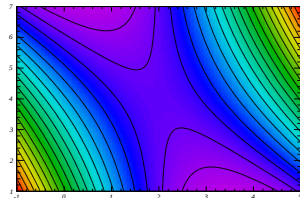


Statistical Methods in Data Analysis II

“classical” statistics

Michael Schmelling - MPI for Nuclear Physics

- Basics
- Error propagation
- Information, inference and confidence level intervals
- Least Squares and Maximum Likelihood
- Sandwich estimator and bootstrap
- Markov Chain Monte Carlo
- COWs and sPlot
- Unfolding



Literature – a non-exhaustive list in alphabetical order. . .

- K. Albertsson et al., <https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf>
- R.J. Barlow, *Statistics*, Wiley
- O. Behnke et al., *Data Analysis in High Energy Physics*, Wiley
- S. Brand, *Data Analysis*, Springer
- G.D. Cowan, *Statistical Data Analysis*, Oxford University Press
- B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap*, CRC Press
- F. James, *Statistical Methods in Experimental Physics*, World Scientific
- D.E. Knuth, *The Art of Computer Programming*, Addison Wesley
- W.T. Press et al., *Numerical Recipes*, Cambridge University Press
- D.S. Sivia, *Data Analysis - A Bayesian Tutorial*, Oxford University Press
- always useful, *Wikipedia*, *Stackoverflow*, *ChapGPT*, . . .

Abbreviations

- CLT: Central Limit Theorem
- EML: Extended Maximum Likelihood
- IID: Independent Identically Distributed
- IIDP: IID with Poisson distributed sample size
- LLN: Law of Large Numbers
- LSQ: Least Squares
- MC: Monte Carlo
- MCMC: Markov Chain Monte Carlo
- ML: Maximum Likelihood

❖ about statistical methods

▣ objectives: data reduction and classification

- ▶ condense a large data set into a single/few numbers
- ▶ identify distinct components of a data set (“signal” vs “background”)

▣ statistical methods are engineered

- ▶ characterised by their properties (more or less “useful” rather than “right” or “wrong”)
- ▶ properties need to be understood to judge applicability and interpret results

▣ requirements for obtaining quantitative results

- ▶ well defined data - usually independent identically distributed (IID)
- ▶ precisely posed questions
- ▶ exact definitions e.g. for “central value” or “width”
- ▶ math (calculus, linear algebra, theory of complex variables . . .)

examples →

Sleeping Beauty (SB), Prince Charming (PC) and a question (Q)



source: Charles Perrault, Histoires ou Contes du Temps passé

→ SB goes to sleep on Sunday; then PC flips a **fair coin**

■ if Head

- ▶ PC wakes up SB on Monday
- ▶ SB answers Q
- ▶ SB goes to sleep until Tuesday

■ if Tail

- ▶ PC wakes up SB on Monday
- ▶ SB answers Q
- ▶ PC erases memory of SB
- ▶ SB goes to sleep
- ▶ PC wakes up SB on Tuesday
- ▶ SB answers Q
- ▶ SB goes to sleep until Wednesday

► Q: What is the probability p for the coin to show Head?

■ answer 1: the probability is $p = 1/2$

► the coin is fair, it is flipped once

■ hence the probability for Head is $1/2$

■ answer 2: the probability is $p = 1/3$

► 3 indistinguishable cases for SB: Monday/Head, Monday/Tail, Tuesday/Tail

■ hence the probability for Head is $1/3$

► solution: consider N experiments with an expected $3N/2$ questions

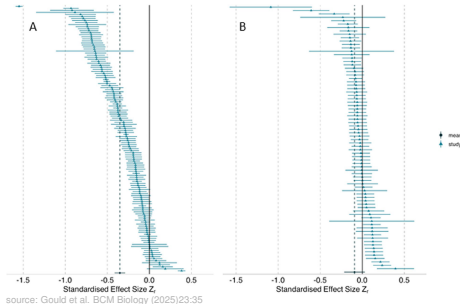
■ both answers are correct - they simply refer to different ensembles

► answer 1: the fraction of the experiments where the coin shows Head

► answer 2: the fraction of questions where the coin shows Head

see e.g.: Winkler, P.: The sleeping beauty controversy, The American Mathematical Monthly, 124, 2017

Gould et al.: same data, different analysts ...



