

Ιστογράμματα

Τα ιστογράμματα αποτελούν ένα εύχρηστο οπτικό τρόπο για να εξάγουμε την κατανομή που ακολουθούν μια σειρά μετρήσεων ενός μεγέθους αλλά και παράλληλα δίνουν τη δυνατότητα για εύκολη στατιστική ανάλυση των μετρήσεων

Αν έχουμε μια σειρά από μεγάλο αριθμό μετρήσεων ενός μεγέθους τότε μπορούμε να ταξινομήσουμε τις μετρήσεις και να βρούμε τη συχνότητα εμφάνισης κάθε τιμής

Συνήθως δεν κρατάμε κάθε μέτρηση ξεχωριστά αλλά χωρίζουμε το εύρος των τιμών σε κατάλληλα ίσα υποδιαστήματα και αθροίζουμε τις τιμές που πέφτουν σε κάθε υποδιάστημα. Με το τρόπο αυτό πέρνουμε τη κατανομή της συχνότητας εμφάνισης κάθε τιμής

Για παράδειγμα έστω ότι μετρήσαμε κάποιο μέγεθος 30 φορές και βρήκαμε:

8.16	8.10	8.12	8.12	8.16	8.16
8.14	8.18	8.12	8.10	8.21	8.14
8.12	8.18	8.17	8.14	8.14	8.17
8.16	8.18	8.06	8.09	8.16	8.18
8.18	8.24	8.10	8.16	8.13	8.21

Οι τιμές του μετρήσεων εμφανίζονται με την ακόλουθη συχνότητα:

Τιμή	#	Τιμή	#	Τιμή	#
8.06	1	8.13	1	8.20	0
8.07	0	8.14	4	8.21	2
8.08	0	8.15	0	8.22	0
8.09	1	8.16	6	8.23	0
8.10	3	8.17	2	8.24	1
8.11	0	8.18	5		
8.12	4	8.19	0		

Δηλαδή κάθε τιμή x_i εμφανίζεται n_i φορές

Η μέση τιμή μπορεί να γραφεί:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{\sum x_k n_k}{\sum n_k}$$

Ιστογράμματα

Ομαδοποιώντας τις μετρήσεις τότε βρίσκουμε ευκολότερα πόσο συνεισφέρει κάθε τιμή στο άθροισμα της μέσης τιμής, $\sum x_k n_k$

Στο παράδειγμα οι τιμές όλες οι τιμές βρίσκονται στο διάστημα 8.00 – 8.30 και άρα η ακριβής τιμή θα βρίσκεται στο διάστημα. Εφόσον οι τιμές 8.12 – 8.18 είναι οι πιο συχνές περιμένουμε ότι και η πραγματική τιμή θα είναι στο υποδιάστημα αυτό. Όντως η μέση τιμή είναι 8.15.

Συνήθως χωρίζουμε το διάστημα σε ίσα υποδιαστήματα κατάλληλου μήκους ώστε στο καθένα να συμπεριλαμβάνεται μεγάλος αριθμός μετρήσεων και βρίσκουμε το ποσοστό των μετρήσεων σε κάθε διάστημα

$$F_k = \frac{n_k}{N}$$

Συχνότητα εμφάνισης μιας μέτρησης x_k .

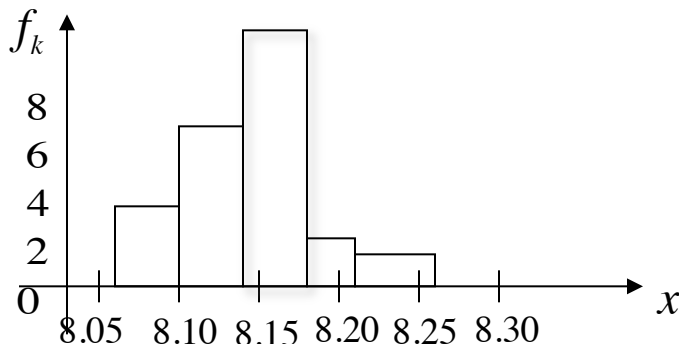
$$\bar{x} = \sum_k F_k x_k$$

$$f_k = \frac{F_k}{\Delta x_k}$$

Πυκνότητα Πιθανότητας μεταβλητής x να βρεθεί στο k -διάστημα Δx , το ποσοστό δηλαδή των μετρήσεων στο διάστημα Δx_k

Το ιστόγραμμα αποτελεί τη γραφική παράσταση της f_k ως προς το x

Η ολική επιφάνεια των ορθογωνίων θα είναι ίση με την μονάδα: $\sum F_k = 1 = \sum f_k \Delta x_k$



Στη περίπτωση αυτή χρησιμοποιούμε $\Delta x = 0.04 \text{ cm}$

Αν αριθμός των μετρήσεων γίνει αρκετά μεγάλος τότε το ιστόγραμμα πέρνει τη μορφή της κατανομής από την οποία προέρχονται οι μετρήσεις.

Η καμπύλη αποτελεί την οριακή κατανομή

Ιστογράμματα – Οριακή κατανομή

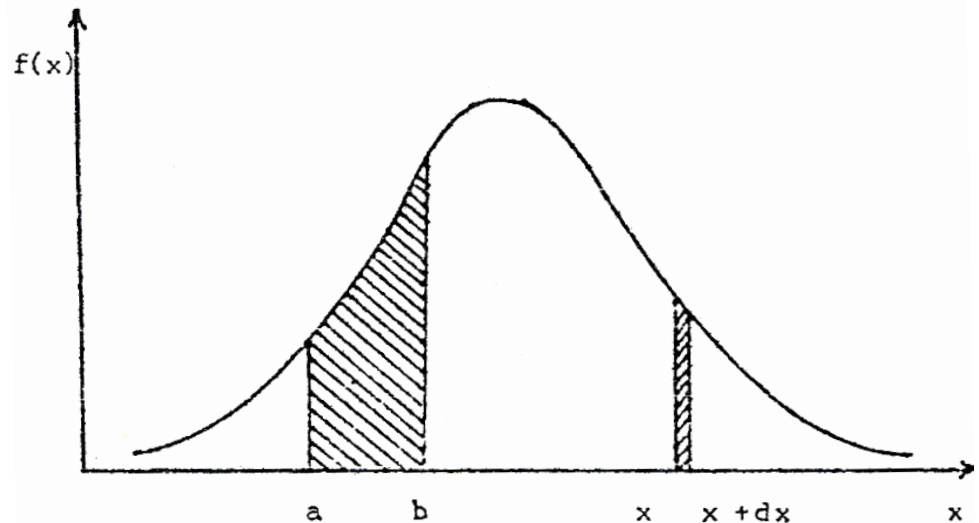
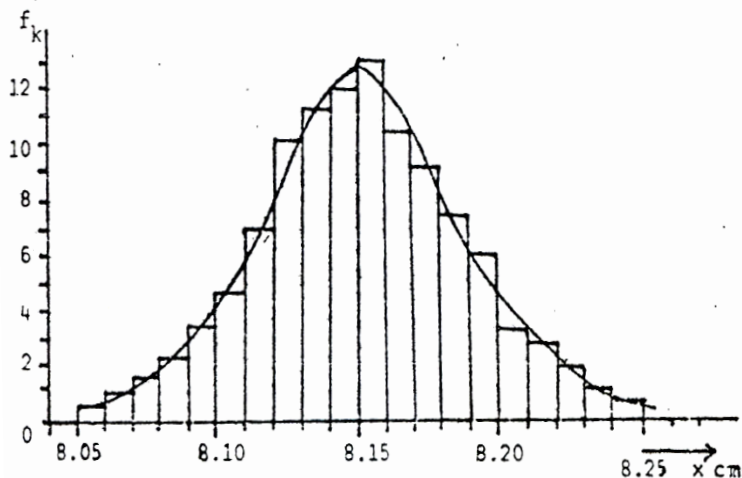
Η οριακή κατανομή είναι μια θεωρητική καμπύλη που δεν μπορεί να προκύψει ποτέ από τις μετρήσεις

Το εμβαδό κάθε παρ/μου του ιστογράμματος είναι ισοδύναμο με τον αριθμό μετρήσεων που περιλαμβάνεται στο διάστημα αυτό.

