Τυχαίες μεταβλητές και πιθανότητες

Διακριτή τ.μ.

Πιθανότητα η τ.μ. X να πάρει τιμή x_i $\qquad P(X = x_i) = \lim_{n\to\infty} \frac{f_i}{n}$

m διακριτές τιμές $x_1, x_2, \ldots, x_m$,

συνάρτηση μάζας πιθανότητας

$$f_X(x_i) \equiv f(x_i) = P(X = x_i)$$

$$f(x_i) \geq 0 \qquad \sum_{i=1}^{m} f(x_i) = 1$$

αθροιστική συνάρτηση κατανομής

$$F_X(x_i) \equiv F(x_i) = P(X \leq x_i) = \sum_{x \leq x_i} f(x)$$

$$x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$$

Συνεχής τ.μ.

Η συνεχής τ.μ. X παίρνει τιμές σ' ένα διάστημα τιμών

Πιθανότητα η τ.μ. X να πάρει μια τιμή x_i **?**

συνάρτηση πυκνότητας πιθανότητας

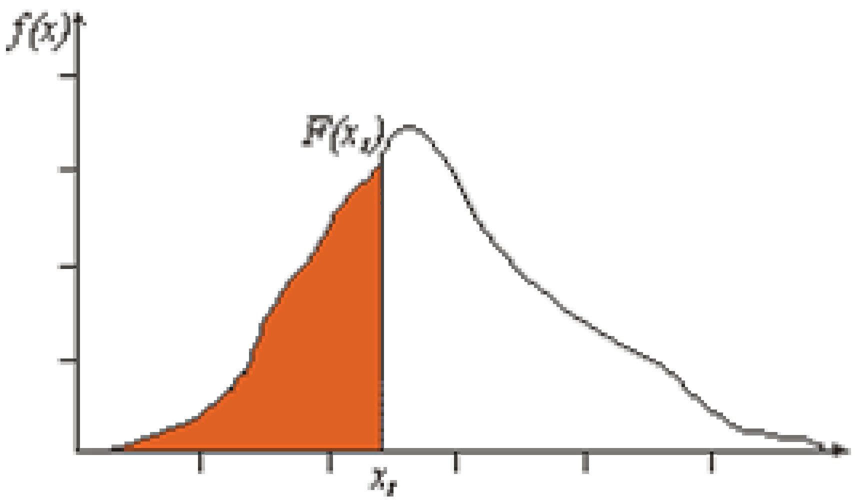
$$f_X(x) = f(x)$$

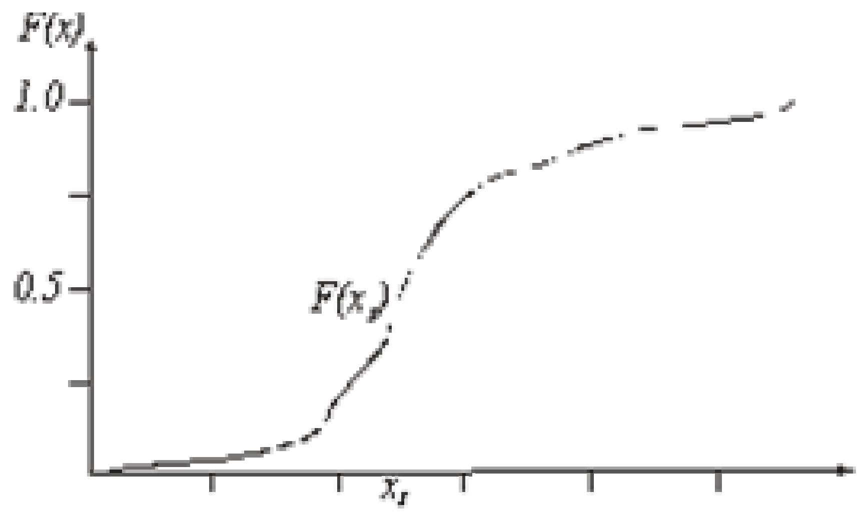
$$f(x) \geq 0 \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