→ different valid methods may give different valid answers!

❖ look for correlations ...

- A: between number of siblings and growth of nestling blue tits
- B: between grass cover and growth of Eucalyptus seedlings

■ meta-analyses

- ▶ all analyses peer reviewed
- ▶ transformation of all results to “standardized effect size”
- ▶ 2σ (95%) error bars
- ▶ assumptions of gaussianity

mathematical basis →

Kolmogorov's axioms on probability

- define **properties** of probabilities - don't care what they are!
- build probability theory on a mapping of **sets** → **real numbers**
- definitions:

Ω : the entire set

E : partial set of Ω

$p(E)$: probability of E

- Kolmogorov's axioms (versus, for illustration, an alternative)

(1) $p(E) \geq 0$ (versus $p(E) \in \mathbb{C}$)

(2) $p(\Omega) = 1$ (versus $p(\Omega)p^*(\Omega) = 1$)

(3) $p(E_1 \cup E_2) = p(E_1) + p(E_2)$ if $E_1 \cap E_2 = \emptyset$

→ it follows $0 \leq p(E) \leq 1$



source: wikipedia.org

(C) MFO / CC BY-SA 2.0 DE

Joint probability $p(E_1, E_2)$ for independent events E_1 and E_2

- most general ansatz

$$p(E_1, E_2) = f(p_1, p_2) = \sum_{k,l=0}^{\infty} a_{kl} p_1^k p_2^l \quad \text{with} \quad p_1 = p(E_1) \quad \text{and} \quad p_2 = p(E_2)$$

- boundary conditions if one of the events is certain: $p_i \in \{0, 1\}$

$$f(p_1, 1) = p_1 \quad \text{implies} \quad a_{kl} = 0 \quad \text{for} \quad k > 1$$

$$f(1, p_2) = p_2 \quad \text{implies} \quad a_{kl} = 0 \quad \text{for} \quad l > 1$$

- ▶ remaining

$$f(p_1, p_2) = a_{00} + a_{10}p_1 + a_{01}p_2 + a_{11}p_1p_2$$

- ▶ furthermore

$$f(0, p_2) = f(p_1, 0) = 0 \quad \text{implies} \quad a_{00} = a_{10} = a_{01} = 0$$

- ▶ finally

$$f(1, 1) = 1 \quad \rightarrow \quad a_{11} = 1 \quad \text{and thus} \quad p(E_1, E_2) = p(E_1)p(E_2)$$

problems and exercises

A deck of 32 cards has been thoroughly shuffled. From the top one card after another is drawn. The process stops when an ace is drawn. There are 4 aces in the deck. What is the most likely number of cards that are drawn?

The probability that a card is an ace is $4/32$, the probability that it is not an ace is $28/32$. So the probability to remove only one card is $4/32$, the probability to remove 2 cards is $(28/32)(4/31)$, to remove three cards is $(28/32)(27/31)(4/30)$ and so on. The most likely number of cards drawn thus is 1.

Conditional probabilities and Bayes' theorem

probability for B and A_k , with A_k from a set of disjoint possibilities $A_i, i = 1, \dots, n$

$$p(A_k, B) = p(B|A_k) p(A_k) = p(A_k|B) p(B)$$

→ Bayes' theorem:
$$p(A_k|B) = \frac{p(B|A_k) p(A_k)}{p(B)}$$

modification of the probability for A_k when B is given

□ alternative form for $p(B)$ when $\sum_i p(A_i) = 1$

$$p(B) = \sum_{i=1}^n p(B|A_i) p(A_i) \quad \text{and thus} \quad p(A_k|B) = \frac{p(B|A_k) p(A_k)}{\sum_i p(B|A_i) p(A_i)}$$



source: wikipedia.org

❖ Bayes' theorem in practice...

MODIFIED BAYES' THEOREM:

$$P(H|X) = P(H) \times \left(1 + P(C) \times \left(\frac{P(X|H)}{P(X)} - 1 \right) \right)$$

H: HYPOTHESIS

X: OBSERVATION

P(H): PRIOR PROBABILITY THAT H IS TRUE

P(X): PRIOR PROBABILITY OF OBSERVING X

P(C): PROBABILITY THAT YOU'RE USING
BAYESIAN STATISTICS CORRECTLY

source: www.xkcd.com

❖ example for applying Bayes' theorem to a hypothetical COVID-19 test

The probability to test positive in case of an infection is $p(+|I) = 0.98$, the probability for a healthy subject to test negative is $p(-|H) = 0.97$. One person in 1000 is actually infected.