$f(x)dx$ πιθανότητα μια μέτρηση να περιλαμβάνεται μεταξύ x και $x+\Delta x$

$\int_a^b f(x)dx$ ποσοστό μετρήσεων μεταξύ $x=a$ και $x=b$, η πιθανότητα δηλαδή μια μέτρηση να «πέσει» στο διάστημα a,b

$\int_{-\infty}^{+\infty} f(x)dx = 1$ Η πιθανότητα μια μέτρηση να βρίσκεται μεταξύ $-\infty$ και $+\infty$
Λέμε τότε ότι η $f(x)$ είναι **κανονικοποιημένη**



Ιστογράμματα – Οριακή κατανομή πολλών μετρήσεων

Αν ξέρουμε την οριακή κατανομή μπορούμε να υπολογίσουμε τη μέση τιμή

$$\bar{x} = \sum x_k F_k \xrightarrow{\lim \Delta x_k \rightarrow 0} \bar{x} = \int_{-\infty}^{+\infty} x f(x) dx \quad \text{και ανάλογα} \quad \sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx$$

Αν οι μέτρησεις επηρεάζονται από πολλές πηγές μικρών τυχαίων σφαλμάτων τότε η οριακή κατανομή προσεγγίζει αυτή της κατανομής Gauss

Για πολύ μεγάλο αριθμό μετρήσεων και για τυχαία μικρά σφάλματα οι τιμές που αποκλίνουν πολύ από τη μέση τιμή κατανέμονται συμμετρικά ως προς τη μέση τιμή τότε στην άθροιση της εύρεσης της μέσης τιμής αυτές οι τιμές αλληλοαναιρούνται και η τελική τιμή πλησιάζει την πιστή τιμή του μεγέθους

$$f(x) = G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \Rightarrow \int_{-\infty}^{+\infty} G(x) dx = 1$$

Η πιθανότητα να πάρουμε μια τιμή x στο διάστημα x_1 και x_1+dx είναι: $P(x_1) \propto \sigma^{-1} e^{-(x_1-\mu)^2/2\sigma^2}$

Η πιθανότητα να πάρουμε μια τιμή x στο διάστημα x_2 και x_2+dx είναι: $P(x_2) \propto \sigma^{-1} e^{-(x_2-\mu)^2/2\sigma^2}$

Αν συνεχίσουμε για όλες τις τιμές x_N τότε η πιθανότητα για την x_N : $P(x_N) \propto \sigma^{-1} e^{-(x_N-\mu)^2/2\sigma^2}$

Μέση τιμή σα καλύτερος υπολογισμός αληθινής τιμής

Η πιθανότητα να παρατηρήσουμε το συγκεκριμένο δείγμα των N μετρήσεων είναι το γινόμενο των πιθανοτήτων:

$$P(x_1, x_2, \dots, x_N) = P(x_1) \times P(x_1) \times \dots \times P(x_N) \Rightarrow P(x_1, x_2, \dots, x_N) \propto \frac{1}{\sigma^N} e^{-\sum (x_i - \mu)^2 / 2\sigma^2}$$

Δεδομένων N παρατηρούμενων τιμών x_1, x_2, \dots, x_N ο καλύτερος υπολογισμός της πραγματικής τιμής είναι αυτός που πέρνουμε αν μεγιστοποιήσουμε την πιθανότητα

Αυτό θα συμβεί όταν το άθροισμα στον εκθέτη είναι ελάχιστο $\sum (x_i - \mu)^2 / 2\sigma^2$

Επομένως θα πρέπει:

$$\frac{d\left(\sum (x_i - \mu)^2 / 2\sigma^2\right)}{d\mu} = 0 \Rightarrow \sum (x_i - \mu) / \sigma^2 = 0 \Rightarrow \mu = \frac{\sum_{i=1}^N x_i}{N}$$

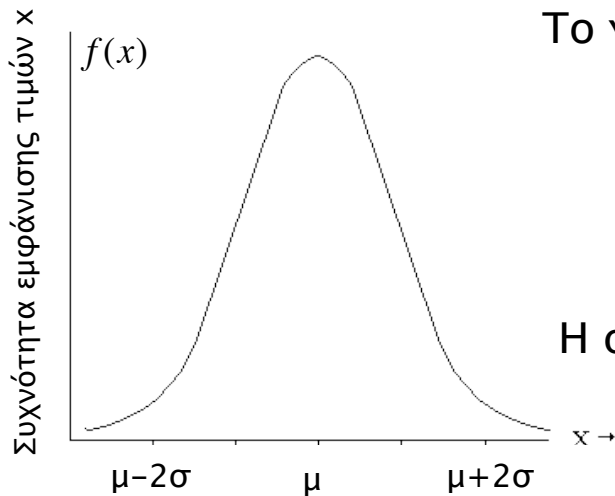
Ανάλογα ο καλύτερος υπολογισμός της τυπικής απόκλισης βρίσκεται ότι είναι:

$$\frac{dP(x_1, x_2, \dots, x_N)}{d\sigma} = 0 \Rightarrow \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

Gaussian κατανομή

Είδαμε ότι τα τυχαία σφάλματα υπάρχουν ακόμα και όταν όλες οι πηγές άλλων σφαλμάτων έχουν εξαλειφθεί. Συχνά χρειάζεται να επαναλάβουμε μετρήσεις αρκετές φορές ώστε να εκτιμήσουμε την πιθανότητα κάποιος που επαναλαμβάνει μια πανομοιότυπη μέτρηση θα βρει διαφορετικό αποτέλεσμα

Οι εκτιμήσεις αυτές στηρίζονται σε κάποιο μαθηματικό μοντέλο το οποίο και περιλαμβάνει τις περισσότερες περιπτώσεις. Το μοντέλο αυτό περιγράφεται από την **Gaussian** ή **κανονική** (**normal**) κατανομή



Το γράφημα του σχήματος δίνεται από τη συνάρτηση

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

Η συνάρτηση περιγράφεται από δύο παραμέτρους, μ και σ

Η μεταβλητή μ περιγράφει που βρίσκεται η κορυφή της καμπύλης στον άξονα x

Η μεταβλητή σ περιγράφει τη διασπορά των τιμών του x , δηλαδή το εύρος της καμπύλης

Ουσιαστικά η καμπύλη αυτή δείχνει την συχνότητα με την οποία θα εμφανίζονταν διάφορες τιμές του x αν επαναλαμβάναμε μια μέτρηση πολλές φορές

Κανονική κατανομή

Η μέση τιμή, μ , μας δίνει ουσιαστικά την περισσότερη πιθανή τιμή των x και αυτή είναι η μέση τιμή. Παραπλήσιες τιμές έχουν μια μικρότερη αλλά αρκετά μεγάλη πιθανότητα

Η κατανομή εκτείνεται σε άπειρες τιμές του x (με μικρή ωστόσο πιθανότητα). Αποδεικνύεται ότι εμβαδό της καμπύλης μεταξύ $\mu - \sigma$ και $\mu + \sigma$ ισούται με το 68.3% του συνολικού εμβαδού της καμπύλης $[-\infty, +\infty]$.

Το εμβαδό της καμπύλης που περικλύεται μεταξύ $\mu - 2\sigma$ και $\mu + 2\sigma$ ισούται με το 95.4% του εμβαδού της καμπύλης. Ενώ για $\pm 3\sigma$ το ποσοστό είναι 99.7%

Όπως γράψαμε την παραπάνω εξίσωση το ολοκλήρωμά της από $-\infty, +\infty$

$$\int_{-\infty}^{+\infty} G(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} dx = 1$$

$$\hat{\mu} = \bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{\mu} - x_i)^2$$

Για N μεγάλο $\hat{\mu} = \mu$
 $\hat{\sigma} = \sigma$

Πολλές φορές χρησιμοποιούμε αντί για την ποσότητα σ , το πλήρες εύρος της κατανομής στο μέσο του μέγιστου ύψους της (FWHM) που συμβολίζεται με Γ .

Από τον ορισμό, $G\left(\mu \pm \frac{\Gamma}{2}\right) = \frac{1}{2}G(\mu)$ έχουμε ότι $\Gamma = 2\sigma\sqrt{2\ln 2} \approx 2.35\sigma$

Εύρεση καλύτερης ευθείας προσαρμογής

Μέθοδος ελαχίστων τετραγώνων (μέθοδος χ^2)

Όπως είδαμε στη γραφική εύρεση της καλύτερης ευθείας προσαρμογής, αυτό που ενδιαφερόμαστε είναι ο καθορισμός των παραμέτρων της συνάρτησης (π.χ. μιας ευθείας) που περιγράφει τα δεδομένα

Δύο μέθοδοι: Ελαχίστων τετραγώνων ή χ^2

Μέγιστης πιθανότητας – maximum likelihood

Υποθέτουμε ότι έχουμε μια συνάρτηση y μιας μεταβλητής x και μια σειρά παραμέτρων $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$

$$y = y(x; \vec{\theta})$$

Έστω ότι μετρήσαμε διάφορες τιμές του y για κάποιες τιμές της ανεξάρτητης μεταβλητής x .