αθροιστική συνάρτηση κατανομής

$$F_X(x) \equiv F(x) = P(X \leq x_i) = \int_{-\infty}^{x} f(u)\,du$$

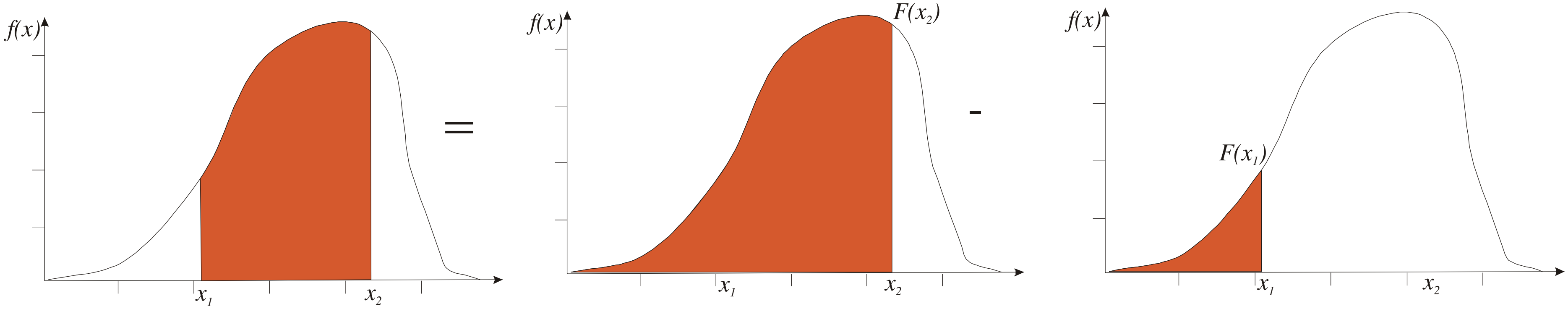
$$x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$$





$$P(x_1 \le X \le x_2) = P(X \le x_2) - P(X \le x_1) = F(x_2) - F(x_1)$$

$$P(x_1 \le X \le x_2) \quad = \quad F(x_2) \quad - \quad F(x_1)$$



Παράμετροι κατανομής της τ.μ. X

Μέση τιμή $\mu_X \equiv \mathrm{E}[X]$

X διακριτή $\mu = \sum_{i=1}^{m} x_i f(x_i)$

X συνεχής $\mu = \int_{-\infty}^{+\infty} x\, f(x)\, dx$

Ιδιότητες:

$$\mathrm{E}[cX] = c\,\mathrm{E}[X]$$

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$$

Αν X, Y ανεξάρτητες $\mathrm{E}[X\,Y] = \mathrm{E}[X]\mathrm{E}[Y]$

Διάμεσος $\tilde{\mu}$

$$F(\tilde{\mu}) = 0.5.$$

Διασπορά $\sigma_X^2 \equiv \mathrm{Var}[X]$

$$\boxed{\sigma^2 \equiv \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2}$$

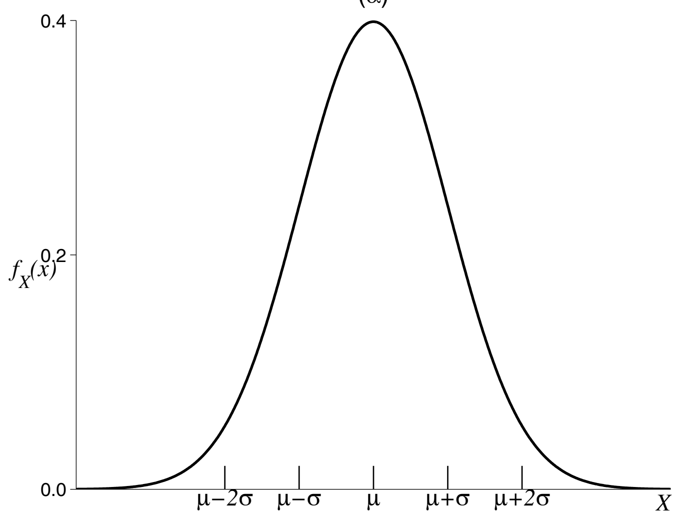
Ιδιότητες:

