνάρτηση κατανομής μιας τυπικής κανονικής κατανομής  $Z$  συμβολίζεται με:

$$\begin{aligned}\Phi(x) &= P(Z < x) = \int_{-\infty}^x f(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx.\end{aligned}$$

Δεν υπάρχει κλειστός τύπος για το πιο πάνω ολοκλήρωμα. Έτσι, μπορεί να υπολογιστεί μόνο με προσεγγιστικές μεθόδους. Για αυτό το λόγο, περιοχές (areas) της κανονικής κατανομής, δηλαδή εμβαδά όπως στο διάγραμμα πιο κάτω, παρουσιάζονται σε πίνακες.



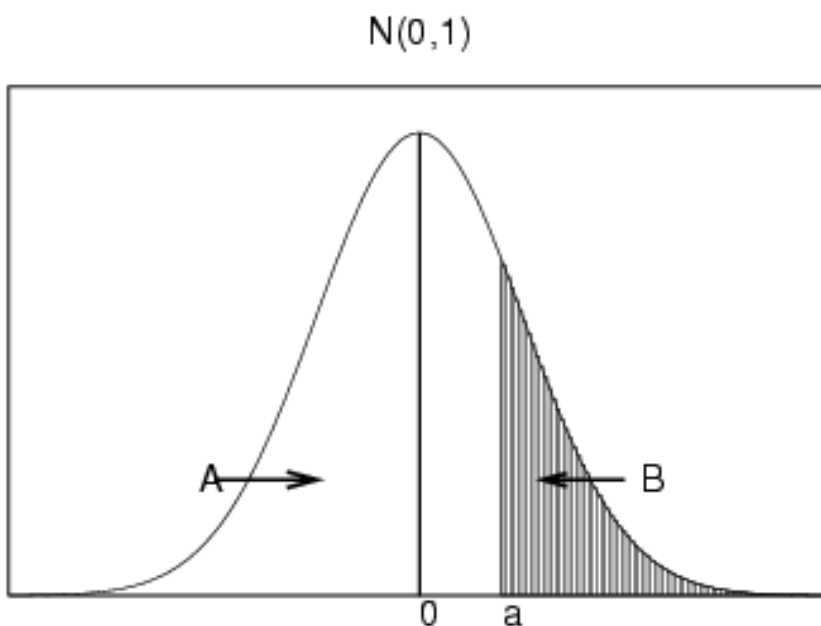
## Παράδειγμα

$$P(Z < 0.1) = \Phi(0.1) = 0.5398.$$

Οι πίνακες Κανονικής κατανομής δίνουν τιμές μόνο για  $\Phi(x)$  όπου  $x \geq 0$ . Οι πιθανότητες  $P(Z < -0.1)$  και  $P(Z \geq 0.3)$  πρέπει να μετασχηματιστούν σε πιθανότητες της μορφής  $P(Z < x)$ , όπου  $x \geq 0$ .

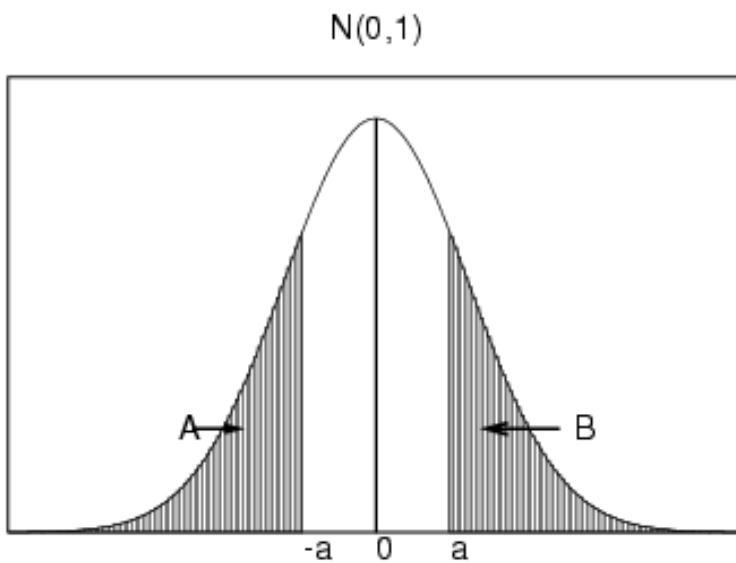
Θεωρούμε  $Z \sim N(0, 1)$  και  $\Phi(x) = P(Z < x)$ . Αν  $a \geq 0$ , τότε

1.  $P(Z > a) = 1 - P(Z < a) = 1 - \Phi(a)$ .



