Υπολογιστική Φυσική Στοιχειωδών Σωματιδίων

4ο Εξάμηνο 2005-2006

Στατιστική
Πιθανότητες-Κατανομές

Διδάσκοντες : Χαρά Πετρίδου

Δήμος Σαμψωνίδης

Πιθανότητες

- Ορισμός 1 -Μαθηματικός: P(A) είναι αριθμός που πληρεί τα αξιώματα του Kolmogorov

$$\boxed{\begin{aligned}&P(A) \geq 0\\&P(A_1 \vee A_2) = P(A_1) + P(A_2)\\&\sum P(A_i) = 1\end{aligned}}$$

- ΔΕΝ δινει πληροφορία ο μαθηματικός ορισμός!

Πιθανότητες

- Ορισμός 2 : Κλασσικός ορισμός

  Η πιθανότητα $P(A)$ είναι η ιδιότητα ενός αντικειμένου που καθορίζει πόσο συχνά συμβαίνει ένα γεγονός A. Δίνεται από την συμμετρία για εξίσου πιθανά γεγονότα. Γεγονότα όχι το ίδιο πιθανά τα θεωρουμε εξισου πιθανα

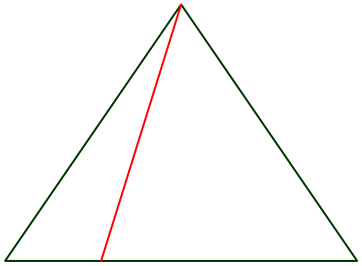
Παράδειγμα : ρίχνουμε ένα νομισμα : P(H) = 1/2

Ρίχνουμε δύο ζάρια : P(8) = 5/36 ? ή P(8) = 6/36 ?

Προβήματα με τον κλασσικό ορισμό

1. Ποτε πρόκειται για εξισου πιθανές περιπτώσεις? .... Αν ρίξουμε δύο νομίσματα έχουμε 3 ή 4 ίσες δυνατότητες?
2. Τι κανουμε σε περιπτώσεις που έχουμε συνεχείς μεταβλητές? .... Πως να χωρίσουμε ένα τρίγωνο σε δύο τυχαία μέρη?

Έχουμε λύση



Δεν Έχουμε λύση

Πιθανότητες - Συχνότητα

- Ορισμός 3 -Frequentist (‘συχνάζουσα’ πιθανότητα)

  Η πιθανότητα $P(A)$ είναι το όριο του λόγου N(A)/N (σε ένα σύνολο N με ιδιότητα A)

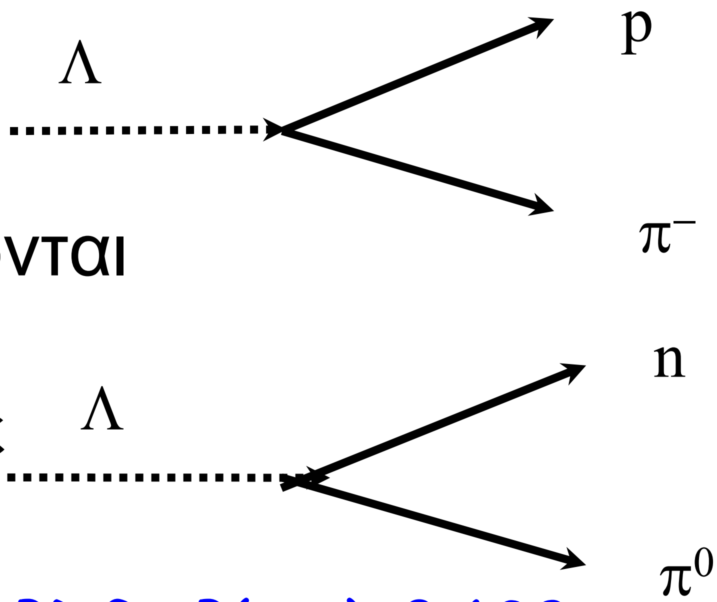
$$P(A) \underset{N\to\infty}{=} {N(A)}\big/{N}$$

Πρόβλημα (περιορισμός) με τον ορισμό της ‘συχνάζουσας’ πιθανότητας

- Η $P(A)$ εξαρτάται και απο το Α και απο το σύνολο Ν. Η πιθανότητα που υπολογίζουμε μπορεί να είναι παραπλανητική αν βασιστούμε στην συχνότητα του γεγονότος : … 10 στους 30 ανθρώπους σε ένα σύνολο είναι ‘πλούσιοι’ => P(πλούσιος) = 1/3 !!!

Συνέπειες για την Κβαντομηχανική (QM)

- Στην QM υπολογίζουμε πιθανότητες
- Οι πιθανότητες εξαρτώνται απο το είδος των γεγονότων που έχουμε (π.χ. Λ^0) και την αλληλεπίδραση



PDG: $P(p\pi^-)=0.639$

$P(n\pi^0)=0.358$

Το πρόβλημα με τον ορισμό της ‘συχνάζουσας’ πιθανότητας

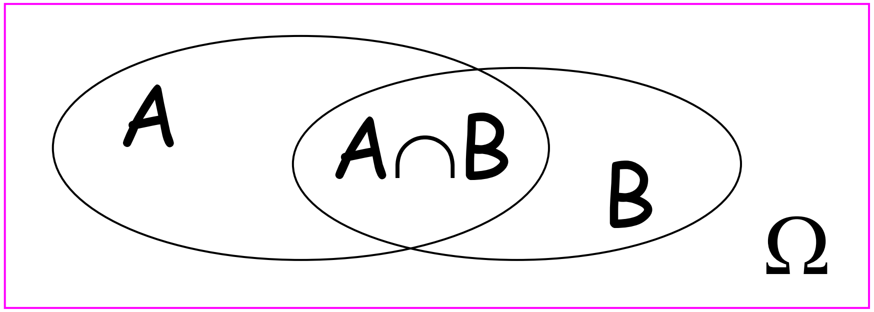
- ΔΕΝ μπορεί να εφαρμοστεί σε μεμονωμένα γεγονότα !
  - … η έκφραση ‘πιθανόν να βρέξει αύριο’ ΔΕΝ είναι επιστημονική !!!

    Ενω : ‘η δήλωση “ πιθανόν να βρέξει αύριο” είναι σωστή’ είναι
  - Η ανακάλυψη ή όχι του Higgs
  - Η ανακάλυψη ή όχι σκοτεινής ύλης

Ορισμός της υποκειμενικής πιθανότητας (Bayesian)

- *P(A)* είναι η πιθανότητα που δίνουμε στην ιδιότητα Α να υπάρξει.
- Το Α μπορει να είναι ο,τιδήποτε. Νεα θεωρία, νέο σωματίδιο, δίασπαση σωματιδίου,… βροχή, … ενα στοίχημα…

Το Θεώρημα του Bayes



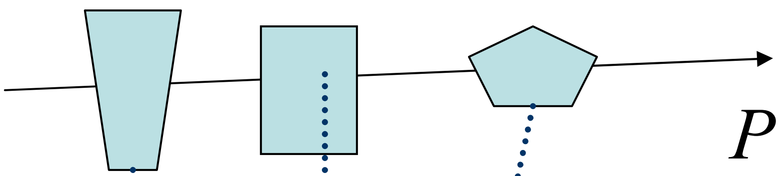
- Υπό όρους πιθανότητα (conditional probability) : P(A|B) είναι η πιθανότητα των κοινών σημείων των A και B σε σχέση με το σύνολο B και η P(A∩B) σε σχέση με το Ω
- P(A∩B) = P(A|B)*P(B)
  = P(B|A)*P(A) =>

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A)$$

Γνωρίζουμε την πιθανότητα να συμβεί B όταν συμβαίνει A και υπολογίζουμε την αντίστροφη πιθανότητα

Εφαρμογή του Θεωρήματος Bayes για την συχνάζουσα πιθανότητα

- Ταυτοποίηση σωματιδίων
- -Τύποι σωματιδίων :e, π, μ, K, p
- -Σήματα απο ανιχνευτές: DCH, RICH, TOF, TRD

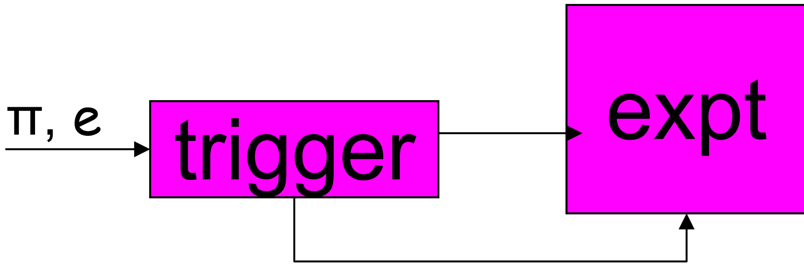


$$P'(e) = P(e \mid DCH) = \frac{P(DCH \mid e)}{P(DCH)} P(e)$$

$$P(DCH) = P(DCH \mid e)P(e) + P(DCH \mid \mu)P(\mu) + P(DCH \mid \pi)P(\pi)\ldots$$

- Αντίστοιχα υπολογίζουμε την P(e|RICH) κλπ (δηλ. σήμα στον ανιχνευτή να οφείλεται σε e)

Παράδειγμα χρήσης του θεωρήματος Bayes με την συχνάζουσα πιθανότητα



- Δεσμη αποτελείται από π’ς & e’ς

  $N_e/N_\pi=10^{-3}$
- Το σύστημα σκανδαλισμού:

  $\varepsilon_e=0.98$, $\varepsilon_\pi=0.03$

  $P(e)=10^{-3}$ $P(\pi)=1-10^{-3}$

  $P(T|e)=0.98$, $P(T|\pi)=0.03$
- $P(e|T)=P(T|e)*P(e)/(P(T|e)*P(e)+P(T|\pi)P(\pi))$ είναι η πιθανότητα το εισερχόμενο σωμάτιο να είναι ηλεκτρόνιο όταν το σύστημα σκανδαλισμού λέει ‘ναι’

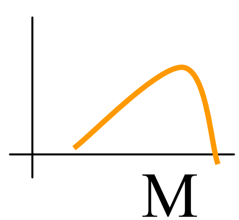
Το θεώρημα Bayes και η υποκειμενική πιθανότητα

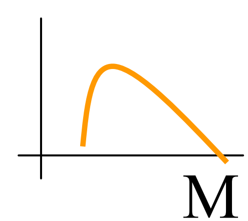
$$P(Theory|Result) = \frac{P(Result|Theory)}{P(Result)} P(Theory)$$

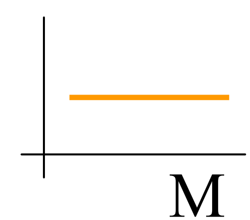
- Η εκ των προτέρων 'πίστη' σε μια θεωρία αλλάζει με βάση τα πειραματικά αποτελέσματα :
- Εαν (P(Result|Theory) = 0, (η Θεωρία απαγορεύει το πειραματικό αποτέλεσμα) =>η Θεωρία δεν ισχύει
- Εαν (P(Result|Theory) = Μεγάλο, ενισχύεται η πίστη μας στη Θεωρία

Πρόβλημα με την υποκειμενική πιθανότητα

- Η πιθανότητα 'μου' $P(A)$ και η πιθανότητα $P(A)$ ενος τρίτου μπορει να διαφέρουν…
- Αλλα οι επιστήμονες θάπρεπε να ειναι αντικειμενικοί!
- Δικαιολογείται η άγνοια να δηλώνεται με $P(A)$=flat? (αν το A παιρνει διακριτές τιμές το ίδιο πιθανές, ναι. Αν συνεχής μεταβλητή, π.χ. M_{higgs} διαφορετική υπόθεση θα δώσει άλλα αποτελέσματα)







Συνάρτηση Πυκνότητας Πιθανότητας-Κατανομές

- Ορίσαμε τη Πιθανότητα
- Όταν μετρούμε ένα παρατηρήσιμο μέγεθος πολλές φορές, το αποτέλεσμα x κατανέμεται σύμφωνα με μιά κατανομή πιθανότητας (probability distribution)
- Το αποτέλεσμα έχει μια διακύμανση είτε λόγω καποιων παραμέτρων (π.χ. ηλετρονικό θόρυβο) είτε λόγω διακύμανσης του φυσικού φαινομένου (π.χ. διάσπαση ή σκέδαση σωματιδίων)
- Οι κατανομές πιθανότητας μπορεί να είναι διακριτές ή συνεχείς.

