Skolkovo Institute of Science and Technology

Stochastic methods in Mathematical Modelling

Lecture 4. Laws of large numbers

Convergence in probability vs convergence in distribution

A sequence $\{X_n\}$ of random variables ***converges in probability*** towards the random variable X if for all ε > 0

$$\lim_{n\to\infty} Pr\left(|X_n - X| > \varepsilon\right) = 0.$$

A sequence X_1, X_2, … of real-valued RVs, with CDFs F_1, F_2, … , is said to ***converge in distribution***, or *converge weakly*, or *converge in law* to a random variable X with CDF F if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every number $x \in R$ at which F is continuous.

Convergence in probability implies convergence in distribution

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X,$$

Skoltech
Skolkovo Institute of Science and Technology

Central Limit Theorem (classical)

Weak law of large numbers

sample average converges in probability towards the expected value

$$\lim_{n \to \infty} \bar{X}_n \xrightarrow{P} \mu$$

$$\lim_{n \to \infty} Pr\left(|\bar{X}_n - \mu| < \epsilon\right) = 1 \ \forall \ \epsilon > 0$$

Central limit theorem

Take n independent and identically distributed (i.i.d) variables $x_1, \cdots, x_n$ from a distribution with mean μ and variance, $\sigma > 0$. Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right) \xrightarrow{d} N(0, \ \sigma^2)$$

Q: How does one prove it?

Weak law of large numbers

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_n}{n} \xrightarrow{P} \mu$$

$E[\overline{X}_n] = \mu$

$Var[\overline{X}_n] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Chebyshev's inequality

$$Pr(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{Var[\overline{X}_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

$\forall \varepsilon \quad n \to \infty \quad \Rightarrow \lim_{n\to\infty} Pr(|\overline{X}_n - \mu| \leq \varepsilon) = 0$

$\Rightarrow \lim_{n\to\infty} Pr(|\overline{X}_n - \mu| > \varepsilon) = 1$

CLT proof

$\bar{X}_n = \frac{\sum_{i=1}^{n} x_i}{n}$ x_i are i.i.d from $P(x)$ with μ, σ

$z_i = \frac{\sqrt{n}(x_i - \mu)}{\sigma} \Rightarrow \bar{z}_n = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$

$\sigma = 1$:

$\mu_{z_n} = \mu_z = 0$, $\sigma_z = \sqrt{n}$ & $\sigma_{z_n} = 1$

The characteristic function for $P_n(z_n)$

$g_n(k) = \langle e^{ikz_n} \rangle = \int dz_n P_n(z_n) e^{ikz_n} =$

$= \int dz_1 \dots dz_n \, p(z_1) \dots p(z_n) e^{ik\frac{z_1 + \dots + z_n}{n}} = \left(\int dz \, p(z) \, e^{\frac{ikz}{n}} \right)^n = G^n\left(\frac{k}{n}\right)$

$G(k) \simeq 1 - \frac{\sigma_z^2 k^2}{2} + O(k^3) \simeq 1 - \frac{nk^2}{2}$

$g_n(k) = G^n\left(\frac{k}{n}\right) = \left(1 - \frac{k^2}{2n}\right)^n \overset{n\to\infty}{\approx} e^{-k^2/2} \Rightarrow P(\bar{z}_n) \xrightarrow{d} N(0,1)$

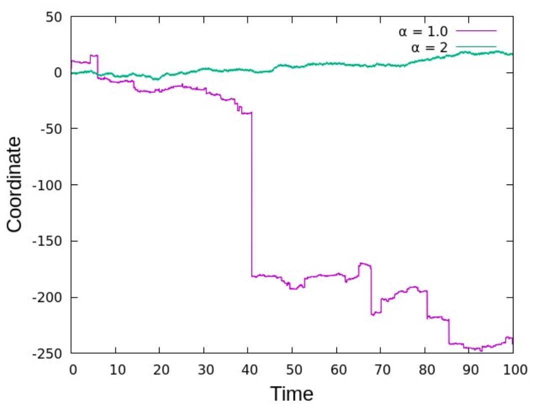
Central Limit Theorem (classical)

Take n independent and identically distributed (i.i.d) variables $x_1, \cdots, x_n$ from a distribution with mean μ and variance, $\sigma > 0$. Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right) \xrightarrow{d} N(0,\ \sigma^2)$$

What happens if a mean or a variance do not exist (i.e. diverge)?

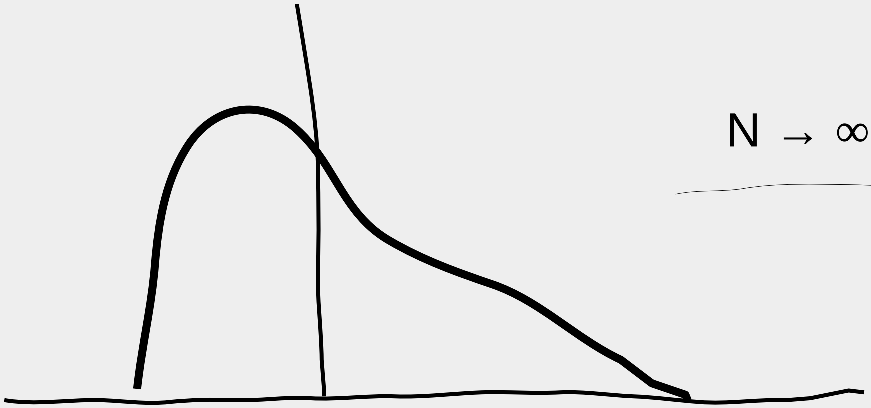
Generalisation of CLT = GLT = generalised central limit theorem: stable laws!