Επομένως θα έχουμε N ζεύγη τιμών $(x_i, y_i \pm \sigma_{y_i})$ όπου $i=1, \dots, N$

Ορίζουμε σαν “ χ^2 ” τη ποσότητα:

$$\chi^2(x; \vec{\theta}) = \sum_{i=1}^N \left[\frac{y_i - y(x_i; \vec{\theta})}{\sigma_i} \right]^2$$

Η μέθοδος των ελαχίστων τετραγώνων δηλώνει ότι η καλύτερη εκτίμηση των παραμέτρων θ_i επιτυγχάνεται όταν βρεθεί μια ομάδα τιμών θ_i για τις οποίες η συνάρτηση χ^2 είναι ελάχιστη

Μέθοδος ελαχίστων τετραγώνων – χ^2

Αν οι αποκλίσεις κάθε μέτρησης, σ_i , είναι ίσες τότε η σχέση απλουστεύεται

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^N [y_i - y(x_i; \vec{\theta})]^2$$

Είναι σημαντικό να προσέξετε ότι η μέθοδος όπως ορίστηκε χρειάζεται τη γνώση των αβεβαιοτήτων σ_i και ότι υποθέτει ότι δεν υπάρχουν αβεβαιότητες στη γνώση της ανεξάρτητης μεταβλητής x .

Αβεβαιότητες στις τιμές x_i μπορούν να αγνοηθούν εφόσον:

$$\frac{\sigma_{x_i}}{x_i} \ll \frac{\sigma_{y_k}}{y_k} \quad \mu\epsilon \quad i = 1, 2, \dots, N, \quad k = 1, 2, \dots, N$$

Αγνοώντας τις αβεβαιότητες, η μέθοδος ελαχίστων τετραγώνων είναι απλά το άθροισμα των αποστάσεων κάθε μέτρησης y_i από τα σημεία της θεωρητικής καμπύλης $(x_i, y_i(x_i))$. Ελαχιστοποιώντας τη συνάρτηση ψάχνουμε τα σημεία της καλύτερης καμπύλης για τα οποία αυτή η απόσταση ελαχιστοποιείται

Η εισαγωγή των αβεβαιοτήτων είναι απαραίτητη αν θέλουμε να κάνουμε στατιστική ανάλυση των αποτελεσμάτων διαφορετικά είναι απλό γεωμετρικό πρόβλημα

Εφαρμογή χ^2 – εύρεση παραμέτρων ευθείας

Ας υποθέσουμε ότι η συνάρτηση την οποία θέλουμε να προσαρμόσουμε στα σημεία των μετρήσεων μας είναι της μορφής

$$y(x) = ax + b$$

Θα υποθέσουμε ακόμα ότι οι αβεβαιότητες σ_i των y_i είναι όλες ίσες μεταξύ τους. Επομένως το πρόβλημα εύρεσης των παραμέτρων a και b έγγυται στην ελαχιστοποίηση της συνάρτησης χ^2 ως προς a και b .

Η ελαχιστοποίηση γίνεται πέρνοντας τη μερική παράγωγο της χ^2 ως προς a και b και εξισώνοντας με μηδέν:

$$\frac{\partial \chi^2}{\partial a} = 0 \quad \text{και} \quad \frac{\partial \chi^2}{\partial b} = 0$$

και λύνοντας το σύστημα των γραμμικών εξισώσεων που προκύπτει προς a και b

$$\left. \begin{aligned} \frac{\partial}{\partial a} \left[\sum_{i=1}^N (y_i - ax_i - b)^2 \right] &= -2 \sum_{i=1}^N (y_i - ax_i - b)x_i = 0 \\ \frac{\partial}{\partial b} \left[\sum_{i=1}^N (y_i - ax_i - b)^2 \right] &= -2 \sum_{i=1}^N (y_i - ax_i - b) = 0 \end{aligned} \right\} \Rightarrow \left. \begin{aligned} \sum_{i=1}^N y_i x_i &= a \sum_{i=1}^N x_i^2 + b \sum_{i=1}^N x_i \\ \sum_{i=1}^N y_i &= a \sum_{i=1}^N x_i + Nb \end{aligned} \right\} \Rightarrow b = \frac{\sum_{i=1}^N y_i}{N} - a \frac{\sum_{i=1}^N x_i}{N}$$

Αντικαθιστούμε στην 1^η εξίσωση και λύνουμε ως προς a :

$$\sum_{i=1}^N y_i x_i = a \sum_{i=1}^N x_i^2 + \frac{1}{N} \sum_{i=1}^N y_i \sum_{i=1}^N x_i - \frac{a}{N} \sum_{i=1}^N x_i \sum_{i=1}^N x_i \Rightarrow a \left[N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 \right] = N \sum_{i=1}^N y_i x_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i$$

Εύρεση παραμέτρων ευθείας με μέθοδο χ^2

Από την τελευταία εξίσωση παίρνουμε το a και αντικαθιστώντας στην εξίσωση του b παίρνουμε το b

$$a = \frac{N \sum_{i=1}^N y_i x_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\left[N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 \right]} = \frac{N \sum xy - \sum x \sum y}{\Delta} \quad \text{και} \quad b = \frac{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i x_i}{\left[N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 \right]} = \frac{\sum x^2 \sum y - \sum x \sum xy}{\Delta}$$

Αν διαιρέσουμε με N^2 τον αριθμητή και παρονομαστή στις παραπάνω σχέσεις

$$a = \frac{\overline{yx} - \bar{x} \bar{y}}{\left(\overline{x^2} - \bar{x}^2 \right)} \quad \text{και} \quad b = \frac{\bar{y} \overline{x^2} - \bar{x} \overline{xy}}{\left(\overline{x^2} - \bar{x}^2 \right)} \quad \text{όπου} \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

Παράδειγμα

Έστω ότι κάποιος φοιτητής θέλει να μετρήσει τις μάζες διαφόρων σωμάτων με μια ζυγαριά ελατηρίου. Θα πρέπει πριν μετρήσει τις μάζες να βαθμονομήσει τη ζυγαριά. Για να το κάνει αυτό χρησιμοποιεί 5 γνωστές μάζες των 2kgr τις οποίες τοποθετεί διαδοχικά πάνω στη ζυγαριά και μετρά κάθε φορά το αντίστοιχο μήκος του ελατηρίου l_k . Υποθέτοντας ότι το ελατήριο υπακούει στο νόμο του Hooke, περιμένει ότι $l = A + Bm$

➡ Η σταθερά A είναι το φυσικό μήκος τους εκκρεμούς και B είναι $B=g/k$, k η σταθερά ελατηρίου

Χρησιμοποιώντας την γραμμική σχέση $l = A + Bm$ μπορεί μετρώντας το μήκος του ελατηρίου για μια άγνωστη μάζα m να βρει τη μάζα. Χρειάζεται επομένως τις τιμές A και B.

