Στατιστική των Επιχειρήσεων Ι

Διάλεξη 2η

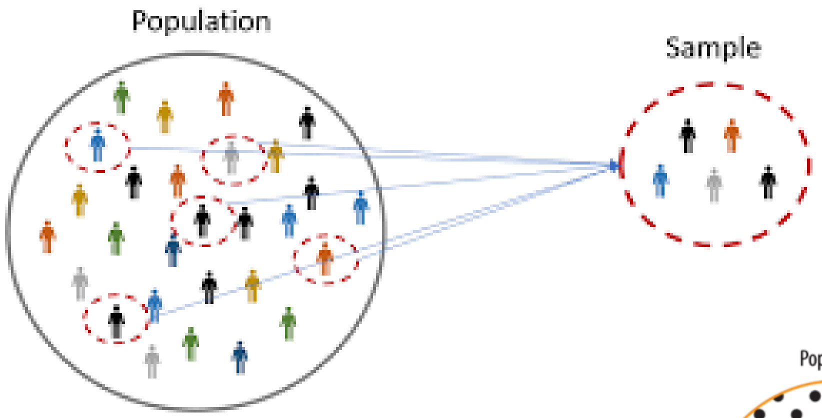
Γιώργος Τσιρογιάννης

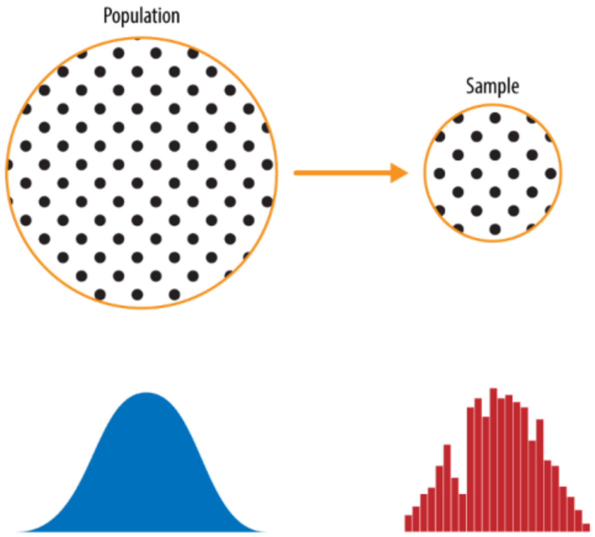
Τμήμα Διοίκησης Επιχειρήσεων Αγροτικών Προϊόντων και Τροφίμων,

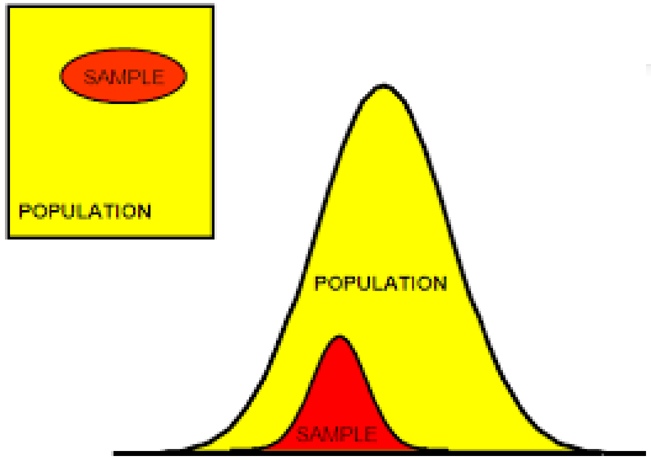
Πανεπιστήμιο Πατρών

Ακαδημαϊκό έτος: 2020-2021

Πληθυσμός vs δείγμα







Τυχαίο δείγμα

- Τυχαίο δείγμα στην στατιστική θεωρείται η βάση το οποίο οδηγούμαστε σε συμπεράσματα από το «μέρος» για το «όλο».
- Το δείγμα αποτελείται από ανεξάρτητες τυχαίες μεταβλητές.

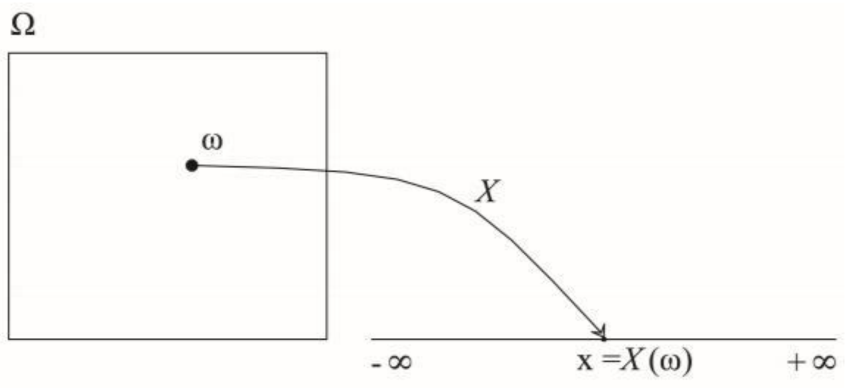
Στατιστικός πληθυσμός

- **Πληθυσμός ή στατιστικός πληθυσμός** (statistical population), είναι οι δυνατές τιμές μιας τυχαίας μεταβλητής, δηλαδή η κατανομή των δυνατών τιμών μιας τυχαίας μεταβλητής.

Πληθυσμός και Τυχαίο Δείγμα

- **Απλό στοιχείο ή πειραματική/δειγματοληπτική μονάδα** είναι κάθε υποκείμενο επί του οποίου μετράμε/παρατηρούμε την τιμή μιας τυχαίας μεταβλητής
- Η κατανομή των τιμών μιας τυχαίας μεταβλητής ονομάζεται πληθυσμός ή **στατιστικός πληθυσμός.**
- **Τυχαίο δείγμα μεγέθους ν από έναν πληθυσμό**, δηλαδή, από την κατανομή των τιμών μιας τυχαίας μεταβλητής X, ονομάζουμε ν ανεξάρτητες τυχαίες μεταβλητές $\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_{3,} \ldots, \mathrm{X}_\nu$ που παίρνουν τιμές από τον πληθυσμό αυτό, που ακολουθούν δηλαδή την ίδια κατανομή, αυτήν της τ.μ. X.
- Οι συγκεκριμένες τιμές $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_{3,} \ldots, \mathrm{x}_\nu$, της X που έχουμε διαθέσιμες για επεξεργασία μετά τη λήψη του δείγματος αποτελούν μια **πραγματοποίηση των $\boldsymbol{X_1, X_2, X_{3,} \ldots, X_\nu}$ και ονομάζονται δεδομένα** ή παρατηρήσεις.

Τυχαίες μεταβλητές



- Μια **συνάρτηση** που αντιστοιχίζει το αποτέλεσμα που εμφανίζεται κατά την εκτέλεση ενός πειράματος τύχης σε έναν **πραγματικό αριθμό**

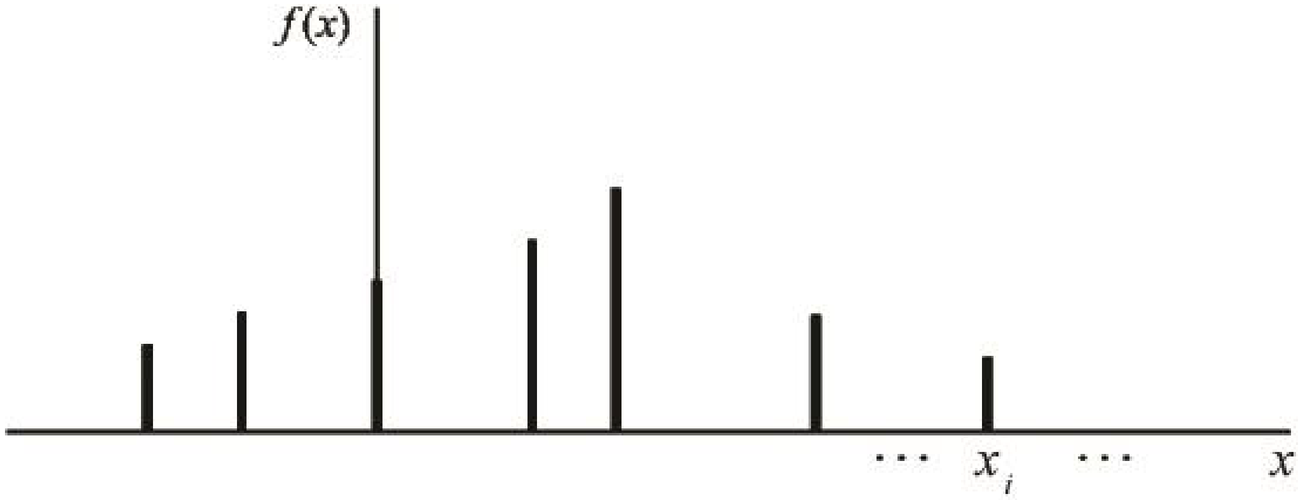
Συνάρτηση (μάζας) πιθανότητας (διακριτές τμ)

- Έστω X διακριτή τυχαία μεταβλητή με σύνολο τιμών R_X. Η πραγματική συνάρτηση f με

$$f(x)=\begin{cases} P(X=x), & \text{για κάθε } x\in R_X \\ 0, & \text{για κάθε } x\notin R_X \end{cases}$$

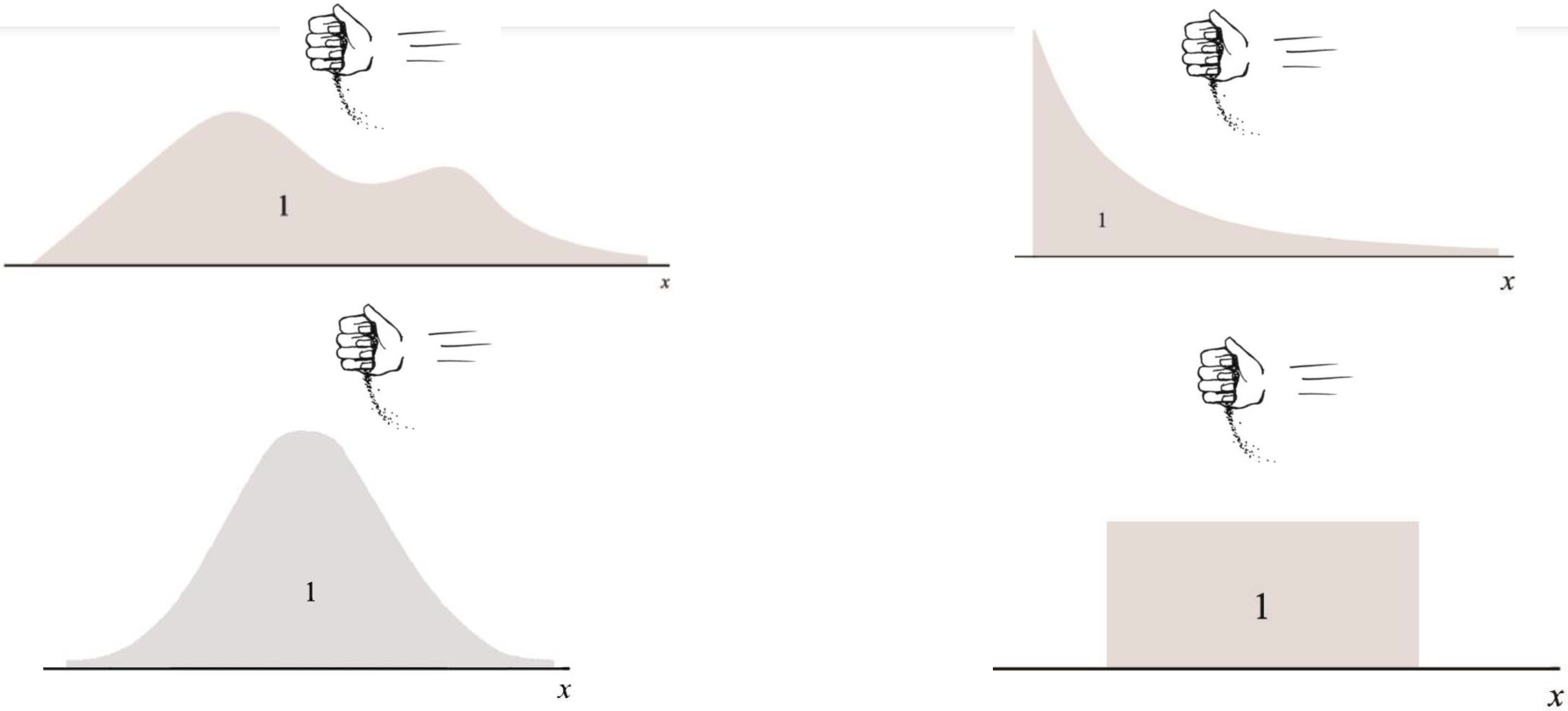
ονομάζεται συνάρτηση πιθανότητας ή συνάρτηση μάζας πιθανότητας της X (probability mass function).