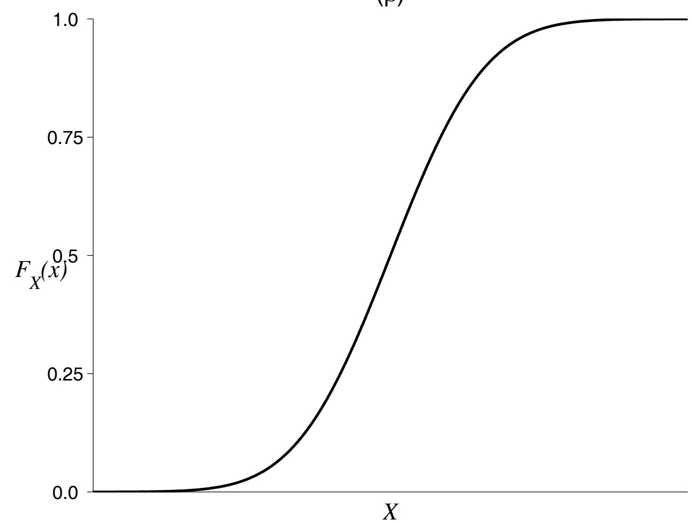
$$\mathrm{Var}[X+c] = \mathrm{Var}[X] \qquad \mathrm{Var}[cX] = c^2\,\mathrm{Var}[X]$$

$$\left.\begin{array}{l}\text{Αν } X,\ Y \text{ ανεξάρτητες} \\ f(x,y) = f_X(x) f_Y(y)\end{array}\right\} \quad \mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Κανονική κατανομή

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty<x<\infty \qquad X\sim \mathrm{N}(\mu,\sigma^2)$$





Τυπική κανονική κατανομή

$Z\sim \mathrm{N}(0,1)$

$$f_Z(z)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}, \qquad -\infty<z<\infty$$

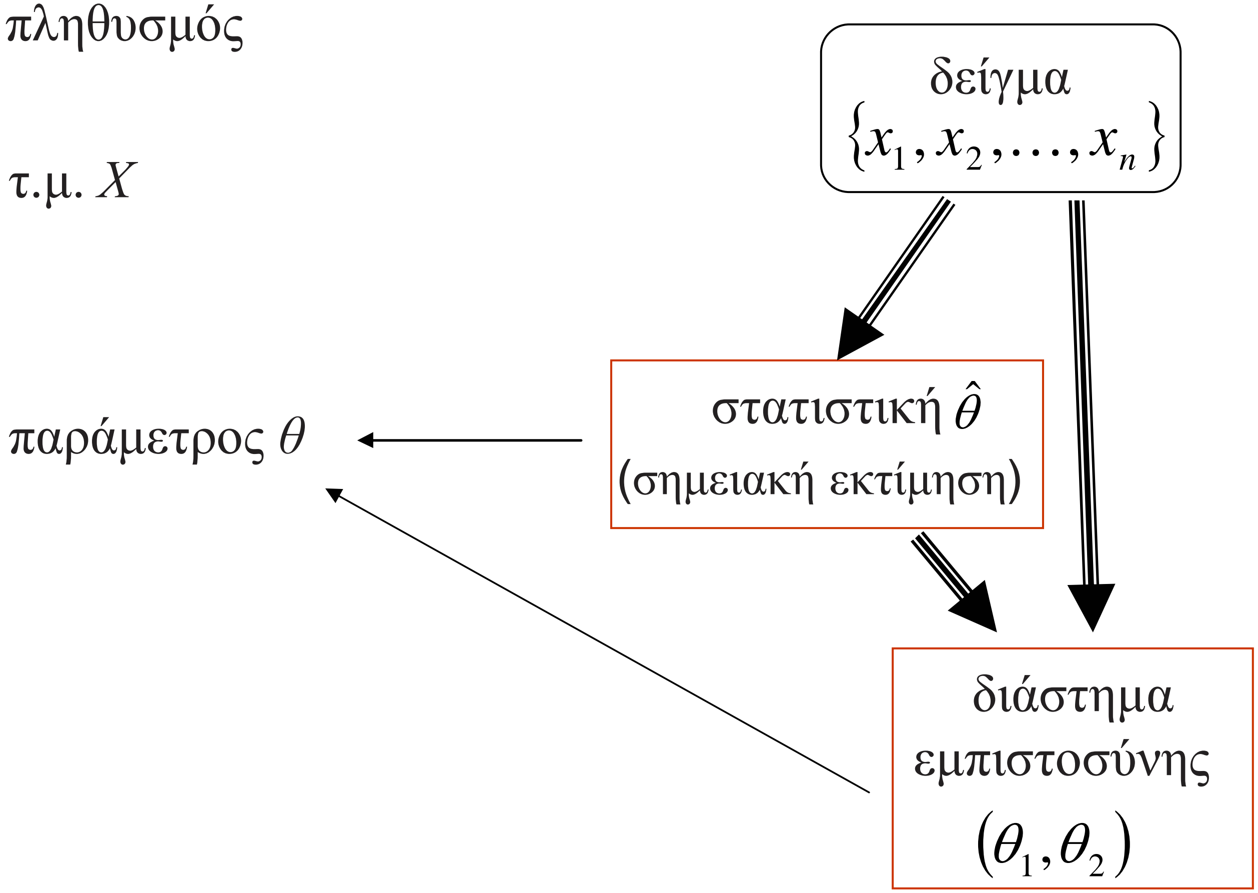
$$\Phi(z)\equiv F_Z(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\mathrm{e}^{-\frac{u^2}{2}}\,du, \qquad -\infty<z<\infty$$

