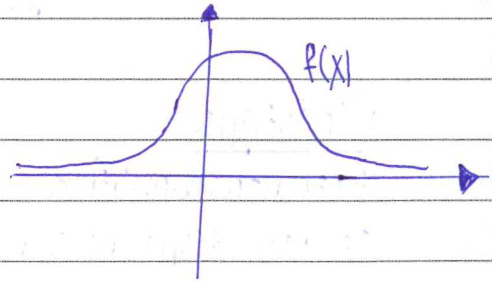
ΚΕΦΑΛΑΙΟ 5 - Κανονική κατανομή και Κεντρικό Οριακό Θεώρημα

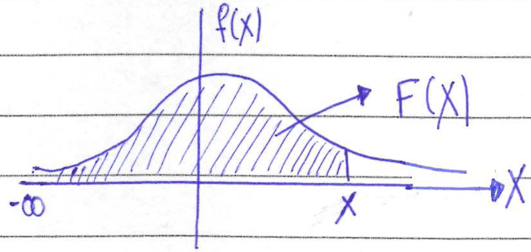
Υπενθύμιση

Τυχαίες Μεταβλητές → Διακριτές
Τυχαίες Μεταβλητές → Συνεχείς → τιμές σε διάστημα (συνήθως όλο το R)



- Έστω X συνεχής τυχαία μεταβλητή (τ.μ.)
- Η f(x) είναι συνεχής

1) $P(X=x)=0$
2) $f(x) \neq P(X=x)$
3) $F(x) = P(X \leq x) =$ Εμβαδόν κάτω από την γραφ. παράσταση της f μέχρι το x



4) Εμβαδόν κάτω από τη γραφ. παράσταση της f είναι ίσο με 1

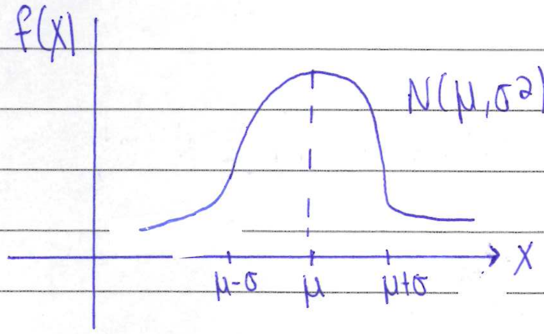
Κανονική Κατανομή

Έστω X συνεχής τ.μ.
Η X ακολουθεί την κανονική κατανομή με μέση τιμή = μ και διασπορά = σ^2
αν έχει συνάρτηση πυκνότητας πιθανότητας

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

Συμβολίζουμε με $X \sim N(\mu, \sigma^2)$

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



Θεώρημα

Αν $X \sim N(\mu, \sigma^2)$ τότε η $Z = \frac{X-\mu}{\sigma}$ ακολουθεί την τυποποιημένη κανονική κατανομή $N(0,1)$. Η συνάρτηση κατανομής της Z συμβολίζεται με Φ

$P(Z \leq z) = \Phi(z)$

Ευρεση πιθανοτήτων ($X \sim N(\mu,\sigma^2)$)

α) τυποποίηση

$$P(a<X\leq b) = P(X\leq b) - P(X\leq a)$$
$$= P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right)$$
$$= P(Z\leq b') - P(Z\leq a')$$

$$P(X\leq b) = P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P(Z\leq b')$$

$$P(X>a) = 1-P(X\leq a) = 1-P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = 1-P(Z\leq a')$$

β) Χρήση της Φ

$$P(a<X\leq b) = P(Z\leq b') - P(Z\leq a') = \Phi(b') - \Phi(a')$$

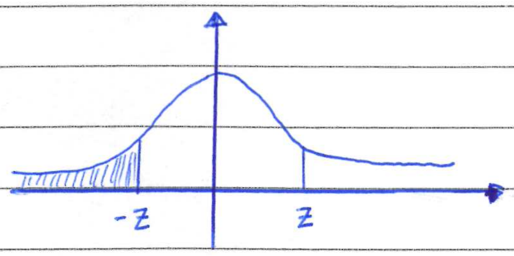
$$P(X\leq b) = 1-\Phi(b')$$
$$P(X>a) = 1-P(Z\leq a') = 1-\Phi(a')$$