Συνάρτηση (μάζας) πιθανότητας (διακριτές τμ)

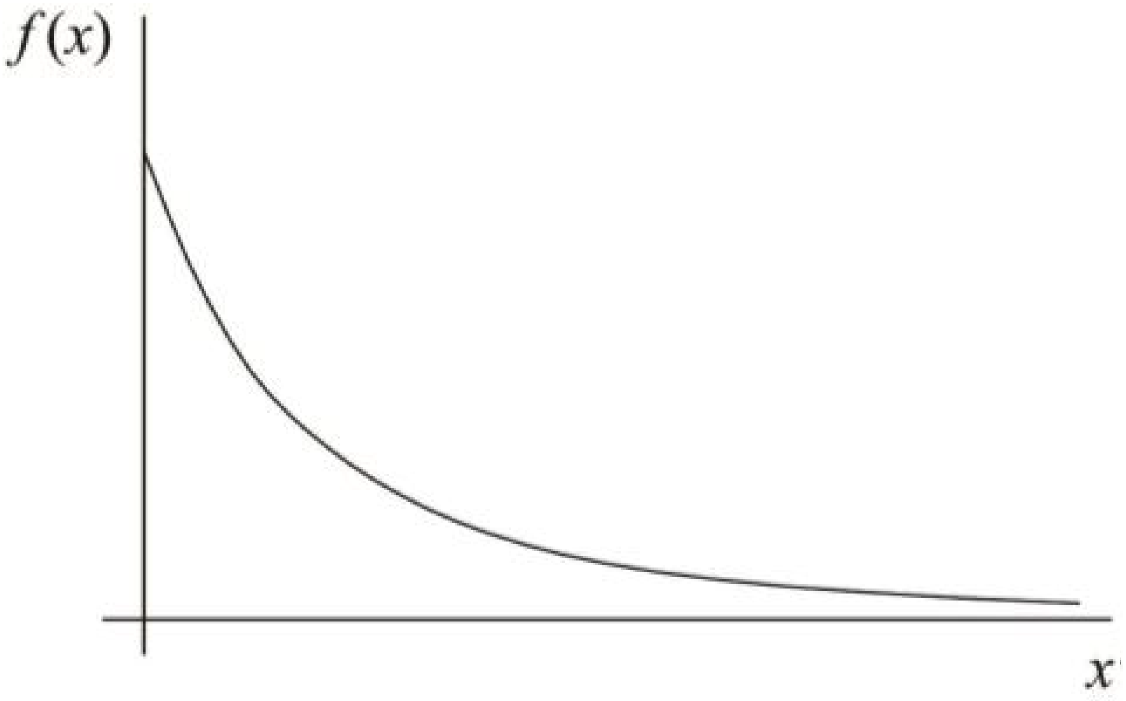


Συνεχείς τυχαίες μεταβλητές

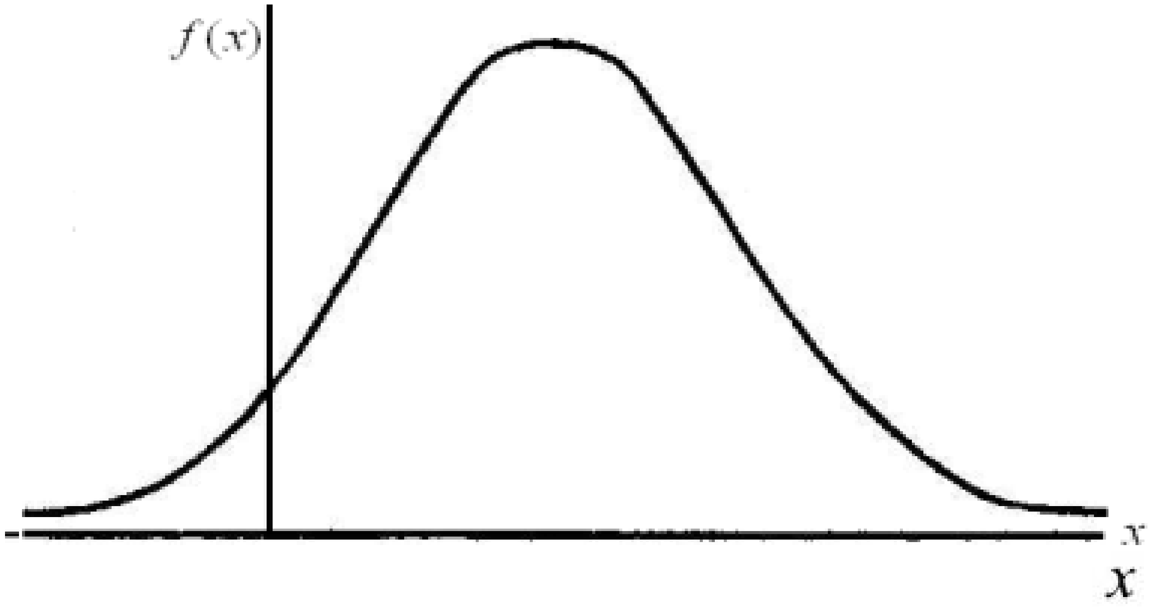
Παραδείγματα εκχώρησης/κατανομής της συνολικής πιθανότητας σε συνεχές διάστημα τιμών



Παράδειγμα συνάρτησης πυκνότητας πιθανότητας

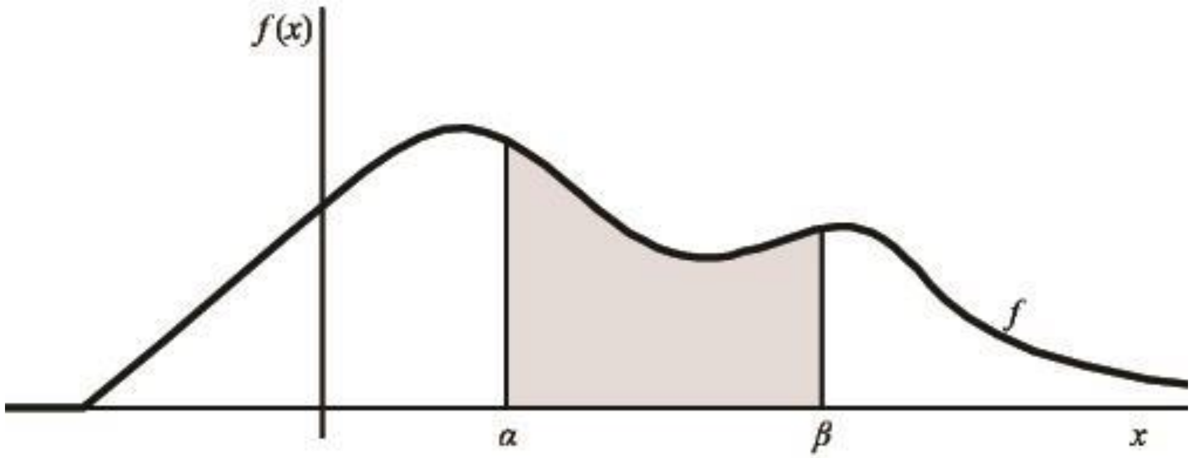


Παράδειγμα συνάρτησης πυκνότητας πιθανότητας



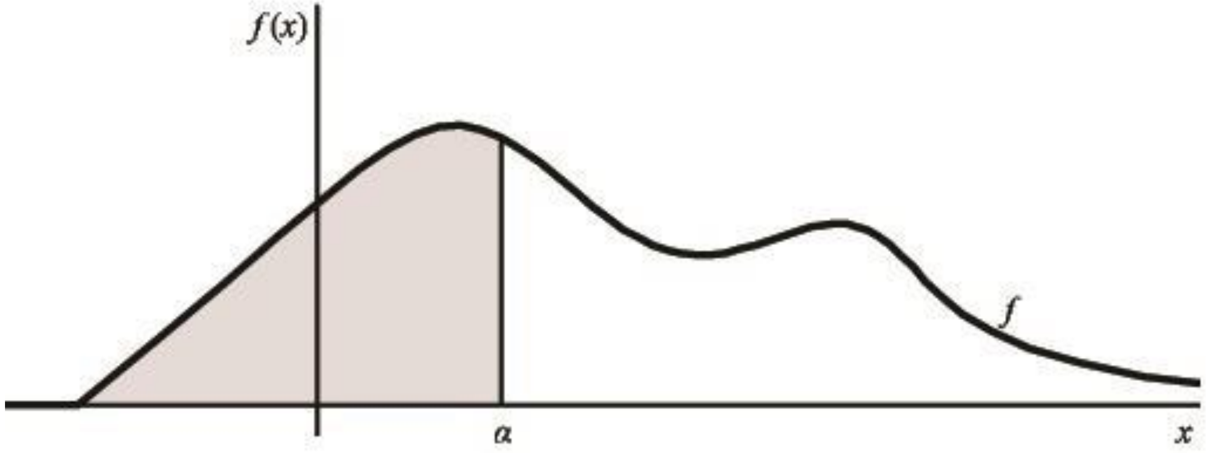
Πιθανότητες ενδεχομένων

$X \in [\alpha, \beta]$



$P(\alpha \leq X \leq \beta)$

$X \in (-\infty, \alpha]$



$P(X \leq \alpha)$

Συνεχής τυχαία μεταβλητή και συνάρτηση πυκνότητας πιθανότητας

- Μια τυχαία μεταβλητή X λέγεται συνεχής αν υπάρχει μια μη αρνητική πραγματική συνάρτηση

$$f : R \to [0, +\infty)$$

τέτοια ώστε για κάθε υποσύνολο A των πραγματικών να ισχύει

$$P(X \in A) = \int_A f(x)dx$$

- Η συνάρτηση f ονομαζεται **συνάρτηση πυκνότητας πιθανότητας** ή απλά συνάρτηση πυκνότητας της X (probability density function, pdf).