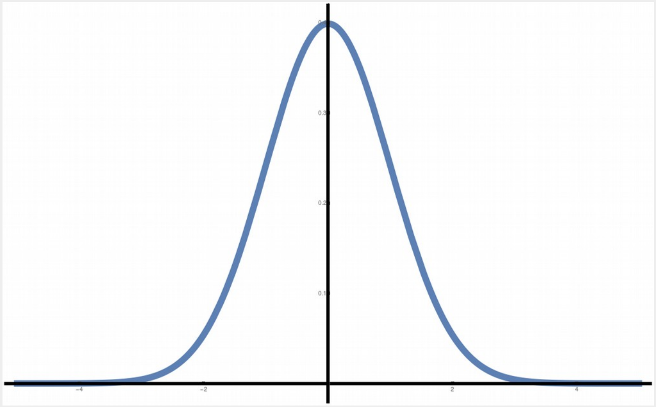
Reminder: a Gaussian process (green) vs a heavy-tailed Lorentz-Cauchy process with $f(x) \sim 1/x^2$ for $x \gg 1$

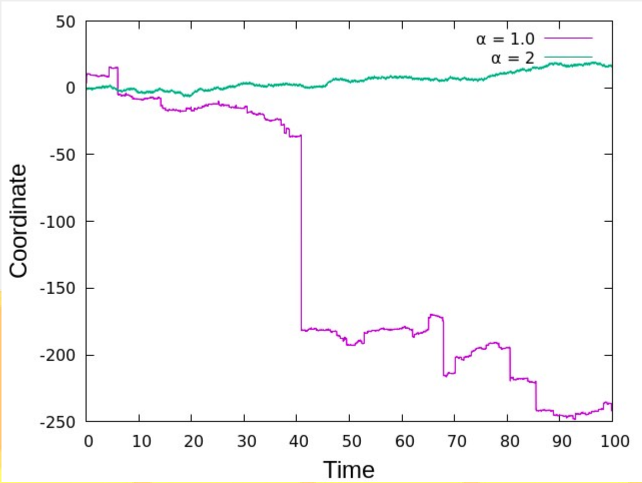
Classical Central Limit Theorem. Issues

1) What happens if a mean or a variance do not exist (i.e. diverge)?
2) Produces exponential tails
3) Produces symmetric distribution



$N \to \infty$



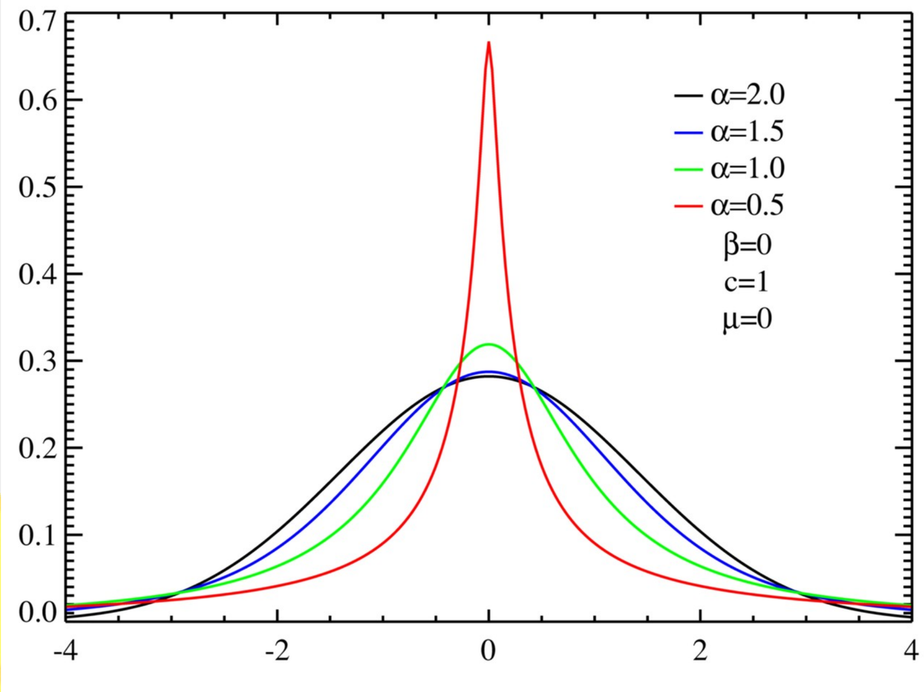


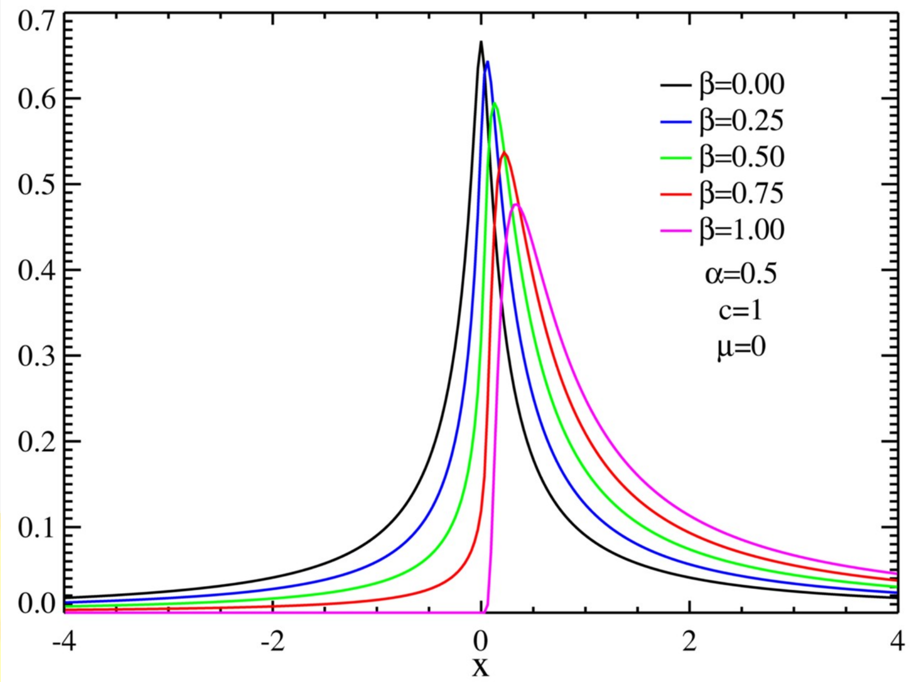
Reminder: a Gaussian process (green) vs a heavy-tailed Lorentz-Cauchy process with $f(x) \sim 1/x^2$ for $x \gg 1$

Stable laws (distributions)

- According to the ***generalised central limit theorem*** they are the limits of sums of i.i.d
- Allow for skewness and heavy-tails
- Useful model since lots of data sets exhibit skewness and heavy-tails

$$\varphi(\omega)=e^{i\omega\mu-a|\omega|^{\alpha}[1-i\beta Sign(\omega)\tan(\pi\alpha/2)]},\alpha\neq 1$$





Stable distributions

For Gaussian variables the sum of two of them is itself a normal random variable.
A consequence of this is that if X is normal, then for X_1 and X_2 independent copies of X and any positive constants *a* and *b*

$$a\,X_1 + bX_2 \stackrel{d}{=} cX + d$$

Definition. Let X_1 and X_2 be independent copies of *X*. Then *X* is said to be ***stable*** if for any constants $a > 0$ and $b > 0$ the random variable $aX_1 + bX_2$ has *the same distribution* as $cX + d$ for some constants $c > 0$ and *d*. The distribution is said to be ***strictly stable*** if this holds with $d = 0$.