Πιθανότητα vs Στατιστικής

- Πιθανότητα: απο την θεωρία στα δεδομένα
  - Υπολογίζουμε όλα τα πιθανά αποτελέσματα (συνέπειες) ενός πειράματος για συγκεκριμένο πρόβλημα
- Στατιστική: απο τα δεδομένα στη θεωρία
  - Λύνουμε το αντίστροφο πρόβλημα: απο τα δεδομένα προσπαθούμε να βρούμε τους κανόνες-νόμους=> ανάλυση των δεδομένων
  - => Υπολογισμός παραμέτρων και του σφάλματος
    => έλεγχος της υπόθεσης: πιστότητα, συμφωνία…

Συναρτήσεις Πυκνότητας Πιθανότητας

- Επανειλημένη μέτρηση -πειραματικά- μιας συνεχούς μεταβλητής x
- Ορισμός : η πιθανότητα P να μετρήσουμε μια τιμή x στο διάστημα (x, x+dx) δίνεται από την συνάρτηση πυκνότητας πιθανότητας : probability density function f(x) (p.d.f.) : P=f(x) dx
  - Είναι το μέτρο του πόσο συχνά η τιμή x εμφανίζεται σένα δείγμα

$$Norm = \int_{x\min}^{x\max} f(x') \cdot dx' = 1$$

Κατανομές-ιδιότητες

- Αναμενόμενη τιμή =Μέση τιμή

$$E[x] \equiv \int_{x\min}^{x\max} x \cdot f(x) \cdot dx = \langle x \rangle = \mu$$

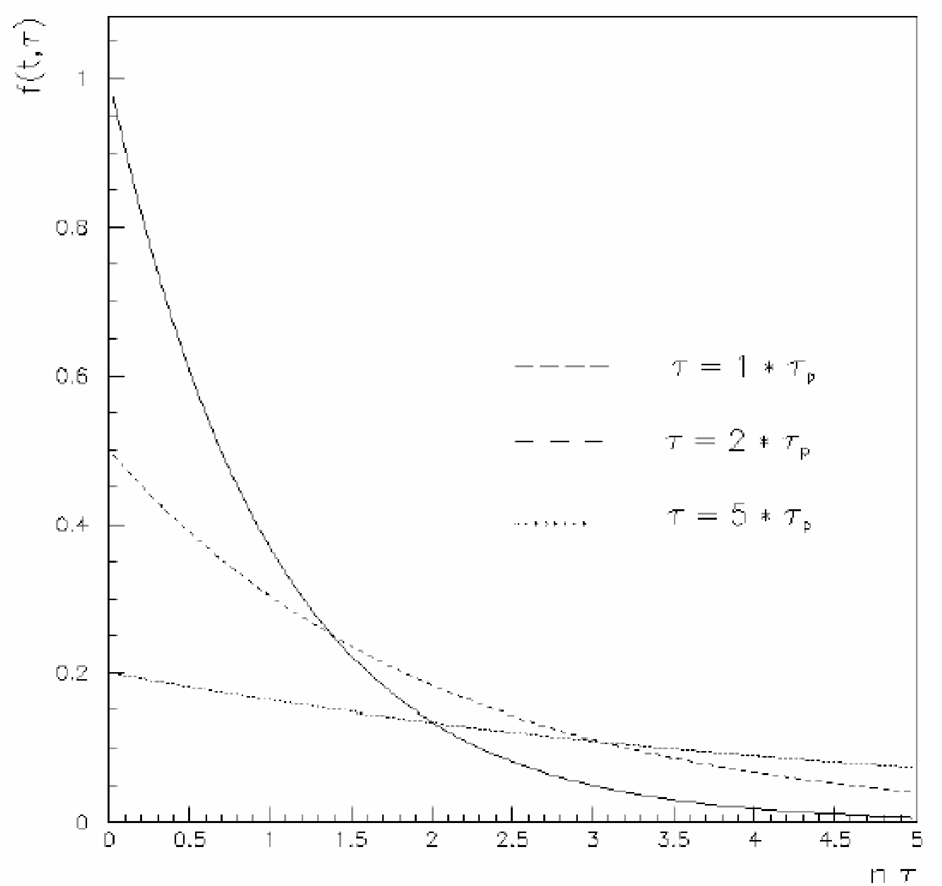
- Διασπορά -Variance = σ² =το τετραγωνο της απόκλισης σ Μετράει την διασπορά του x σε σχέση με την μέση τιμή

$$V[x] = E((x-\mu)^2) \equiv \int_{x\min}^{x\max} (x-\mu)^2 f(x) \cdot dx = \sigma^2$$

$$\equiv <(x-\mu)^2> \equiv <x^2> - \mu^2$$

Παραδείγματα

- Ο χρόνος ζωής t σωματιδίου στο σύστημα ηρεμίας του π.χ. Το πιόνιο : μέσος χρόνος ζωής $\tau_\pi = 2.6\ 10^{-8}$ sec

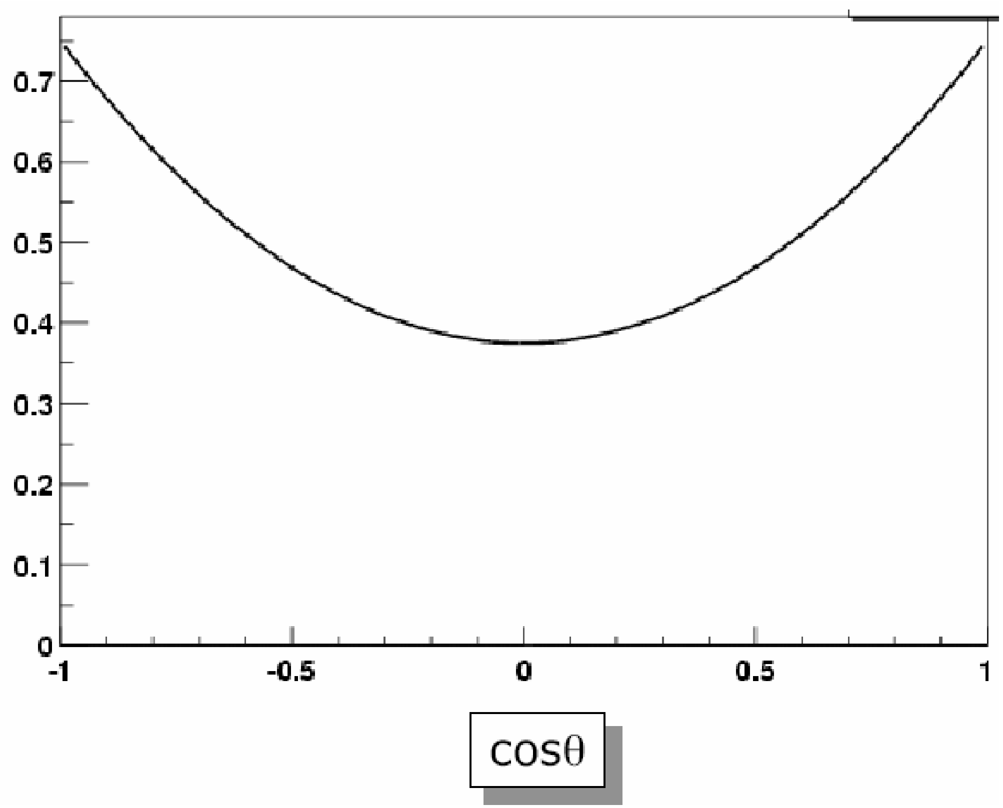


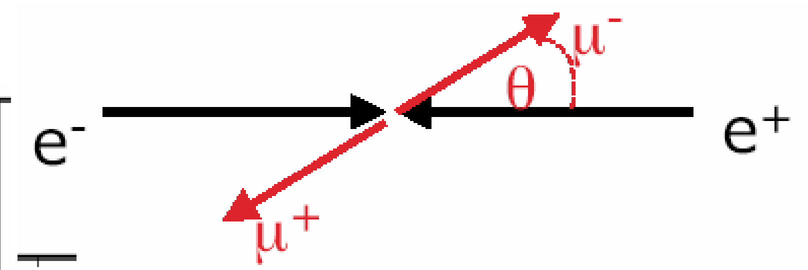
$$f(t;\ \tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

$$E[t] = \tau$$
$$V[t] = \tau^2$$

Παραδείγματα

- Κατανομή της πολικής γωνίας θ του μιονίου στην σκέδαση ee→μμ





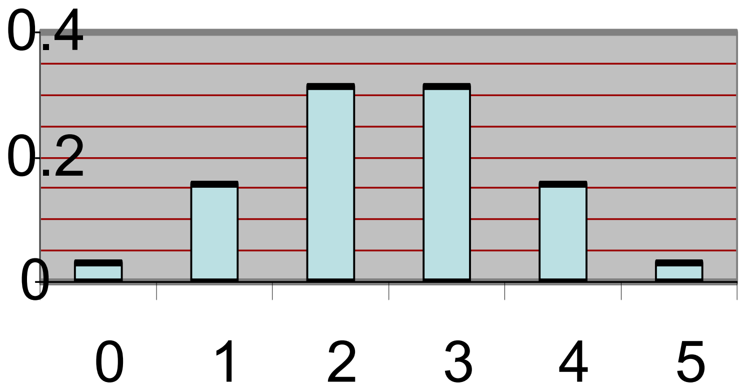
$$\frac{1}{\sigma_{tot}}\frac{d\sigma}{d\cos\theta} = $$