Πρόταση

- Αν X συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής F και συνάρτηση πυκνότητας f, τότε:

$$\alpha)\ \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\beta)\ F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

$$\gamma)\ f(x) = F'(x)$$ (στα σημεία συνέχειας της f)

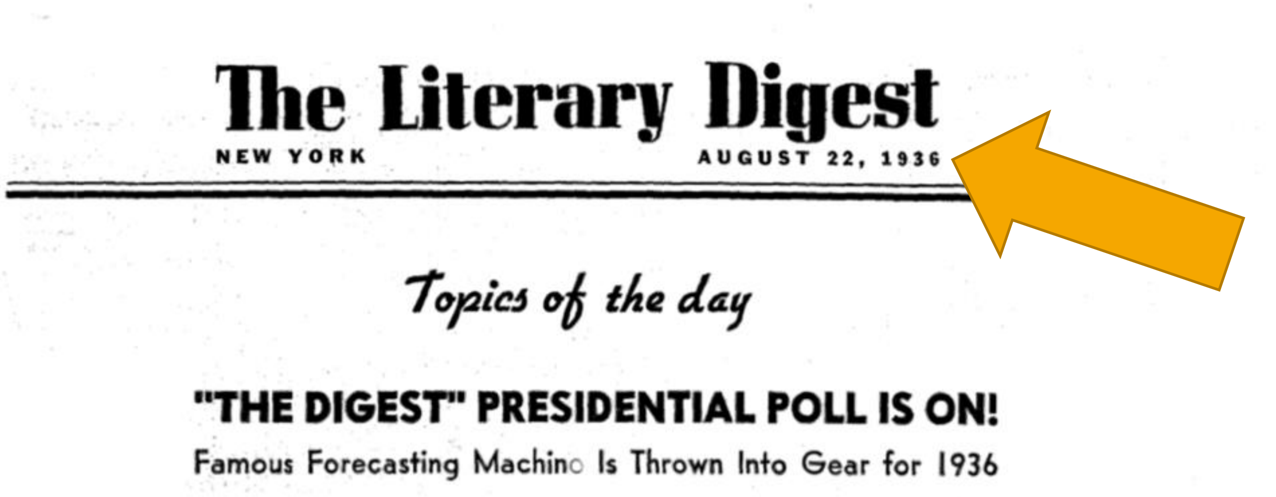
$\delta)$ *Για οποιουσδήποτε πραγματικούς* α, β *με* $\alpha \le \beta$

$$(i)\ \ P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)dx = F(\beta) - F(\alpha).$$

$$(ii)\ \ P(\alpha < X < \beta) = P(\alpha \le X < \beta) = P(\alpha < X \le \beta) = P(\alpha \le X \le \beta)$$

Σχόλια

- Η πιθανότητα μια συνεχής τυχαία μεταβλητή X να πάρει μια συγκεκριμένη τιμή α , είναι ίση μηδέν για οποιοδήποτε α.
- Αυτό δεν σημαίνει ότι το ενδεχόμενο X=α είναι το αδύνατο ενδεχόμενο ∅.
- **Σημαίνει ότι αυτό το ενδεχόμενο α είναι απίθανο να συμβεί και όχι ότι είναι απραγματοποίητο.**


The Literary Digest

NEW YORK AUGUST 22, 1936

Topics of the day

"THE DIGEST" PRESIDENTIAL POLL IS ON!

Famous Forecasting Machine Is Thrown Into Gear for 1936

Η αξία της δειγματοληψίας

Η αξία της δειγματοληψίας

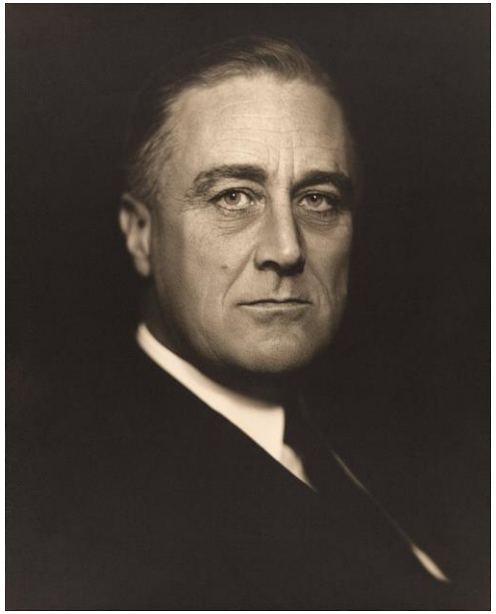
Alf Landon

Δείγμα: 10000000

Sample bias

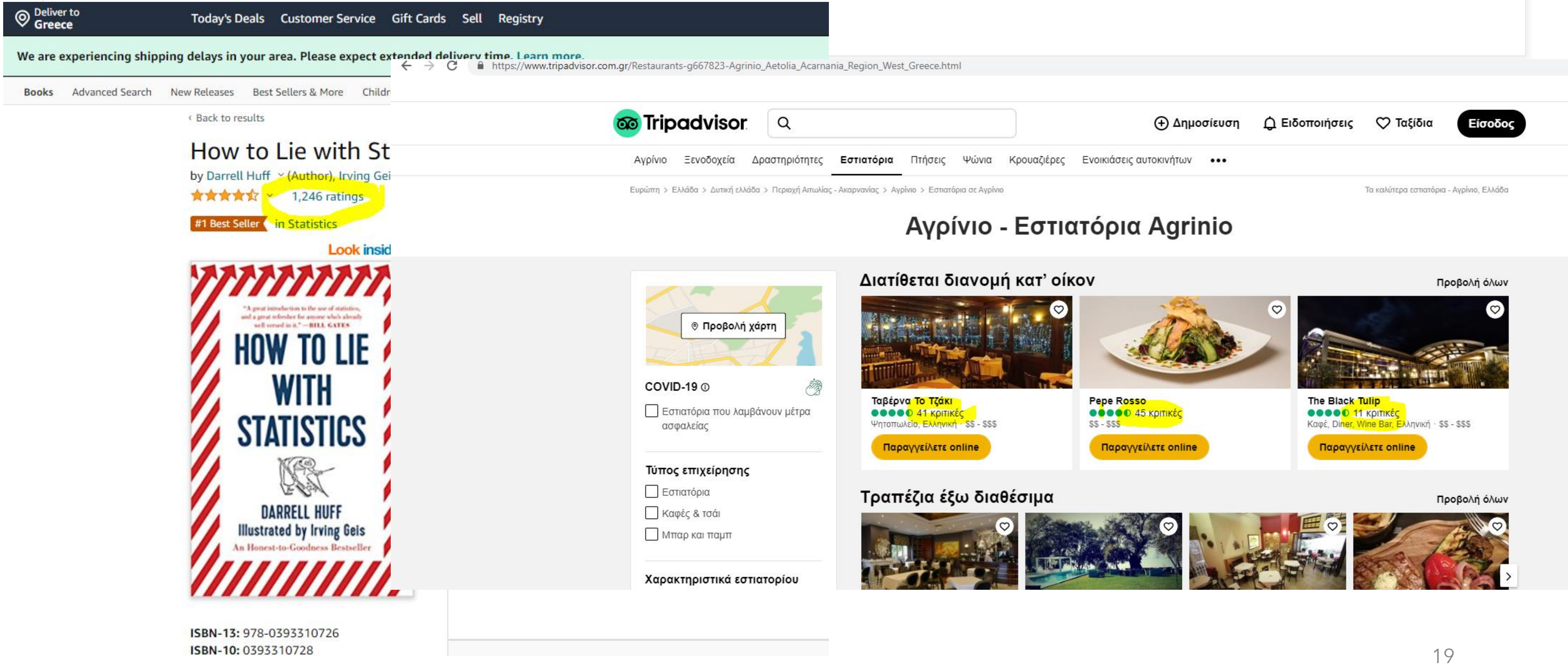
George Gallup

Πολλαπλά δείγματα: 2000



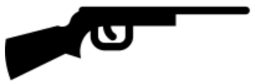
Franklin D. Roosevelt

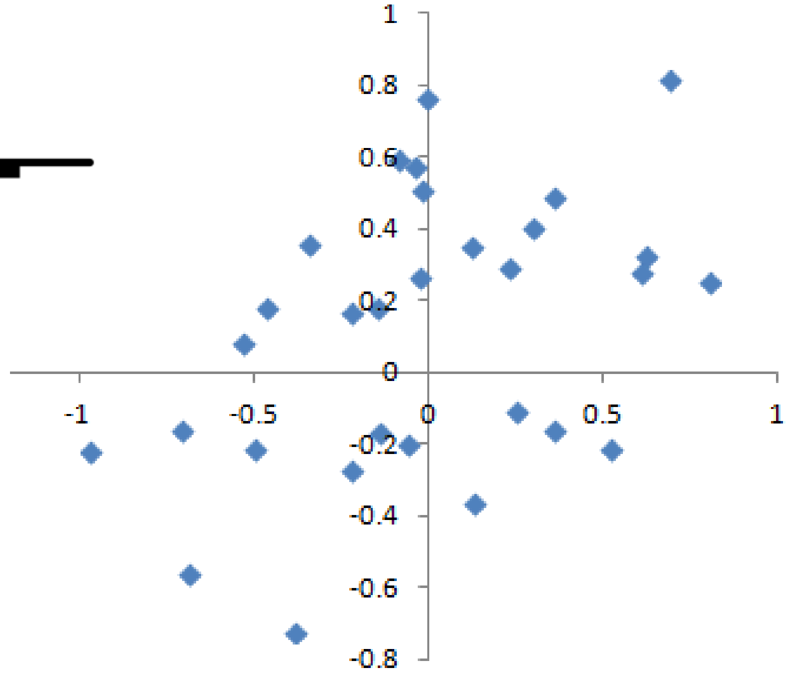
self-selection bias



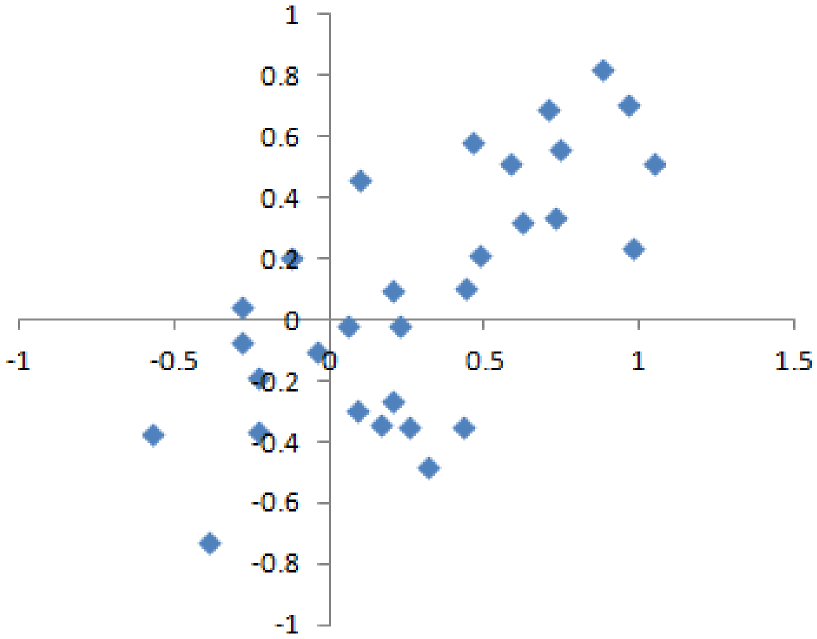
Bias (μεροληψία/προτίμηση)

- Η στατιστική bias αναφέρεται σε σφάλματα μέτρησης ή δειγματοληψίας που είναι συστηματικά και προκύπτουν από τον τρόπο μέτρησης και διαδικασία δειγμαληψίας.





unbiased



biased

Τυχαία δειγματοληψία

- Ομοιόμορφος τρόπος επιλογής του δείγματος ώστε να είναι αντιπροσωπευτικός
- Είναι απαιτητικό ώστε να γίνει σωστά

Αριθμητικά περιγραφικά μέτρα (numerical descriptive measures)

- Ποσοτικά μεγέθη που βοηθούν στην περιγραφή της κατανομής ενός δείγματος ή στην περιγραφή ενός πληθυσμού
- Γνώριμα από την θεωρία πιθανοτήτων:
  - Αν αφορούν έναν πληθυσμό, ονομάζονται **παράμετροι (parameters)**
  - Αν αφορούν ένα δείγμα από έναν πληθυσμό ονομάζονται **στατιστικά (statistics)**