→ determine the probability for a person that tested positive to be infected

probabilities:	$p(I)$	=	0.001	$p(H)$	=	0.999
	$p(+ I)$	=	0.980	$p(- I)$	=	0.020
	$p(+ H)$	=	0.030	$p(- H)$	=	0.970

The rows sum up to unity. Application of Bayes' theorem:

$$p(I|+) = \frac{p(+|I)p(I)}{p(+|I)p(I) + p(+|H)p(H)} \approx 0.032$$

- ▶ sweets for all patients diagnosed “infected” will yield 97% “healing rate”
- ▶ a really good test with $p(I|+) = 1$ has $p(+|H) = 0$

problems and exercises

Assume a PID system that discriminates between muons and pions. The probability to identify a muon as a muon is $p(\mu_{\text{id}}|\mu) = 0.999$, the probability to identify a pion as a pion is $p(\pi_{\text{id}}|\pi) = 0.95$. In an experiment the number of pions is 99 times larger than the number of muons. What is the probability that a track identified as muon is actually a muon?

$$p(\mu|\mu_{\text{id}}) = \frac{p(\mu_{\text{id}}|\mu)p(\mu)}{p(\mu_{\text{id}}|\mu)p(\mu) + p(\mu_{\text{id}}|\pi)p(\pi)} = \frac{0.999 \cdot 0.01}{0.999 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.168$$

Probability Density Functions (PDFs)

- non-negative, normalizable functions in n dimensions, e.g. $n = 2$

$$f(x, y) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x, y) = 1$$

- probability for (x, y) to be in $[x, x + dx] \times [y, y + dy]$

$$p(x, x + dx; y, y + dy) = \int_x^{x+dx} dx \int_y^{y+dy} dy f(x, y) \approx f(x, y) dx dy$$

- variables are independent if the PDF factorizes

$$f(x, y) = g_1(x) \cdot g_2(y)$$

❖ definition

subset of a function's domain where the function is not mapped to zero

■ examples

- ▶ gaussian: $x \in [-\infty, \infty]$
- ▶ exponential: $x \in [0, \infty]$
- ▶ uniform: $x \in [0, 1]$
- ▶ Poisson: $x \in \mathbb{N}_0$

■ remarks

- ▶ since $f(x) = 0$ outside the support, integrals can always be over the entire domain
- ▶ for a discrete support the PDF is a sum of delta-functions

$$f(x) = \sum_{k=0}^{\infty} p_k \delta(x - k) \quad \text{with discrete probabilities } p_k$$

Some common distributions

■ discrete probability distributions

► binomial: $p(n, N, p) = \binom{N}{n} p^n (1-p)^{N-n}$

► Poisson: $p(n, \mu) = e^{-\mu} \frac{\mu^n}{n!}$

■ continuous PDFs

► uniform: $f(x) = 1$ for $x \in [0, 1]$

► exponential: $f(x, \mu) = \frac{1}{\mu} e^{-x/\mu}$ for $x > 0$

► gaussian: $f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

► χ^2 : $f(x, n) = \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2} x^{n/2-1}$

number of successes n for N trials and success probability p

binomial limit $N \rightarrow \infty, p \rightarrow 0, Np = \mu$; independent events n recorded in time T with $\mu \propto T$

most basic random numbers

lifetime of an unstable particle

limit of a sum of many random variables

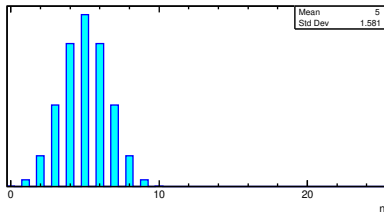
sum of n squares of $\mu = 0, \sigma = 1$ gaussian random variables

problems and exercises

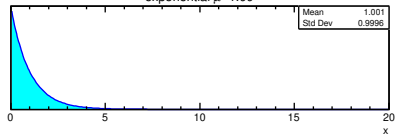
■ generate probability distributions

- ▶ a binomial distribution for $N = 20$ and $p = 1/2$
- ▶ Poisson distributions with mean values $\mu = 1$, $\mu = 10$ and $\mu = 100$
- ▶ exponential distributions with $\mu = 1$ and $\mu = 4$
- ▶ gaussian distributions with $\mu = 0$, $\mu = 1$, $\sigma = 1$ and $\sigma = 1/2$
- ▶ χ^2 -distributions with $\text{ndf} = 1$ and $\text{ndf} = 6$