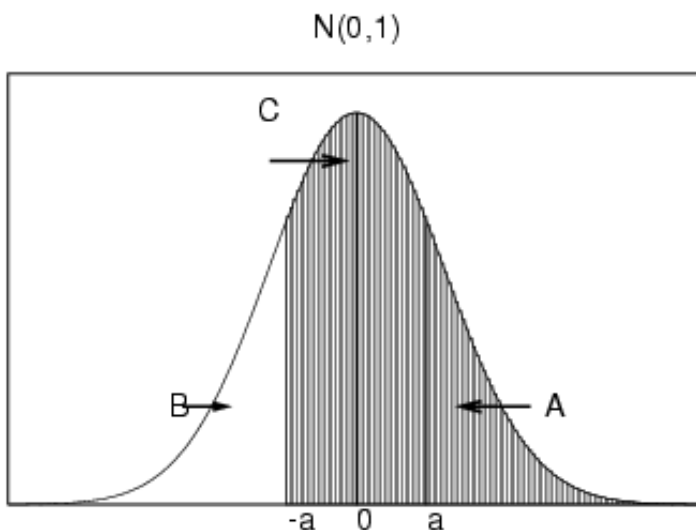
$$P(Z > a) = \mathbf{Area(B)} = \mathbf{Total Area - Area(A)} = 1 - \Phi(a).$$

$$2. P(Z < -a) = \Phi(-a) = 1 - \Phi(a).$$



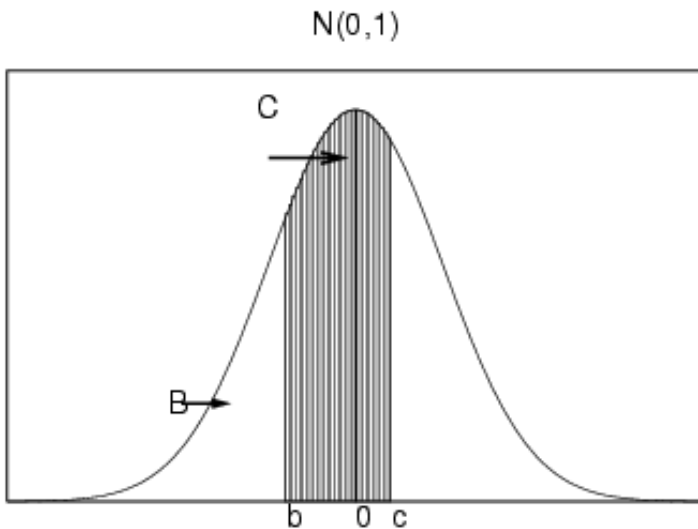
$$\Phi(-a) = \mathbf{Area(A)} = \mathbf{Area(B)} = P(Z > a) = 1 - \Phi(a).$$

$$3. P(Z > -a) = \Phi(a).$$



$$\begin{aligned} P(Z > -a) &= \mathbf{Area(C)} + \mathbf{Area(A)} \\ &= \mathbf{Area(C)} + \mathbf{Area(B)} = P(Z < a) = \Phi(a). \end{aligned}$$

4. Αν  $b$  και  $c$  είναι οποιοδήποτε θετικοί ή αρνητικοί αριθμοί έτσι ώστε  $b \leq c$ , τότε  
 $P(b < Z < c) = \Phi(c) - \Phi(b)$ .



$$\begin{aligned} P(b < Z < c) &= \mathbf{Area(C)} = \mathbf{Area(B+C)} - \mathbf{Area(B)} \\ &= P(Z < c) - P(Z < b) = \Phi(c) - \Phi(b). \end{aligned}$$

Παράδειγμα 1

Η τυχαία μεταβλητή  $X \sim N(\mu, \sigma^2)$  συμβολίζει τον αριθμό των απαιτήσεων ανά χρονιά, όπου  $\mu = 100$  και  $\sigma = 4$ . Βρέστε τις πιθανότητες ο αριθμός των απαιτήσεων να είναι:

1. Μικρότερος από 90.
2. Μεγαλύτερος από 108.
3. Μεταξύ 96 και 104 (συμπεριλαμβανομένου).

Λύση

1.

$$P(X < 90) = P\left(\frac{X - 100}{4} < \frac{90 - 100}{4}\right)$$

$$P(Z < -2.5) \quad \text{αφού} \quad (X - \mu)/\sigma \sim N(0, 1)$$

$$1 - \Phi(2.5) = 0.0062$$

2.