Επομένως:

$$\Delta = N \sum m^2 - (\sum m)^2$$

$$\Rightarrow \Delta = 5 \times 220 - 30^2 = 200$$

Η σταθερά A είναι:

$$A = \frac{\sum m^2 \sum l - \sum m \sum l}{\Delta}$$

$$A = \frac{220 \times 256.6 - 30 \times 1622}{200}$$

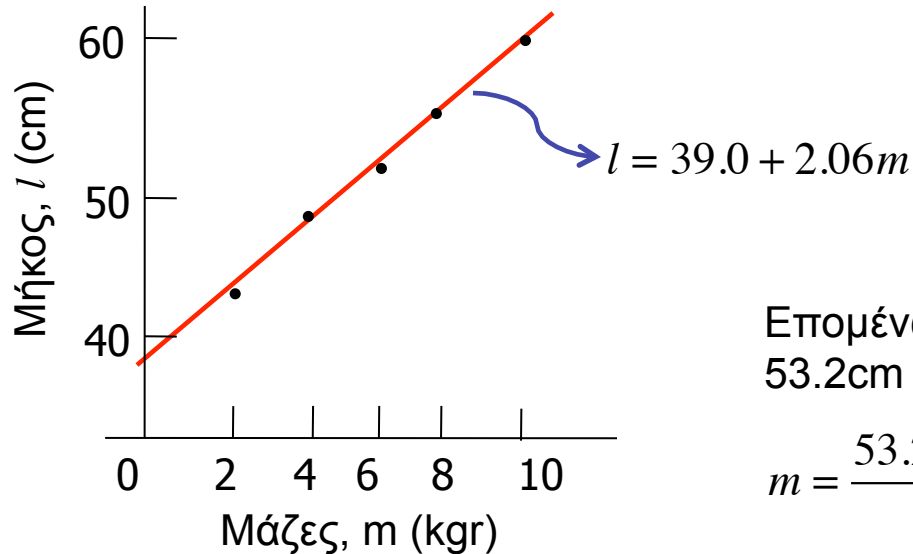
$$A = 39.0 \text{ cm}$$

Μέτρηση	Μάζα m	Μήκος, l	m^2	$m_i l_i$
1	2	42.0	4	84
2	4	48.4	16	194
3	6	51.3	36	308
4	8	56.3	64	450
5	10	58.6	100	586
N=5	$\sum m_i = 30$	$\sum l_i = 256.6$	$\sum m_i^2 = 220$	$\sum m_i l_i = 1622$

$$\text{Η σταθερά B θα είναι } B = \frac{N \sum ml - \sum m \sum l}{\Delta} = \frac{5 \times 1622 - 30 \times 256.6}{200} \Rightarrow B = 2.06 \text{ cm / kgr}$$

Γραφική παράσταση

Το γράφημα των προηγούμενων μετρήσεων θα είναι:



Επομένως αν κάποια μάζα επιμηκύνει το ελατήριο κατά 53.2cm τότε σύμφωνα με την εξίσωση της χ^2 ευθείας:

$$m = \frac{53.2 - 39.0}{2.06} = 6.9 \text{ kg}$$

Χρειάζεται να υπολογίσουμε τις αβεβαιότητες των A και B.

Αρχικά όμως ποια είναι η αβεβαιότητα των μετρήσεων y?

Έχουμε 5 μετρήσεις αλλά η διασπορά τους δεν μας δίνει την αβεβαιότητα τους. Ωστόσο κάθε μέτρηση περιμένουμε να κατανέμεται σύμφωνα με την Gaussian κατανομή γύρω από την αληθινή τιμή $y = A + Bx_i$ με εύρος σ_i

Επομένως όλες οι αποκλίσεις $y_i - A - Bx_i$ θα είναι κατανεμημένες Gaussian με κεντρική τιμή 0 και το ίδιο εύρος σ_y .

Το εύρος σ_y δίνεται από την σχέση

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - A - Bx_i)^2}$$

Εύρεση αβεβαιότητας παραμέτρων ευθείας

Οι αβεβαιότητες των a και b βρίσκονται από εφαρμογή διάδοσης σφαλμάτων και ...πολύ άλγεβρα

$$a = \sum_{i=1}^N \left[\frac{(x_i - \bar{x}) y_i}{(\overline{x^2} - \bar{x}^2)} \right] \Rightarrow \sigma_a^2 = \sum_{i=1}^N \sigma_{y_i}^2 \left(\frac{\partial a}{\partial y_i} \right)^2 \Rightarrow \sigma_a^2 = \sigma_y^2 \sum_{i=1}^N \left[\frac{x_i - \bar{x}}{N(\overline{x^2} - \bar{x}^2)} \right]^2 \Rightarrow \sigma_a^2 = \sigma_y^2 \sum_{i=1}^N \left[\frac{x_i^2 + \bar{x}^2 - 2\bar{x}x_i}{N^2(\overline{x^2} - \bar{x}^2)^2} \right]$$

$$\text{Αλλά } \sum_i (x_i^2 + \bar{x}^2 - 2\bar{x}x_i) = \sum_{i=1}^N x_i^2 + \sum_{i=1}^N \bar{x}^2 - 2\bar{x} \sum_{i=1}^N x_i = \sum_{i=1}^N x_i^2 + N\bar{x}^2 - 2\bar{x} N\bar{x} = \sum_{i=1}^N x_i^2 - N\bar{x}^2$$

Αντικαταστήσουμε την τελευταία σχέση στον αριθμητή της εξίσωσης για σ_a^2 οπότε

$$\sigma_a^2 = \sigma_y^2 \frac{\sum_{i=1}^N x_i^2 - N\bar{x}^2}{N^2(\overline{x^2} - \bar{x}^2)^2} = \sigma_y^2 \frac{(N\overline{x^2} - N\bar{x}^2)}{N^2(\overline{x^2} - \bar{x}^2)^2} = \sigma_y^2 \frac{N(\overline{x^2} - \bar{x}^2)}{N^2(\overline{x^2} - \bar{x}^2)^2} \Rightarrow \sigma_a^2 = \sigma_y^2 \frac{1}{N(\overline{x^2} - \bar{x}^2)}$$

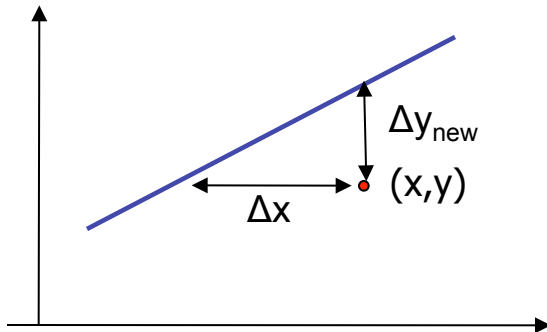
Ανάλογα βρίσκουμε και την αβεβαιότητα του b

$$\sigma_b^2 = \sigma_y^2 \frac{\bar{x}^2}{N(\overline{x^2} - \bar{x}^2)}$$

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - A - Bx)^2}$$

Περίπτωση με σφάλμα σε x και y

Αν οι μετρήσεις έχουν σφάλμα σε x , σ_x και y , σ_y , τότε μπορούμε να βρούμε ένα ισοδύναμο σφάλμα στην διεύθυνση y που προκαλεί το ίδιο αποτέλεσμα



Υποθέτοντας αρχικά ότι $\sigma_y = 0$

$$\Delta y_{new} = \frac{dy}{dx} \Delta x \Rightarrow \sigma_{y_{new}} = \frac{dy}{dx} \sigma_x$$

Αν σ_y δεν είναι 0 τότε το νέο σφάλμα μπορεί να γραφεί

$$\sigma_{y_{new}} = \sqrt{\sigma_y^2 + \left(\frac{dy}{dx} \sigma_x \right)^2}$$

αφού οι δύο αβεβαιότητες είναι ανεξάρτητες μεταξύ τους

Περίληψη – Μέθοδος χ^2

Οπότε συνοψίζοντας τη μέθοδο των ελαχίστων τετραγώνων για την περίπτωση της ευθείας και υποθέτοντας ότι τα σφάλματα των N επιμέρους μετρήσεων, y_i , είναι ίσα μεταξύ τους ότι :

κλίση

$$a = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\Delta}$$

$$\sigma_a^2 = \sigma_y^2 \frac{N}{\Delta}$$

όπου:

$$\Delta = N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2$$

τεταγμένη

$$b = \frac{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i - \sum_{i=1}^N x_i \sum_{i=1}^N x_i y_i}{\Delta}$$

$$\sigma_b^2 = \sigma_y^2 \frac{\sum_{i=1}^N x_i^2}{\Delta}$$

αβεβαιότητα μετρήσεων:

$$\sigma_y = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - A - Bx)^2}$$

Η συσχέτιση μεταξύ των x και y :

$$r^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\left(\sum_{i=1}^N x_i y_i - N \bar{x} \bar{y} \right)^2}{\left(\sum_{i=1}^N x_i^2 - N \bar{x}^2 \right) \left(\sum_{i=1}^N y_i^2 - N \bar{y}^2 \right)}$$

Μέθοδος ελαχίστων τετραγώνων – σ_{y_i} άνισα

Αν τα σφάλματα των μετρήσεων είναι διαφορετικά μεταξύ τους τότε η συνάρτηση που θα πρέπει να ελαχιστοποιήσουμε είναι

$$\chi^2 = \sum_{i=1}^N \left[\frac{y_i - ax_i^2 - b}{\sigma_i} \right]^2$$

Η ελαχιστοποίηση δίνει τις ίδιες εξισώσεις μόνο που στη περίπτωση αυτή οι μέσες τιμές των μεγεθών αντιστοιχούν σε αυτές που προκύπτουν με το να ζυγίσουμε τις τιμές με τα ανάλογα βάρη ($1/\sigma_i^2$) και η κανονικοποίηση δεν γίνεται ως προς N αλλά ως προς το συνολικό βάρος $\sum_{i=1}^N 1/\sigma_i^2$