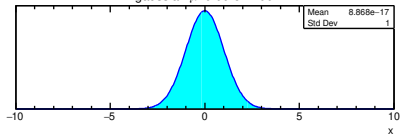
binomial $N=10$ $p=0.50$



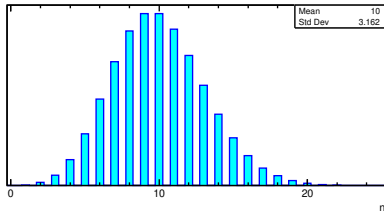
exponential $\mu=1.00$



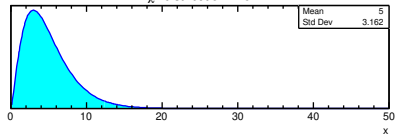
gaussian $\mu=0.00$ $\sigma=1.00$



Poisson distribution $\mu=10.00$



χ^2 -distribution $n=5$



Characterizing the location of a 1-dim distribution

- maximum

- ▶ not necessarily unique and possibly at one edge of the distribution

- median value m

$$\int_{-\infty}^m dx f(x) = \int_m^{\infty} dx f(x)$$

- ▶ insensitive to tails
- ▶ for discrete distributions usually not in the support

- mean value $\langle x \rangle$

$$\langle x \rangle = \int dx x f(x) \quad \text{or} \quad \langle n \rangle = \sum_{n=0}^{\infty} n p_n$$

- ▶ well defined for continuous and discrete distributions

- all three “typical values” coincide for symmetric uni-modal distributions

Characterising the width of a 1-dim distribution

- **FWHM**: full width at half the maximum value or
- **central q quantile $[a, b]$** , with, e.g., $q = 68.3\%$, 90% or 95%

$$\int_{-\infty}^a dx f(x) = \int_b^{\infty} dx f(x) = \frac{1}{2}(1 - q)$$

- ▶ both quantities are insensitive to tails
- ▶ for discrete distributions the interval end points are usually not in the support

- **standard deviation σ and variance σ^2**

$$\sigma^2 = \int_{-\infty}^{\infty} dx f(x) (x - \langle x \rangle)^2 \quad \text{or} \quad \sigma^2 = \sum_{n=0}^{\infty} (n - \langle n \rangle)^2 p_n$$

- ▶ well defined for continuous and discrete distributions
- ▶ sensitive to tails and may even diverge

Mean value and standard deviation

a generic measure for the scatter s of x with PDF $f(x)$ around a point a is

$$s^2 = \int dx (x - a)^2 f(x)$$

for s to characterize $f(x)$, a should be chosen to minimize s

$$\frac{\partial s^2}{\partial a} = -2 \int dx (x - a) f(x) \stackrel{!}{=} 0 \quad \text{i.e.} \quad a = \int dx x f(x) = \langle x \rangle$$

■ findings

- ▶ the mean value $\langle x \rangle$ is the location parameter that minimizes s
- ▶ the standard deviation σ or RMS-width is the minimal s

Expectation values

❖ integral transform of a PDF $f(x)$

$$\langle a \rangle = \int_{-\infty}^{\infty} dx \, a(x) f(x)$$

- mapping function $f(x) \rightarrow$ real number
- different functions $a(x)$ probe different properties of $f(x)$, e.g.
 - ▶ $a(x) = x$: expectation value = mean value
 - ▶ $a(x) = (x - \langle x \rangle)^2$: expectation value = variance
- an expectation value is a linear operator

$$\langle \alpha A + \beta B \rangle = \alpha \langle A \rangle + \beta \langle B \rangle$$

- ▶ application

$$\langle (x - \langle x \rangle)^2 \rangle = \langle x^2 - 2x \langle x \rangle + \langle x \rangle^2 \rangle = \langle x^2 \rangle - 2 \langle x \rangle \langle x \rangle + \langle x \rangle^2 = \langle x^2 \rangle - \langle x \rangle^2$$