Παρατηρήσεις

1) Σε συνεχή τ.μ X

- $P(X<x) = P(X\leq x)$
- $P(X>x) = P(X\geq x)$

2) $\Phi(-z) = 1-\Phi(z)$



Παράδειγμα

Ο χρόνος που χρειάζεται ένα ασθενοφόρο για να φτάσει από ένα κέντρο υγείας στο πλησιέστερο νοσοκομείο ακολουθεί κανονική κατανομή με μέση τιμή $\mu = 17$ λεπτά και τυπική απόκλιση $\sigma = 3$ λεπτά.

Να βρεθεί η πιθανότητα ο χρόνος που θα χρειαστεί να είναι:

α) Το πολύ 15 λεπτα

β) περισσότερο από 22 λεπτά

γ) Του[illegible]ιστον 13 και το πολύ 21 λεπτά

X = χρόνος σε λεπτά που χρειάζεται το ασθενοφόρο

$X \sim N(17, 9) \quad (\sigma^2 = 9)$

1) $P(X \leq 15) = P\left(\frac{X-17}{3} \leq \frac{15-17}{3}\right) = P(Z \leq -0{,}67) = \Phi(-0{,}67)$

$= 1 - \Phi(0{,}67) = 1 - 0{,}7486 = \boxed{0{,}2519}$

2) $P(X > 22) = 1 - P(X \leq 22) = 1 - P\left(\frac{X-17}{3} \leq \frac{22-17}{3}\right) = 1 - P(Z \leq 1{,}67)$ *

* $= 1 - \Phi(1{,}67) = 1 - 0{,}9525 = \boxed{0{,}0475}$

3) $P(13 \leq X \leq 21) = P(X \leq 21) - P(X < 13)$

$= P(X \leq 21) - P(X \leq 13) =$

$= P\left(\frac{X-17}{3} \leq \frac{21-17}{3}\right) - P\left(\frac{X-17}{3} \leq \frac{13-17}{3}\right) =$

$= \Phi(1{,}33) - \Phi(-1{,}33) = \Phi(1{,}33) - (1 - \Phi(1{,}33))$

$= 2\Phi(1{,}33) - 1$

$= 2 \cdot 0{,}9082 - 1 = \boxed{0{,}8164}$

4) Πόσο χρόνο θα χρειαστεί για να είναι 97,5% σίγουρο ότι θα φτάσει στο νοσοκομείο;

Εστω x αυτός ο χρόνος

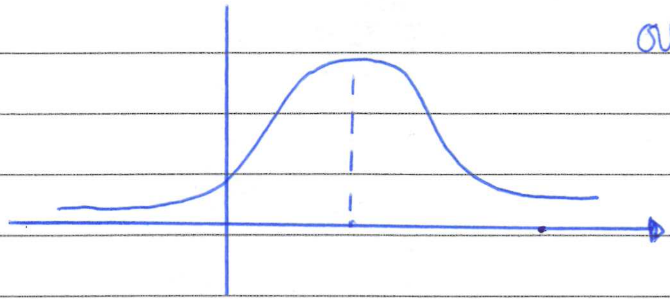
$P(X \leq x) = 0{,}975 \Rightarrow P\left(\frac{X-17}{3} \leq \frac{x-17}{3}\right) = 0{,}975$

$\Rightarrow \Phi\left(\frac{x-17}{3}\right) = 0{,}975 \Rightarrow \Phi\left(\frac{x-17}{3}\right) = \Phi(1{,}96)$

$\Rightarrow \frac{x-17}{3} = 1{,}96 \Rightarrow x-17 = 5{,}88$ $\boxed{x = 22{,}88 \text{ λεπτά}}$

Εύρεση τιμών με δεδομένη πιθανότητα

Θέτουμε την άγνωστη τιμή ίση με κάποιον άγνωστο x
- Εκφράζουμε και τυποποιούμε
- Χρησιμοποιούμε τη Φ και καταλήγουμε σε Φ(συνάρτηση του αγνώστου) = γνωστό
- Βρίσκουμε τη γνωστή τιμή από τον πίνακα της Φ
- Λύνουμε την εξίσωση



συμμετρ.