$$\boxed{X\sim \mathrm{N}(\mu,\sigma^2) \;\Rightarrow\; Z\equiv\frac{X-\mu}{\sigma}\sim \mathrm{N}(0,1)}$$

$$\mathrm{P}(a\le X\le b)=\mathrm{P}\left(\frac{a-\mu}{\sigma}\le Z\le\frac{b-\mu}{\sigma}\right)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\Phi(-z)=\mathrm{P}(Z\le -z)=\mathrm{P}(Z>z)=1-\mathrm{P}(Z\le z)=1-\Phi(z)$$

Εκτίμηση Παραμέτρων



Παράμετρος	**Στατιστική**

<u>αριθμητικά δεδομένα</u>

μέση τιμή	δειγματική μέση τιμή
$\mu \equiv \mathrm{E}[X]$	$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$
διασπορά	δειγματική διασπορά
$\sigma^2 \equiv \mathrm{Var}[X] = \mathrm{E}[X^2] - \mu^2$	$s^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$

<u>κατηγορικά δεδομένα</u>

αναλογία	δειγματική αναλογία
p	$\hat{p} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ όπου $x_i = 0$ ή 1

Εκτιμήτρια $\hat{\theta}$

Καλύτερη μέθοδος εκτίμησης της θ:

μέθοδος της μεγίστης πιθανοφάνειας

Κριτήρια καλής εκτιμήτριας

- ▶ Η $\hat{\theta}$ είναι αμερόληπτη: $\boxed{\mathrm{E}[\hat{\theta}] = \theta}$
- ▶ Η $\hat{\theta}$ έχει μικρή διασπορά

$\bar{x}$ και s^2 : *αμερόληπτες εκτιμήτριες με τη μικρότερη διασπορά*

$$\tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

δεν είναι αμερόληπτη εκτιμήτρια του σ^2

Κατανομές πιθανοτήτων εκτιμητριών

$\hat{\theta}$ αλλάζει από δείγμα σε δείγμα

$\{x_1, x_2, \ldots, x_n\}$ τ.μ. $\Rightarrow$ $\hat{\theta}$ είναι τ.μ.

$\hat{\theta}$ δίνεται από κάποια $f_\theta(\theta)$

- μέση τιμή $\mu_{\hat{\theta}} \equiv \mathrm{E}[\hat{\theta}]$
- διασπορά $\sigma_{\hat{\theta}}^2 \equiv \mathrm{Var}[\hat{\theta}]$
- τυπική απόκλιση $\sigma_{\hat{\theta}}$

  τυπικό σφάλμα της $\hat{\theta}$

Κατανομή δειγματικής μέσης τιμής $\theta \rightarrow \mu$

$$\mu_{\bar{x}} = \mathrm{E}\left[\frac{x_1 + x_2 + \cdots x_n}{n}\right] = \frac{\mathrm{E}[x_1 + x_2 + \cdots x_n]}{n} =$$

$$\frac{\mathrm{E}[x_1] + \mathrm{E}[x_2] + \cdots \mathrm{E}[x_n]}{n} = \frac{n\mu}{n} = \mu$$

$x_1, x_2, \ldots, x_n$ είναι ανεξάρτητες

$$\sigma_{\bar{x}}^2 = \mathrm{Var}\left[\frac{x_1 + x_2 + \cdots x_n}{n}\right] = \frac{\mathrm{Var}[x_1 + x_2 + \cdots x_n]}{} =$$

$$\frac{\mathrm{Var}[x_1] + \mathrm{Var}[x_2] + \cdots + \mathrm{Var}[x_n]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