$$\equiv f(\cos\theta) = \frac{3}{8}\left[1+(\cos\theta)^2\right]$$

Κατανομές-Διακριτές κατανομές

- Διώνυμη κατανομή: η τυχαία μεταβλητή εχει δύο δυνατότητες

n προσπάθειες και r επιτυχίες p η μεμονωμένη πιθανότητα για επιτυχία



$$P(r\ ;n,p)=\frac{n!}{r!(n-r)!}p^{r}(1-p)^{n-r}$$

Mean

$\mu=\langle r\rangle=\sum rP(r)$

$= np$

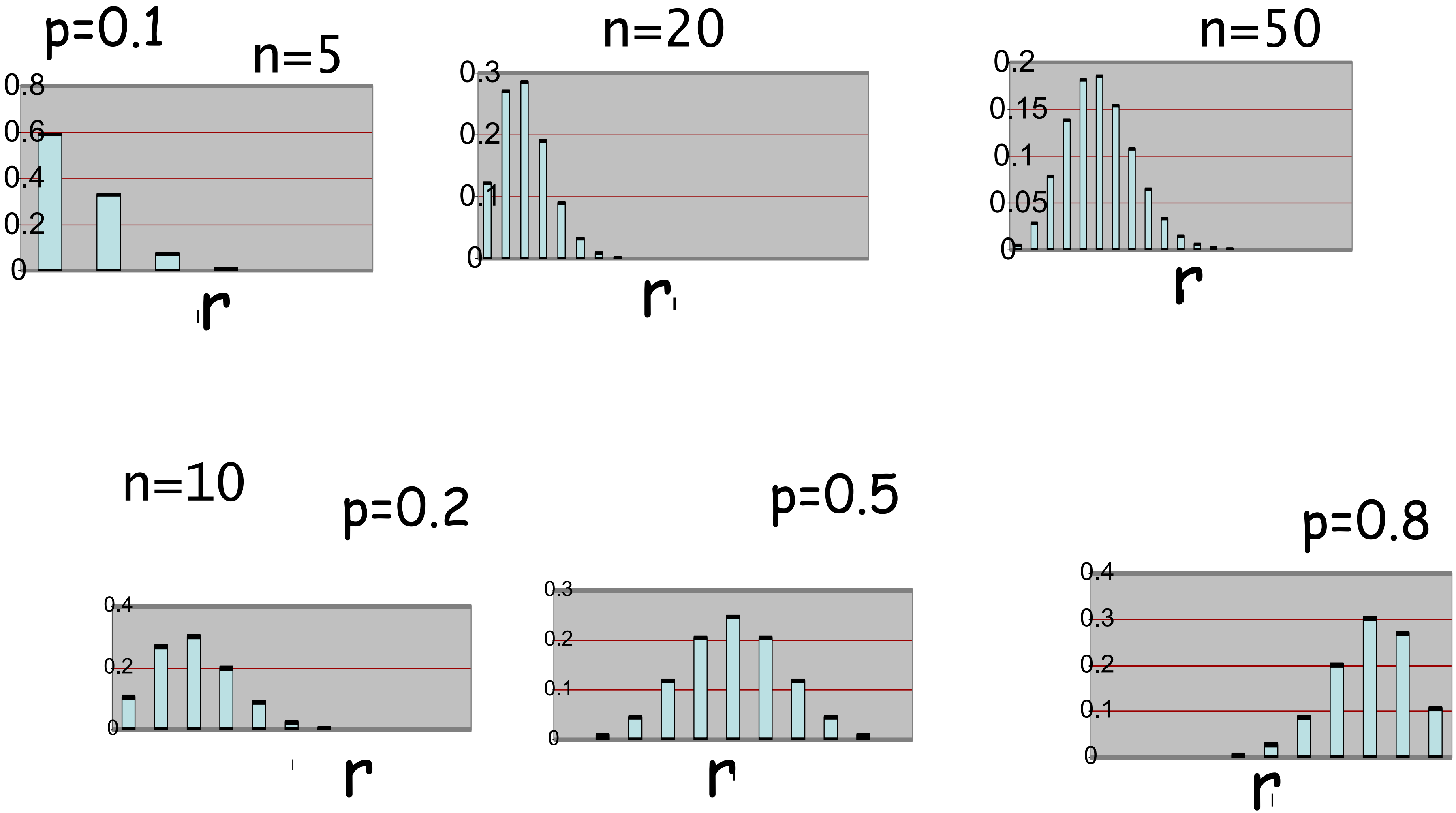
Variance

$V\equiv\sigma^2=\langle(r-\mu)^2\rangle=\langle r^2\rangle-\langle r\rangle^2$

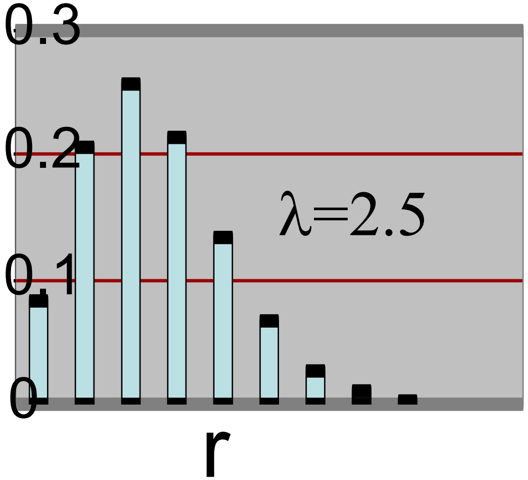
$=np(1-p)$

$1\text{-}p \equiv \bar{p} \equiv q$

Παραδείγματα διώνυμης κατανομής



Κατανομή Poisson



- Δίνει την πιθανότητα να συμβούν r γεγονότα όταν ο ρυθμός των γεγονότων κατά μέσον όρο είναι λ (είναι το όριο της διωνυμικής κατανομής όταν n→∞ p→0, np=λ)

$$P(r;\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$$

Mean

$\mu = \langle r \rangle = \sum rP(r)$

$= \lambda$

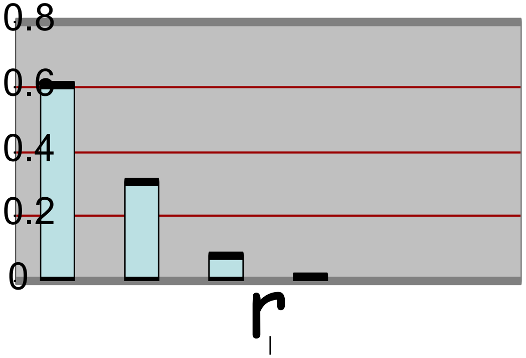
Variance

$V \equiv \sigma^2 = \langle (r-\mu)^2 \rangle = \langle r^2 \rangle - \langle r \rangle^2$

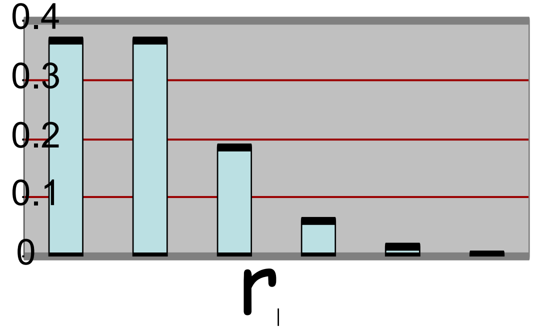
$= \lambda$

Παραδείγματα κατανομής Poisson

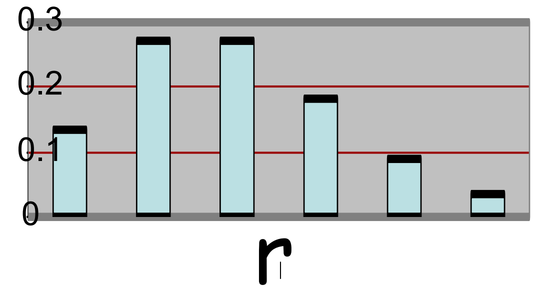
$\lambda=0.5$



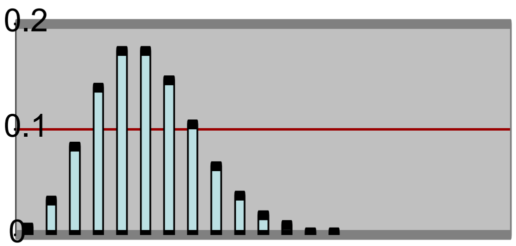
$\lambda=1.0$



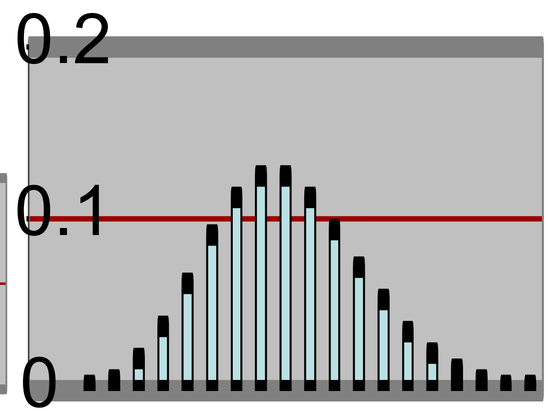
$\lambda=2.0$



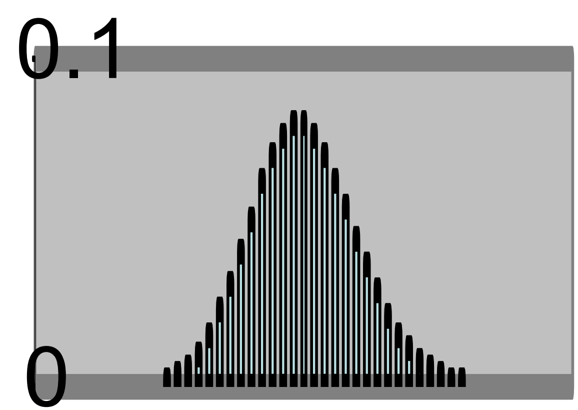
$\lambda=5.0$



$\lambda=10$



$\lambda=25$

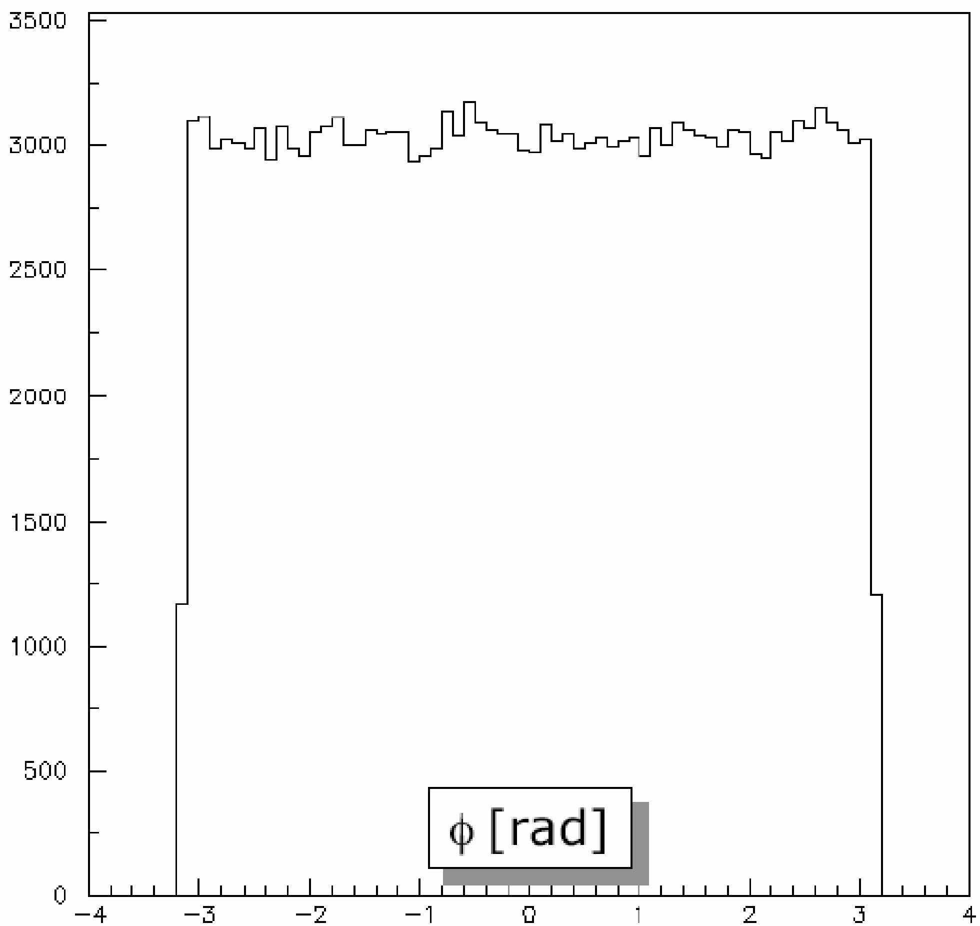


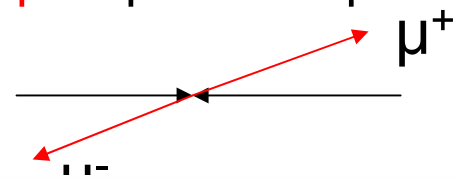
Κατανομές Poisson

- Το πλήθος των ραδιενεργών διασπάσεων σε ορισμένο χρόνο t, μικρό σε σχέση με τον χρόνο διάσπασης
- Το πλήθος συγκεκριμένου τύπου σωματιδίων σε σκέδαση σωματιδίου-σωματιδίου όταν ο συνολικός αριθμός των γεγονότων είναι πολύ μεγάλος και η συγκεκριμένη διαδικασία σπάνια
- Η πιθανότητα παρατήρησης n γεγονότων σε χρόνο t, όταν ο μέσος ρυθμός είναι μ : $\lambda=\mu t$. Ημέση τιμή λ-αναμενόμενη τιμή- και η διακύμανση : variance στην κατανομή Poisson είναι ισες!
- …. Απο δώ προκύπτει ο τύπος : $n \pm \sqrt{n}$ που χρησιμοποιείται στα στατιστικά σφάλματα όταν μετρούμε σε συγκεκριμένο χρονικό διάστημα

Ομοιόμορφες Κατανομές (Uniform Distributions)

- Η πιθανότητα είναι σταθερή σε ένα διάστημα : πχ η κατανομή των μιονίων κατα την αζιμουθιακή γωνία φ στη σκέδαση : ee→μμ





$$f(x;\ \alpha,\beta) = \begin{cases} \dfrac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & \text{otherwise} \end{cases}$$