Example: Gaussian variables

Q: What about Cauchy-Lorentz distribution?

$$P(x) = \frac{1}{\pi(x^2+1)}$$

Addition rule for Gaussian case

P_{X+Y} $\quad \tilde{y} = aX + bY$

$$\langle e^{iky} \rangle = \langle e^{ikaX + ikbY} \rangle = \langle e^{ikaX} \rangle \langle e^{ikbY} \rangle =$$

$$= e^{ia\mu_x k + \frac{a^2\sigma_x^2 k^2}{2}} e^{ib\mu_y k + \frac{b^2\sigma_y^2 k^2}{2}} = e^{ik(a\mu_x + b\mu_y) + \frac{k^2\sigma^2}{2}}$$

$$N(a\mu_x; a\sigma_x^2),\ N(b\mu, (b\sigma)^2) \qquad N(c\mu + d, (c\sigma)^2)$$

$\mu_x = \mu_y$

$\sigma_x = \sigma_y$

$$\begin{cases} c\mu + d = a\mu + b\mu \\ c^2 = a^2 + b^2 \end{cases} \to \underline{d, c}$$

Cauchy

$$P(x) = \frac{1}{\pi\gamma\left(1 + \frac{x^2}{\gamma}\right)}$$

$$\langle e^{ik\tilde{y}} \rangle = \langle e^{ikaX_1} \rangle \langle e^{ikbX_2} \rangle = e^{-|ka| - |kb|}$$

$$= e^{-|k|(|a| + |b|)}$$

$\tilde{y} = aX_1 + bX_2$

$$\begin{cases} c = |a| + |b| \\ d = 0 \end{cases}$$

Remark

CF for Cauchy

$$\varphi(\omega) = e^{-N\gamma|\omega|} \xrightarrow{\text{inverse Fourier}} f_N(x) = \frac{1}{\gamma N \pi\left(1 + \left(\frac{x}{N\gamma}\right)^2\right)}$$

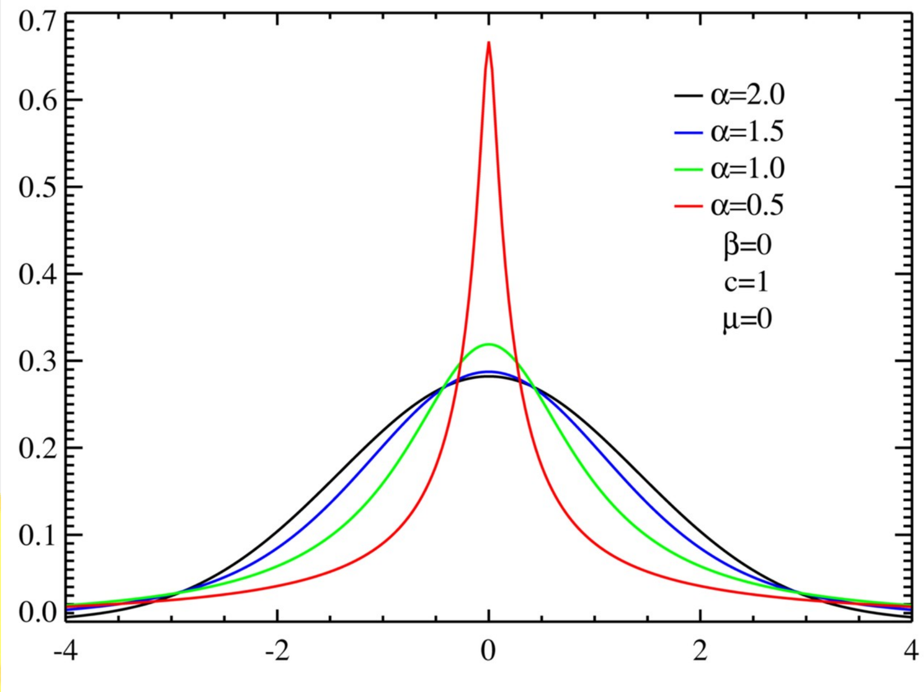
Generalised CLT

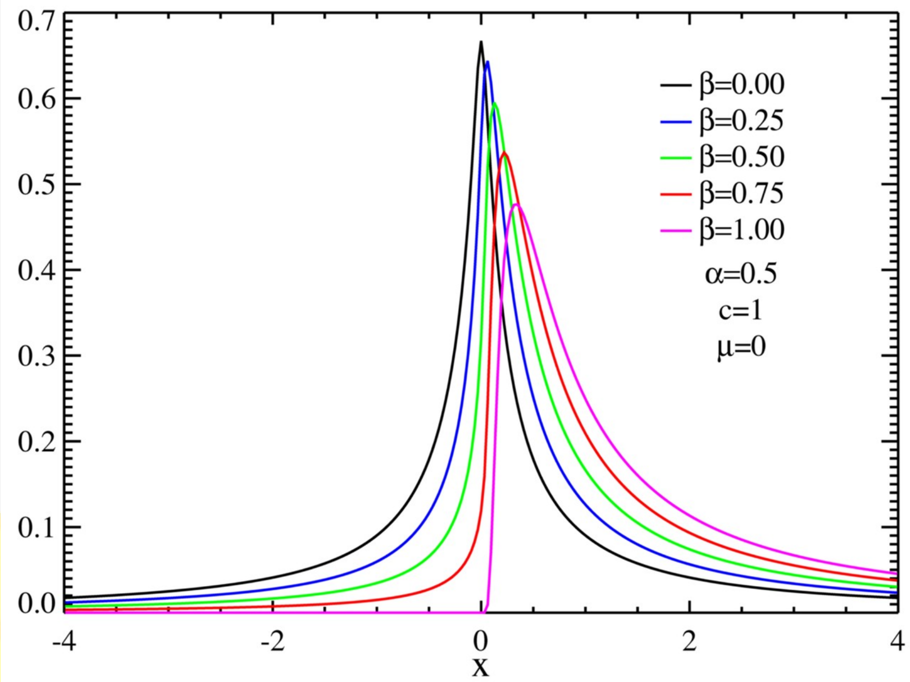
A non-degenerate random variable Z is α-stable for some $0 < \alpha \leq 2$ if and only if there is an independent, identically distributed sequence of random variables $X_1, X_2, X_3, \ldots$ and constants $a_n > 0$, $b_n \in R$ with

$$a_n(X_1 + X_2 + \ldots + X_n) - b_n \rightarrow Z \sim S(\alpha, \beta, 1, 0; x)$$