Covariance and correlation

look for expectation values that are sensitive to dependencies between two variables

0th order $\langle 1 \rangle$

1st order $\langle x \rangle, \langle y \rangle$

2nd order $\langle x^2 \rangle, \langle xy \rangle, \langle y^2 \rangle$

The **lowest order** term probing dependencies is $\langle xy \rangle$. For **independent** variables one finds

$$\begin{aligned}\langle xy \rangle &= \int dx \int dy (x y) g_1(x) g_2(y) \\ &= \left(\int dx x g_1(x) \right) \left(\int dy y g_2(y) \right) = \langle x \rangle \langle y \rangle\end{aligned}$$

→ correlation-sensitive: $C_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$ the “covariance” between x and y

lowest order correlation-sensitive expression, zero for independent variables

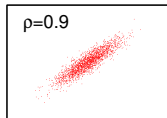
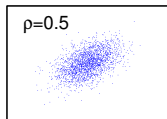
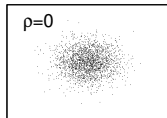
❖ dimensionless measure of correlation: the correlation coefficient

$$\rho = \frac{C_{xy}}{\sigma_x \sigma_y} = \frac{C_{xy}}{\sqrt{C_{xx} C_{yy}}} = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \sqrt{\langle y^2 \rangle - \langle y \rangle^2}}$$

- invariance under $x \rightarrow a_x x + b_x$ and $y \rightarrow a_y y + b_y$
- not invariant under rotations
- $-1 \leq \rho \leq 1$, $|\rho|$ grows with the tightness of the correlation
- for independent variables x and y one has $\rho = 0$
- $\rho = 0$ is not sufficient for independence of x and y

► example: pairs $\{x_i, y_i\}$ with $y_i = x_i^k$ and $\langle x \rangle = \langle x^3 \rangle = 0$

$$\rho = \frac{\langle x^{k+1} \rangle - \langle x^k \rangle \langle x \rangle}{\sqrt{\langle x^2 \rangle - \langle x \rangle^2} \sqrt{\langle x^{2k} \rangle - \langle x^k \rangle^2}} \rightarrow \begin{cases} k=1 \rightarrow \rho=1 \\ k=2 \rightarrow \rho=0 \end{cases}$$



❖ the covariance matrix

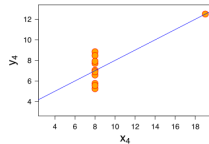
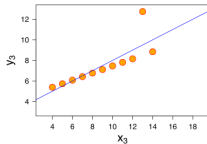
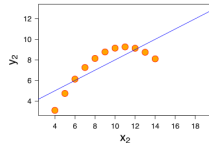
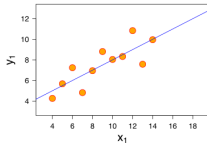
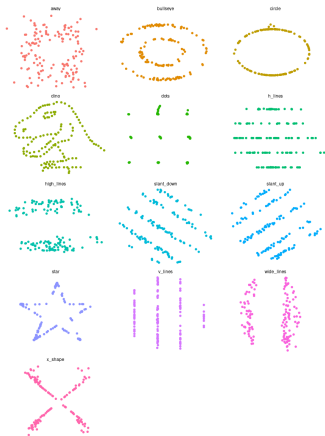
array of all variances and covariances of an n -dim PDF

$$C = \langle x x^T \rangle - \langle x \rangle \langle x \rangle^T \quad \text{or} \quad C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = \rho_{ij} \cdot \sigma_i \sigma_j$$

➔ notation: M = entire matrix object, M_i, M_{ij} = individual components

- diagonal terms C_{ii} : variances of the individual variables
- off-diagonal terms C_{ij} : covariances between pairs of variables
- properties of the covariance matrix:
 - ▶ symmetric and positive definite
 - ▶ can be diagonalized by rotation in the space of the variables
 - ▶ describes extension and orientation of a PDF

not all information is contained in mean values and variances ...