E : αναμενόμενη τιμή

$$E[x] = \frac{1}{2}(\alpha+\beta)$$

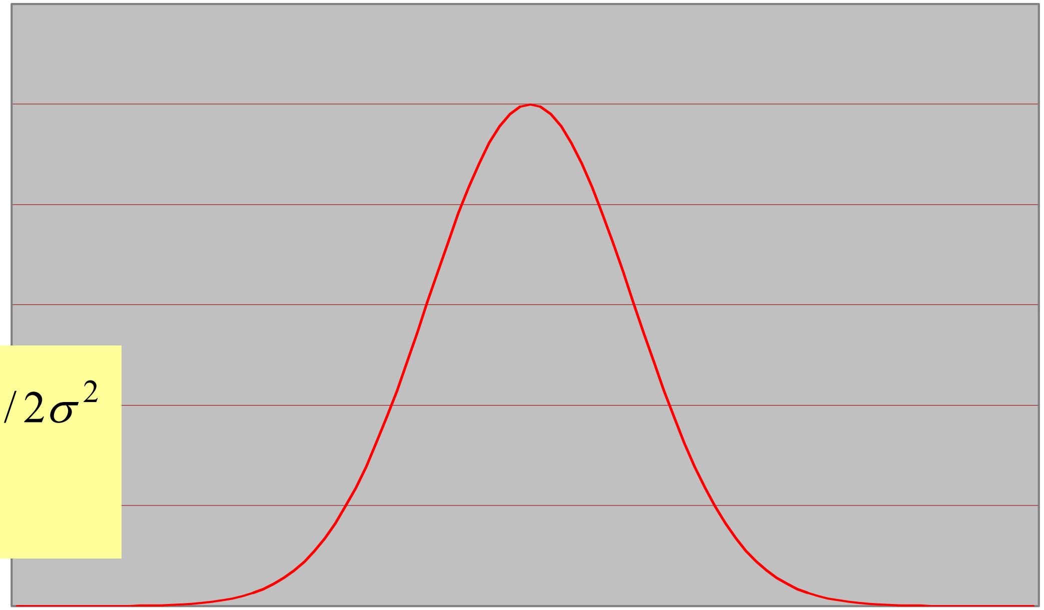
V : variance

$$V[x] = \frac{1}{12}(\beta-\alpha)^2$$

Κατανομή Gauss ή Normal

Πυκνότητα πιθανότητας

$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$



Mean

$\mu = \langle x \rangle = \int x P(x)\, dx$

$= \mu$

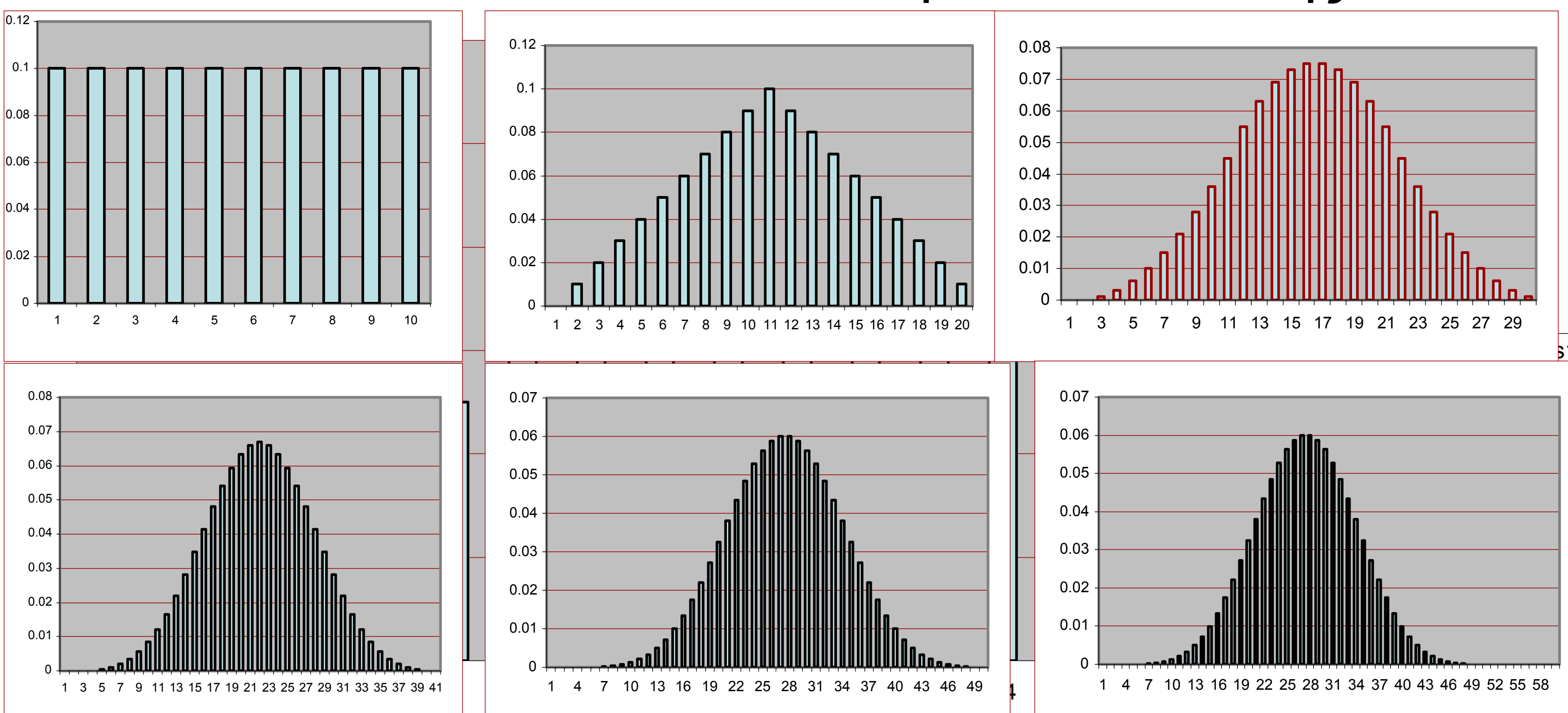
Variance

$V \equiv \sigma^2 = \langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2$

$= \sigma^2$

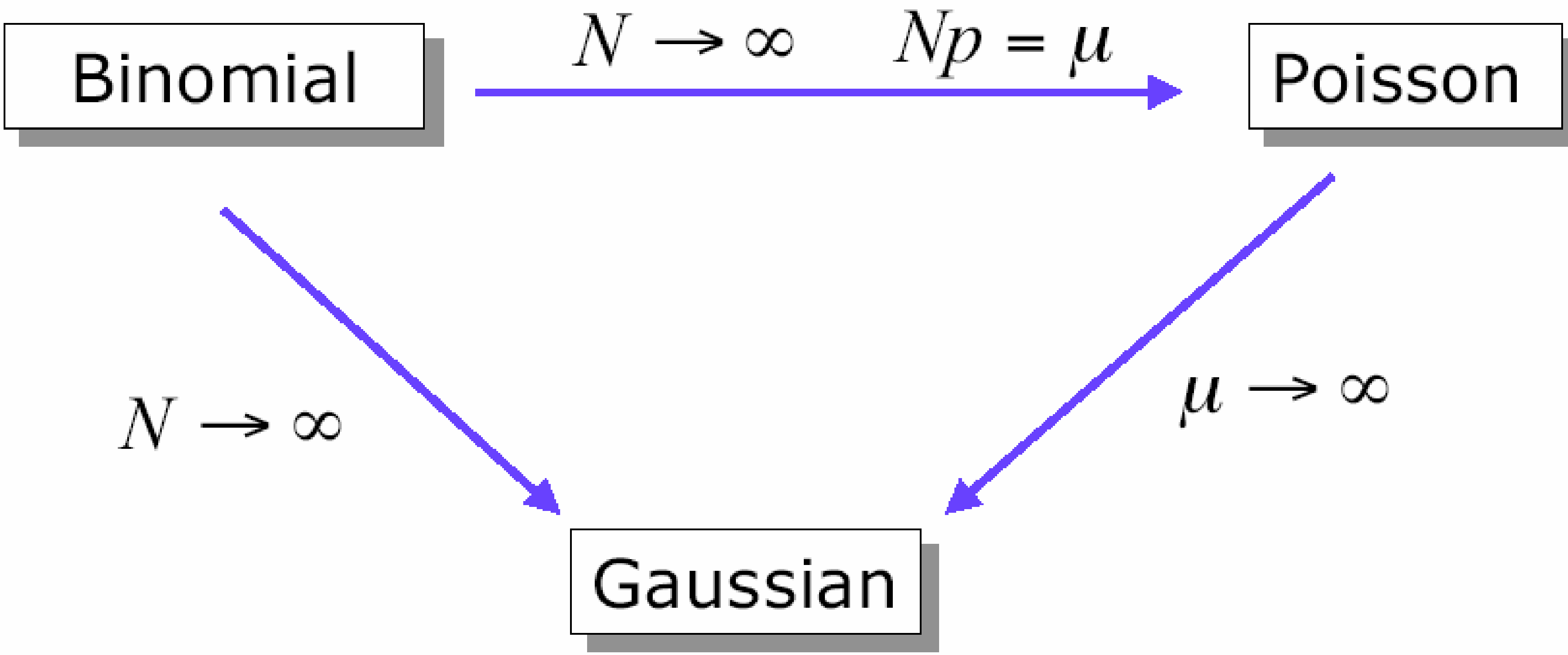
Central Limit Theorem (CLT)

- Το άθροισμα y, n τυχαίων, ανεξάρτητων αριθμών x_i είναι gaussian ($n \to \infty$) όποια κι αν είναι η κατανομή των x_i
- Convolute uniform distribution με τον εαυτό της



Σχέσεις μεταξύ κατανομών

Διώνυμη : N=το πλήθος των δοκιμών, p=η πιθανότητα επιτυχίας σε κάθε δοκιμή



Τύποι κατανομών Gauss

Μια μόνο κατανομή υπαρχει!

68.27% within 1σ

95.45% within 2σ

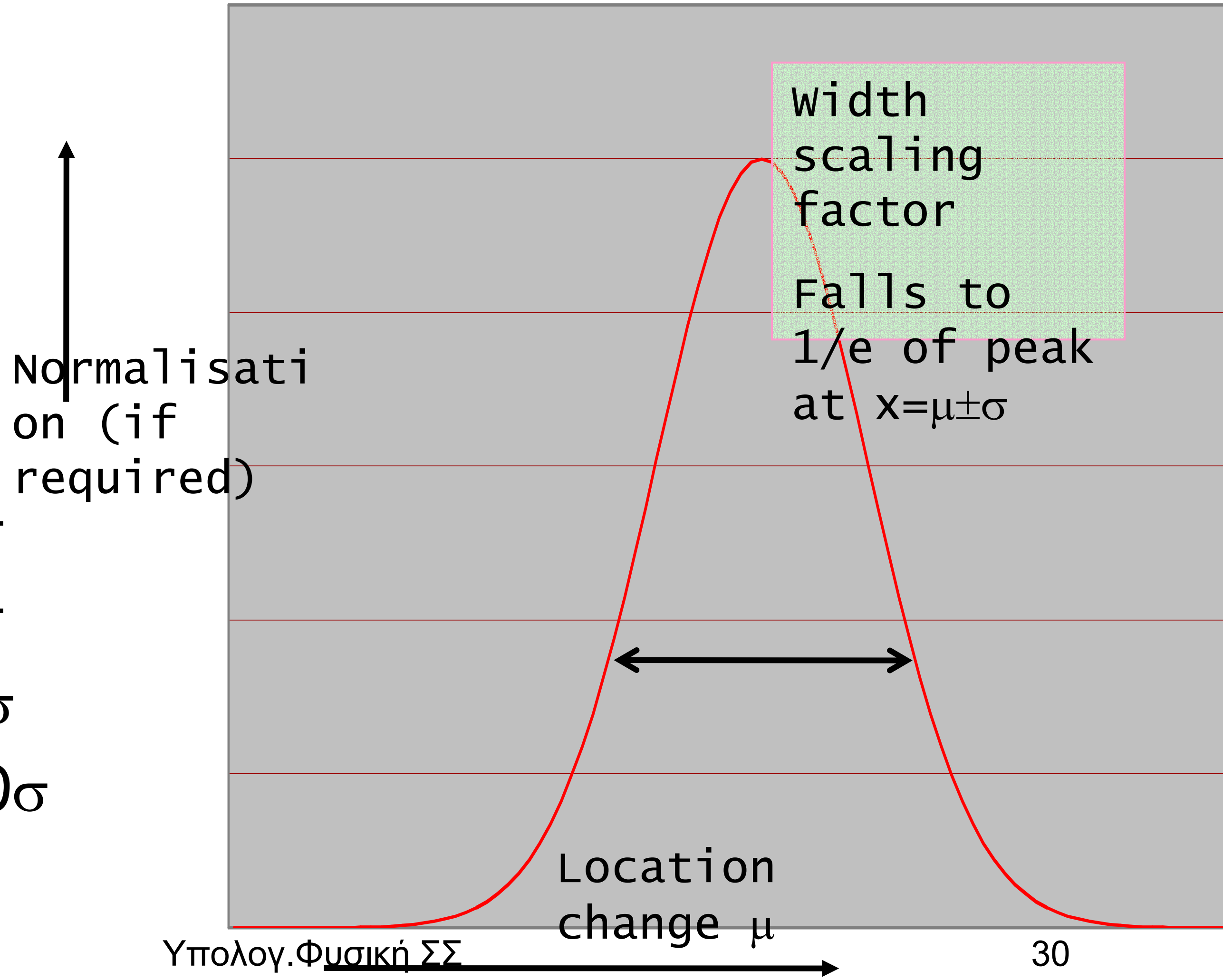
99.73% within 3σ

90% within 1.645 σ

95% within 1.960 σ

99% within 2.576 σ

99.9% within 3.290σ



Κατανομή Chi-Squared (χ^2)