$$P(X > 108) = P\left(\frac{X - 100}{4} > \frac{108 - 100}{4}\right)$$

$$P(Z > 2.0) = 1 - \Phi(2.0)$$

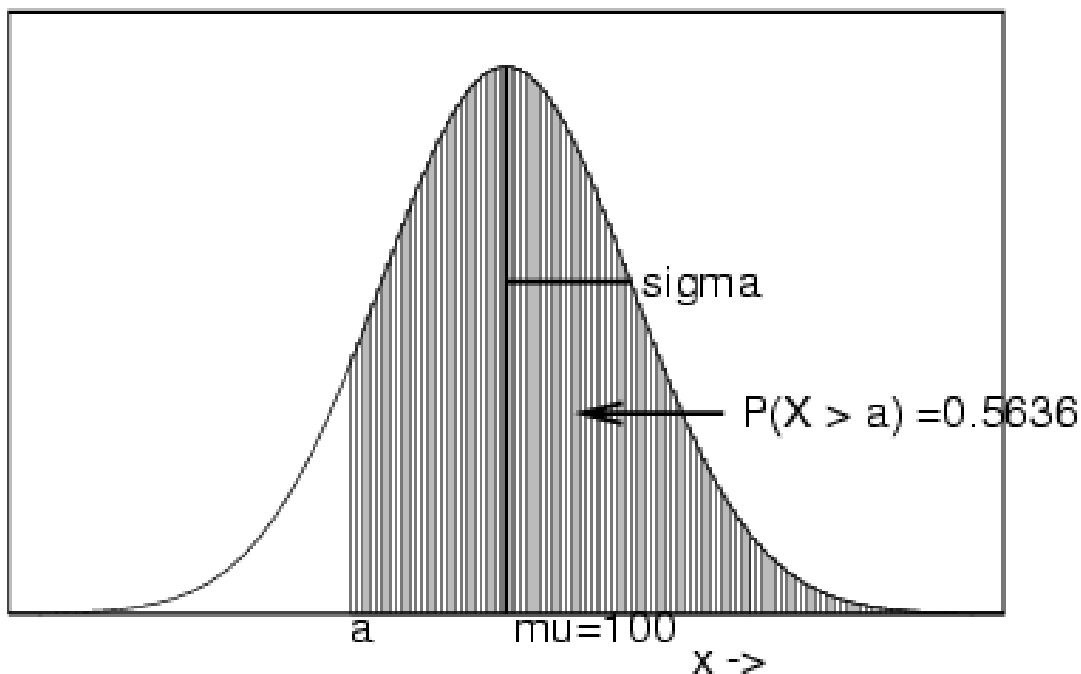
$$1 - 0.9772 = 0.0228.$$

3.

$$\begin{aligned}
 P(95 < X < 105) &= P\left(\frac{95 - 100}{4} < \frac{X - 100}{4} < \frac{105 - 100}{4}\right) \\
 &= P(-1.25 < Z < 1.25) \\
 &= \Phi(1.25) - \Phi(-1.25) \\
 &= 2 \times \Phi(1.25) - 1 = 0.7888.
 \end{aligned}$$

Παράδειγμα 2

Η πιθανότητα να έχουμε περισσότερες από  $a$  απαιτήσεις είναι 0.5636. Ποια είναι η τιμή  $a$ , όταν  $X \sim N(100, 16)$ ?

Normal PDF  $N(100, 16)$ 

Δεδομένου ότι η πιθανότητα που δίνεται είναι

μεγαλύτερη από 0.5 τότε το  $a$  πρέπει να είναι μικρότερο από το μέσο  $\mu = 100$ .

Η πιθανότητα,  $P(X > a) = 0.5636$  και άρα,

$$P\left(\frac{X - 100}{4} > \frac{a - 100}{4}\right) = 0.5636$$

$$\text{ή } P\left(Z > \frac{a - 100}{4}\right) = 0.5636.$$

Αφού το  $a$  είναι μικρότερο από το μέσο, τότε  $(a - 100)/4 < 0$ . Σαν αποτέλεσμα,

$$\begin{aligned} P\left(Z > \frac{a - 100}{4}\right) &= P\left(Z > -\left(\frac{100 - a}{4}\right)\right) \\ &= \Phi\left(\frac{100 - a}{4}\right). \end{aligned}$$

Συμπερασματικά,

$$\Phi\left(\frac{100 - a}{4}\right) = 0.5636$$

Από τους πίνακες συμπεραίνουμε ότι

$$\frac{100 - a}{4} = 0.16 \quad \text{και άρα,} \quad a = 99.36.$$

### Παράδειγμα 3

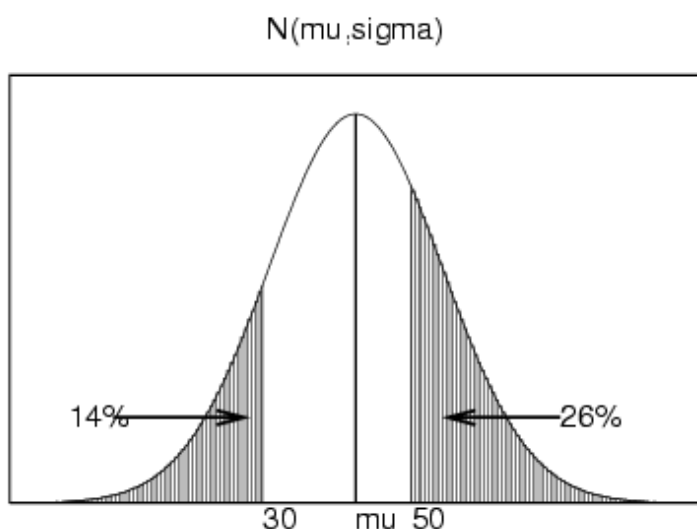
Ένας ασφαλιστής έχει μεγάλο αριθμό ασφαλιστήριων συμβολαίων. Αν γνωρίζουμε ότι τα συμβόλαια ακολουθούν κανονική κατανομή και

1. 14% των συμβολαίων ανέρχονται σε λιγότερο από 30 απαιτήσεις·
2. 26% ανέρχονται σε περισσότερο από 50 απαιτήσεις.

Βρείτε το μέσο και τη διακύμανση των απαιτήσεων.