Άσκηση: Εστω $X \sim N(\mu, \sigma^2)$

Να βρεθεί η πιθανότητα η X να απέχει:

α) 1 τυπική απόκλιση από το μ
β) 2 τυπικές αποκλ. " " "
γ) 3 τυπικές αποκλ. " " "

69

α) $P(|X-\mu| \leq \sigma)$

$= P(-\sigma \leq X-\mu \leq \sigma)$

$= P\left(\frac{-\sigma}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{\sigma}{\sigma}\right) = P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z < -1)$ *

$* \; P(Z \leq 1) - P(Z \leq -1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1-\Phi(1))$

$= 2\Phi(1) - 1$

$= 2 \cdot 0{,}8413 - 1$

$= 0{,}68 = \boxed{68\%}$

β) $P(|X-\mu| \leq 2\sigma)$

$= P(-2\sigma \leq X-\mu \leq 2\sigma)$

$= P(-2 \leq Z \leq 2) \quad = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 2 \cdot 0{,}9772 - 1 \cong \boxed{0{,}954}$

γ) $P(|X-\mu| \leq 3\sigma) = P(-3\sigma \leq X-\mu \leq 3\sigma) = P(-3 \leq Z \leq 3)$

$= \Phi(3) - \Phi(-3) = 2\Phi(3) - 1 = 2 \cdot 0{,}9987 - 1 = \boxed{0{,}997}$

Συνέχεια Μαθήματος (κανονική κατανομή)

Θεώρημα.

Έστω $x_1, x_2 \dots x_n$ ανεξάρτητες τ.μ που ακολουθούν κανονική κατανομή με μέσο μ και διασπορά σ^2

- Ο δειγματικός μέσος $\bar{X} = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ ακολουθεί κανονική κατανομή με μέσο μ και διασπορά $\frac{\sigma^2}{n}$.

- Το μερικό άθροισμα $S_n = X_1 + X_2 + \dots + X_n$ ακολουθεί κανονική κατανομή με μέσο $n\mu$ και διασπορά $n\sigma^2$

Παράδειγμα

Το βάρος X των χαλικιών μιας παροχίας ακολουθεί κανονική κατανομή $N(30,10$

i) Αν πάρουμε τυχαία ένα χαλίκι

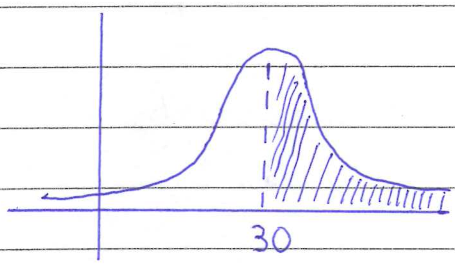
α) Ποια η πιθανότητα να έχει βάρος $X > 30$ gr;

β) » » » » $28 < X < 40$ gr;

ii) Πήραμε 10 χαλίκια. Ποια η πιθανότητα το συνολικό βάρος τους να είναι πάνω από 350 gr;

iii) Αν πάρουμε τυχαία 4 χαλίκια, ποια η πιθανότητα 2 ακριβώς να έχουν βάρος $28 < X < 40$ gr;

iα) $P(X > 30)$



$$P(X>30) = 1 - P(X \le 30)$$
$$= 1 - P\left(\frac{X-30}{10} \le \frac{30-30}{10}\right)$$
$$= 1 - P(Z \le 0) = 1 - \Phi(0)$$
$$= 1 - 0{,}5 = \boxed{0{,}5}$$