$$z = \chi^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

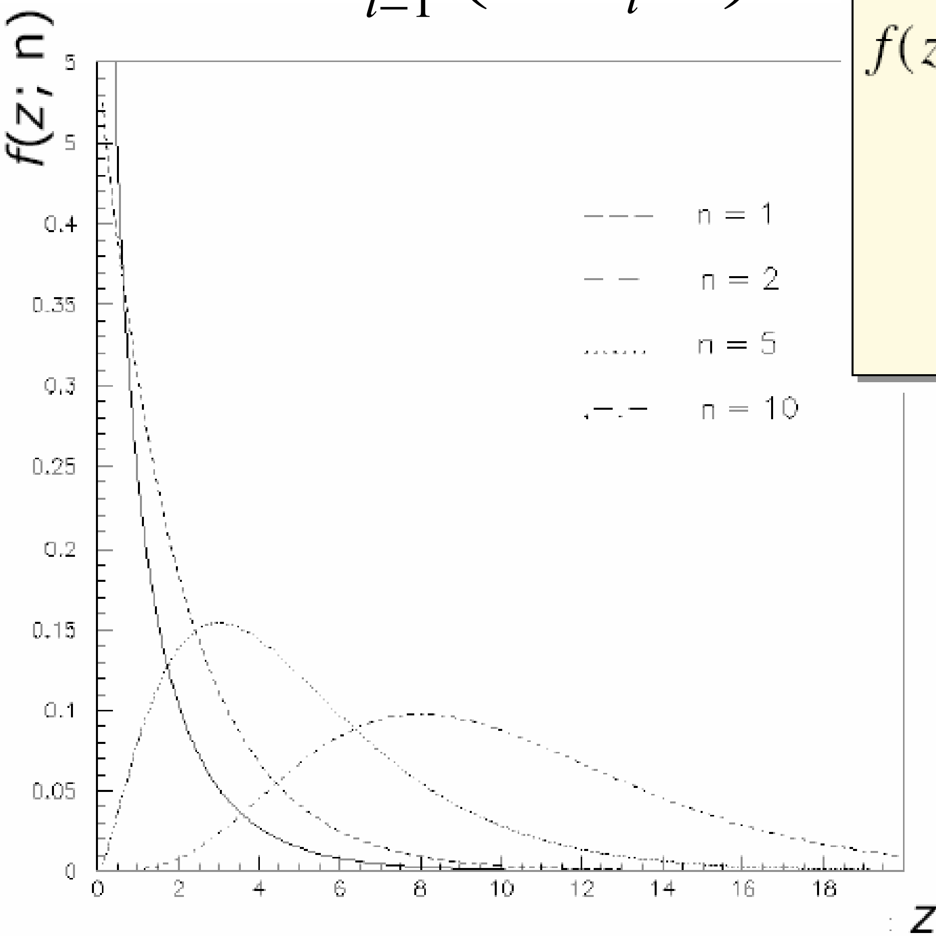
x_i : ανεξαρτητες μεταβλητές με κατανομή Gauss

$$f(z;\ n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \quad ; \quad n = 1,2,\ldots$$

$$E[z] = n$$

$$V[z] = 2n$$

Το πλήθος των βαθμών ελευθερίας



$$\Gamma(x) = e^{-Cx} \frac{1}{x} \prod_{n=1}^{+\infty} \frac{e^{x/n}}{1 + x/n}$$

$$C = \lim_{n \to \infty} 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \ln(n)$$

Υπολογιστική Φυσική Στοιχειωδών Σωματιδίων

4ο Εξάμηνο 2005-2006

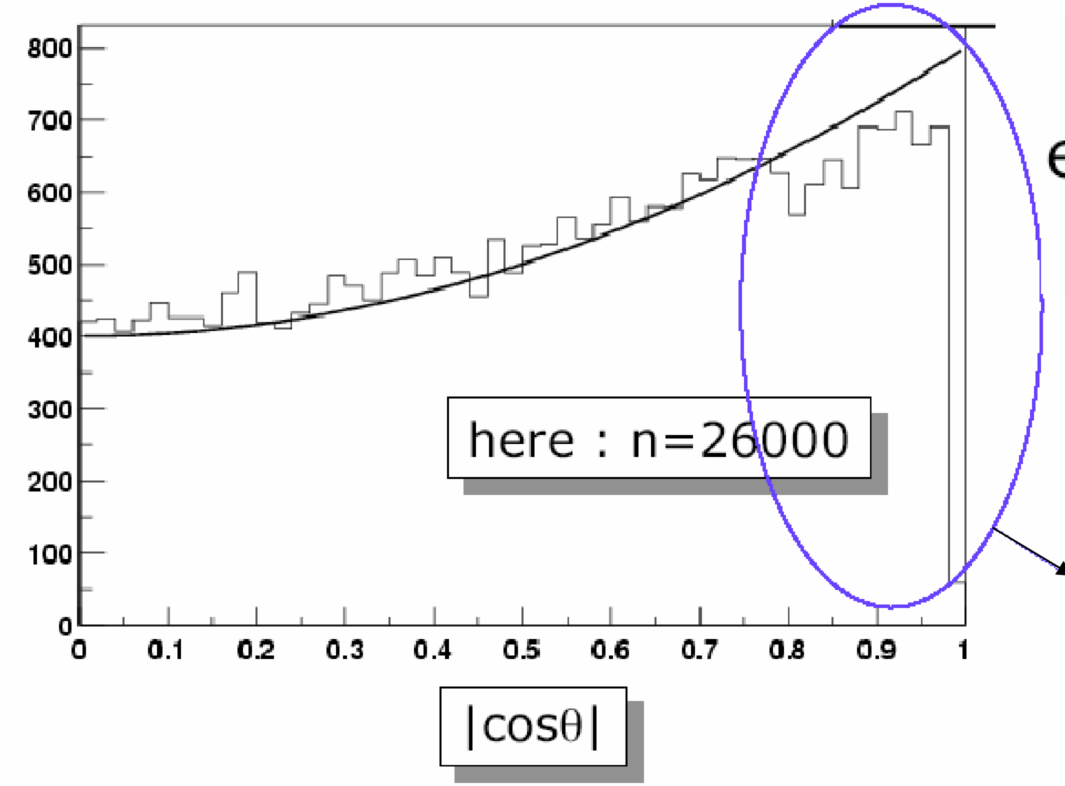
Ιστογράμματα (Histograms)

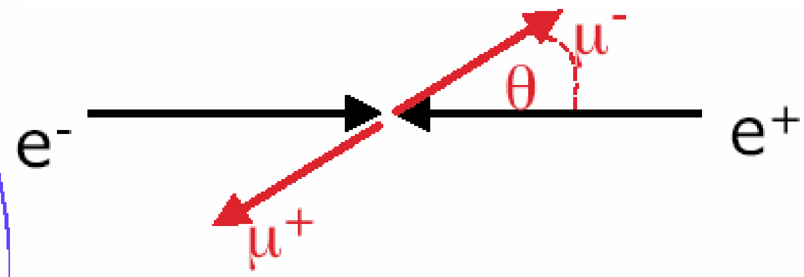
Διδάσκοντες : Χαρά Πετρίδου

Δήμος Σαμψωνίδης

Ιστογράμματα

- Συνήθως οι μεταβλητές που χρησιμοποιούμε είναι διακριτές μεταβλητές x_1, x_2,… x_n που προέρχονται από πειραματικές μετρήσεις
- Παράδειγμα: κατανομή πολικής γωνίας θ

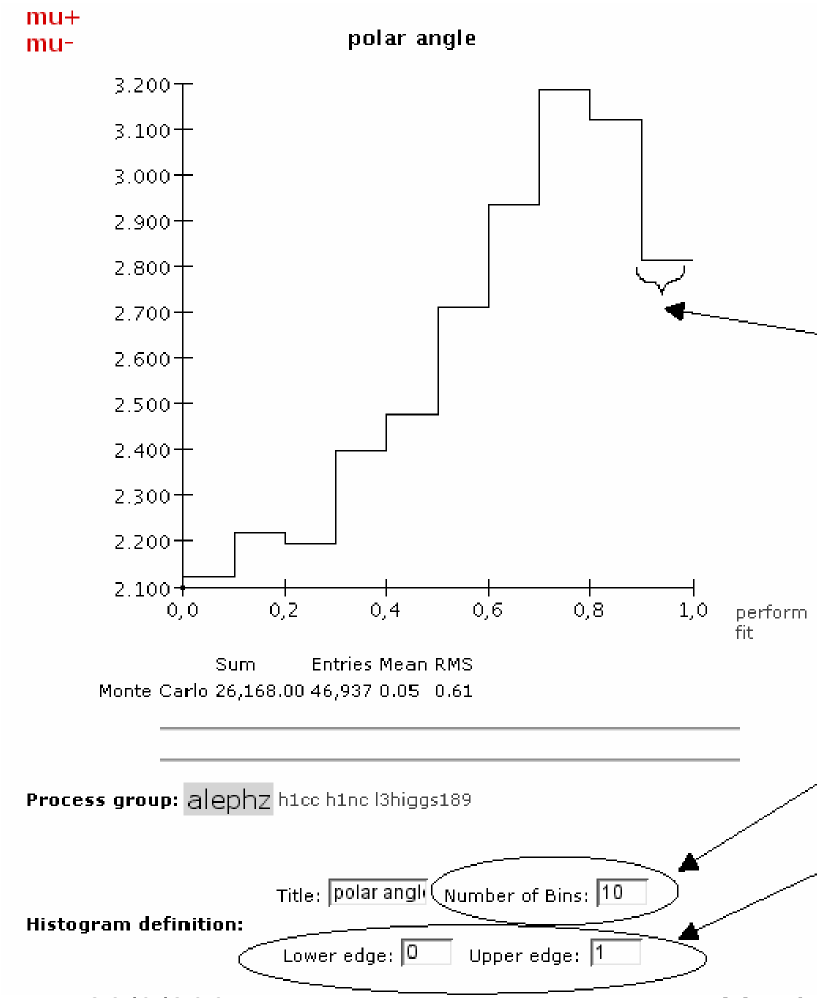




Σύγκριση των δεδομένων

με ομαλή κατανομή

Ιστογράμματα

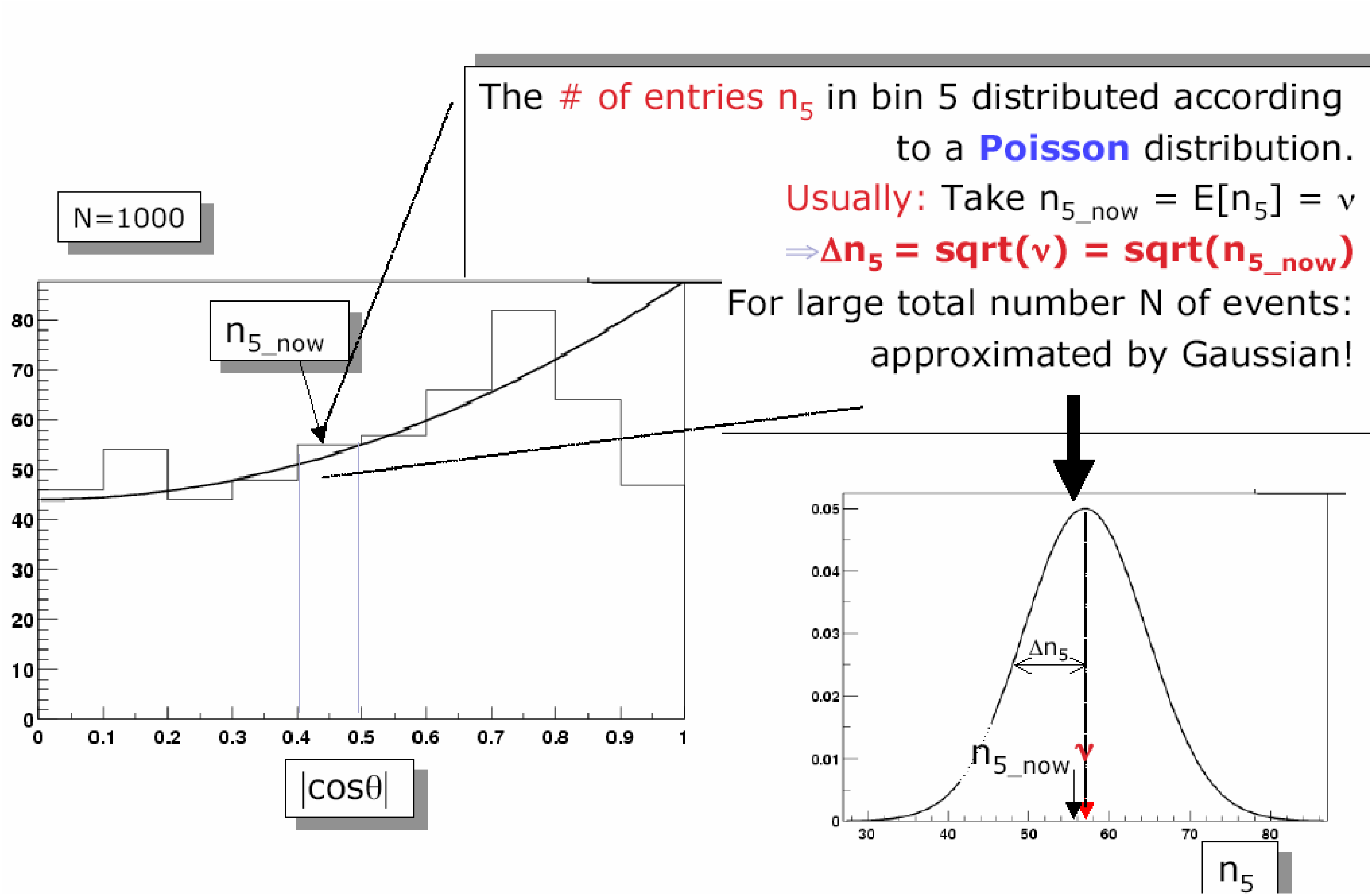


- Entries, normalization,…
- Bin width(s) (στο παράδειγμα ισαπέχουσες bins)
- Πλήθος των bins
- Περιοχή (Range)