### Απάντηση

Θεωρούμε  $X$  τις απαιτήσεις, δηλαδή.  $X \sim N(\mu, \sigma^2)$ . Οι πληροφορίες που δίνονται από το (1) και (2) μπορούν να γραφούν ως εξής  $P(X < 30) = 0.14$  και  $P(X > 50) = 0.26$ . Γραφικά αυτό μπορεί να αναπαρασταθεί ως εξής:



Η  $P(X < 30) = 0.14$  μπορεί να γραφτεί ως

$$P\left(\frac{X - \mu}{\sigma} < \frac{30 - \mu}{\sigma}\right) = 0.14$$

$$\text{ή } P\left(Z < -\left(\frac{\mu - 30}{\sigma}\right)\right) = 0.14$$

$$\text{ή } \Phi\left(-\left(\frac{\mu - 30}{\sigma}\right)\right) = 0.14$$

$$\text{ή } 1 - \Phi\left(\frac{\mu - 30}{\sigma}\right) = 0.14 \quad \text{ή } \Phi\left(\frac{\mu - 30}{\sigma}\right) = 0.86.$$

Από τους πίνακες συνεπάγεται ότι  $\Phi(1.08) = 0.8599$ .

Έτσι,

$$\frac{\mu - 30}{\sigma} = 1.08 \quad \text{ή} \quad \mu - 1.08 \times \sigma = 30. \quad (1)$$

Προφανώς, από την  $P(X > 50) = 0.26$  συνεπάγεται ότι

$$P\left(Z > \frac{50 - \mu}{\sigma}\right) = 0.26, \quad \text{όπου} \quad Z = \frac{X - \mu}{\sigma}.$$

Έτσι,

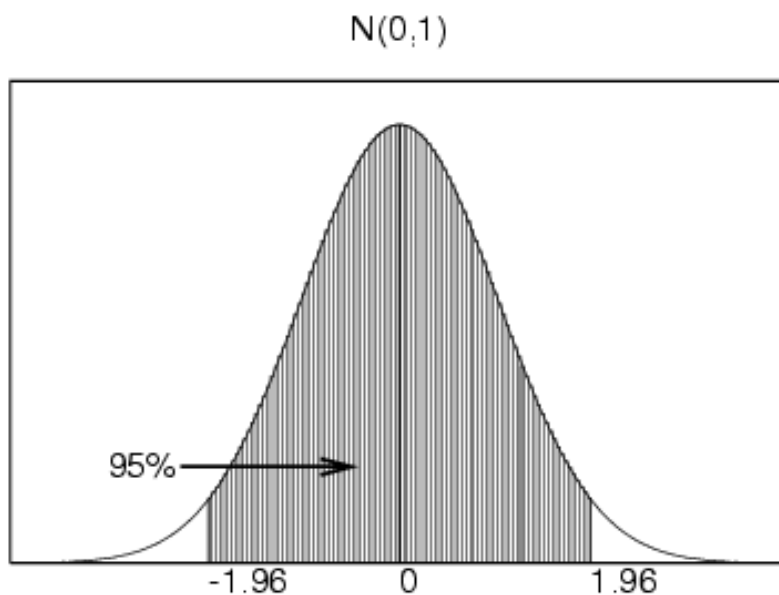
$$1 - \Phi\left(\frac{50 - \mu}{\sigma}\right) = 0.26 \quad \text{ή} \quad \Phi\left(\frac{50 - \mu}{\sigma}\right) = 0.74.$$

Από τους πίνακες συνεπάγεται ότι

$$\frac{50 - \mu}{\sigma} = 0.643 \quad \text{ή} \quad \mu + 0.643 \times \sigma = 50. \quad (2)$$

Από (1) και (2) συνεπάγεται ότι ο μέσος  $\mu = 42.54$  και η διακύμανση  $\sigma^2 = (11.61)^2 = 134.8$ .

Το κεντρικό 95% της τυπικής κανονικής κατανομής βρίσκεται μεταξύ των ορίων  $\pm 1.96$ . Εναλλακτικά, το κεντρικό 95% οποιασδήποτε κανονικής κατανομής βρίσκεται εντός 1.96 τυπικές αποκλίσεις από το μέσο.



Το κεντρικό 99% (99.8%) μιας τυπικής κανονικής κατανομής εμπίπτει μεταξύ των ορίων  $\pm 2.58$  ( $\pm 3.09$ ).

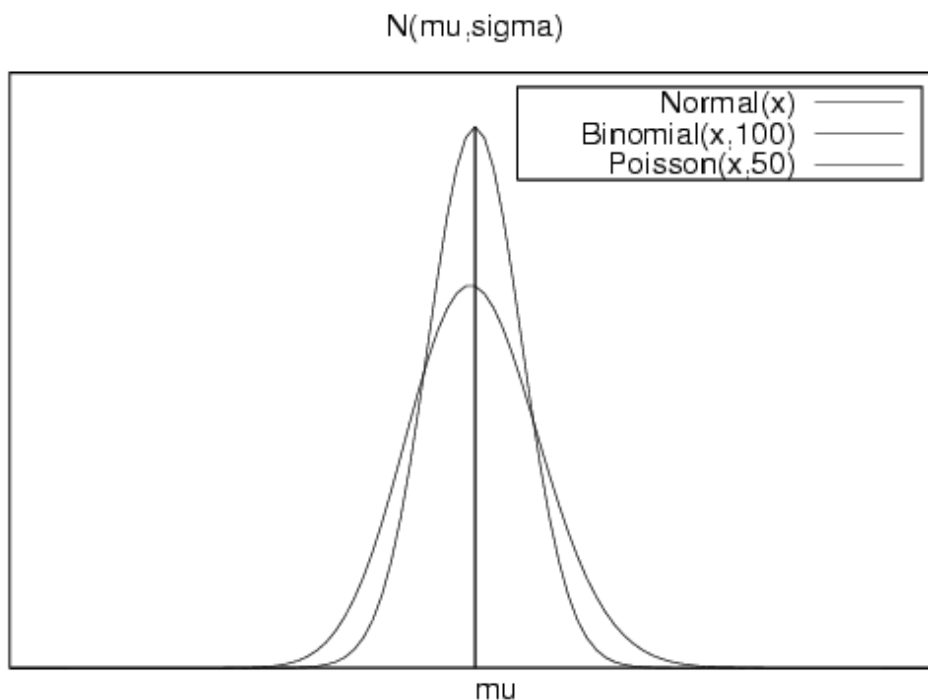
**Προσέγγιση Κανον. Καταν. από Διων. και Poisson**