β) $P(28 < X < 40) = P(X \le 40) - P(X \le 28)$

$$= P\left(\frac{X-30}{10} \le \frac{40-30}{10}\right) - P\left(\frac{X-30}{10} \le \frac{28-30}{10}\right)$$

$$= P(Z \le 1) - P(Z \le -0{,}2) = \Phi(1) - \Phi(-0{,}2) = \Phi(1) - [1 - \Phi(0{,}2)]$$

$$= 0{,}8413 - (1 - 0{,}5793) = \boxed{0{,}4206}$$

ii) Έστω $X_1, X_2, \dots X_{10}$ το βάρος του κάθε χαλικιού
Για κάθε i, $X_i \sim N(30, 100)$

$X_1 + X_2 + \dots X_{10} = S_{10} \sim N(10 \cdot 30, 10 \cdot 100)$
δηλ. $S_{10} \sim N(300, 1000)$ $\sqrt{1000} \cong 31{,}62$

$P(S_{10} > 350) = 1 - P(X \leq 350)$
$= 1 - P\left(\frac{S_{10} - 300}{31{,}62} \leq \frac{350-300}{31{,}62}\right) = 1 - P(Z \leq 1{,}58)$
$= 1 - \Phi(1{,}58) = 1 - 0{,}9429 = \boxed{0{,}05711}$

iii) Y = # χαλικιών (από τα 4) που έχει βάρος > 28 και < 40

$Y \sim Bin(4, 0{,}4206)$

$P(Y=2) = \binom{4}{2} \cdot 0{,}4206^2 \cdot 0{,}5794^2 = \boxed{0{,}3563}$

Κεντρικό Οριακό Θεώρημα

Έστω $X_1, X_2 \dots X_n$ ανεξάρτητες τ.μ από οποιοδήποτε πληθυσμό (διακριτό ή συνεχή) έτσι ώστε για κάθε i

$E(X_i) = \mu$ και $V(X_i) = \sigma^2$
Τότε, για μεγάλα n, $\bar{X} = \frac{X_1 + X_2 + \dots X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$S_n = X_1 + X_2 + \dots X_n \sim N(n\mu, n\sigma^2)$
(πρακτικά για $n \geq 30$)

Ειδική περίπτωση: κάθε X_i είναι δοκιμή Bernoulli με πιθ. επιτυχίας p
δηλ. διωνυμική (n, p)

Άρα

Αν $X \sim Bin(n,p)$ τότε προσεγγιστικά για $n \geq 30$:

$X \sim N(np, npq)$

Διόρθωση συνέχειας

Αν η διακριτή τ.μ. X προσεγγίζεται από συνεχή, τότε κάνουμε διορθ. συνέχειας

$$P(X=k) = P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right) \quad \bullet \; P(a \leq X \leq b) = P\left(a - \frac{1}{2} < X < b + \frac{1}{2}\right)$$

Παράδειγμα

Το 10% των προϊόντων που παράγει μια μηχανή είναι ελαττωματικά. Επιλέγουμε 400. Ποια η πιθανότητα να είναι ελαττωματικά

α) το πολύ 30

β) από 30 ως 50

γ) 55 ή περισσότερα

X = # ελαττωματικών (από τα 400) $\quad X \sim Bin(400, 0{,}1)$

α) $P(X \leq 30) = ;$

Εφόσον $400 \geq 30$ την προσεγγίζουμε από κανονική κατανομή

$X \sim N(400 \cdot 0{,}1 \,,\, 400 \cdot 0{,}1 \cdot 0{,}9) = X \sim N(40, 36)$

$$P(X \leq 30) \xrightarrow{\text{διορθ. συνεχ.}} P(X \leq 30{,}5) = P\left(\frac{X-40}{6} \leq \frac{30{,}5-40}{6}\right)$$

$= P(Z \leq -1{,}58)$

$= \Phi(-1{,}58)$

$= 1 - \Phi(1{,}58)$

$= 1 - 0{,}9429 = 0{,}0571$

β) $P(30 \leq X \leq 50) \overset{\text{διορθ. συνεχ.}}{=} P(29{,}5 < X < 50{,}5)$