Οι σχέσεις που δίνουν τις παραμέτρους a και b γίνονται ($y = ax + b$):

$$b = \frac{\sum wx^2 \sum wy - \sum wx \sum wxy}{\Delta}$$

$$a = \frac{\sum w \sum wxy - \sum wx \sum wy}{\Delta}$$

$$\sigma_b = \sqrt{\frac{\sum wx^2}{\Delta}}$$

$$\sigma_a = \sqrt{\frac{\sum w}{\Delta}}$$

$$\Delta = \sum w \sum wx^2 - \left(\sum wx \right)^2$$

Μετρήσεις ίδιου μεγέθους με διαφορετικά σφάλματα

Θεωρούμε κάποιο μέγεθος x το οποίο μετρήθηκε από διαφορετικούς παρατηρητές ή από τον ίδιο παρατηρητή κάνοντας διαφορετικά set μετρήσεων

Οι μετρήσεις οι οποίες βρέθηκαν ήταν : $x = x_A \pm \sigma_A$ και $x = x_B \pm \sigma_B$

Αφού έχουμε 2 πληροφορίες για το μέγεθος αυτό, θέλουμε να εξάγουμε από αυτές τις μετρήσεις το καλύτερο αποτέλεσμα για το μέγεθος x . **Αλλά πως?**

Αν η διαφορά των 2 τιμών $|x_A - x_B|$ είναι περισσότερο από $\pm 1 \max(\sigma_A, \sigma_B)$ τότε κάποια από τις μετρήσεις υποφέρει από κάποιο σφάλμα (συνήθως συστηματικό) το οποίο και θα πρέπει να απαλοιψουμε πριν συγκριθούν και συνδυαστούν

Από τη στιγμή που οι δύο μετρήσεις έχουν διαφορετικό σφάλματα δεν μπορούν να συνδυαστούν απευθείας για την εύρεση της μέσης τιμής. Θα δίνουμε την ίδια βαρύτητα και στις δυο μετρήσεις.

Η λογική λύση είναι να συνδυαστούν με τέτοιο τρόπο ώστε η τιμή με την καλύτερη ακρίβεια συνεισφέρει περισσότερο από τη λιγότερο ακριβή μέτρηση

Θεωρώντας ότι οι δυο μετρήσεις προκύπτουν από κατανομές Gauss και με με εύρος σ_A και σ_B αντίστοιχα και οι κατανομές έχουν αναμενόμενη μέση τιμή την άνωστη τιμή X του μεγέθους που θέλουμε να μετρήσουμε τότε οι πιθανότητες να μετρήσουμε x_A και x_B είναι

$$P(x_A) \propto \frac{1}{\sigma_A} e^{-\frac{(x_A - X)^2}{2\sigma_A^2}} \quad \text{και} \quad P(x_B) \propto \frac{1}{\sigma_B} e^{-\frac{(x_B - X)^2}{2\sigma_B^2}}$$

Μετρήσεις ίδιου μεγέθους με διαφορετικά σφάλματα

Η πιθανότητα για τα δύο αποτελέσματα ταυτόχρονα είναι

$$P(x_A, x_B) = P(x_A)P(x_B) \propto \frac{1}{\sigma_A} e^{-\frac{(x_A - X)^2}{2\sigma_A^2}} \frac{1}{\sigma_B} e^{-\frac{(x_B - X)^2}{2\sigma_B^2}} = \frac{1}{\sigma_A \sigma_B} e^{-\frac{1}{2} \left[\frac{(x_A - X)^2}{\sigma_A^2} + \frac{(x_B - X)^2}{\sigma_B^2} \right]} = \frac{1}{\sigma_A \sigma_B} e^{-\frac{y^2}{2}}$$

Αυτή η πιθανότητα θέλουμε να είναι μέγιστη  y^2 πρέπει να είναι ελάχιστο

Αλλά y^2 ελάχιστο όταν $\frac{dy^2}{dX} = 0$

Επομένως παραγωγίζοντας και εξισώνοντας με 0 έχουμε:

$$\frac{d}{dX} \left[\frac{(x_A - X)^2}{\sigma_A^2} + \frac{(x_B - X)^2}{\sigma_B^2} \right] = \frac{-2(x_A - X)}{\sigma_A^2} + \frac{-2(x_B - X)}{\sigma_B^2} = -2 \left[\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2} - X \left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} \right) \right] = 0 \Rightarrow X = \frac{\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2}}$$

Θέτοντας $w_A = \frac{1}{\sigma_A^2}$ και $w_B = \frac{1}{\sigma_B^2}$ σημαντικότητα ή βάρος της κάθε μέτρησης

$$X = \frac{x_A w_A + x_B w_B}{w_A + w_B}$$

Με το τρόπο αυτό η συμβολή της πιο ακριβής μέτρησης είναι μεγαλύτερη από της λιγότερο ακριβής

Για πολλές διαφορετικές μετρήσεις γράφουμε:

$$X = \frac{\sum x_i w_i}{\sum w_i}$$

$$\text{όπου } w_i = \frac{1}{\sigma_i^2}$$

Αβεβαιότητα της τιμής X συνδυασμού μετρήσεων

Χρησιμοποιώντας το γενικό τύπο διάδοσης σφαλμάτων για τις μετρήσεις x_i έχουμε

$$\sigma_X = \sqrt{\left(\frac{\partial X}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial X}{\partial x_2}\right)^2 \sigma_2^2 + \dots + \left(\frac{\partial X}{\partial x_N}\right)^2 \sigma_N^2} \quad \text{αλλά} \quad X = \frac{\sum x_i w_i}{\sum w_i} \quad \text{οπότε θα έχουμε:}$$

$$\sigma_X = \sqrt{\frac{w_1^2 \sigma_1^2}{(\sum w_i)^2} + \frac{w_2^2 \sigma_2^2}{(\sum w_i)^2} + \dots + \frac{w_N^2 \sigma_N^2}{(\sum w_i)^2}} \quad \text{αλλά} \quad w_i = \frac{1}{\sigma_i^2} \Rightarrow w_i^2 = \frac{1}{\sigma_i^4} \Rightarrow w_i^2 \sigma_i^2 = \frac{1}{\sigma_i^2}$$

$$\sigma_X = \sqrt{\frac{1}{\sigma_1^2 (\sum w_i)^2} + \frac{1}{\sigma_2^2 (\sum w_i)^2} + \dots + \frac{1}{\sigma_N^2 (\sum w_i)^2}} \Rightarrow \sigma_X = \frac{1}{\sum w_i} \sqrt{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2}}$$

$$\Rightarrow \sigma_X = \frac{1}{\sum w_i} \sqrt{w_1 + w_2 + \dots + w_N} \Rightarrow \sigma_X = \frac{(\sum w_i)^{1/2}}{\sum w_i} \Rightarrow \sigma_X = (\sum w_i)^{-1/2}$$

Επομένως έχουμε: $X = \sigma_X^2 \sum x_i w_i$ με αβεβαιότητα

$$\sigma_X^2 = \frac{1}{\sum w_i} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2}}$$

Θα μπορούσαμε να γράψουμε:

$$\frac{1}{\sigma_X^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2}$$

Η αβεβαιότητα της συνδυασμένης τιμής X μοιάζει σα το τύπο της ολικής αντίστασης N αντιστάσεων συνδεδεμένων παράλληλα

Παράδειγμα

Έστω ότι μετρήθηκε μια ωμική αντίσταση πολλές φορές από τρία άτομα και βρέθηκε $R_1=(11\pm1)\Omega$, $R_2=(12\pm1)\Omega$ και $R_3=(10\pm3)\Omega$. Ποιά η καλύτερη τιμή της R και ποια η αβεβαιότητα της;

Θα έχουμε $w_1 = \frac{1}{\sigma_1^2} = 1\Omega^{-1}$ $w_2 = \frac{1}{\sigma_2^2} = 1\Omega^{-1}$ και $w_3 = \frac{1}{\sigma_3^2} = \frac{1}{9}\Omega^{-1}$