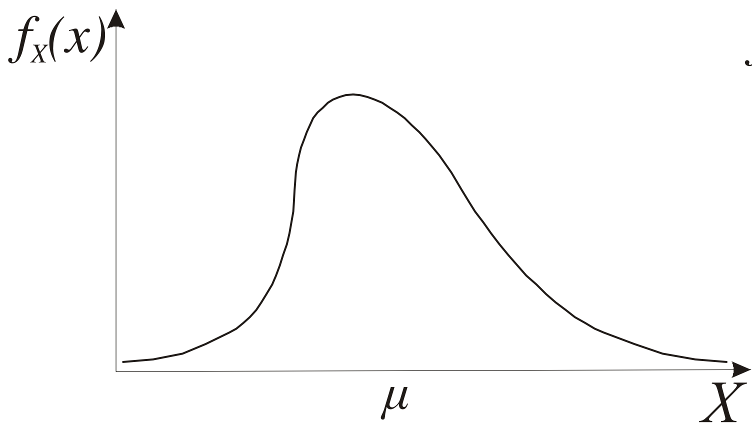
$\boxed{n > 30}$

σ^2 γνωστό

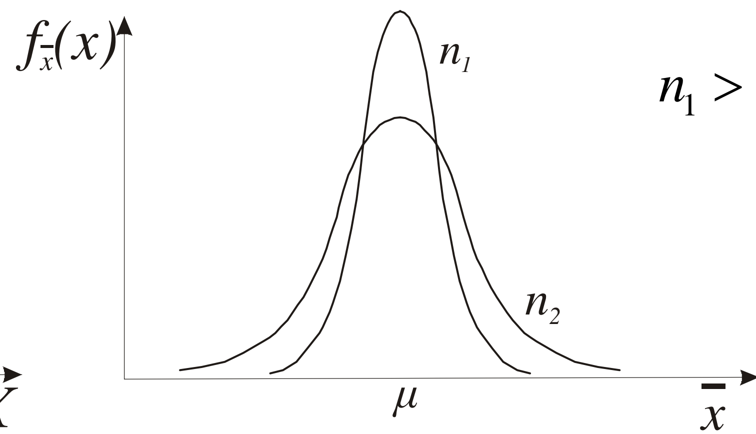
σ^2 άγνωστο

$$\boxed{n > 30 \quad \Rightarrow \quad \bar{x} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \acute{\eta} \quad \bar{x} \sim \mathrm{N}\left(\mu, \frac{s^2}{n}\right)}$$

(α)



(β)



$n_1 > n_2$

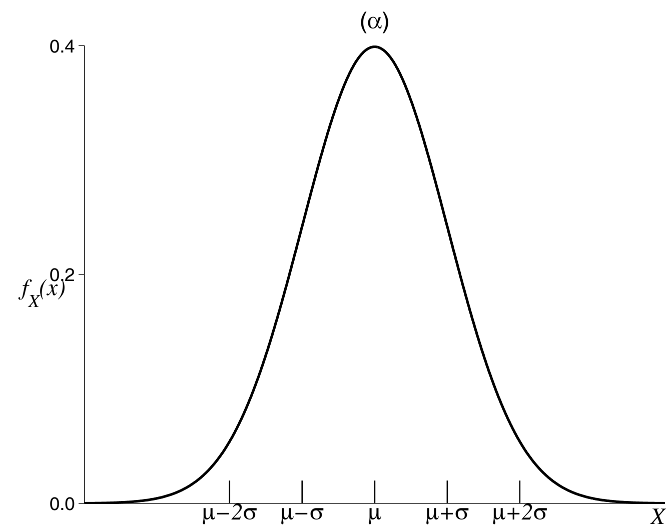
$$z \equiv \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim \mathrm{N}(0,1)$$

$$\mathrm{P}(a \leq \bar{x} \leq b) = \mathrm{P}\left(\frac{a-\mu}{\sigma_{\bar{x}}} \leq z \leq \frac{b-\mu}{\sigma_{\bar{x}}}\right) = \Phi\left(\frac{b-\mu}{\sigma_{\bar{x}}}\right) - \Phi\left(\frac{a-\mu}{\sigma_{\bar{x}}}\right)$$

$n < 30$

σ^2 γνωστό

$$n < 30 \text{ και } X \sim N(\mu, \sigma^2) \;\Rightarrow\; z \equiv \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



Κατανομή Student

σ^2 άγνωστο

$$n < 30 \text{ και } X \sim N(\mu, \sigma^2) \;\Rightarrow\; t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$