- two examples for groups of different data sets with the same mean values and variances

The Law of Large Numbers (LLN)

❖ the average of $a(x_i)$ with x_i drawn from a PDF $f(x)$ converges to $\langle a \rangle$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N a(x_i) = \langle a \rangle = \int dx a(x) f(x)$$

► proof

- consider infinitesimal intervals (bins) $[x_{\text{bin}}, x_{\text{bin}} + \Delta x]$
- probability for an entry in a bin: $p(x) \approx \Delta x f(x)$
- number of entries: $n_{\text{bin}} \approx N p(x_{\text{bin}}) \approx N \Delta x f(x_{\text{bin}})$
- average:

$$\frac{1}{N} \sum_{i=1}^N a(x_i) \approx \frac{1}{N} \sum_{\text{bins}} a(x_{\text{bin}}) n_{\text{bin}} \approx \sum_{\text{bins}} \Delta x a(x_{\text{bin}}) f(x_{\text{bin}})$$

- for $N \rightarrow \infty$ and $\Delta x \rightarrow 0$ the approximations become exact → proof

→ basis of MC method: estimate an integral by an expectation value

The Central Limit Theorem (CLT)

► for $N \rightarrow \infty$ the sum of N random variables is a gaussian distributed random variable

- holds under very general conditions
- main requirement: **finite variance** of all random variables
- simplest case: x_i with $\langle x_i \rangle = 0$ and $\sigma^2(x_i) = 1$

$$y = \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i \quad \text{has the limiting PDF} \quad \lim_{N \rightarrow \infty} f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

- $\langle x_i \rangle = 0$ ensures that the mean of y is finite
- $\sigma(x_i) = 1$ ensures the y has a gaussian PDF with unit variance
- not required: IID variables x_i

problems and exercises

Do a numerical test of the central limit theorem with

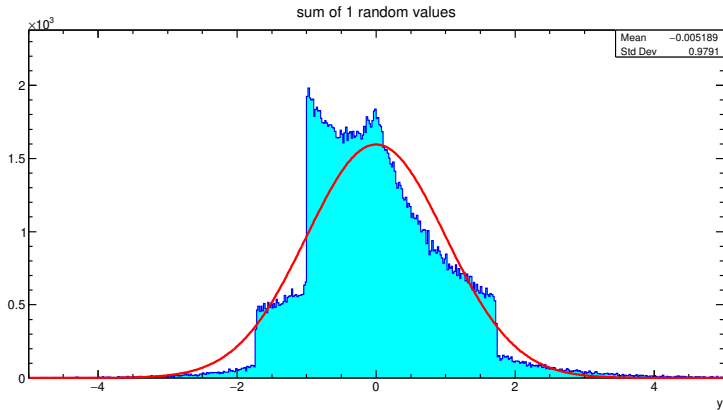
- uniformly distributed random numbers in the range $\pm\sqrt{3}$
- exponentially distributed random numbers $f(x) = e^{-x-1}$ for $x \geq -1$
- Laplace-distributed random numbers drawn from $f(x) = e^{-|x|\sqrt{2}}/\sqrt{2}$

and determine the distribution of

$$y = \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i \quad \text{for } N = 1, 10, 100, 1000 .$$

- ▶ a) do the test separately for the three types
- ▶ b) do a test with randomly picking from the three types

numerical study



The Bienaymé-Chebycheff inequality

❖ general result about the probability content in the tails of a PDF

- PDFs can be characterised by mean μ and standard deviation σ
- usually small probability content beyond few σ around μ , e.g. gaussian distribution:
 - ▶ outside $\pm 1\sigma$: probability content $p_1 = 0.317$
 - ▶ outside $\pm 2\sigma$: probability content $p_2 = 0.054$
 - ▶ outside $\pm 3\sigma$: probability content $p_3 = 0.004$
- there is a limit that holds for every PDF:

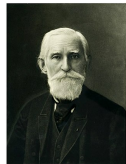
$$p_k = P(|x - \mu| > k\sigma) \leq \frac{1}{k^2}$$

- ▶ consequence of the fact that a PDF is positive and normalized



source: wikipedia.org

Irénée Jules Bienaymé

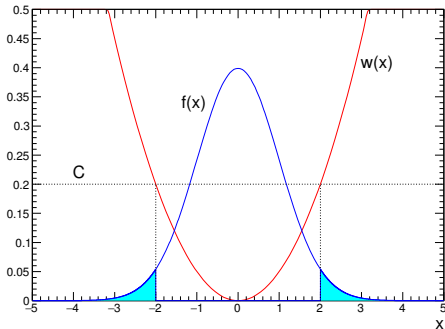


Павлутий Чебышев
source: wikipedia.org

Pafnuty Chebyshev

→ proof: consider a PDF $f(x)$, a function $w(x) \geq 0$ and an x -region with $w(x) \geq C$

$$P(w(x) \geq C) = \int_{w(x) \geq C} dx f(x) \leq \int_{w(x) \geq C} dx \frac{w(x)}{C} f(x) \leq \int dx \frac{w(x)}{C} f(x) = \frac{\langle w \rangle}{C}$$