- Αν  $X$  κατανέμεται διωνυμικά από τις παραμέτρους  $n$  και  $p$ , δηλαδή  $X \sim Bin(n, p)$ , τότε για μεγάλο αριθμό  $n$  και όχι πολύ μικρό (ή πολύ μεγάλο)  $p$  μπορούμε να θεωρήσουμε  $X \sim N(np, np(1 - p))$ .

Ο αριθμός των δοκιμών  $n$  πρέπει να είναι, γενικά,  $n > 50$  και  $p \approx 0.5$ .

- Αν  $X$  ακολουθεί την κατανομή Poisson με παράμετρο  $\mu$ , δηλαδή  $X \sim Po(\mu)$ , τότε για μεγάλο  $\mu$ , μπορούμε να θεωρήσουμε  $X \sim N(\mu, \mu)$ .

Γενικά η παράμετρος  $\mu > 20$  δίνει καλές προσεγγίσεις.



### Κεντρικό Οριακό Θεώρημα

Αν  $X_1, X_2, \dots, X_n$  είναι ένα τυχαίο δείγμα μεγέθους  $n$  από οποιαδήποτε κατανομή με μέσο  $\mu$  και διασπορά  $\sigma^2$ , τότε ο πληθυσμιακός μέσος  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  προσεγγίζεται από την κανονική κατανομή με μέσο  $\mu$  και διακύμανση  $\sigma^2/n$ . Δηλαδή, κατά προσέγγιση

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

Η προσέγγιση είναι καλύτερη για μεγάλα  $n$ .

#### Παράδειγμα

Θεωρούμε  $X_1, X_2, \dots, X_n$  ένα τυχαίο δείγμα απαιτήσεων από Poisson κατανομή με παράμετρο  $\mu = 3$ , όπου  $n = 20$ . Χρησιμοποιώντας το κεντρικό οριακό θεώρημα βρέστε την προσέγγιση του πληθυσμιακού μέσου έτσι ώστε να είναι μεγαλύτερος από 4.

#### Λύση

Αν  $X$  ακολουθεί Poisson με παράμετρο 3, τότε  $E(X) = \mu = 3$  και  $\text{Var}(X) = \sigma^2 = \mu = 3$ . Από το κεντρικό οριακό θεώρημα συνεπάγεται ότι  $\bar{X} \sim N(\mu, \mu/n) = N(3, 0.15)$  κατά προσέγγιση.

Έτσι,

$$\begin{aligned} P(\bar{X} > 4) &= P\left(\frac{\bar{X} - 3}{\sqrt{0.15}} > \frac{4 - 3}{\sqrt{0.15}}\right) \\ &= P(Z > 2.58) \quad (\text{όπου } Z = (\bar{X} - 3)/\sqrt{0.15}) \\ &= 1 - \Phi(2.58) = 0.005. \end{aligned}$$

### Παράδειγμα

Το σύνολο των ασφαλισμένων πελατών έχει μέσο όρο απαιτήσεων (σε Ευρό) ύψους  $\mu = 69$  και τυπική απόκλιση  $\sigma = 3.22$ . Εάν επιλεγεί ένα τυχαίο δείγμα  $n = 10$  ασφαλιζομένων από το σύνολο, τότε ποια είναι η πιθανότητα ο μέσος  $\bar{X}$  να απέχει € 2 από τον δειγματικό μέσο  $\mu$ ;

### Λύση

Θέλουμε να βρούμε ότι

$$\begin{aligned} P(|\bar{X} - \mu| < 2) &= P(-2 < \bar{X} - \mu < 2) \\ &= P(67 < \bar{X} < 71) \\ &= 1 - P(\bar{X} < 67) - P(\bar{X} > 71). \end{aligned}$$

Βάση του κεντρικού οριακού θεωρήματος,

$$\bar{X} \sim N(\mu, \sigma^2/n) = N(69, 1.04).$$

Θεωρούμε  $Z = (\bar{X} - 69)/\sqrt{1.02}$ , έτσι ώστε μετά την τυποποίηση (standardization) η  $P(|\bar{X} - \mu| < 2)$  μπορεί να γραφτεί ως

$$\begin{aligned} P(|\bar{X} - \mu| < 2) &= 1 - P(Z < -1.96) - P(Z > 1.96) \\ &= 1 - 0.025 - 0.025 \\ &= 0.95. \end{aligned}$$

Επομένως, υπάρχει 95% πιθανότητα ο δειγματικός μέσος (sample mean) να απέχει € 2 από τον πληθυσμιακό.