CF of α-stable distribution:

$$\varphi(\omega) = e^{i\omega\mu - a|\omega|^{\alpha}[1 - i\beta Sign(\omega)\tan(\pi\alpha/2)]}, \alpha \neq 1$$





Prob. distribution of sums

x from $f(x)$, y from $g(y)$. Then for $z = x + y$ the PDF P(z) reads

$$P(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$$

For their characteristic functions

$$\tilde{P}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)$$

If $X = x_1 + x_2 + \ldots + x_n$

$$\tilde{P}(\omega) = \tilde{f}_1(\omega)\ldots\tilde{f}_n(\omega)$$

N.B. The same property holds for Laplace transforms of the PDFs of positive variables

Sketch of GCLT proof.

1) Partial sum

$S_n = \sum x_n$

$$\xi_n = \frac{S_n - b_n}{a_n}$$

$b_n = \langle X \rangle n$ if mean is finite

2) $\lim_{n\to\infty} P(\xi_n = \xi)$ exists

$\varphi(\omega)$ is CF of scaled variable ξ

S_N: $\Phi_N(\omega) = e^{ib_N\omega}\varphi(a_N\omega)$

Renormalisation group trick

$N = n \cdot m$ $\qquad m = \frac{N}{n}$

$$\Phi_n(\omega) = e^{i b_n \omega} \varphi(a_n \omega)$$

$$f(\omega) = e^{i m b_n \omega} \left(\varphi(a_n \omega)\right)^m$$

$\downarrow$ independent of n

$$\frac{\partial}{\partial n}\left(e^{i \frac{N}{n} b_n \omega + \frac{N}{n} \ln[\varphi(a_n \omega)]}\right) = 0 \qquad c_n = \frac{b_n}{n}$$

$$\Downarrow$$

$$i N \omega \frac{\partial c_n}{\partial n} - \frac{N}{n^2} \ln \varphi + \frac{N}{n} \frac{\varphi'}{\varphi} \frac{\partial a_n}{\partial n} \omega = 0$$

Reworking this equation

$$\frac{\varphi'}{\varphi} - \frac{C_1 \ln \varphi(z)}{z} = i C_2, \quad z = a_n \omega$$

$$\frac{\varphi'}{\varphi} - \frac{c_1 \ln \varphi(z)}{z} = i c_2$$

homogeneous part

$$\ln \ln \varphi = c_1 \ln z + D \Rightarrow \varphi = e^{A|z|^{c_1}}$$

non hom. part if $c_1 \neq 1$ $\varphi_{part} = D z$

$$\begin{cases} \varphi(z) = e^{A|z|^{c_1} + Dz}, & c_1 \neq 1 \\ \varphi(-z) = \varphi^*(z) \end{cases}$$

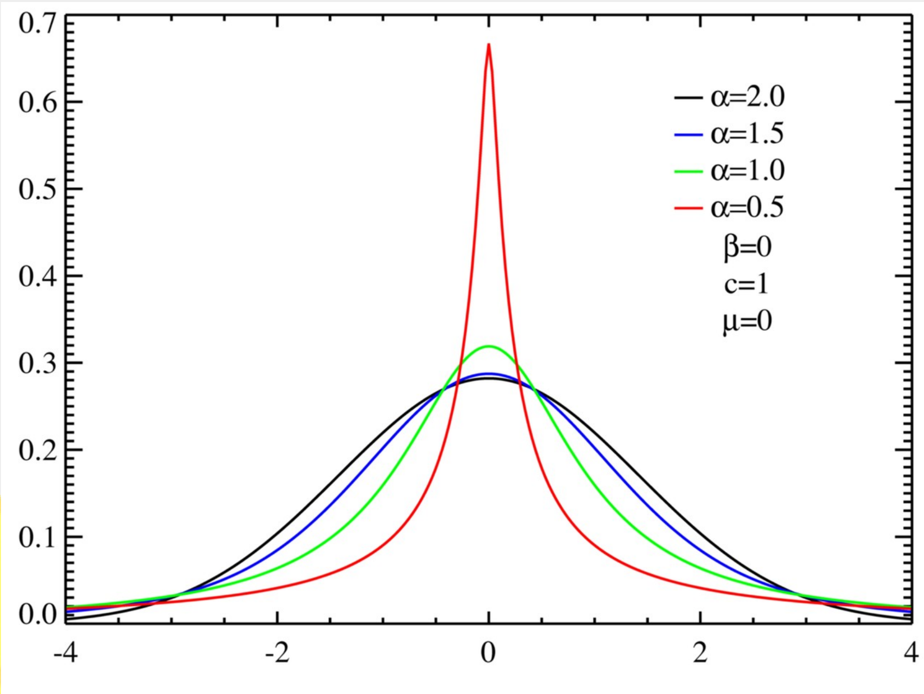
$$\Rightarrow \varphi = \begin{cases} e^{A\omega^{c_1}}, & \omega > 0 \\ e^{A^*|\omega|^{c_1}}, & \omega < 0 \end{cases} \Rightarrow \varphi = e^{-a|\omega|^{\alpha}\left[1 - \beta \operatorname{sgn}(\omega) \tan\left(\frac{\pi\alpha}{2}\right)\right]}$$