Διαφορές

- Οι παράμετροι ενός πληθυσμού είναι συγκεκριμένοι/μοναδικοί αριθμοί (γνωστοί άγνωστοι).
- Τα στατιστικά, για συγκεκριμένη πραγματοποίηση $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_\nu}$ ενός δείγματος $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}, \ldots, \boldsymbol{X_\nu}$ μπορούμε να τα υπολογίσουμε και επομένως μας είναι γνωστά
- Σε μια άλλη πραγματοποίηση του δείγματος η τιμή τους μεταβάλλεται, δηλαδή, τα στατιστικά είναι τυχαίες μεταβλητές.
- Κάθε στατιστικό, είναι τυχαία μεταβλητή και για αυτό συμβολίζεται με κεφαλαίο γράμμα όπως οι τυχαίες μεταβλητές

Δειγματικός μέσος (sample mean/arithmetic mean/average)

- Ορίζεται από τον τύπο: $\overline{X} = \frac{1}{\nu}\sum_{i=1}^{\nu} X_i$
- Η συγκεκριμένη τιμή του $\overline{X}$, που υπολογίζεται για μια πραγματοποίηση $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_\nu}$ του τυχαίου δείγματος $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}, \ldots, \boldsymbol{X_\nu}$ συμβολίζεται με $\bar{x}$, δηλαδή $\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i$
- Ευαίσθητο σε *ακραίες-έκτροπες (outlying* τιμες*)*
    - αποκρύπτει τις έκτροπες τιμές

Παλινδρόμηση στον μέσο (regression to the mean)

- Αναφέρεται στο φαινόμενο όπου όταν λαμβάνουμε επαναλαμβανόμενα δείγματα από ένα πληθυσμό (δηλαδή μιας τμ), ακραίες τιμές τείνουν να ακολουθούνται από περισσότερες κεντρικές

Παράδειγμα

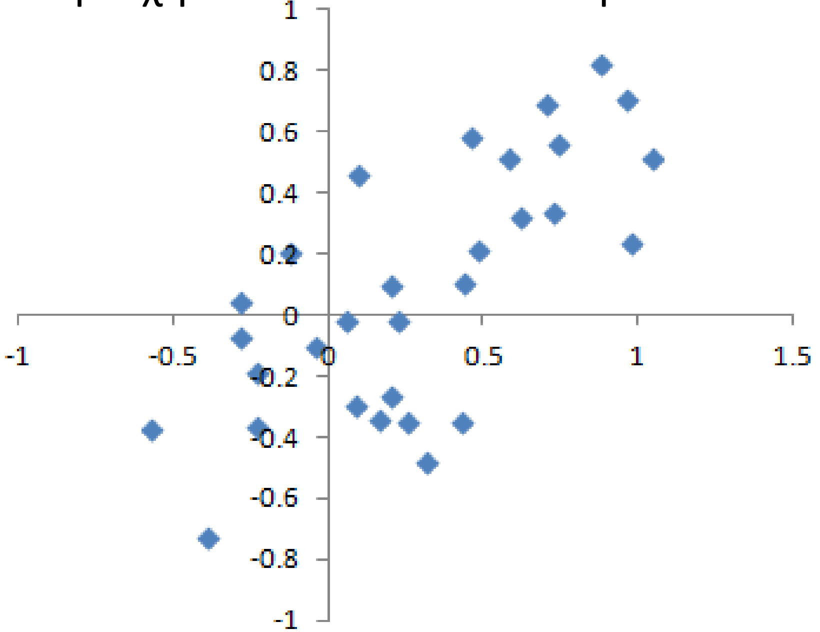
Το φαινόμενο του νεοφώτιστου “rookie of the year, sophomore slump:

- Τις περισσότερες χρονιές εμφανίζεται κάποιος νέος αθλητής με απόδοση κορυφαίων αθλητών
- Γιατί δεν συνεχίζει η εντυπωσιακή πορεία τα επόμενη χρόνια;
  - Δεξιότητες
  - Τύχη

Παράδειγμα (συν)

Η επιλογή του "rookie of the year" είναι συνήθως αποτέλεσμα του selection bias.
Το πιθανότερο οι δεξιότητες θα παραμείνουν
Αλλά η τύχη δεν θα είναι ευνοϊκή

Η μέση απόδοση των επόμενων χρόνων θα παλινδρομήσει γύρω από την μέση τιμή των συμπαικτών του



Συνάρτηση δειγματοληψίας στατιστικού

- Αναφέρεται στην κατανομή που ακολουθεί το στατιστικό ενός αρκούντως μεγάλου δείγματος του πληθυσμού.
- Γιατί είναι σημαντική:
  - Πολλές φορές προσπαθούμε μέσω του στατιστικού να βγάλουμε συμπέρασμα ή να πάρουμε απόφαση

Παράδειγμα

Φορτωση των βιβλιοθηκών

[2]:
```
from pathlib import Path
import pandas as pd
import numpy as np
from scipy import stats
import seaborn as sns
import matplotlib.pylab as plt
```

Καθορισμός πως θα εμφανίζονται τα σχήματα

[3]:
```
%matplotlib inline
```

Καθορισμός της τοποθεσίας των δεδομένων

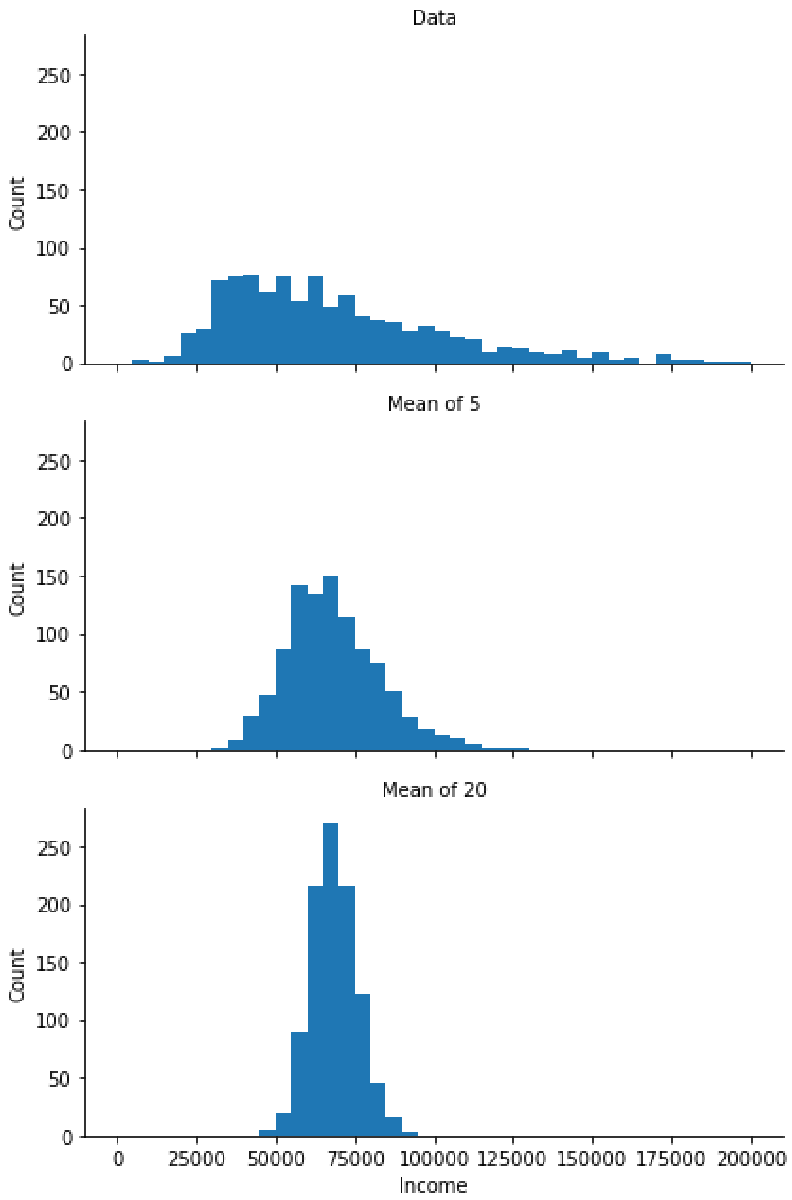
[7]:
```
LOANS_INCOME_CSV = '../data/loans_income.csv'
```

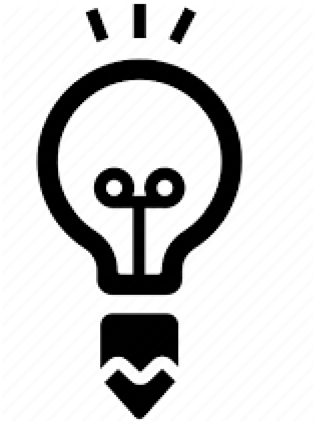
Παράδειγμα (συν)

Προσέγγιση Συναρτηση δειγματοληψίας στατιστικού

[12]:
```
loans_income = pd.read_csv(LOANS_INCOME_CSV, squeeze=True)
sample_data = pd.DataFrame({
    'income': loans_income.sample(1000),
    'type': 'Data',})
sample_mean_05 = pd.DataFrame({
    'income': [loans_income.sample(5).mean() for _ in range(1000)],
    'type': 'Mean of 5',})
sample_mean_20 = pd.DataFrame({
    'income': [loans_income.sample(20).mean() for _ in range(1000)],
    'type': 'Mean of 20',})
results = pd.concat([sample_data, sample_mean_05, sample_mean_20])
g = sns.FacetGrid(results, col='type', col_wrap=1,
                  height=5, aspect=2)
g.map(plt.hist, 'income', range=[0, 200000], bins=40)
g.set_axis_labels('Income', 'Count')
g.set_titles('{col_name}')
plt.tight_layout()
plt.show()
```

Παράδειγμα (συν)





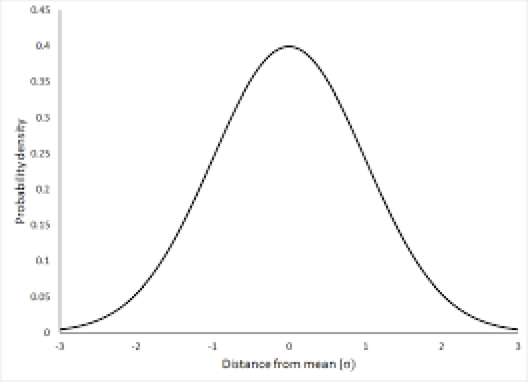
Κεντρικό Οριακό Θεώρημα (Central Limit Theorem)

- Αν $\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_{3,} \ldots, \mathrm{X}_\nu$ ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ίδια κατανομή με $\mathrm{E}(X_i) = \mu$ και $Var(X_i) = \sigma^2$ για $i = 1, \ldots, n$ για πολύ μεγάλα ν (θεωρητικά $\nu \to \infty$) ισχύει

$$\overline{X} = \frac{\sum_{i=1}^{\nu} X_i}{\nu} \sim N(\mu, \frac{\sigma^2}{\nu}) \quad \text{και} \quad S_\nu = \sum_{i=1}^{\nu} X_i \sim N(\nu\mu, \nu\sigma^2)$$

Πόσο μεγάλο πρέπει
να είναι το ν; Τουλάχιστον 30

Κανονική κατανομή



- Η συνάρτηση πυκνότητας της κανονικής κατανομής δίνεται από τον τύπο

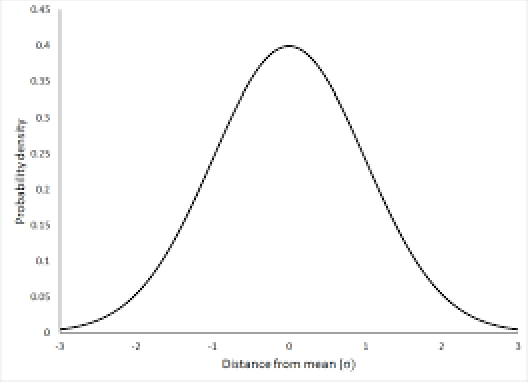
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < +\infty$$