Ιστογράμματα

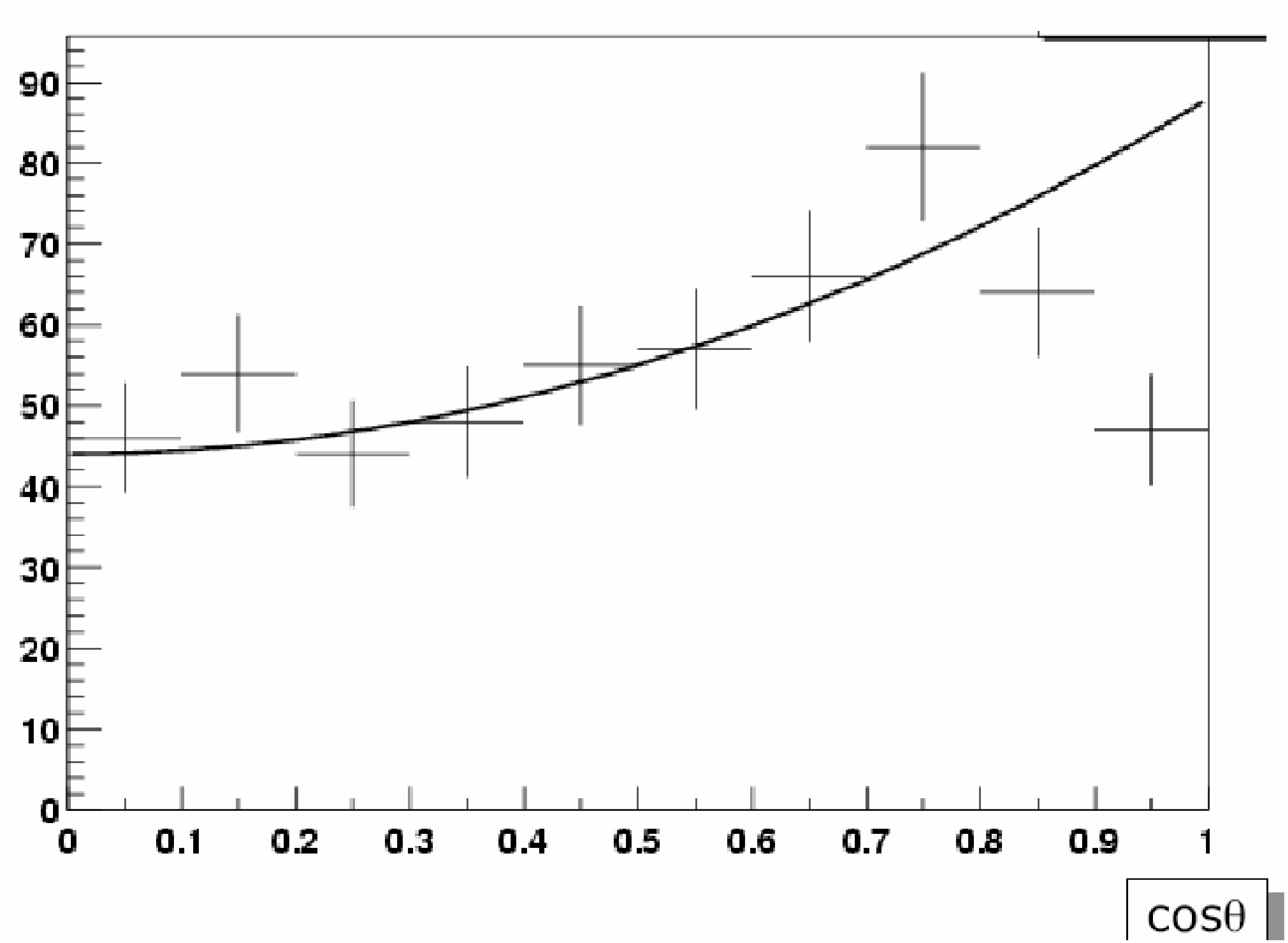
- Εαν N (entries) μεγάλος αριθμός το ιστόγραμμα αντιπροσωπεύει κατανομή πιθανότητας (pdf)
- Ενδιαφέρει το πλήθος των γεγονότων σε κάθε bin και το σφάλμα
- Η κανονικοποίηση του ιστογράμματος
- Η μέση τιμή του ιστογράμματος

Σφάλματα στο Ιστόγραμμα



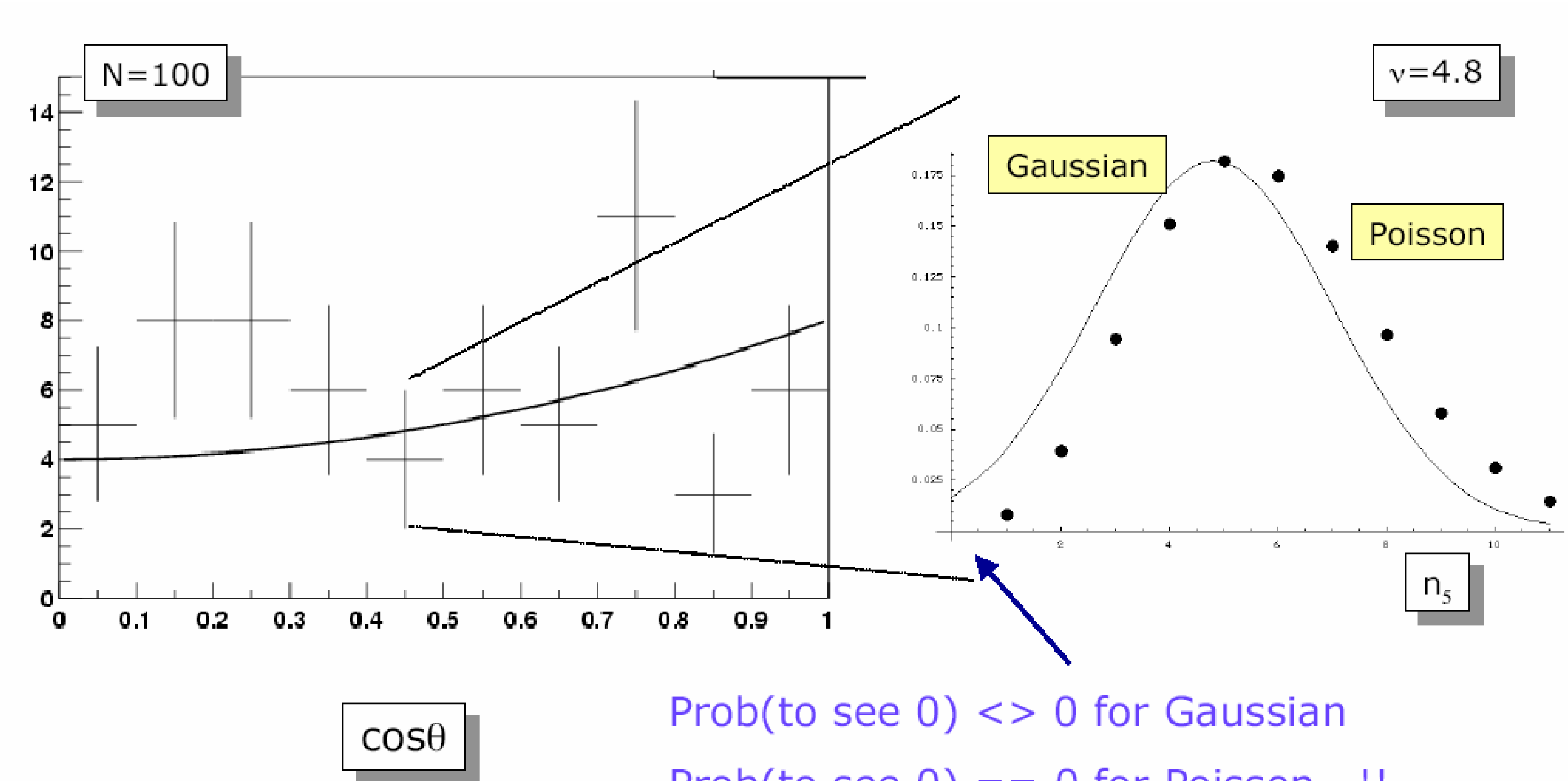
Σφάλματα στο Ιστόγραμμα

- Προσθέτουμε error bars στο ιστόγραμμα



Σφάλματα στο Ιστόγραμμα

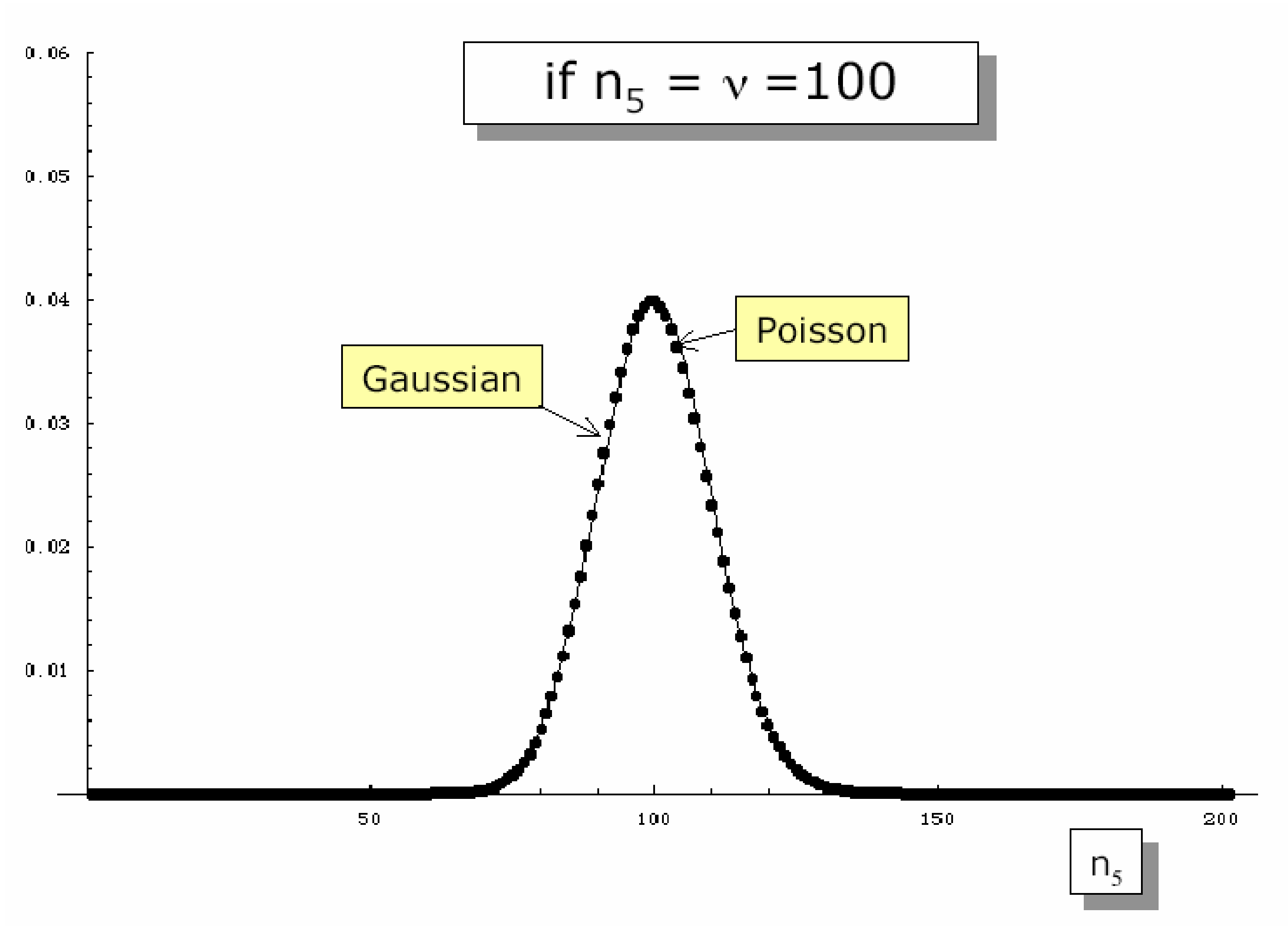
- Προσοχή! Περίπτωση μικρού πλήθους γεγονότων