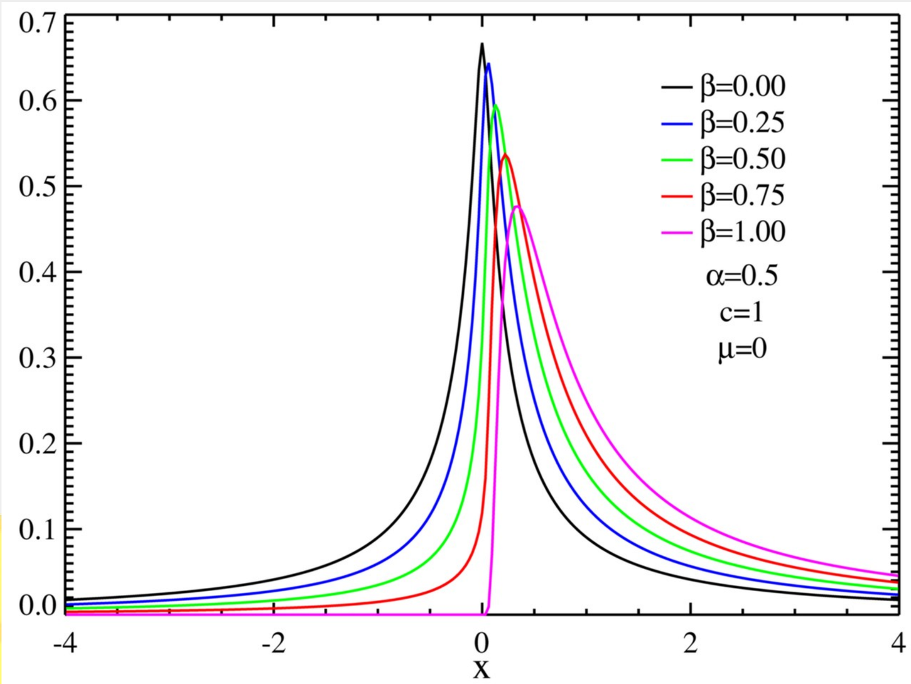
α-stable distributions

$$\varphi(\omega)=e^{i\omega\mu-a|\omega|^{\alpha}[1-i\beta Sign(\omega)\tan(\pi\alpha/2)]},\alpha\neq 1$$

$$\varphi(\omega)=e^{i\omega\mu-a|\omega|[1+i\beta\pi Sign(\omega)\ln|\omega|]},\alpha=1$$

The asymmetry vanishes when α → 2





Skoltech
Skolkovo Institute of Science and Technology

Tauberian theorem

Links behaviour at large *x* with behaviour at small *ω*

$$f(x) = \frac{A_+}{x^{1+\mu}}$$

For x>x*>0

For 0 < μ < 1, small ω

$$\varphi(\omega) \approx 1 - C\omega^{\mu}$$

For 1 < μ < 2, small ω

$$\varphi(\omega) = 1 + iC\,\omega^{\mu} + iB\,\omega$$

For the reference α-stable CF:

$$\varphi(\omega) = e^{i\omega\mu - a|\omega|^{\alpha}[1 - i\beta Sign(\omega)\tan(\pi\alpha/2)]}, \alpha \neq 1$$

Skoltech

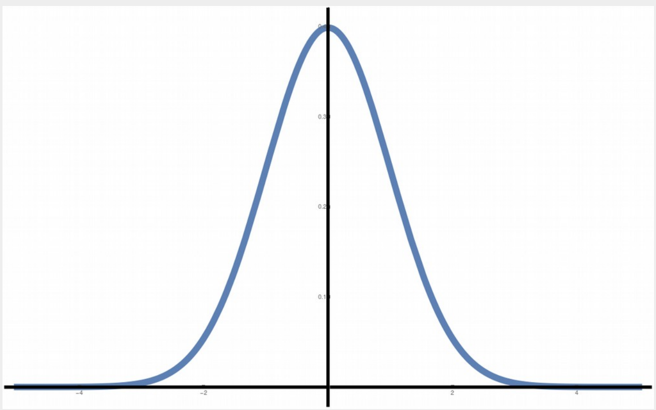
Skolkovo Institute of Science and Technology