όπου, σ > 0 η *τυπική απόκλιση* και μ ∈(−∞,+ ∞) η *μέση τιμή* της κατανομής

Ιδιότητες κανονικής κατανομής

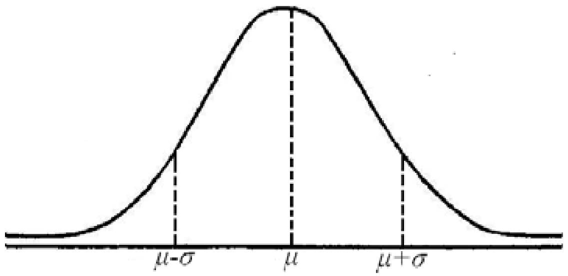
- Η κανονική καμπύλη είναι συμμετρική και οι «ουρές» της πλησιάζουν τον οριζόντιο άξονα ομαλά (ασυμπτωτικά).
- Η μέση τιμή και η διάμεσος ταυτίζονται.
- Η κορυφή ταυτίζεται με τη μέση τιμή και τη διάμεσο.
- Η περιοχή που παρουσιάζει τη μεγαλύτερη πυκνότητα, βρίσκεται και αυτή στο μέσο της κατανομής:
  - γύρω από τη μέση τιμή τους υπάρχουν σχετικά πολλές τιμές ενώ μακριά από τη μέση τιμή βρίσκονται σχετικά λίγες τιμές

Ιδιότητες κανονικής κατανομής



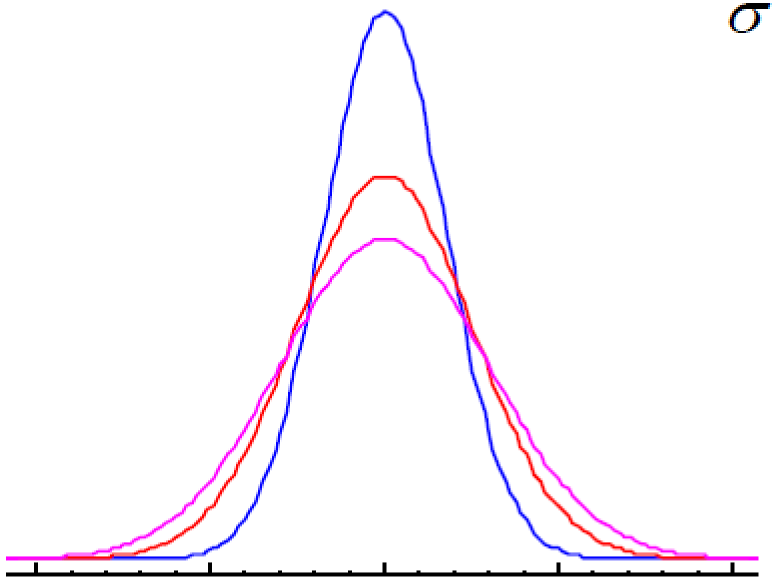
- Στη θέση x = μ παρουσιάζει μέγιστη τιμή, ίση με $\frac{1}{\sigma\sqrt{2\pi}} = \frac{0.399}{\sigma}$

- Στις θέσεις x = μ −σ και x = μ +σ παρουσιάζει σημεία καμπής

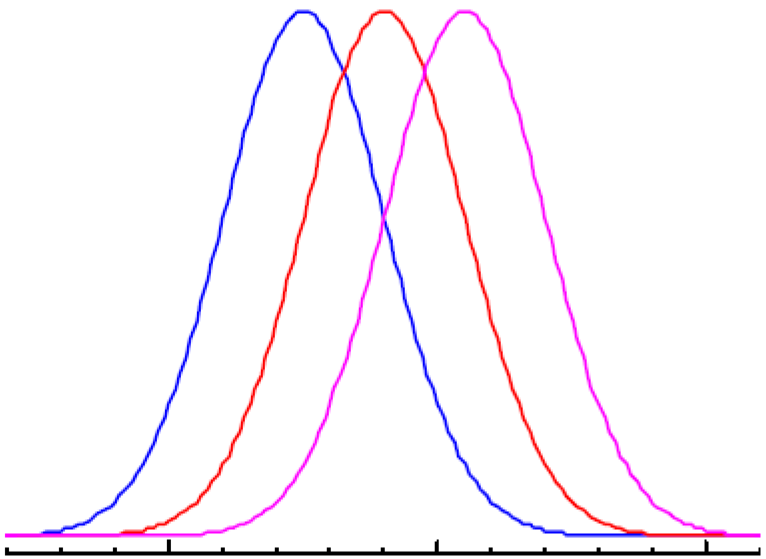


Ο ρόλος της μέσης τιμής και διακύμανσης

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



$$N(\mu,\sigma^2)$$

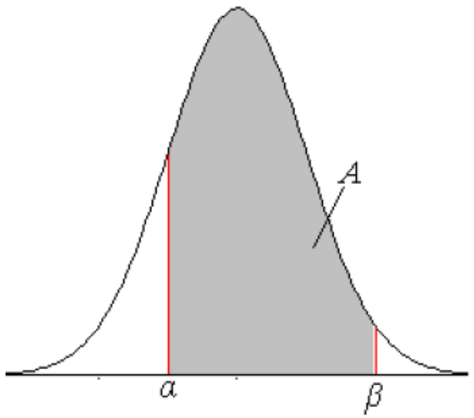


Ίδια μέση τιμή και διαφορετική τυπική απόκλιση

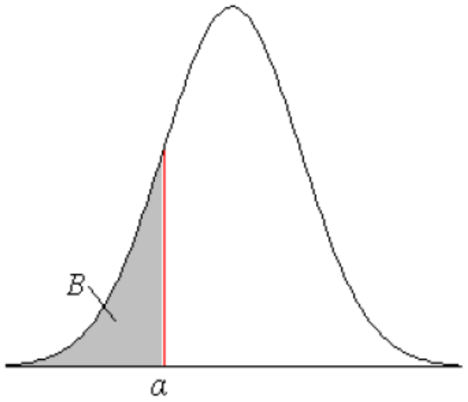
Ίδια τυπική απόκλιση και διαφορετική μέση τιμή

Όσο μικρότερη είναι η τυπική απόκλιση, τόσο ψηλότερη και τόσο πιο στενή είναι η κανονική καμπύλη,

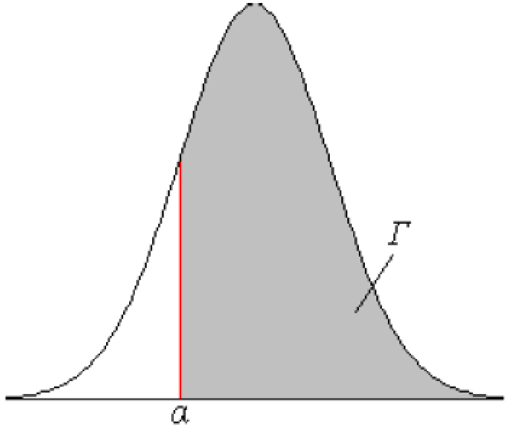
Υπολογισμός πιθανοτήτων



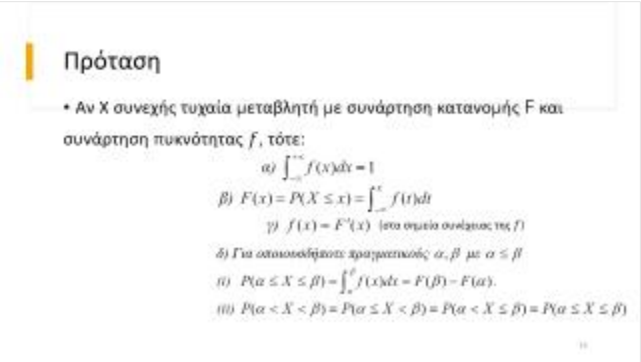
$$A = P(\alpha \leq X \leq \beta)$$



$$B = P(X \leq \alpha)$$



$$\Gamma = P(X \geq a)$$



Πρόταση

- Αν X συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής F και συνάρτηση πυκνότητας f, τότε:

$$\alpha)\ \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\beta)\ F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

$$\gamma)\ f(x) = F'(x) \quad \text{(στα σημεία συνέχειας της } f\text{)}$$

δ) Για οποιουσδήποτε πραγματικούς α, β με $\alpha \leq \beta$

$$(i)\ P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x)dx = F(\beta) - F(\alpha).$$

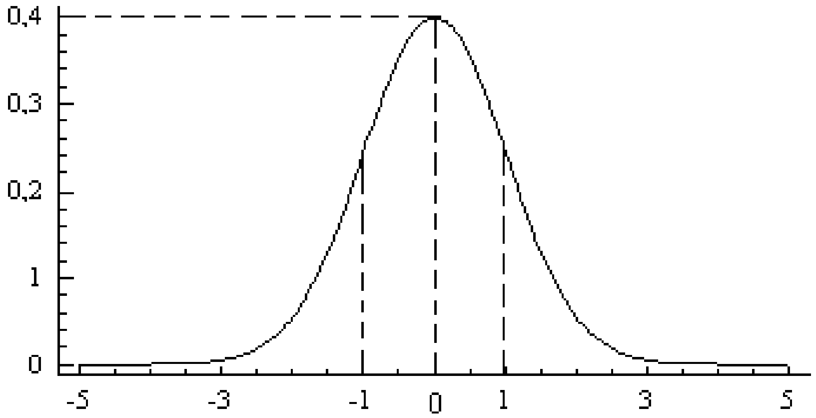
$$(ii)\ P(\alpha < X < \beta) = P(\alpha \leq X < \beta) = P(\alpha < X \leq \beta) = P(\alpha \leq X \leq \beta)$$

Τυποποιημένη κανονική κατανομή

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

- Είναι η κανονική κατανομή που έχει μέση τιμή 0 και τυπική απόκλιση 1 με συνάρτηση πυκνότητας

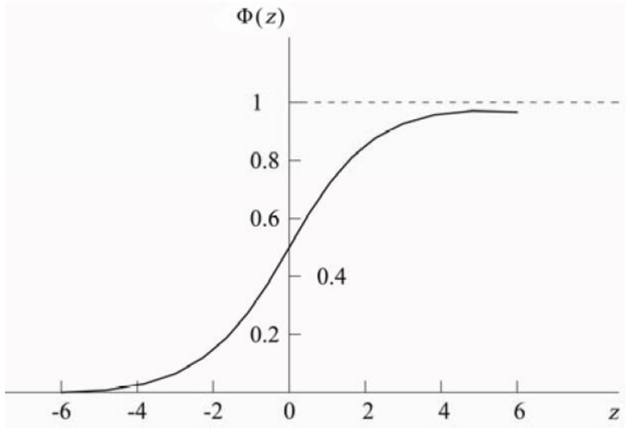
$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}, \quad -\infty < z < +\infty$$

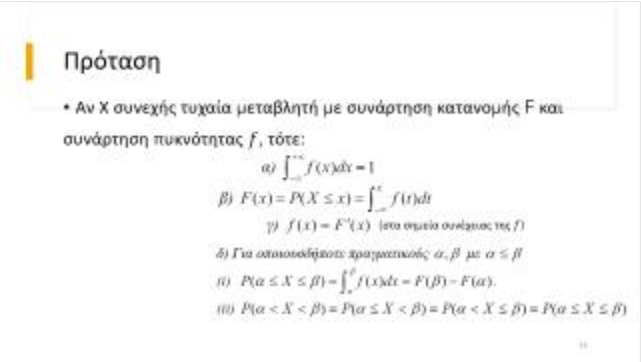


Τυποποιημένη κανονική κατανομή

- Αθροιστική συνάρτησης κατανομής

$$\Phi(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt, \ -\infty < z < +\infty$$





Πρόταση

- Αν X συνεχής τυχαία μεταβλητή με συνάρτηση κατανομής F και συνάρτηση πυκνότητας f, τότε:

$$\alpha)\ \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\beta)\ F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