$P\left(\frac{29{,}5-40}{6} \leq \frac{X-40}{6} \leq \frac{50{,}5-40}{6}\right)$

$= P(-1{,}75 \leq Z \leq 1{,}75) = P(Z \leq 1{,}75) - P(Z \leq -1{,}75)$

$= \Phi(1{,}75) - \Phi(-1{,}75) = \Phi(1{,}75) - (1-\Phi(1{,}75))$

$= 2\Phi(1{,}75) - 1 = 2 \cdot 0{,}9599 - 1 = \boxed{0{,}9198}$

γ) $P(X \geq 55) \overset{\text{δ. συνεχ}}{=} P(X > 55{,}5) = P\left(\frac{X-40}{6} > \frac{55{,}5-40}{6}\right)$

$= P(Z \geq 2{,}42) = 1 - \Phi(2{,}42) = 1 - 0{,}9922 = \boxed{0{,}0078}$

Υπεργεωμετρική

Παρατήρηση

Αν $X \sim H(N, r, n)$ και $n \geq 30$ τότε προσεγγιστικά:

$X \sim N\left(\frac{nr}{N}, \frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}\right)$ (και εδώ χρειάζεται διορθωση συνέχειας

Παράδειγμα

Το ποσοστό p των ατόμων που πάσχουν από μια σπάνια ασθένεια είναι μικρότερο του 3%. Επιλέγουμε n άτομα τυχαία. Πόσο πρέπει να είναι το n ώστε το ποσοστό των ατόμων του δείγματος που πάσχουν από την ασθένεια να απέχει από το p λιγότερο από 1% με πιθανότητα τουλάχιστον 95%. (Θεωρούμε ότι $n \geq 30$)

X = # ατόμων του δείγματος που πασχει από την ασθένεια

$X \sim Bin(n, p)$

$P\left(\left|\frac{X}{n} - p\right| \leq 0{,}01\right) \geq 0{,}95 \Rightarrow P\left(-0{,}01 \leq \frac{X}{n} - p \leq 0{,}01\right) \geq 0{,}95$

Προσεγγιστικά $X \sim N(np, npq)$

$$P\left(-0{,}01 \le \frac{X-np}{n} \le 0{,}01\right) \ge 0{,}95$$

$$\Rightarrow P\left(-0{,}01 \cdot \frac{n}{\sqrt{npq}} \le \frac{X-np}{\sqrt{npq}} \le 0{,}01 \cdot \frac{n}{\sqrt{npq}}\right) \ge 0{,}95$$

$$\Rightarrow P(\ldots \le Z \le \ldots) \ge 0{,}95 \Rightarrow P(Z \le \ldots) - P(Z \ge \ldots) \ge 0{,}95$$

$$\Rightarrow \Phi\left(0{,}01 \cdot \frac{n}{\sqrt{npq}}\right) - \Phi\left(-0{,}01 \cdot \frac{n}{\sqrt{npq}}\right) \ge 0{,}95$$

$$\Rightarrow \Phi\left(0{,}01 \cdot \frac{n}{\sqrt{npq}}\right) - \left(1 - \Phi\left(0{,}01 \cdot \frac{n}{\sqrt{npq}}\right)\right) \ge 0{,}95$$

$$\Rightarrow 2\Phi\left(\frac{0{,}01 \cdot n}{\sqrt{npq}}\right) - 1 \ge 0{,}95$$

$$\Rightarrow \Phi\left(\frac{0{,}01 \cdot n}{\sqrt{npq}}\right) \ge 0{,}975 \Rightarrow \Phi\left(\frac{0{,}01 \cdot n}{\sqrt{npq}}\right) \ge \Phi(1{,}96)$$

$$\Rightarrow \frac{0{,}01 \cdot n}{\sqrt{npq}} \ge 1{,}96 \Rightarrow \sqrt{n} \ge \frac{1{,}96}{0{,}01} \cdot \sqrt{pq} \Rightarrow n \ge 38416p(1-p)$$

Εφόσον $p < 0{,}3 \Rightarrow n \ge 38410 \cdot 0{,}3 \cdot 0{,}7 \cong 1118$