Three analytically tractable stable distributions

Gaussian distribution

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

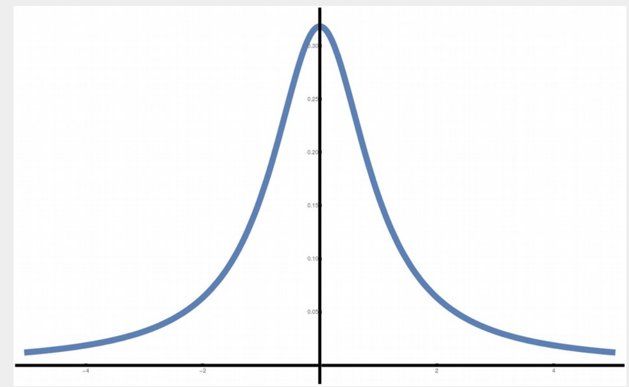
$$G(k)=\exp(ik\mu-\sigma^2k^2)$$



Cauchy-Lorentz distribution

$$P(x)=\frac{\gamma}{\pi\left((x-x_0)^2+\gamma^2\right)}$$

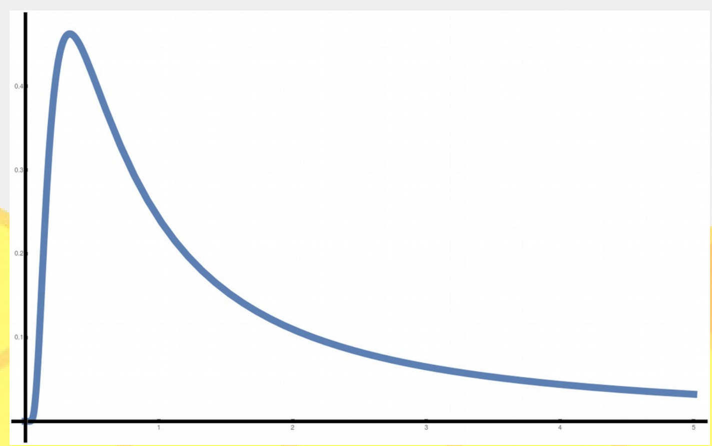
$$G(k)=\exp(ikx_0-\gamma|k|)$$



Lévy(-Smirnoff) distribution

$$P(x)=\sqrt{\frac{C}{2\pi}}\frac{e^{-C/(2x)}}{x^{3/2}},x\geq 0$$

$$G(k)=e^{-\sqrt{-2iCk}}$$

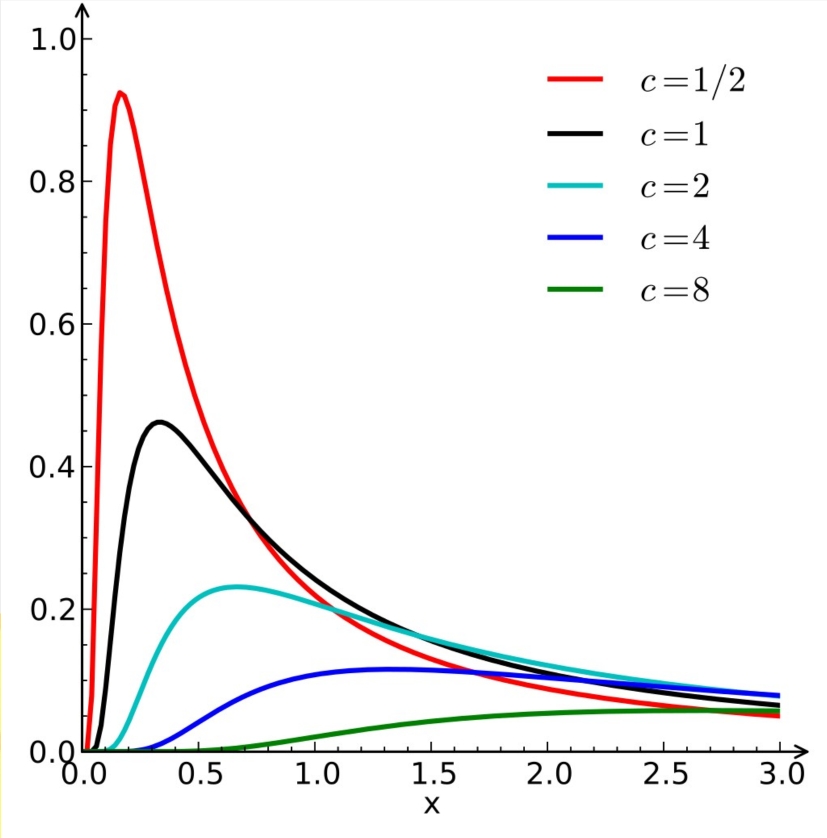


Lévy(-Smirnoff) distribution

A nice example of one-sided alpha-stable distribution with an analytical expression in x-space, $\alpha = 1/2$, $\beta = 1$.

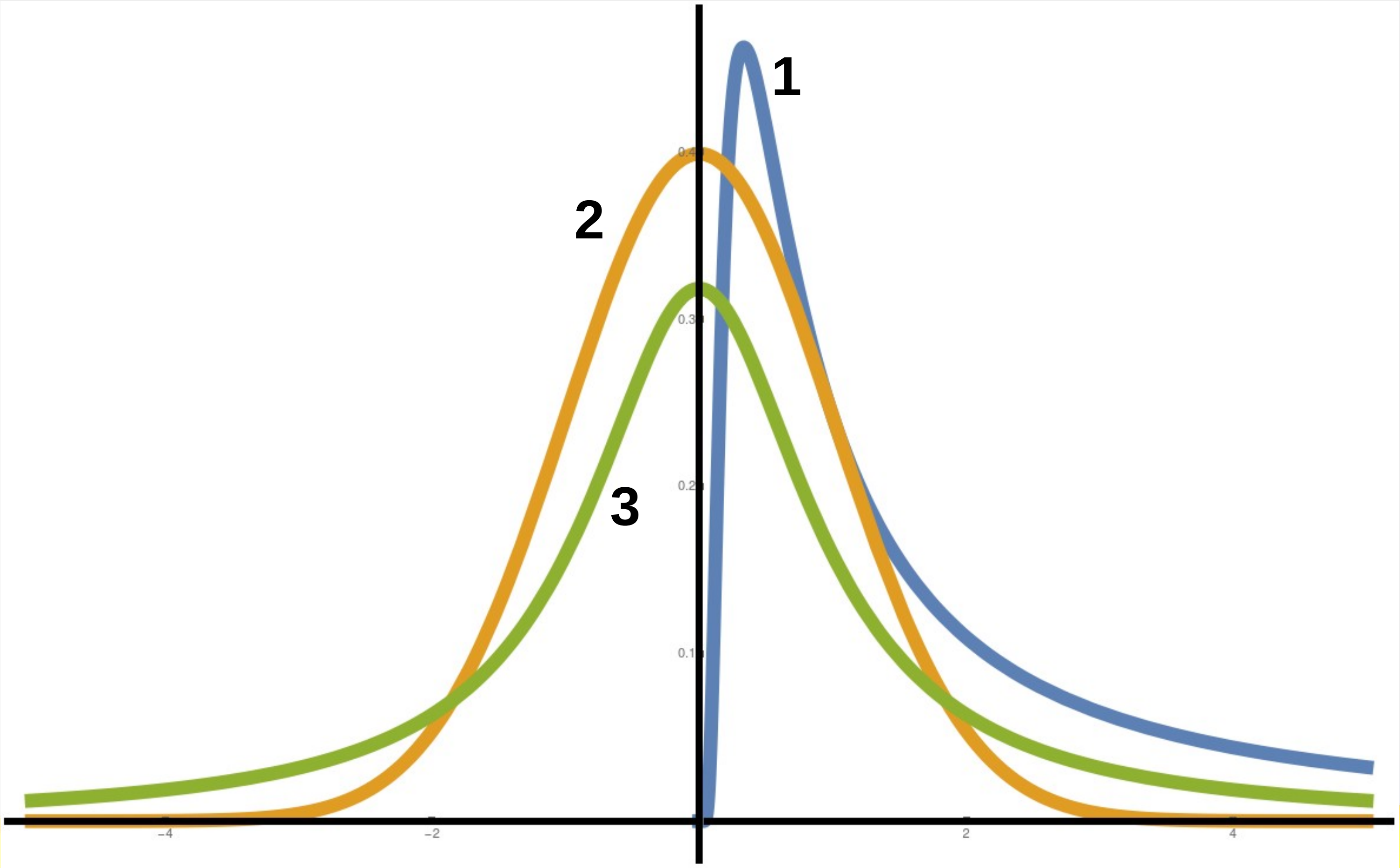
$$P(x) = \sqrt{\frac{C}{2\pi}} \frac{e^{-C/(2x)}}{x^{3/2}}, x \geq 0$$

The characteristic function $G(k) = e^{-\sqrt{-2iCk}}$



Three analytically tractable stable distributions

Quiz. What are the numbers corresponding to Gaussian, Cauchy and Lévy-Smirnoff distributions?