Οπότε εφαρμόζοντας το τύπο: $X = \frac{\sum x_i w_i}{\sum w_i}$ παίρνουμε

$$X = \frac{11 \times 1 + 11 \times 1 + 10 \times \frac{1}{9}}{1 + 1 + \frac{1}{9}} = \frac{24.1}{2.1} = 11.4\Omega$$

Η αβεβαιότητα θα είναι: $\frac{1}{\sigma_X^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_N^2} \Rightarrow \frac{1}{\sigma_X^2} = \frac{1}{1} + \frac{1}{1} + \frac{1}{9} \Rightarrow \frac{1}{\sigma_X^2} = 2.1$

$$\Rightarrow \sigma_x = \sqrt{\frac{1}{2.1}} = 0.69\Omega$$

Έπομένως το τελικό αποτέλεσμα είναι: $X = 11.4 \pm 0.7 \Omega$

Αγνοώντας τη τρίτη μέτρηση με το μεγαλύτερο σφάλμα βρίσκουμε $X = 11.5 \pm 0.7 \Omega$

Δηλαδή η τρίτη και χειρότερη μέτρηση ουσιαστικά δεν συνεισφέρει στη τελική τιμή του μεγέθους

Κριτήριο απόρριψης τιμών

Έστω ότι κάποιος μετρώντας ένα μέγεθος 7 φορές βρήκε τις ακόλουθες τιμές

32.8 29.2 33.5 30.7 31.8 28.8 και 49.3

Η τελευταία μέτρηση φαίνεται αρκετά μεγαλύτερη από όλες τις υπόλοιπες και έξω από το διάστημα το οποίο περιέχει τις περισσότερες μετρήσεις

Στην περίπτωση αυτή ελέγχουμε για κάποιο λάθος. Αν δεν μπορεί να αποδοθεί λάθος μπορούμε να την απορρίψουμε σύμφωνα με το ακόλουθο κριτήριο

Αν πληρούνται οι προϋποθέσεις για κάνουν την κατανομή των μετρήσεών μας Gaussian τότε μπορούμε να υπολογίσουμε την πιθανότητα να προκύψει κάποια τιμή πολύ διαφορετική από τις υπόλοιπες.

Θεωρούμε όλες τις μετρήσεις αξιόπιστες και υπολογίζουμε τη μέση τιμή τους και τυπική απόκλιση

$$\bar{x} = 33.7 \quad \text{και} \quad \sigma_x = 7$$

Η μέτρηση που δεν συμφωνεί με τις υπόλοιπες διαφέρει από τη μέση τιμή κατά

$$\frac{|49.3 - \bar{x}|}{\sigma} = \frac{15.6}{7} = 2.23$$

Η πιθανότητα να προκύψει μια μέτρηση από μια Gaussian κατανομή με τιμή ίση ή και μεγαλύτερη από την μέση τιμή κατά $t=2.23\sigma$ δίνεται από:

$$1 - \int_{\bar{x}-t\sigma}^{\bar{x}+t\sigma} G(x)dx$$

Το ολοκλήρωμα της Gaussian βρίσκεται από πίνακες.

Στη περίπτωσή μας η πιθανότητα αυτή είναι 2.57%. Δηλαδή Αν είχαμε 100 μετρήσεις τότε 2.57 θα διέφεραν από τη μέση τιμή τουλάχιστον 2.23σ. Για 7 μετρήσεις έχουμε πιθανότητα 0.18.

Κριτήριο απόρριψης τιμών

Θεωρώντας αυθαίρετα σα κριτήριο απορριψιμότητας την πιθανότητα 0.5 ή μικρότερη μπορούμε να απορρίψουμε τη συγκεκριμένη μέτρηση.

Στη περίπτωση αυτή αφού απορρίψουμε τη μέτρηση θα πρέπει να υπολογίσουμε και πάλι τη μέση τιμή και απόκλιση των υπόλοιπων μετρήσεων

Μπορούμε να επαναλάβουμε τη διαδικασία για τις υπόλοιπες τιμές χρησιμοποιώντας την νέα μέση τιμή και απόκλιση. Συνήθως όμως προσπαθούμε να αποφύγουμε τη μεταγενέστερη απόρριψη τιμών γιατί μπορούμε να κάνουμε το σφάλμα μας τεχνικά πολύ μικρό

Πιθανότητες – ορολογία

Σε όλες τις περιπτώσεις που είναι αδύνατη η πρόβλεψη με βεβαιότητα το αποτέλεσμα ενός μοναδικού πειράματος καταφεύγουμε σε μια στατιστική περιγραφή του συστήματος η οποία αποτελείται από την εύρεση

(α) όλων των δυνατών εκδοχών

(β) της αντίστοιχης πιθανότητας εμφάνισης καθεμιάς από αυτές

Για την πιθανότητα εμφάνισης θεωρούμε μεγάλο αριθμό N από πανομοιότυπα συστήματα με αυτό που μελετούμε

Αν N_r είναι ο αριθμός των συστημάτων στα οποία εμφανίζεται η εκδοχή r

ορίζουμε σα **πιθανότητα εμφάνισης** $p_r = \frac{N_r}{N}$ για $N \rightarrow \infty$

Δηλαδή αν ο αριθμός των πειραμάτων γίνει πολύ μεγάλος περιμένουμε η επανάληψη της ίδιας παρατήρησης να οδηγεί σε αυξανόμενη δυνατότητα αναπαραγωγής σύμφωνα με το λόγο p_r και παύει να είναι αβέβαιος για $N \rightarrow \infty$

Οι τιμές της p_r μπορεί να είναι μεταξύ 0 και 1

Τυχαία θεωρούμε μια μεταβλητή της οποίας κάθε τιμή αντιστοιχεί σε κάθε δυνατή εκδοχή του πειράματος.

Διακεκριμένη αν οι τιμές που μπορεί να πάρει είναι αριθμήσιμες

Συνεχής αν μπορεί να πάρει οποιαδήποτε τιμή σε ένα διάστημα

Πιθανότητες – ορολογία

Συνάρτηση κατανομής (distribution function) ονομάζουμε τη συνάρτηση $P(n)$ που αντιστοιχεί κάθε τιμή που μπορεί να πάρει η διακεκριμένη τυχαία μεταβλητή x στη πιθανότητα να προκύψει η συγκεκριμένη αυτή τιμή

Συνάρτηση πυκνότητας πιθανότητας (distribution density function) ονομάζουμε τη συνάρτηση, $f(x)$, για την οποία η πιθανότητα να πάρει η συνεχής τυχαία μεταβλητή x , μια τιμή στο διάστημα $(x, x+dx)$ είναι $f(x)dx$.

Παραδείγματα

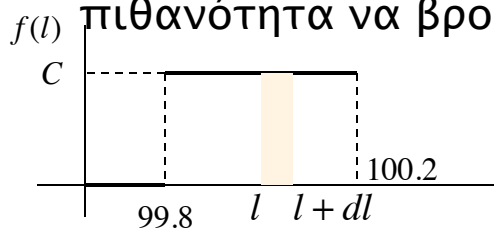
(1) Η στατιστική περιγραφή ενός νομίσματος περιγράφεται από

n	0	1
P(n)	0.5	0.5

(2) Η στατιστική περιγραφή ενός ζαριού περιγράφεται από

n	1	2	3	4	5	6
P(n)	1/6	1/6	1/6	1/6	1/6	1/6

(3) Έστω ότι μετριέται μια απόσταση $l = 100\text{cm}$ μεταξύ 2 σημείων και είμαστε σίγουροι ότι η πραγματική βρίσκεται μεταξύ 99.8 και 100.2cm. Αν θεωρήσουμε ότι κάθε τιμή μεταξύ 99.8 και 100.2cm έχει την ίδια πιθανότητα να προκύψει τότε η πυκνότητα πιθανότητας θα έχει τη μορφή του σχήματος και η πιθανότητα να βρούμε τη τιμή l μεταξύ $(l, l+dl)$ είναι το εμβαδό του παρ/μου



Επειδή η απόσταση l είναι μεταξύ 99.8 και 100.2cm

πρέπει:
$$\int_{99.8}^{100.2} f(l)dl = \int_{99.8}^{100.2} Cdl = 1 \Rightarrow C = \frac{1}{0.4} = 2.5$$

Πιθανότητες – σχέσεις

Το άθροισμα των πιθανοτήτων όλων των δυνατών αποτελεσμάτων είναι ίσο με τη μονάδα (συνθήκη κανονικοποίησης)

$$\sum_{i=1}^a p_i = 1 \quad \text{όπου } a \text{ είναι το πλήθος των δυνατών εκβάσεων}$$