### Παράδειγμα

Υποθέστε ότι οι βαθμολογίες ενός μεγάλου συνόλου φοιτητών σε ένα μάθημα ποσοτικών μεθόδων είναι κανονικά κατανεμημένες γύρω από μέσο 72 με τυπική απόκλιση 9.

1. Βρέστε την πιθανότητα ένας μαθητής που επιλέγηκε τυχαία από το δείγμα να έχει βαθμολογία μεγαλύτερη του 80.

2. Βρέστε την πιθανότητα ένα τυχαίο δείγμα 10 φοιτητών να έχει μέση βαθμολογία μεγαλύτερη του 80. Πώς αλλάζει η πιθανότητα όταν ο πληθυσμός ΔΕΝ είναι κανονικά κατανεμημένος;

### Λύση

1. Θεωρούμε ότι  $X$  υποδηλώνει τις βαθμολογίες των φοιτητών, όπου  $\mu = 72$  και  $\sigma^2 = 81$ . Έτσι,  $X \sim N(72, 81)$ .

Θέλουμε να βρούμε την πιθανότητα  $P(X > 80)$ .

Θεωρούμε  $Z = (X - 72)/9$ . Σημειώστε ότι  $(80 - 72)/9 = 0.89$ . Έτσι, μετά την τυποποίηση  $P(X > 80) = P(Z > 0.89) = 0.187$ .

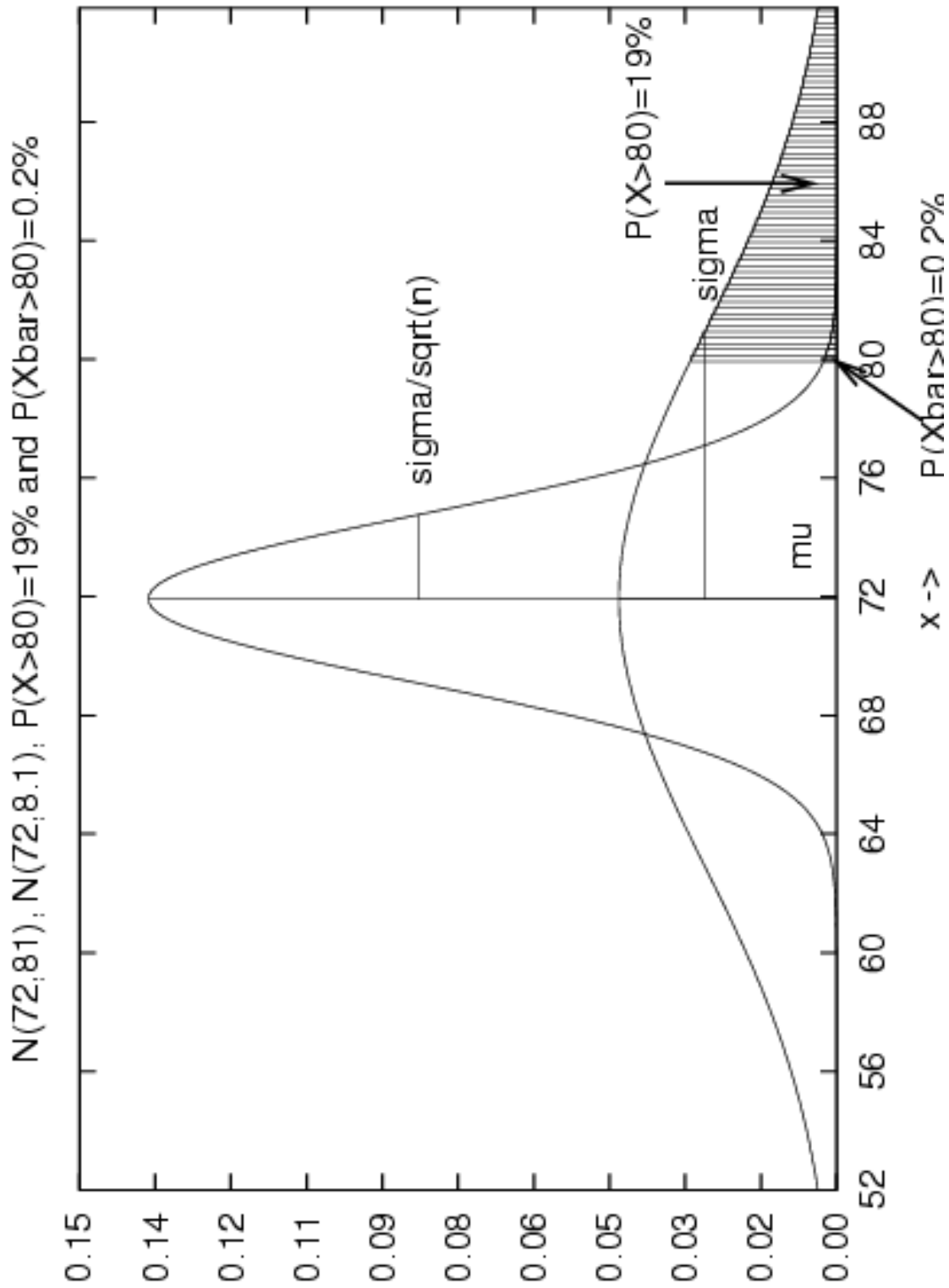
2. Από το κεντρικό οριακό θεώρημα ισχύει ότι  $\bar{X} \sim N(\mu, \sigma^2/n) = N(72, 8.1)$ , όπου  $\bar{X}$  συμβολίζει τη μέση βαθμολογία ενός τυχαίου δείγματος 10 φοιτητών.