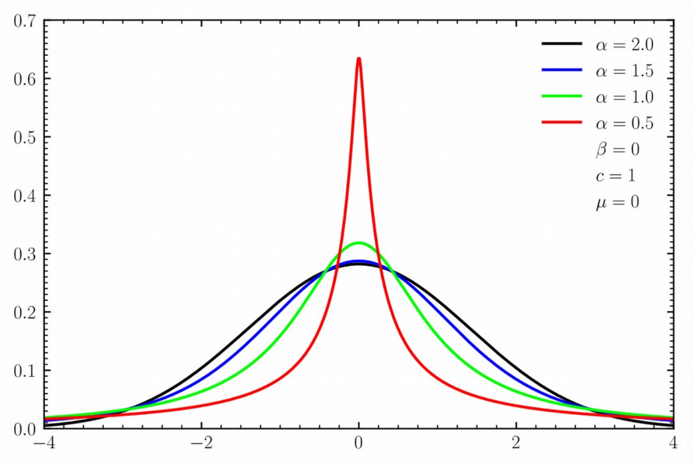
Other cases

Can be represented through Meijer-G and Fox H-functions

Holtsmark distribution
(model for the fluctuating fields in plasma due to chaotic motion of charged particles)

$$G(k) = \exp\left(ik\mu - |c\,k|^{3/2}\right)$$

$$f(x;0,1) = \frac{1}{\pi}\Gamma\left(\frac{5}{3}\right)\,{}_2F_3\left(\frac{5}{12},\frac{11}{12};\frac{1}{3},\frac{1}{2},\frac{5}{6};-\frac{4x^6}{729}\right) - \frac{x^2}{3\pi}\,{}_3F_4\left(\frac{3}{4},1,\frac{5}{4};\frac{2}{3},\frac{5}{6},\frac{7}{6},\frac{4}{3};-\frac{4x^6}{729}\right) + \frac{7x^4}{81\pi}\Gamma\left(\frac{4}{3}\right)\,{}_2F_3\left(\frac{13}{12},\frac{19}{12};\frac{7}{6},\frac{3}{2},\frac{5}{3};-\frac{4x^6}{729}\right),$$



Skoltech
Skolkovo Institute of Science and Technology

Multivariate Gaussian distributions

We consider *M* zero-mean random variables sampled i.i.d.

Joint distribution

$$P(x_1, \ldots, x_M) = \frac{1}{N} \exp\left(-\frac{x_i A_{ij} x_j}{2}\right)$$

A_{ij} is a positive definite matrix

Normalisation prefactor

One can diagonalise A and transform the joint PDF into a product of different Gaussians

$$N = \frac{(2\pi)^{M/2}}{\sqrt{\det A}}$$

Skoltech
Skolkovo Institute of Science and Technology

Joint, conditional, marginal

Joint distribution

$$P(x_1, x_2)$$

Conditional distribution

$$P(x_1 | x_2) = \frac{P(x_1, x_2)}{\int_{\Sigma_1} P(x_1, x_2)\, dx_1}$$

Marginal distribution

$$P(x_1) = \int_{\Sigma_2} P(x_1, x_2)\, dx_2$$

Bayes' theorem

$$P(x|y)=\frac{P(x,y)}{P(y)} \qquad \& \qquad P(y|x)=\frac{P(x,y)}{P(x)}$$

consequently

$$P(x|y)P(y)=P(y|x)P(x)$$

$$\boxed{P(x|y)=\frac{P(y|x)P(x)}{P(y)}}$$

Bayes' theorem

$$P(x|y)=\frac{P(x,y)}{P(y)} \qquad \& \qquad P(y|x)=\frac{P(x,y)}{P(x)}$$

consequently

$$P(x|y)P(y)=P(y|x)P(x)$$

P(x) is a "prior" probability
The degree of initial "belief" in *x*

$$\boxed{P(x|y)=\frac{P(y|x)P(x)}{P(y)}}$$

The ratio P(y|x)/P(y) is a "support knowledge" *y* provides for *x*

P(x|y) is a "posterior" probability
The degree of belief that we have accounted for y

An illustration of conditional probability can be found here
https://setosa.io/ev/conditional-probability/

Multivariate Gaussian distributions

Pair moments (correlation functions)

$$\mathbf{E}[x_i x_j] = A^{-1}_{ij}$$

The higher order moments can be expressed in terms of second moments

$$\begin{aligned} \mathbf{E}[x_1 x_2 \dots x_{2n}] &= \sum \prod \mathbf{E}[x_i x_j] \\ \mathbf{E}[x_1 x_2 \dots x_{2n+1}] &= 0, \end{aligned}$$

Sum over all pairings p for {1,..,n}

In particular, for the fourth order the moment transforms into

$$\mathbf{E}[x_i x_j x_k x_m] = \mathbf{E}[x_i x_j]\mathbf{E}[x_k x_m] + \mathbf{E}[x_i x_k]\mathbf{E}[x_j x_m] + \mathbf{E}[x_i x_m]\mathbf{E}[x_j x_k]$$

(Wick's or Isserlis' theorem)

Isserlis' theorem

$$E[x_1 \cdot x_2 \cdot \ldots x_{2n}] = \underbrace{\sum \prod}_{\text{all pairings}} E[x_i x_j]$$

$$E[x_1 \cdot x_2 \cdot \ldots x_{2n+1}] = 0$$

Odd case

$-X = \{-x_1, -x_2, \ldots, -x_n\}$ has the same distr. as X

$$E[X_1 \cdot \ldots X_{2m+1}] = E[(-X_1) \cdot \ldots \cdot (-X_{2m+1})] = 0$$

$$= (-1)^{2m+1} E[X_1 \cdot \ldots X_{2m+1}] =$$

$$= -E[X_1 \cdot \ldots \cdot X_{2m+1}]$$

Even case

$$\underline{CF} \quad G(k_1,\dots,k_n) = \langle e^{\bar{k}^T \bar{x}} \rangle = \int dx_1 \dots dx_n\, P(\bar{x})\, e^{\bar{k}^T \cdot \bar{x}} =$$

$$= \frac{1}{N} \int dx_1 \dots dx_n \exp\left(-\frac{1}{2} \sum_{ij} x_i A_{ij} x_j + \frac{1}{2}\left(\bar{k}^T \cdot \bar{x}\right) + \frac{1}{2}\left(\bar{x}^T \cdot \bar{k}\right)\right) =$$

$$= \frac{1}{N} \int \prod_e dx_e \left(-\frac{1}{2}\left(\bar{x}^T - \underbrace{\bar{k}^T \hat{A}^{-1}}\right) \hat{A}\left(\bar{x} - \hat{A}^{-1} \bar{k}\right) + \frac{1}{2} \bar{k}^T \hat{A}^{-1} \bar{k}\right) =$$

$$= \left| \hat{x} = \bar{x} - \hat{A} \bar{k} \right| = \exp\left(\frac{1}{2} \bar{k}^T A^{-1} \bar{k}\right)$$

$$\langle x_{i_1} \dots x_{i_{2m}} \rangle = \frac{\partial^{2m}}{\partial k_{i_1} \dots \partial k_{i_{2m}}} \exp\left(\frac{1}{2} \bar{k}^T \hat{A}^{-1} \bar{k}\right)\Big|_{k=0}$$