Για συνεχή τυχαία μεταβλητή που μεταβάλλεται στο διάστημα $[a, b]$ η πιο πάνω συνθήκη είναι

$$\int_a^b f(x) dx = 1$$

Αν η πιθανότητα εμφάνισης μια εκδοχής r είναι p_r και η πιθανότητα εμφάνισης μιας εκδοχής s είναι p_s και οι εκδοχές είναι στατιστικά ανεξάρτητες τότε η πιθανότητα συνδυασμένης εμφάνισης και των δυο εκδοχών θα είναι

$$p_{rs} = p_r p_s$$

Αναμενόμενη τιμή μιας τυχαίας συνεχούς μεταβλητής x ορίζεται το ολοκλήρωμα

$$E[x] = \int_{-\infty}^{+\infty} x f(x) dx$$

Ενώ για τυχαία διακριτή μεταβλητή θα έχουμε:

$$E[x] = \sum_{i=1}^n x_i p(x_i)$$

Όπως φαίνεται από τον ορισμό της αναμενόμενης τιμής διακριτών μεταβλητών η ανεμενόμενη τιμή παριστά τη μέση τιμή των τιμών εμφάνισης της x

Διασπορά και Αναμενόμενη τιμή ομοιόμορφης κατανομής

Ορίζουμε σα **διασπορά** (variance) την ανεμενόμενη τιμή του τετραγώνου της διαφοράς μεταξύ της τυχαίας μεταβλητής x και της αναμενόμενης τιμής

$$V[x] = E[(x - E[x])^2] = \int (x - \mu)^2 p(x) dx$$

Από τον ορισμό έχουμε ότι $V[x] = E[x^2] - E[x]^2$

Συνήθως συμβολίζουμε τη διασπορά με σ και ορίζουμε $\sigma^2 = V[x]$

Για μια ομοιόμορφα κατανεμημένη τυχαία μεταβλητή στο διάστημα $[a, b]$ η συνάρτηση πυκνότητας πιθανότητας είναι

$$p(x) = \begin{cases} 1/(b-a) & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

Η ανεμενόμενη τιμή της x θα είναι επομένως:

$$E[x] = \int_a^b x p(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} \Rightarrow E[x] = \frac{a+b}{2}$$

Η διασπορά της θα είναι:

$$V[x] = E[x^2] - E[x]^2 = \int_a^b x^2 \frac{1}{b-a} dx - E[x]^2 = \frac{1}{3} \frac{b^3 - a^3}{b-a} - \left(\frac{a+b}{2} \right)^2 = \frac{1}{3} \frac{(b-a)(b^2 + a^2 + ab)}{b-a} - \frac{(a+b)^2}{4}$$

$$V[x] = \frac{4b^2 + 4a^2 + 4ab - 3a^2 - 3b^2 - 6ab}{12} = \frac{a^2 + b^2 - 2ab}{12} \Rightarrow V[x] = \frac{(a-b)^2}{12}$$

Τυχαίες μεταβλητές ομοιόμορφα κατανεμημένες

Αν υποθέσουμε ότι έχουμε μετρήσει μια φυσική ποσότητα N φορές και πάντοτε πέρνουμε την ίδια τιμή x_0 . Στην περίπτωση αυτή δεν μπορούμε να δούμε πως κατανέμεται στατιστικά το x . Μια λογική υπόθεση ωστόσο θα είναι ότι είναι κατανεμημένη ομοιόμορφα στο διάστημα $[x_0 - \delta x/2, x_0 + \delta x/2]$ όπου δx η ακρίβεια του οργάνου.

Κάτω από αυτές τις προϋποθέσεις η καλύτερη εκτίμηση της αβεβαιότητας στο x είναι

$$\sigma_x = \sqrt{V[x]} = \frac{b - a}{\sqrt{12}} = \frac{\delta x}{\sqrt{12}}$$

Αυτή είναι η περίπτωση που η **στατιστική αβεβαιότητα** δεν μπορεί να υπολογιστεί και χρησιμοποιούμε την αβεβαιότητα του οργάνου

Είναι καθαρό ότι η αναμενόμενη τιμή, διασπορά εξαρτάται από την συνάρτηση πυκνότητας πιθανότητας που ακολουθεί η μεταβλητή x

Κατανομές – Διωνυμική (binomial) κατανομή

Ας υποθέσουμε ότι κάνουμε κάποιο πείραμα το οποίο επιτρέπει να πάρουμε 2 αποτελέσματα/γεγονότα A και \tilde{A} . Το πρώτο έχει πιθανότητα $P(A) = p$ και το άλλο αναγκαστικά έχει πιθανότητα $P(\tilde{A}) = 1 - p$

Αν επαναλάβουμε το πείραμα N φορές τότε η πιθανότητα να έχουμε k φορές το γεγονός A (και N-k φορές το \tilde{A}) δίνεται από

$$P_{N,p}(k) = \binom{N}{k} p^k (1-p)^{N-k} \quad 0 \leq p \leq 1 \quad \text{όπου} \quad \binom{N}{k} = \frac{N!}{(N-k)!k!}$$

Η παραπάνω κατανομή αποτελεί τη διωνυμική κατανομή ή κατανομή Bernoulli

Η αναμενόμενη τιμή του k και η διασπορά του είναι

$$E[k] = Np$$

$$V[k] = Np(1-p)$$

Για μεγάλες τιμές του N ($N > 20$ και ιδιαίτερα για $p \sim 1/2$) $P(k)$ προσεγγίζεται από την gaussian κατανομή

Για τιμές $N \rightarrow \infty$ η διωνυμική κατανομή περιγράφεται από τη κατανομή Poisson

Η Διωνυμική κατανομή

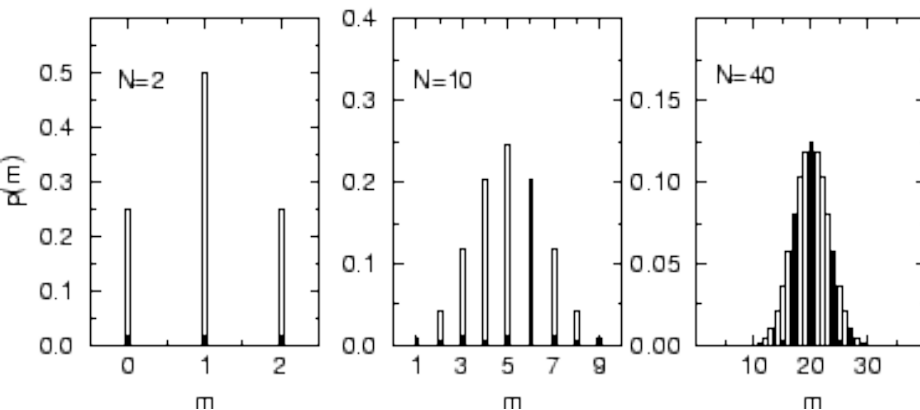
Έστω ότι ρίχνουμε ένα κέρμα N φορές. Ποια η πιθανότητα να πάρουμε m φορές τη μια όψη του?

Η πιθανότητα αυτή δίνεται από τη διωνυμική κατανομή η οποία λέει ότι:

Αν μια προσπάθεια η οποία δίνει επιτυχία με πιθανότητα p , τεθεί σε δοκιμή N φορές, η πιθανότητα $p(m)$ να βρούμε ακριβώς m επιτυχίες δίνεται από τη σχέση:

$$p(m) = {}^N C_m p^m (1-p)^{(N-m)} \quad \text{με} \quad {}^N C_m = \frac{N!}{m!(N-m)!}$$