Αναζητούμε  $P(\bar{X} > 80)$ .

Θεωρούμε,  $Z = (\bar{X} - 72)/2.85$  και σημειώνουμε ότι  $(80 - 72)/2.85 = 2.81$ . Έτσι μετά από τυποποίηση:

$$P(\bar{X} > 80) = P(Z > 2.81) = 0.002.$$

Το ΚΟΘ (CLT) υπονοεί ότι για ένα επαρκές μεγάλο δείγμα  $n$ , ισχυει ότι  $\bar{X} \sim N(\mu, \sigma^2/n)$  (κατα προσέγγιση) όποια και αν είναι η αρχική κατανομή του πληθυσμού. Έτσι,  $P(\bar{X} > 80)$  είναι κατα προσέγγιση 0.2%.



## Εκτίμηση διαστήματος

Η εκτίμηση διαστημάτων έχει στοχο τη δημιουργία τιμών ορίων μέσα στα οποία αναμένεται να βρίσκονται, με μια δεδομένη πιθανότητα, οι τιμές των πληθυσμιακών παραμέτρων με βάση τα αποτελεσμάτα από ένα τυχαίο δείγμα.

Ένα  $B\%$  διάστημα εμπιστοσύνης (*confidence interval*) (C.I.) μιας άγνωστης παραμέτρου  $\theta$  είναι ένα διάστημα που δημιουργείται βάση των αποτελεσμάτων ενός τυχαίου δείγματος έτσι ώστε η πιθανότητα το  $\theta$  να βρίσκεται σε αυτό το διάστημα να είναι  $B/100$ .

Το πιο συνηθισμένο διάστημα είναι το 95% C.I. Αν  $(a, b)$  αποτελεί 95% C.I. για μια παράμετρο  $\theta$ , τότε η πιθανότητα δίνεται από :  $P(a \leq \theta \leq b) = 0.95$ .

Η κατασκευή των διαστημάτων με τα οποία θα ασχοληθούμε, θα βασίζεται στις τιμές (που έχουμε πάρει από ένα δείγμα) αμερόληπτων εκτιμητών για τις παραμέτρους που ενδιαφερόμαστε.

Ο μέσος (γνωστή διασπορά)

Αν  $\bar{X}$  είναι ο μέσος ενός τυχαίου δείγματος μεγέθους  $n$  από μια κανονική κατανομή με γνωστή διασπορά  $\sigma^2$ , το 95% C.I. για το  $\mu$ , που είναι ο πληθυσμιακός μέσος, δίνεται από  $\bar{X} \pm 1.96\sigma/\sqrt{n}$ . Έτσι,

$$P\left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95. \quad (3)$$

Άσκηση

Τα αποτελέσματα ενός πειράματος κατέγραψαν το ύψος ενός plan σε cm

{12.3, 11.8, 11.6, 12.6, 13.4, 12.8, 11.1, 12.2, 14.8, 13.1}

Δίνεται ότι το ύψος των plans είναι κατά προσέγγιση κανονικά κατανεμημένο με διακύμανση  $1.44\text{cm}$ .

Υπολογίστε το 95% C.I. για το μέσο  $\mu$  του ύψους του πληθυσμού.

Λύση

$$P(11.83 \leq \mu \leq 13.31) = 0.95.$$

### Ο μέσος με άγνωστη διακύμανση (μικρά δείγματα)

Θεωρούμε  $\bar{X}$  και  $S^2$  ότι υποδηλώνουν το μέσο και τη διακύμανση ενός τυχαίου δείγματος μεγέθους  $n$  από κανονικό πληθυσμό με άγνωστο μέσο  $\mu$  και άγνωστη διακύμανση  $\sigma^2$ . Το  $B\%$  C.I. το για  $\mu$  δίνεται από  $\bar{X} \pm t S/\sqrt{n-1}$ , όπου  $t$  είναι τέτοιο που το διάστημα  $(-t, t)$  εμπερικλείει το  $B\%$  μιας κατανομής  $T(n-1)$ . Έτσι,

$$P\left(\bar{X} - t \frac{S}{\sqrt{n-1}} \leq \mu \leq \bar{X} + t \frac{S}{\sqrt{n-1}}\right). \quad (4)$$

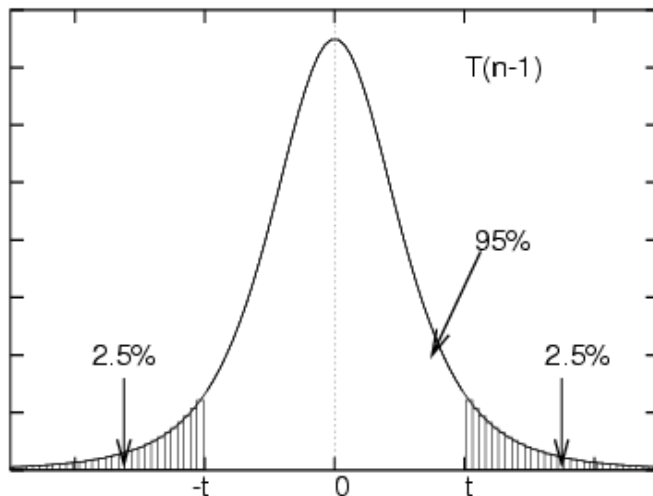
Σημειώστε ότι η πιο πάνω μορφή χρησιμοποιείται για δείγματα που είναι κατα προσέγγιση κανονικά κατανεμημένα.