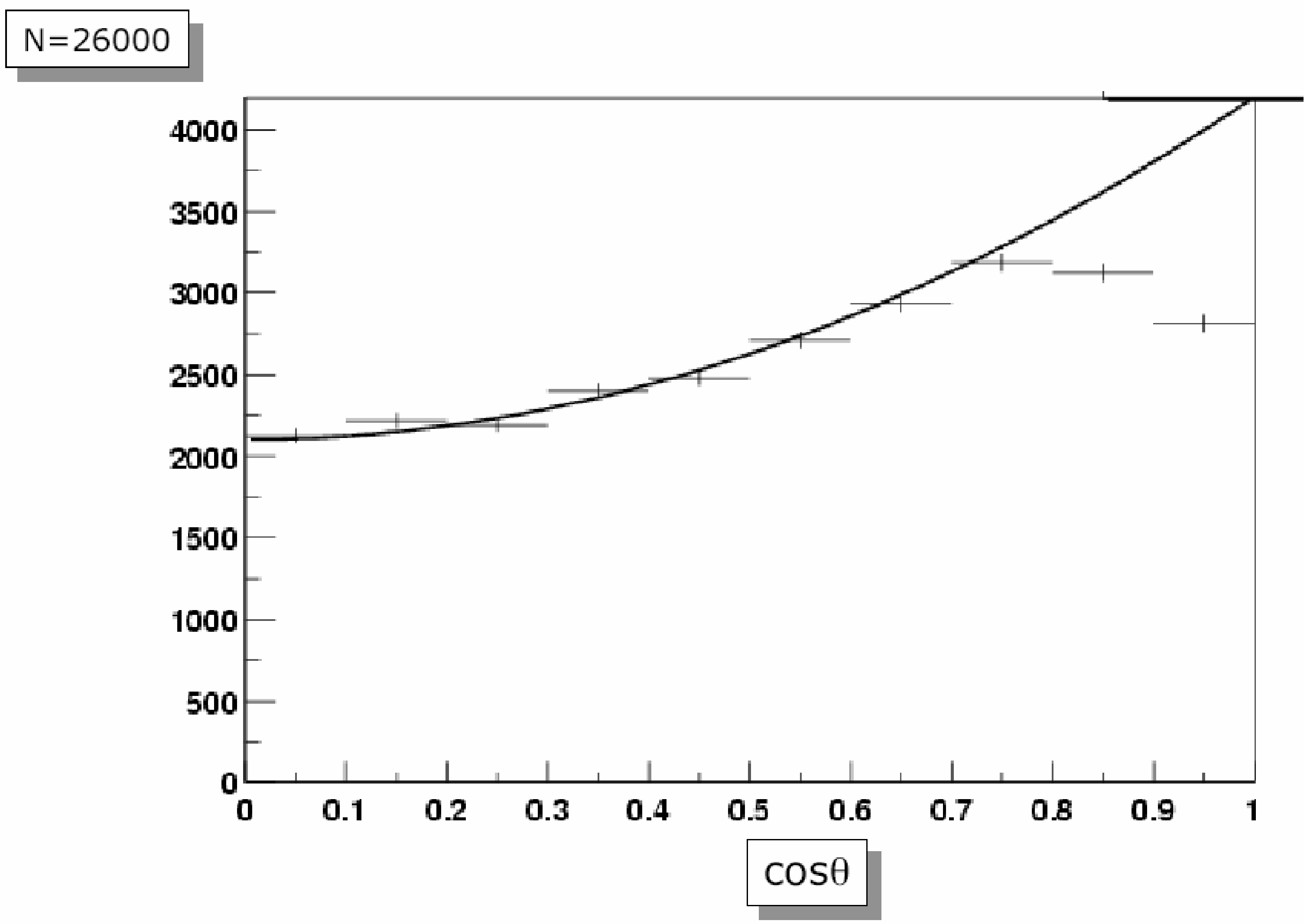
Prob(to see 0) <> 0 for Gaussian

Prob(to see 0) == 0 for Poisson !!

Poisson vs Gaussian



Σφάλματα στο Ιστόγραμμα

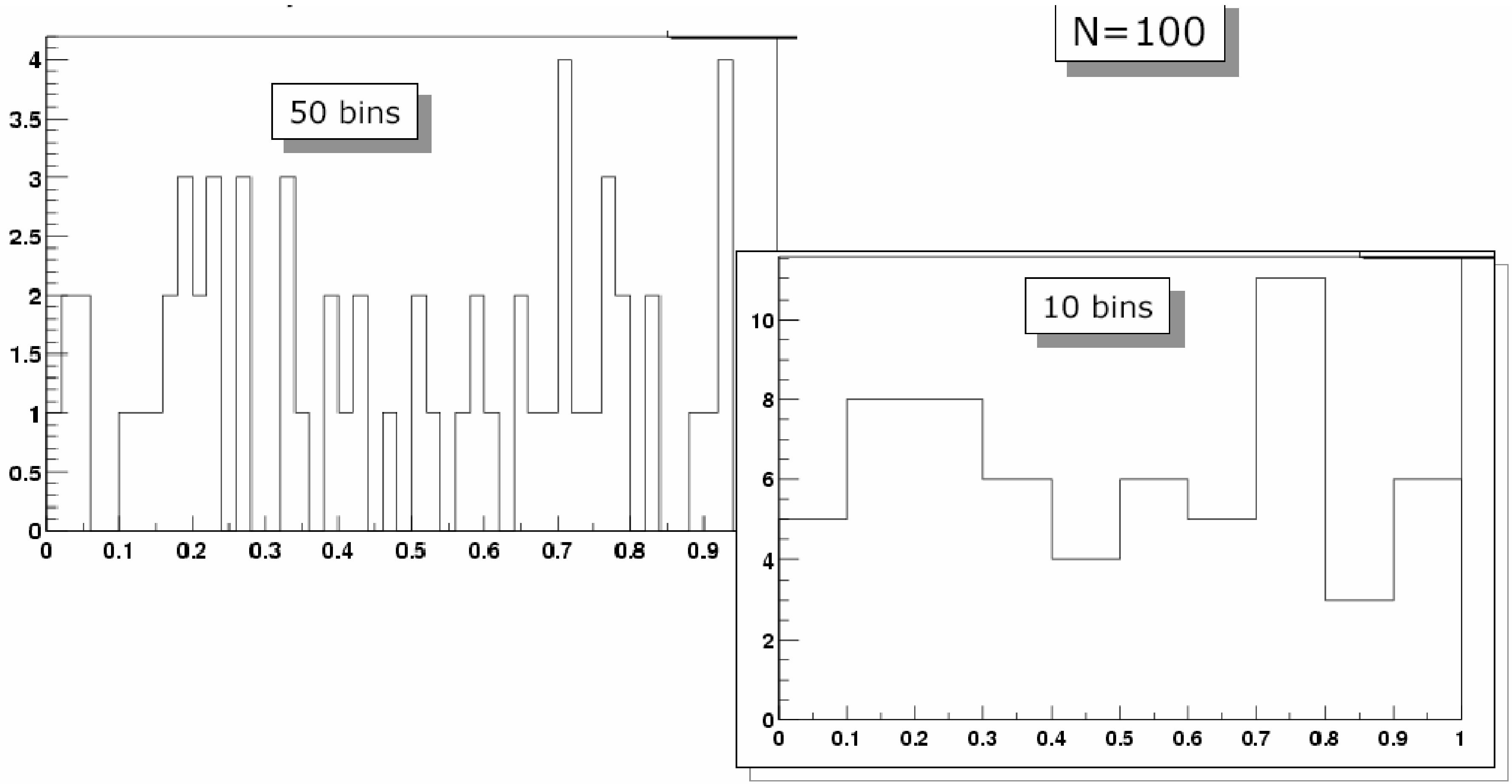


Ιστογράμματα
Τι θα πρέπει να προσέχουμε?

- Το μέγεθος του bin (bin size)
- Γεγονότα εκτός των ορίων του ιστογράμματος (underflows, overflows)
- Κατανομές με γρήγορες μεταβολές (steeply falling/fast varying functions)

Επιλογή του bin width

- Επιλέγουμε την υποδιαίρεση της κλίμακας (bin width) έτσι ώστε το πλήθος των γεγονότων ανά υποδιαίρεση να είναι 'λογικό'

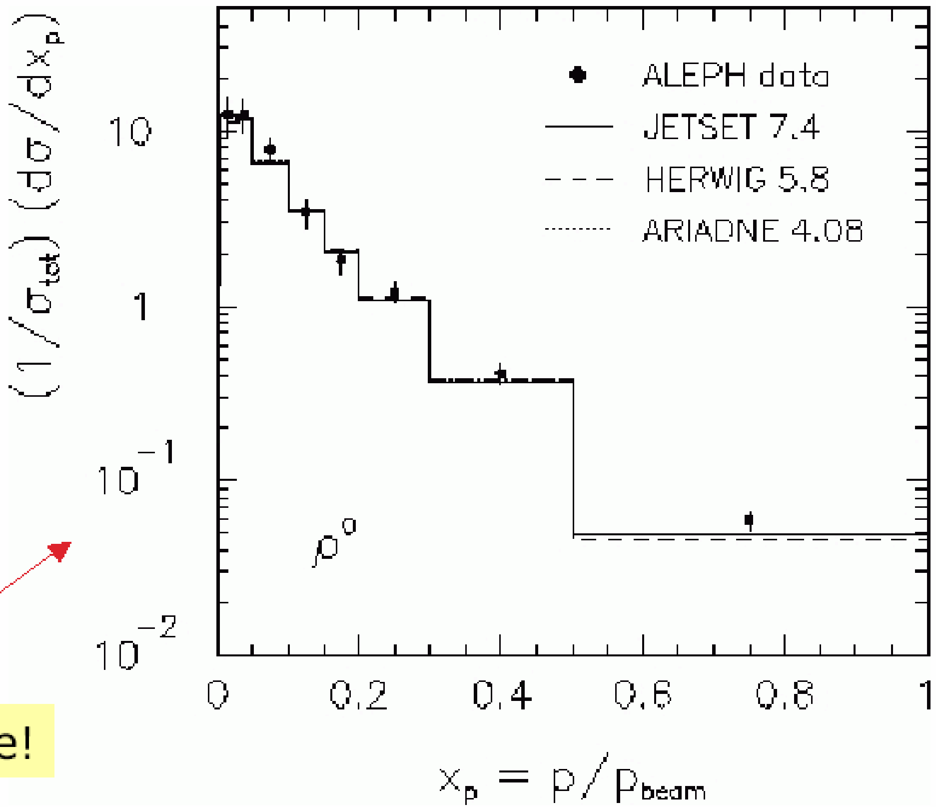


Επιλογή του bin width

- Ο αριθμός των γεγονότων που μετατοπίζονται εντός και εκτός της υποδιαίρεσης πρέπει να είναι παρόμοιος (bin migration-σταθερότητα και καθαρότητα του δείγματος)
- Το bin width πρέπει να ταιριάζει με την πειραματική διακριτική ικανότητα στην συγκεκριμένη μεταβλητή (experimental resolution)
- Να υπάρχει αρκετή στατιστική για κάθε υποδιαίρεση (bin statistics)

Επιλογή του bin width

- Σημαντικό για συναρτήσεις με απότομη κλίση
- Παράδειγμα: η κατανομή της ορμής του σωματιδίου-συντονισμού ρ



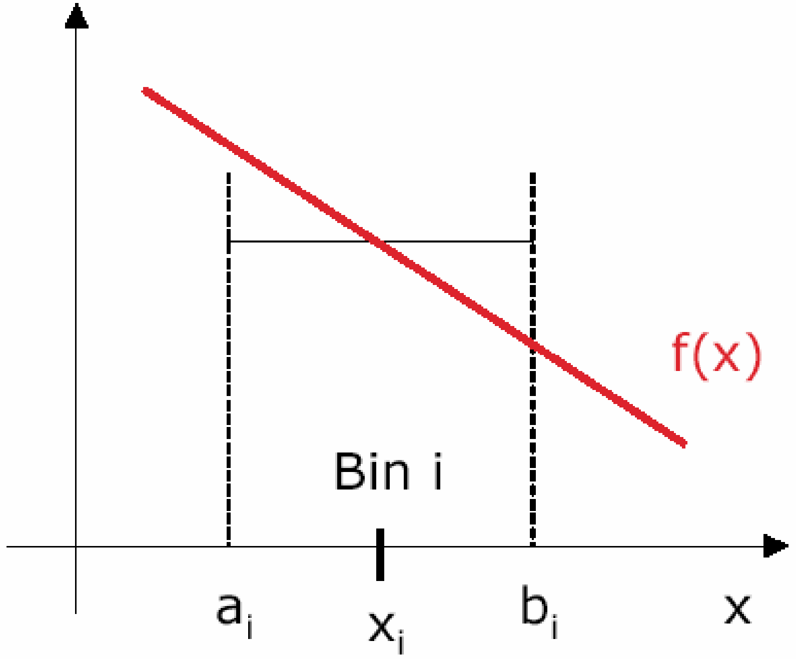
logarithmic scale!