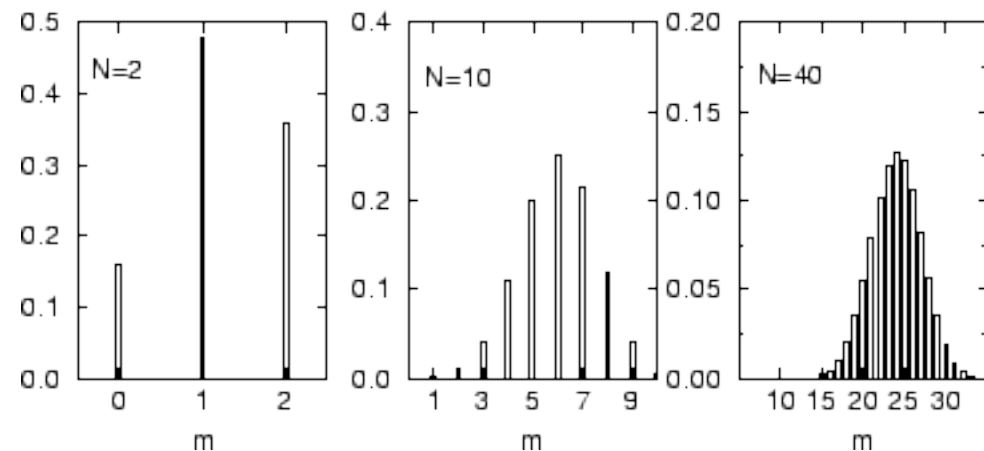
BINOMIAL DISTRIBUTION: $p=q=0.5$



Κανονικό νόμισμα

Κάλπικο νόμισμα

BINOMIAL DISTRIBUTION: $p=0.6, q=0.4$



Ιδιότητες της διονυμικής κατανομής

μέση τιμή: $\langle m \rangle \equiv \sum_{m=0}^N mp(m) = Np$

διασπορά: $V[m] \equiv \langle \Delta m^2 \rangle \equiv \sum_{m=0}^N (m - \langle m \rangle)^2 p(m) = Npq$

τυπική απόκλιση: $\sigma \equiv \sqrt{V[m]} = \sqrt{Npq}$

Η διονυμική κατανομή προσδιορίζεται από δυο παραμέτρους N και p.

Η μέση τιμή αυξάνει γραμμικά με τον αριθμό των προσπαθειών ενώ η τυπική απόκλιση μόνο συναρτήσει της τετραγωνικής ρίζας N

Κατανομές – Κατανομή Poisson

Αν θέσουμε στη διωνυμική κατανομή τις ακόλουθες συνθήκες

$$N \rightarrow \infty$$

$$p \rightarrow 0$$

$$Np = m = \text{σταθ.}$$

παίρνουμε τη λεγόμενη κατανομή Poisson $P_m(k) = \frac{m^k}{k!} e^{-m}$

Η αναμενόμενη τιμή και διασπορά της τιμής k είναι

$$E[k] = m \quad V[k] = m$$

Η κατανομή Poisson αντιπροσωπεύει τη πιθανότητα να μετρήσουμε ένα γεγονός k φορές όταν ο αριθμός των μετρήσεων N είναι πολύ μεγάλος ($N \gg m$)

Για $m < 20$ η $P_m(k)$ είναι αρκετά ασυμμετρική και η αναμενόμενη τιμή δεν συμπίπτει με το μέγιστο της κατανομής

Για $m > 20$ η $P_m(k)$ είναι αρκετά συμμετρική, η αναμενόμενη τιμή είναι κοντά στο μέγιστο της κατανομής και η κατανομή μπορεί να προσεγγιστεί με την gaussian κατανομή της οποίας $\mu = m$ και $\sigma = \sqrt{m}$

Αν k_1, k_2, \dots, k_n είναι οι μετρήσεις του k , μια καλή τιμή για εκτίμηση της m είναι η μέση τιμή των μετρήσεων και η εκτιμώμενη αβεβαιότητα σε μια μέτρηση

$$m = \frac{1}{n} \sum_{i=1}^n k_i \quad \sigma_k = \sqrt{m} \quad \text{Και η αβεβαιότητα του } m \text{ είναι} \quad \sigma_m = \sqrt{\frac{m}{n}}$$

Μέθοδος Likelihood

Έστω ότι έχουμε την ανεξάρτητη μεταβλητή x η οποία έχει σαν συνάρτηση πυκνότητας πιθανότητας την συνάρτηση f η οποία εξαρτάται από μια ομάδα μεταβλητών θ_i

$$f = f(x; \vec{\theta}) \quad \text{όπου} \quad \vec{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$$

Δεδομένων N ανεξάρτητων δειγμάτων του x , $\vec{x} = (x_1, x_2, \dots, x_N)$ ορίζουμε

$$L(\vec{x}; \vec{\theta}) = \prod_{i=1}^N f(x_i; \vec{\theta}) \quad \text{Συνάρτηση πιθανότητας της } f$$

Η συνάρτηση L είναι ανάλογη της πιθανότητας του να πάρουμε την ομάδα των δειγμάτων \vec{x} υποθέτοντας ότι τα N δείγματα είναι ανεξάρτητα μεταξύ τους

Η αρχή της μέγιστης πιθανότητας δηλώνει ότι η καλύτερη εκτίμηση των παραμέτρων $\vec{\theta}$ είναι η ομάδα των τιμών που μεγιστοποιούν την $L(\vec{x}; \vec{\theta})$

Εν γένει η μεγιστοποίηση της συνάρτησης δεν είναι εύκολη αναλυτικά και για το λόγο αυτό χρησιμοποιούνται διάφορες αριθμητικές μέθοδοι σε υπολογιστή

Μέθοδος Likelihood – Εκτίμηση σ και \bar{x} μεταβλητής με Gaussian κατανομή

Έστω ότι έχουμε N ανεξάρτητα δείγματα μιας τυχαίας μεταβλητής x η οποία κατανέμεται σύμφωνα με την Gaussian κατανομή αλλά δεν ξέρουμε την σ και μέση τιμή.

Πειραματικά η κατάσταση αυτή συναντιέται όταν μετράμε ένα φυσικό μέγεθος με το **ίδιο** όργανο μέτρησης.

Στη περίπτωση αυτή η συνάρτηση likelihood είναι:
$$L(\vec{x}; \vec{\theta}) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N e^{-\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{2\sigma^2}}$$

Λογαριθμίζουμε την προηγούμενη σχέση και παίρνουμε την παράγωγο της λογαριθμισμένης συνάρτησης ως προς x και σ

$$\ln L(\vec{x}; \vec{\theta}) = -N \ln \sqrt{2\pi}\sigma - \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{2\sigma^2}$$

Οπότε οι παράγωγοι ως προς x και σ

$$\left. \begin{aligned} \frac{\partial}{\partial \bar{x}} \left[-\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{2\sigma^2} \right] &= 0 \\ \frac{\partial}{\partial \sigma} \left[-N \ln \sigma - \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{2\sigma^2} \right] &= 0 \end{aligned} \right\} \begin{array}{l} \text{Συνθήκες} \\ \text{ελαχιστοποίησης} \end{array} \left\{ \begin{array}{l} \sum_{i=1}^N (x_i - \bar{x}) = 0 \\ \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma^3} - \frac{N}{\sigma} = 0 \end{array} \right.$$

Λύνοντας την 1^η εξίσωση ως προς \bar{x} έχουμε
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Αντικατάσταση στη 2^η πέρνουμε:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Σχέση μεθόδου χ^2 και likelihood

Υποθέτουμε ότι έχουμε ένα δείγμα δεδομένων το οποίο αποτελείται από N σημεία (x_i, y_i) . Τα x είναι γνωστά ακριβώς αλλά τα y είναι μετρημένα με μια αβεβαιότητα σ_i . Υποθέτουμε ακόμα ότι οι τιμές του y προέρχονται από μια συνάρτηση $f(x)$ και η οποία εξαρτάται από μια (ή περισσότερες) παράμετρο a την οποία και θέλουμε να εκτιμήσουμε. Αυτό θα δώσει το ιδανικό y το οποίο ωστόσο έχει αποκλίνει από τη αληθινή τιμή του εξαιτίας διαφόρων τυχαίων σφαλμάτων.

Από το θεώρημα της κεντρικής τιμής βρίσκουμε ότι η πιθανότητα για ένα συγκεκριμένο y που προκύπτει από κάποια δεδομένη τιμή του x είναι

$$P(y_i; a) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-[y_i - f(x; a)]^2 / 2\sigma_i^2}$$

Ο λογάριθμος της likelihood του δείγματος όπως προκύπτει από την παραπάνω σχέση δίνεται από

$$\ln L = -\frac{1}{2} \sum \left[\frac{y_i - f(x_i)}{\sigma_i} \right]^2 - \sum \ln \sigma_i \sqrt{2\pi}$$

Για να μεγιστοποιήσει τη likelihood χρειάζεται να ελαχιστοποιήσουμε τη $\sum \left[\frac{y_i - f(x_i; a)}{\sigma_i} \right]^2$ η οποία αποτελεί και τον ορισμό της συνάρτησης χ^2

Επομένως $\chi^2 = -2 \ln L$