■ specialisation:

► $w(x) = (x - \mu)^2$

► $\langle w \rangle = \sigma^2$

► $C = k^2 \sigma^2$

■ result:

$$P((x - \mu)^2 > k^2 \sigma^2) \leq \frac{1}{k^2}$$

Transformations of 1-dim PDFs

- given: $F(\alpha)$
- change of variable: $\alpha = g(\beta)$
- wanted: $F_g(\beta)$

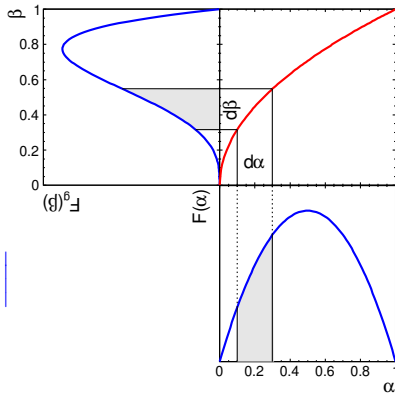
→ conservation of probability:

$$F_g(\beta) |d\beta| = F(\alpha) |d\alpha|$$

it follows, with $\alpha = g(\beta)$,

$$F_g(\beta) = F(g(\beta)) \left| \frac{d\alpha}{d\beta} \right| = F(g(\beta)) \left| \frac{dg(\beta)}{d\beta} \right|$$

$$\left| \frac{dg(\beta)}{d\beta} \right| = |\text{determinant of Jacobian}|$$



Transformation of multi-dimensional PDFs

❖ given $f(x_1, x_2)$ determine the PDF of $y = h(x_1, x_2)$

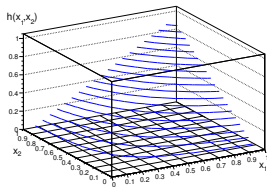
■ starting point: probability $G(Y)$ for $y = h(x_1, x_2) < Y$

$$\begin{aligned} G(Y) &= \int dx_1 dx_2 f(x_1, x_2) \theta(Y - h(x_1, x_2)) \\ &= \int_{-\infty}^Y dy g(y) \end{aligned}$$

■ get $g(y)$ by differentiation with respect to the upper limit Y :

$$g(y) = \left. \frac{d}{dY} G(Y) \right|_{Y=y} = \int dx_1 dx_2 f(x_1, x_2) \delta(y - h(x_1, x_2))$$

- ▶ integrate over input PDF with δ -function for the transformed variable
- ▶ generalizes trivially to any dimension



Numerical transformation of PDFs

- given the n -dim PDF $f(x_1, x_2, \dots, x_n)$, determine the PDF of $y = h(x_1, x_2, \dots, x_n)$
- solution:
 - ▶ generate n -tuples $\{x_1, x_2, \dots, x_n\}$
 - ▶ histogram $y = h(x_1, x_2, \dots, x_n)$
- easy to realize if $f(x_1, x_2, \dots, x_n)$ factorizes into simple 1-dim distributions

problems and exercises

Determine the PDF $h(v)$ of $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$ when v_x , v_y and v_z are gaussian distributed random numbers with zero mean value and unit variance.

❖ analytical solution - transforming from cartesian to polar coordinates

$$h(v) = \frac{1}{(\sqrt{2\pi})^3} \int dv_x dv_y dv_z e^{-v_x^2/2 - v_y^2/2 - v_z^2/2} \delta\left(v - \sqrt{v_x^2 + v_y^2 + v_z^2}\right)$$
$$= \frac{1}{(\sqrt{2\pi})^3} \int d\cos\theta d\phi du u^2 e^{-u^2/2} \delta(v - u) = \sqrt{\frac{2}{\pi}} v^2 e^{-v^2/2}$$

❖ numerical solution