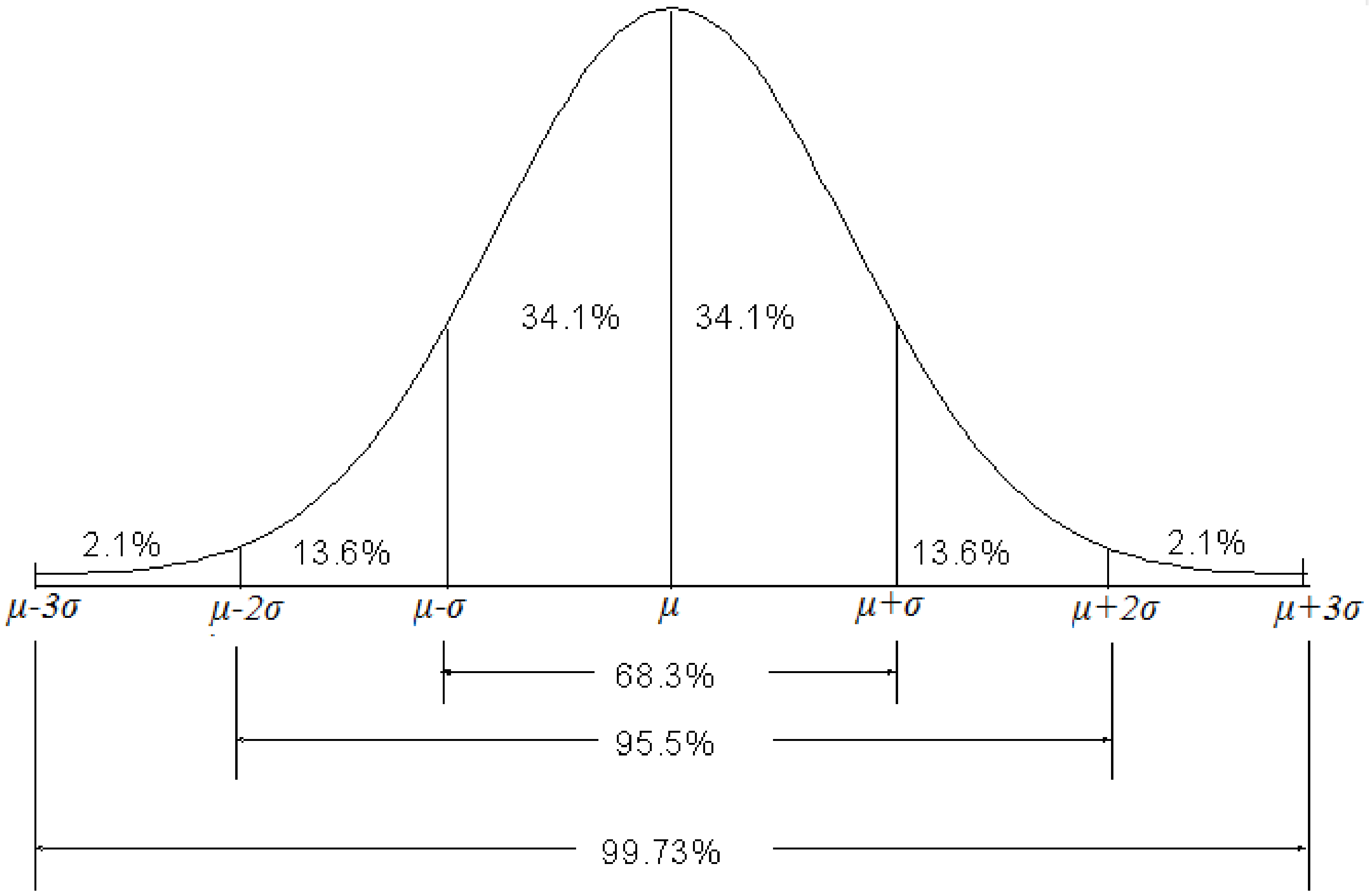
$$\gamma)\ f(x) = F'(x) \text{ (στα σημεία συνέχειας της } f)$$

δ) Για οποιουσδήποτε πραγματικούς α, β με $\alpha \le \beta$

$$(i)\ P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)dx = F(\beta) - F(\alpha).$$

$$(ii)\ P(\alpha < X < \beta) = P(\alpha \le X < \beta) = P(\alpha < X \le \beta) = P(\alpha \le X \le \beta)$$

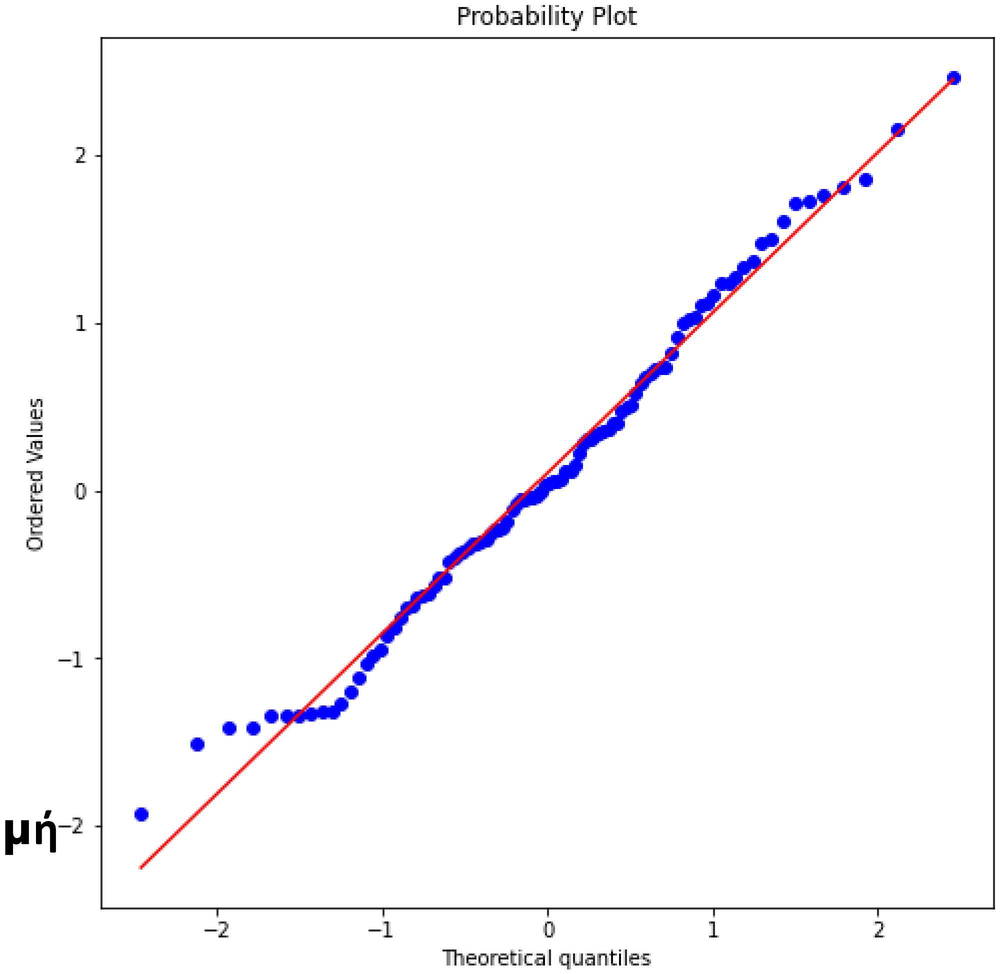
Πιθανότητες σε σχέση της τιμής του σ



$$P(\mu - \sigma \leq X \leq \mu + \sigma) = 2\Phi(1) - 1 = 0.6826 \cong 68.3\%$$
$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 2\Phi(2) - 1 = 0.9544 \cong 95.5\%$$
$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 2\Phi(3) - 1 = 0.9974 \cong 99.7\%$$

Εξομοιώνοντας την δειγματοληψία από την κανονικη κατανομή

```
18]: fig, ax = plt.subplots(figsize=(8, 8))
     norm_sample = stats.norm.rvs(size=100)
     stats.probplot(norm_sample, plot=ax)
```

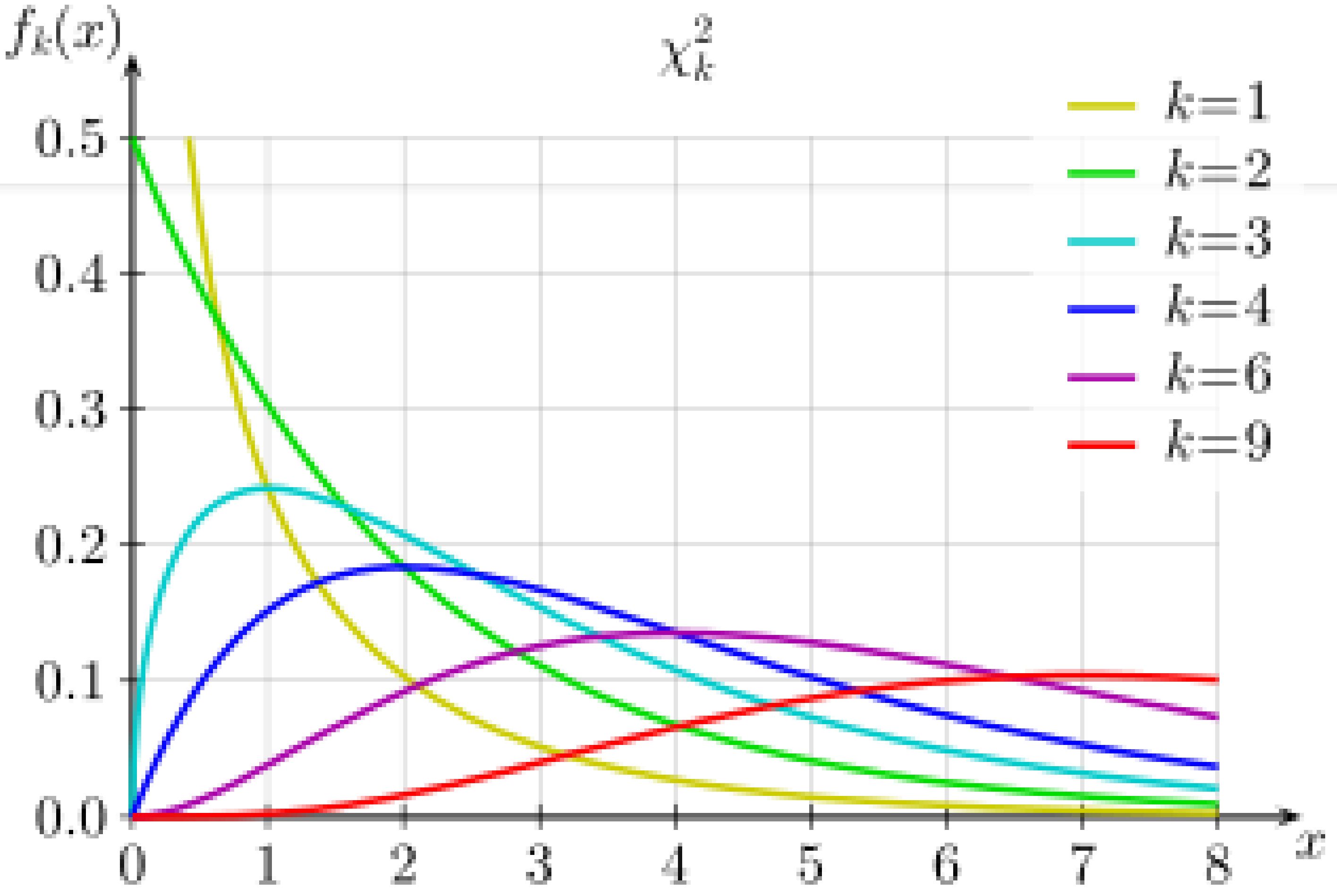


Ενώ τα δεδομένα δεν ακολουθούν πάντα την κανονική κατανομή τα σφάλματα συνήθως την ακολουθούν ως μέσοι οροί και συνολικά αποτελέσματα μετρήσεων σε μεγάλα δείγματα