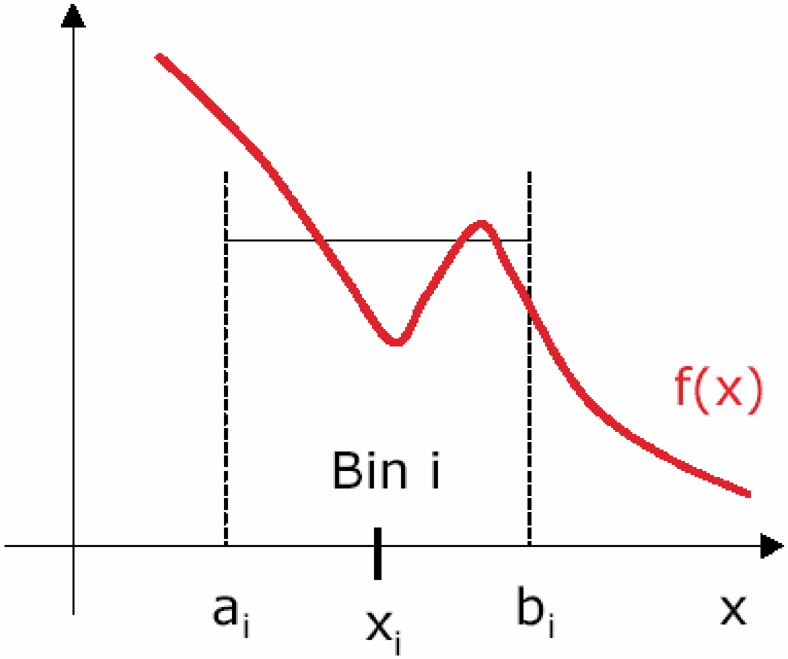
Επιλογή της κλίμακας (Histogram range)

- Κακή επιλογή κλίμακας μπορεί να οδηγήσει σε παραπλανητικά συμπεράσματα
- =>Πρέπει να ελέγχουμε να μην υπάρχουν overflows και underflows
- Η κλίμακα είναι σημαντική και στην περίπτωση κανονικοποίησης

Σύγκριση ιστογράμματος με ομαλή συνάρτηση

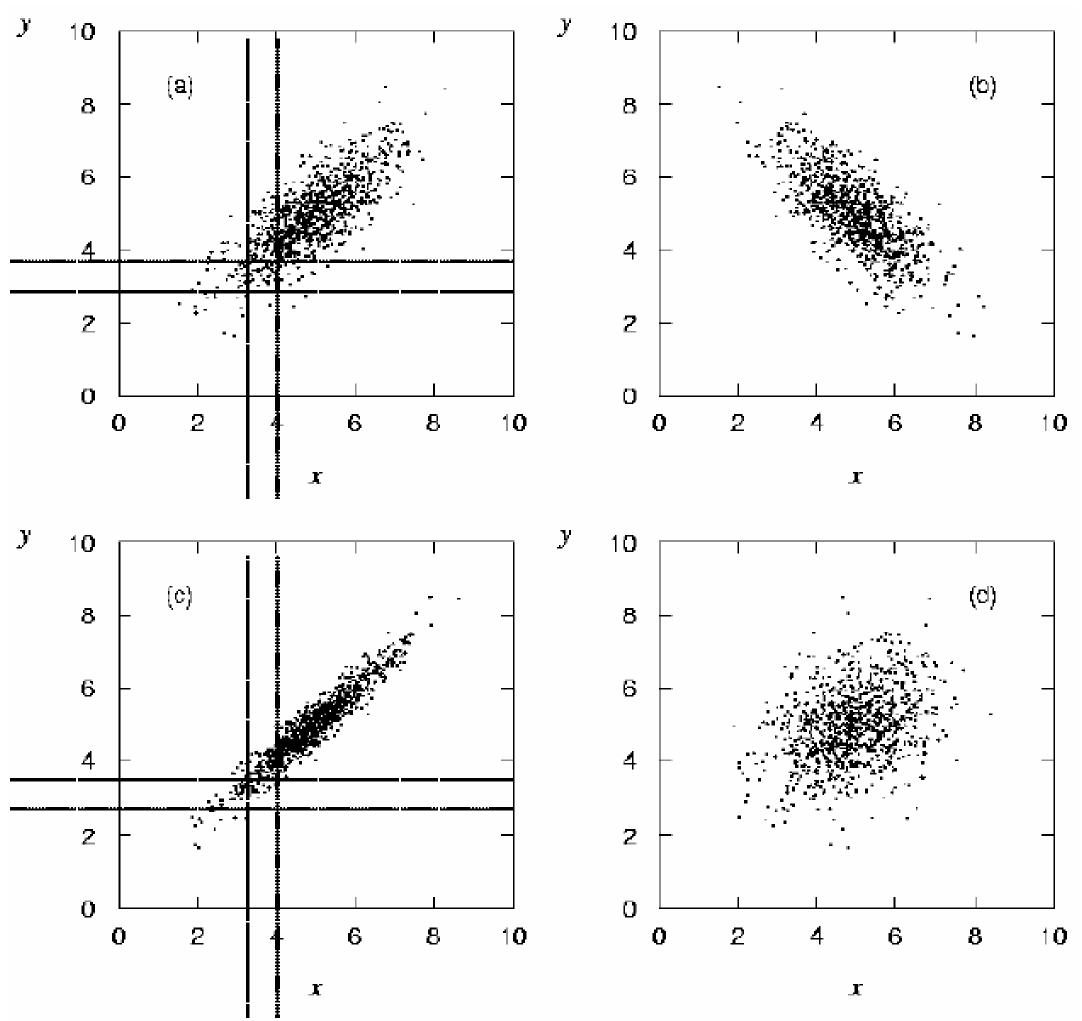
- Προσοχή στις συναρτήσεις με απότομες ή γρήγορες μεταβολές





Συσχετίσεις-Συμμεταβολή (correlations-covariance)

- Ιστογράμματα 2-διαστάσεων (scaterplots)
- =>Χρήσιμο εργαλείο για να μελετήσουμε συσχετισμούς δύο μεταβλητών (correlations)



Ιδιότητες-Συσχετισμοί (correlations)

- Για συναρτήσεις με διάφορες μεταβλητές μπορεί να υπάρχουν συσχετίσεις μεταξύ των μεταβλητών
- Παράδειγμα : για συναρτήσεις δύο τυχαίων μεταβλητών x, y η συμμεταβολή (covariance) V(x,y) (ή cov(x,y)) -error matrix- ορίζεται ως:

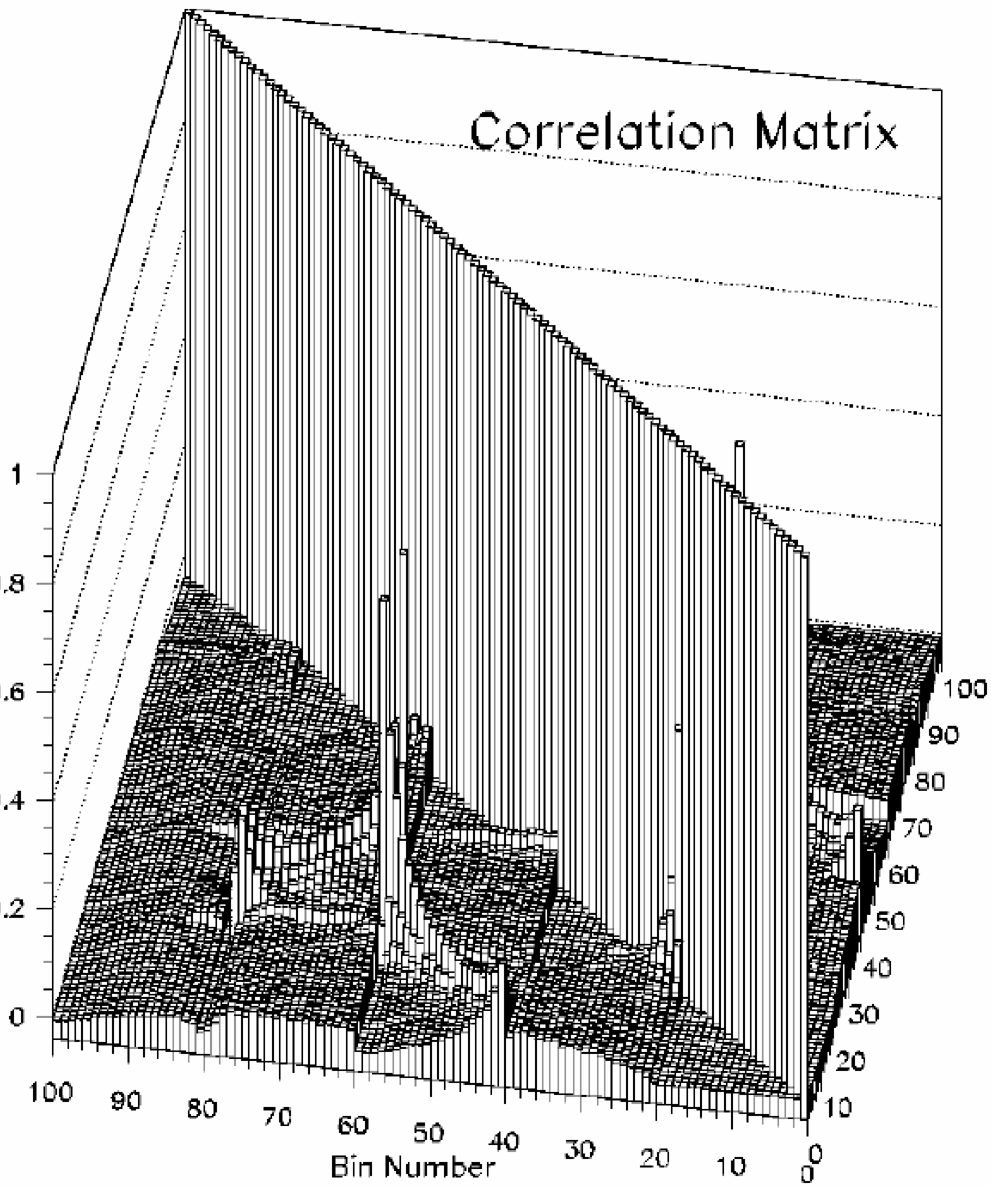
$$V[xy] = E((x-\mu_x)(y-\mu_y)) \equiv E[xy] - \mu_x\mu_y$$

$$E[xy] = \int_{y\,\min}^{y\,\max}\int_{x\,\min}^{x\,\max} x \cdot y \cdot f(x,y)\cdot dx \cdot dy$$

$$V(x,y) = \begin{pmatrix} \sigma_x^2 & V[xy] \\ V[yx] & \sigma_y^2 \end{pmatrix}$$

$$\rho_{xy} = \frac{V[xy]}{\sigma_x \sigma_y} \qquad -1 \le \rho_{xy} \le +1$$

ρ=1 για y=ax+b


Correlation Matrix
1
0.8
0.6
0.4
0.2
0
100
90
80
70
60
50
40
30
20
10
0
Bin Number
100
90
80
70
60
50
40
30
20
10
0

Ιδιότητες-Συσχετισμοί (correlations)

- Υπολογισμός του σφάλματος για συναρτήσεις με συσχετιζόμενες μεταβλητές: παίρνουμε υπ'όψη την συσχέτηση

$$\sigma_f^2 = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \cdot \sigma_{x_i}^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left(\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \right) \cdot \mathrm{cov}(x_i, x_j)$$