### Παράδειγμα

Οκτώ χρεόγραφα μιας εταιρείας περιορισμένης ευθύνης δίνονται πιο κάτω (σε €)

{10.6, 11.2, 10.4, 12.2, 11.3, 10.2, 10.3, 12.5}.

Βρέστε το 95% διάστημα εμπιστοσύνης για τη μέση τιμή ενός δείγματος των εταιρικών χρεογράφων εάν υποθέσουμε ότι οι τιμές προέρχονται από κανονική κατανομή.

Λύση

Τα όρια εμπιστοσύνης έχουν τη μορφή  $\bar{X} \pm t S / \sqrt{n - 1}$ , όπου  $(-t, t)$  είναι το διάστημα μιας  $T(n - 1)$  κατανομής που εμπερικλείει το κεντρικό 95% της κατανομής. Από τους πίνακες της  $T$  για  $n = 7$  παίρνουμε  $t = 2.365$ . Έτσι,  $T_{2.5\%}(7) = 2.365$ .

Από τις τιμές των χρεογράφων ισχύει

$$\bar{X} = (\sum X_i) / n = 88.7 / 8 = 11.09 \text{ και } \sum X_i^2 = 988.7.$$

Έτσι,  $S^2 = \frac{1}{n} \sum X_i^2 - \bar{X}^2 = 0.68$  και  $S = 0.82$ .

Από τα πιο πάνω ισχύει ότι 95% C.I. είναι

$$11.09 \pm \frac{2.365 \times 0.82}{\sqrt{7}} = 11.09 \pm 0.73 = (10.36, 11.82).$$

Συμπερασματικά, με πιθανότητα 95% η μέση τιμή των μετοχών της εταιρείας βρίσκεται στο διάστημα 10.36 και 11.82.

## Ο μέσος με άγνωστη διασπορά (Προσέγ. μεγάλων δειγμάτων)

Για μεγάλα δείγματα  $n$ , το  $T(n)$  προσεγγίζει την τυπική Κανονική κατανομή, δηλαδή  $N(0, 1)$ . Για παράδειγμα  $Z_{2.5\%} = 1.96$  που συνεπάγεται ότι  $(-1.96, 1.96)$  εμπεριέχει το κεντρικό 95% μιας  $N(0, 1)$  κατανομής. Κατά παρόμοιο τρόπο, για  $P = 2.5\%$  και  $n = 120$  από τους T-πίνακες παίρνουμε  $t = 1.98$ . Έτσι, το κεντρικό 95% μιας  $T(120)$  κατανομής εμπίπτει στο διάστημα  $(-1.98, 1.98)$ .

Θεωρούμε  $\bar{X}$  και  $S^2$  ότι υποδηλώνουν το μέσο και τη διασπορά ενός τυχαίου δείγματος μεγέθους  $n$  (μεγάλο) από κανονικό πληθυσμό με άγνωστο μέσο  $\mu$  και άγνωστη διασπορά  $\sigma^2$ . Το κεντρικό  $B\%$  διάστημα εμπιστοσύνης για το  $\mu$  δίνεται (προσεγγιστικά) από  $\bar{X} \pm z S/\sqrt{n}$ , όπου  $z$  είναι το  $0.5(100 - B)\%$  σημείο μιας  $N(0, 1)$  κατανομής. Έτσι,

$$P\left(\bar{X} - z \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z \frac{S}{\sqrt{n}}\right) = \frac{B}{100}.$$

Σημειώστε ότι για  $n \geq 30$  θα θεωρείται ένα επαρκώς μεγάλο δείγμα.

### Παράδειγμα

Εκατό μετοχές από εταιρείες που εκτελούν παρόμοιες λειτουργίες, επιλέγονται τυχαία. Οι τιμές τους έχουν μέσο 69 και διακύμανση 7. Βρέστε το 95% και το 98% διάστημα εμπιστοσύνης των μέσων τιμών των μετοχών.

### Λύση

Το μεγάλο δείγμα  $n = 100$  έχει μέσο  $\bar{X} = 69$  και διακύμανση  $S^2 = 7$ . Καθώς το  $n$  είναι μεγάλο υπολογίζουμε το  $\bar{X} \pm z S/\sqrt{n}$ . Για το 95% C.I., το  $z = 1.96$  με όρια:  $69 \pm 1.96(\sqrt{7/100}) = 69 \pm 0.52$ .

Έτσι, το 95% C.I. δίνεται από (68.48, 69.52).

Για το 98% C.I., το  $z = 2.33$ . Έτσι, .....