Κατανομή χ^2

- Αν $Z_1, Z_2, Z_{3,} \ldots, Z_\nu$ είναι τυποποιημένες κανονικές μεταβλητές με $Z_i \sim N(0,1)$ τότε η κατανομή της τμ $X = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$ ακολουθεί την κατανομή χ-τετράγωνο με ν βαθμούς ελευθερίας (chi-square distribution)
- E(X)=ν και Var(X)=2ν

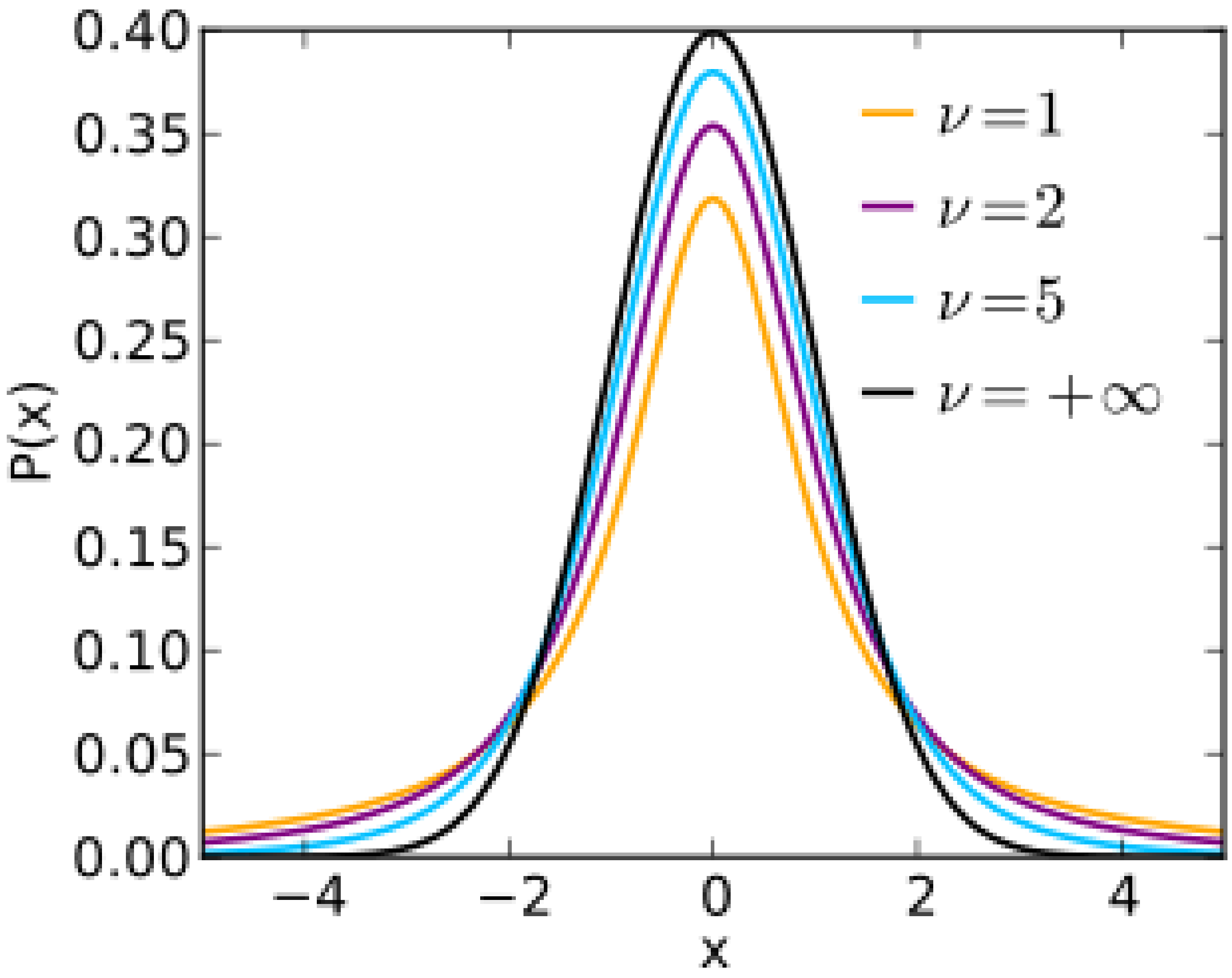
Κατανομή χ^2



Κατανομή t (student)

- Έστω Z τμ η οποία ακολουθεί την τυποποιημένη κατανομή (Z~N(0,1)) και S_ν μια ανεξάρτητη τυχαία μεταβλητή που ακολουθεί την χ-τετράγωνο κατανομή με ν βαθμούς ελευθερίας.
- Η τμ $\mathrm{T} = \frac{\mathrm{Z}}{\sqrt{S_\nu/\nu}}$ ακολουθεί την κατανομή t με ν βαθμούς ελευθερίας (t distribution, student t-distribution) και συμβολίζεται με $T{\sim}t_\nu$.
- E(T)=0 για ν>0 και Var(T)=ν/(ν-2) για ν>2

Κατανομή t (student)



Κατανομή Bernoulli

- Πείραμα τύχης με μόνο δύο, αμοιβαίως αποκλειόμενα, δυνατά αποτελέσματα.
- Συνήθως το ένα αποτέλεσμα ονομάζεται επιτυχία (success) και το άλλο αποτυχία (failure).
- Σε μια *δοκιμή Bernoulli* ο αριθμός των επιτυχιών προφανώς είναι ή μια ή καμία.
- Αν η πιθανότητα επιτυχίας είναι p και η πιθανότητα αποτυχίας είναι q, προφανώς $p + q = 1$ ή $q = 1 - p$

Κατανομή Bernoulli

- Έστω X η διακριτή τυχαία μεταβλητή που εκφράζει τον αριθμό των επιτυχιών σε μια δοκιμή Bernoulli με πιθανότητα επιτυχίας p. Η κατανομή της X ονομάζεται κατανομή Bernoulli (Bernoulli distribution) με παράμετρο p και συμβολίζεται με X~b(p) .
- Συνάρτηση πιθανότητας

$$f(x) = P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$ ή $$f(x) = P(X = x) = p^x (1 - p)^{1-x}, \ x = 0, 1.$$

Μέση τιμή και διακύμανση

- Η μέση τιμή μ της X είναι

$$\mu = E(X) = 0 \cdot f(0) + 1 \cdot f(1) = p$$

- Η διακύμανση σ^2

$$\sigma^2 = E(X^2) - [E(X)]^2 = 0^2 \cdot f(0) + 1^2 \cdot f(1) - p^2 = p - p^2 = p(1-p) = pq$$

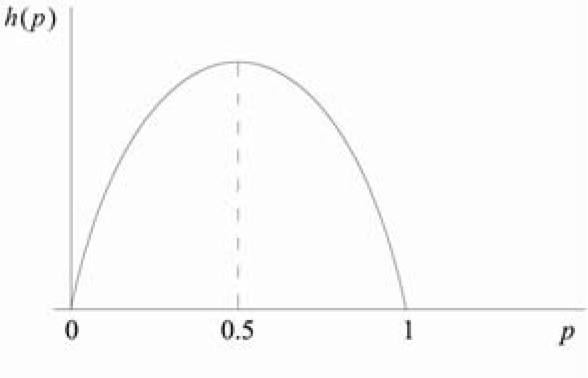
Διωνυμική κατανομή

- Αν X ο αριθμός των επιτυχιών σε μια ακολουθία ν ανεξάρτητων δοκιμών Bernoulli με σταθερή πιθανότητα επιτυχίας p σε όλες τις δοκιμές, τότε η κατανομή της τυχαίας μεταβλητής X ονομάζεται διωνυμική κατανομή (binomial distribution) με παραμέτρους ν και p και συμβολίζεται με B(ν , p) ή X~ B(ν , p).
- Η συνάρτηση πιθανότητας είναι:

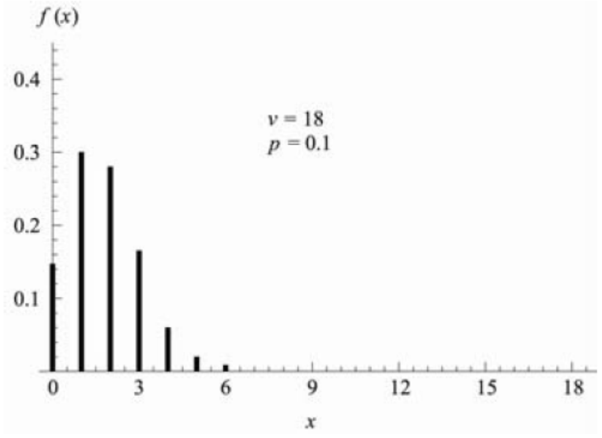
$$f(x) = P(X = x) = \binom{\nu}{x} p^x (1-p)^{\nu - x}, \quad x = 0,1,2,\ldots,\nu$$

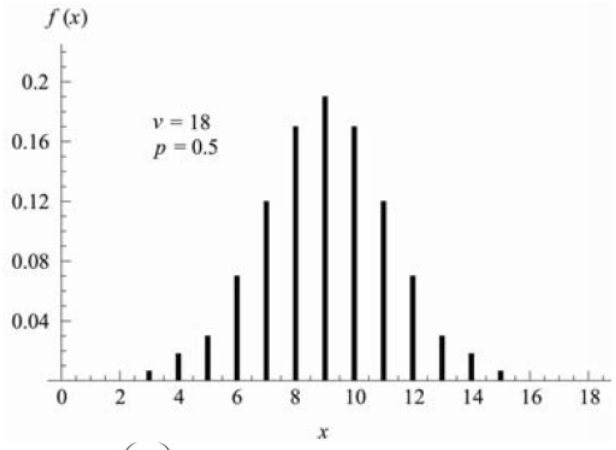
$$(\alpha+\beta)^\nu = \sum_{x=0}^{\nu} \binom{\nu}{x} \alpha^x \beta^{\nu-x}, \quad \nu = 1,2,\ldots$$

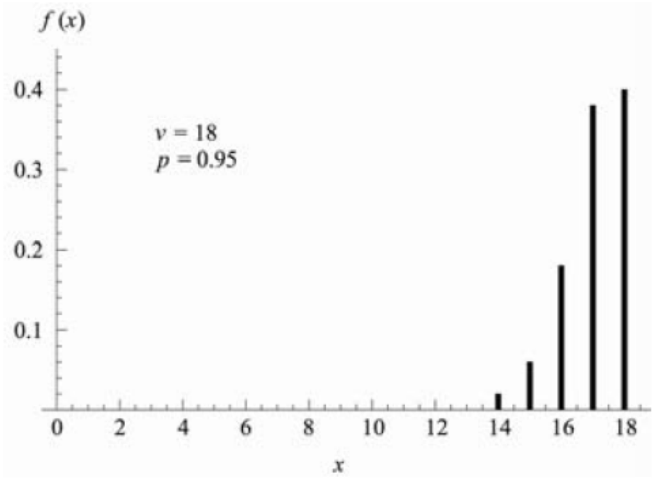
Μέση τιμή και η διακύμανση



- Η μέση τιμή και η διακύμανση είναι: $\mu = \nu p$ και $\sigma^2 = \nu p(1-p) = \nu pq$
- *Η διακύμανση ελαττώνεται όσο το p πλησιάζει το 0 ή το 1 και μεγιστοποιείται για p* = 1/2







$$f(x) = P(X = x) = \binom{\nu}{x} p^x (1-p)^{\nu - x}, \quad x = 0,1,2,\ldots,\nu$$

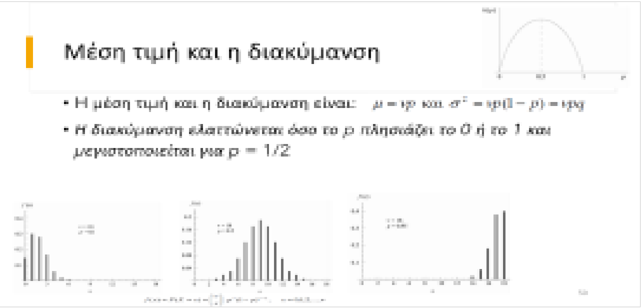
Οριακό θεώρημα Poisson

- Έστω ότι η τυχαία μεταβλητή Χ ακολουθεί τη διωνυμική κατανομή Β(ν , p) με συνάρτηση πιθανότητας

$$P(X=x)=\binom{\nu}{x}\ p^{x}(1-p)^{\nu-x}, \qquad x=0,1,2,\ldots,\nu$$

Αν για ν ⟶ +∞ το p ⟶0 έτσι ώστε η μέση τιμή της Χ να συγκλίνει προς μια θετική σταθερά λ , δηλαδή, έτσι ώστε νp ⟶λ , τότε

$$\lim_{\nu\to+\infty}\binom{\nu}{x}\ p^{x}(1-p)^{\nu-x}=e^{-\lambda}\frac{\lambda^{x}}{x!}, \quad x=0,1,2,\ldots$$



Η κατανομή Poisson

- Έστω X μια διακριτή τυχαία μεταβλητή με συνάρτηση πιθανότητας

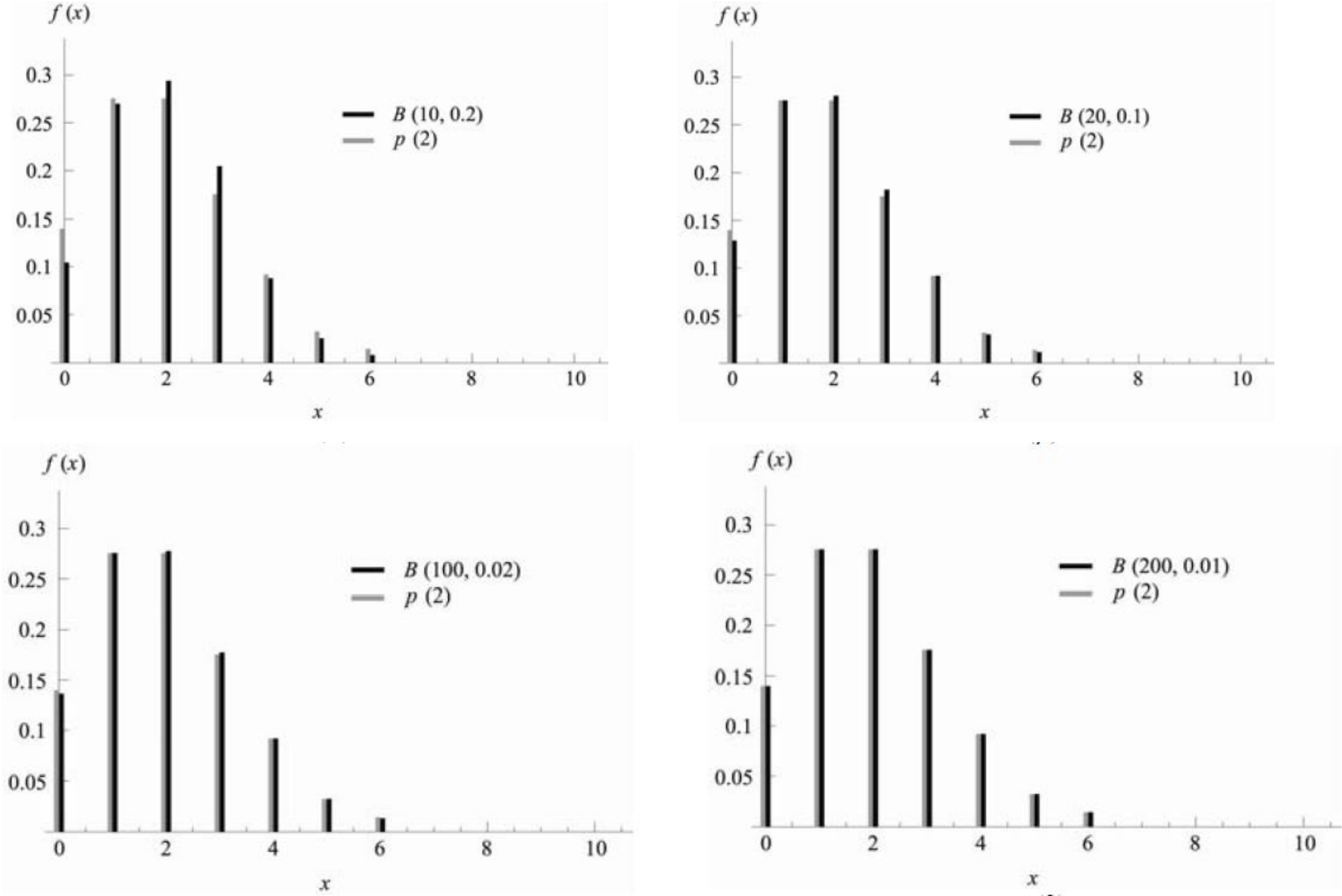
$$f(x) = P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0,1,2,\ldots$$

  όπου λ > 0 . Η κατανομή της τυχαίας μεταβλητής X ονομάζεται κατανομή Poisson με παράμετρο λ και συμβολίζεται με P(λ).

- Η **διωνυμική κατανομή προσεγγίζεται από την κατανομή Poisson** αν για ν μεγάλο (θεωρητικά $\nu \to \infty$), η πιθανότητα επιτυχίας p συγκλίνει στο 0 (δηλ. $p \to 0$) έτσι ώστε η μέση τιμή της να συγκλίνει σε μια θετική σταθερά λ ($\nu p \to \lambda$).

Σχόλια

- Πόσο μεγάλο πρέπει να είναι το ν και πόσο μικρό το p;
- ν ≥ 20 και p≤10/ν
- λ =νp = 2 όταν ν = 10,20,100,200 καιp = 0.2, 0.1, 0.02, 0.01 αντίστοιχα



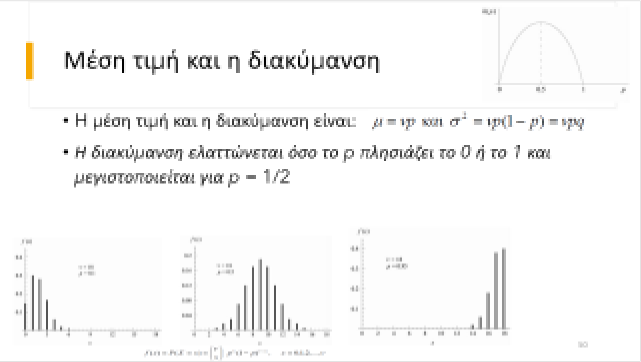
Μέση τιμή και διακύμανση

- Η μέση τιμή και η διακύμανση της κατανομής Poisson με παράμερο λ , αντίστοιχα είναι

$$\mu = E(X) = \lambda$$

και

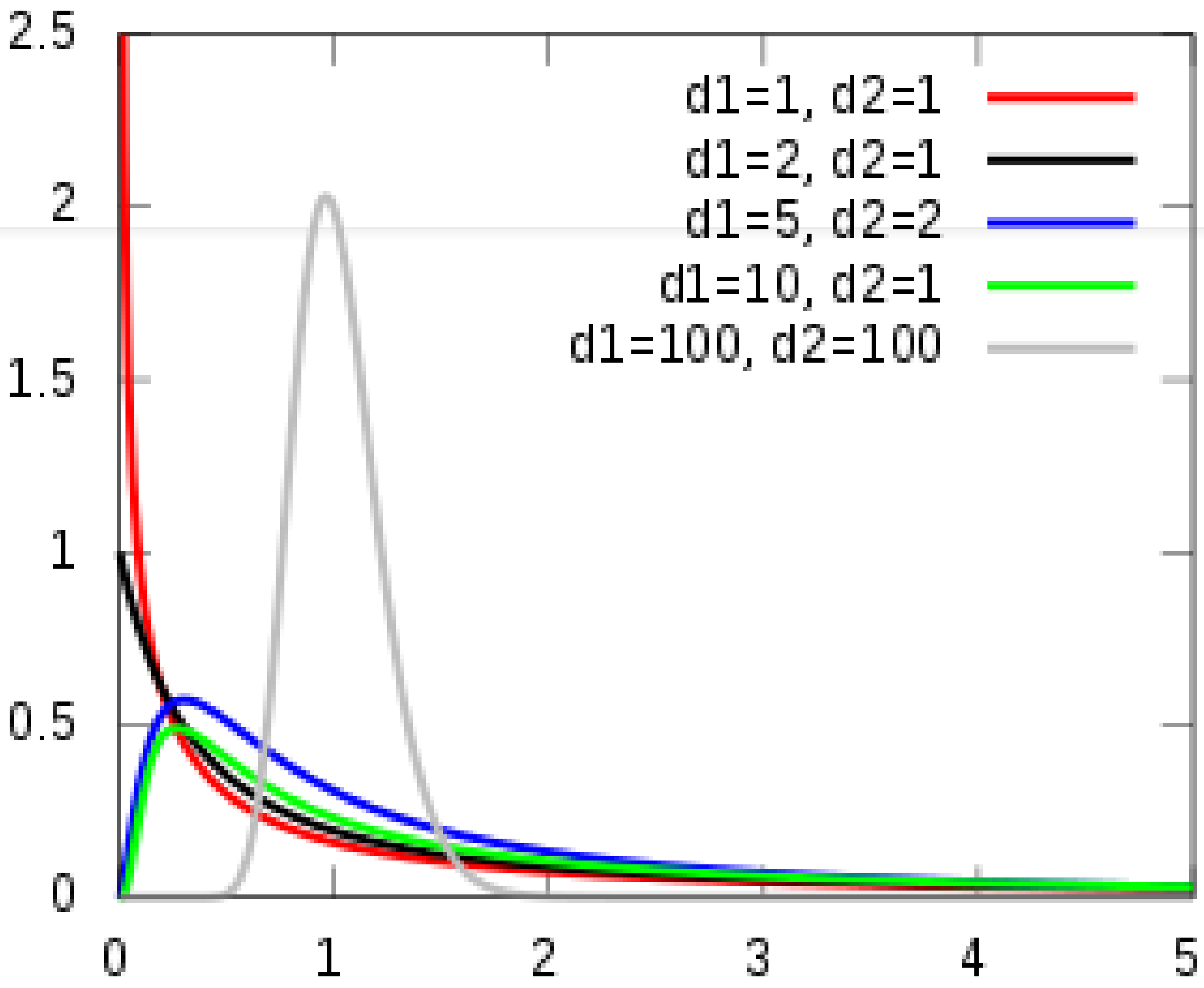
$$\sigma^2 = Var(X) = \lambda$$



Κατανομή F

- Έστω S_ν και S_m δύο ανεξάρτητες τμ που ακολουθούν την κατανομή χ-τετράγωνο με ν και m βαθμούς ελευθερίας αντίστοιχα.
- Η τμ $\mathrm{F} = \frac{S_\nu/\nu}{S_m/m} = \frac{m}{n}\frac{S_\nu}{S_m}$ ακολουθεί την κατανομή F με ν και m βαθμούς ελευθερίας (F-distribution, $F_{\nu,m}$)
- $E(F) = \frac{m}{m-2}$ για m>2 και $Var(F) = \frac{2m^2(\nu+m-2)}{\nu(m-2)^2(m-4)}$ για m>4.

Κατανομή F



d1=1, d2=1
d1=2, d2=1
d1=5, d2=2
d1=10, d2=1
d1=100, d2=100
2.5
2
1.5
1
0.5
0
0
1
2
3
4
5