Two-point correlation

$$E[x_i x_j] = A^{-1}_{ij}$$

$$\langle x_{i_1} x_{i_2} \rangle = \frac{\partial^2}{\partial k_{i_1} \partial k_{i_2}} e^{\frac{1}{2}(\bar{k}^T \hat{A} \bar{k})}\Big|_{\bar{k}=0} = \frac{\partial^2}{\partial k_{i_1} \partial k_{i_2}} \exp\left(\frac{1}{2}\sum_{r,l} k_r A^{-1}_{rl} k_l \right)\Big|_{\bar{k}=0} =$$

$$= \frac{\partial}{\partial k_{i_2}} \underbrace{A^{-1}_{i_1 l}}\, \underbrace{k_l}\, \exp\left(\frac{1}{2}\sum_{l,r} k_r A^{-1}_{rl} k_l\right)\Big|_{\bar{k}=0} = A^{-1}_{i_1 i_2}$$

Four-point correlation

$$\langle x_{i_1} x_{i_2} x_{i_3} x_{i_4} \rangle = \frac{\partial^4}{\partial k_{i_1} \dots \partial k_{i_4}} \exp\left(\frac{1}{2}\sum_{r,l} k_r A^{-1}_{rl} k_l\right) =$$

$$= \frac{\partial^2}{\partial k_{i_3} \partial k_{i_4}} \left[A^{-1}_{i_1 i_2} \exp\left(\frac{1}{2}\sum_{m,n} k_m A^{-1}_{mn} k_n\right) + \left(\sum_l A^{-1}_{i_1 l} k_l\right)\left(\sum_n A^{-1}_{i_2 n} k_n\right) \exp\left(\frac{1}{2}\sum \right) \right]$$

$$E[x_1 x_2]$$

$$= \overline{A^{-1}_{i_1 i_2}} A^{-1}_{i_3 i_4} + A^{-1}_{i_1 i_3} A^{-1}_{i_2 i_4} + A^{-1}_{i_1 i_4} A^{-1}_{i_2 i_3}$$

$$P(x_1,x_2)=\frac{1}{\mathcal{N}}\exp(-x_1^2-x_1x_2-x_2^2)$$

1) $\mathcal{N}=\frac{(2\pi)^{M/2}}{\sqrt{\det A}}$ $\qquad A=\begin{pmatrix}2&1\\1&2\end{pmatrix}\Rightarrow \det A=3$

$M=2$

$\mathcal{N}=\frac{2\pi}{\sqrt{3}}$ $\qquad A^{-1}=\begin{pmatrix}2/3&-1/3\\-1/3&2/3\end{pmatrix}\rightarrow$ $\langle x_1x_2\rangle=-1/3$, $\langle x_1^2\rangle=\langle x_2^2\rangle=2/3$

2) $$P(x_1)=\int_{-\infty}^{+\infty}P(x_1,x_2)\,dx_2=\frac{\sqrt{3}}{2\pi}\int_{-\infty}^{+\infty}e^{-x_1^2-x_1x_2-x_2^2}\,dx_2=$$

$$=\frac{\sqrt{3}}{2\pi}e^{-\frac{3x_1^2}{4}}\underbrace{\int_{-\infty}^{+\infty}e^{-(x_2+\frac{x_1}{2})^2}\,dx_2}_{\sqrt{\pi}}=\frac{\sqrt{3}}{2\sqrt{\pi}}e^{-\frac{3x_1^2}{4}}$$

Check: $$\int_{-\infty}^{+\infty}P(x_1)\,dx_1=\frac{\sqrt{3}}{2\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{-\frac{3x_1^2}{4}}\,dx_1=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty}e^{-y^2}\,dy=1$$

3) $P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)} = \frac{\frac{\sqrt{3}}{2\pi} e^{-x_1^2 - x_1 x_2 - x_2^2}}{\frac{\sqrt{3}}{2\sqrt{\pi}} e^{-3\frac{x_2^2}{4}}} = \frac{1}{\sqrt{\pi}} e^{-x_1^2 - x_1 x_2 - \frac{x_2^2}{4}}$

4) $E[x_1^2 x_2^2] = \frac{1}{N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1^2 x_2^2 e^{-x_1^2 - x_1 x_2 - x_2^2} dx_1 dx_2$

$E[x_1^2 x_2^2] = E[x_1 x_1 x_2 x_2] = E[x_1^2] E[x_2^2] + 2E[x_1 x_2] E[x_1 x_2]$

$\langle x_1^2 \rangle = \frac{1}{N} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1^2 e^{-x_1^2 - x_1 x_2 - x_2^2} dx_1 dx_2 = \frac{1}{N} \int_{-\infty}^{+\infty} e^{-\frac{3x_1^2}{4}} x_1^2 dx_1 \int_{-\infty}^{+\infty} e^{-\left(\frac{x_1}{2} + x_2\right)^2} dx_2 =$

$= \frac{\sqrt{3}}{2\sqrt{\pi}} \frac{4}{3\sqrt{3}} \int_{-\infty}^{+\infty} y^2 e^{-y^2} dy \overset{y=t^{1/2}}{=} \frac{4}{3\sqrt{\pi}} \underbrace{\int_0^{\infty} t^{1/2} e^{-t} dt}_{} = \frac{4}{3\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = 2/3$

$$\langle x_1 x_2 \rangle = \frac{1}{\sqrt{N}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1 x_2 \, e^{-x_1^2 - x_1 x_2 - x_2^2} dx_1 \, dx_2 =$$

$$= \frac{1}{N} \int_{-\infty}^{+\infty} dx_1 \, x_1 \, e^{-\frac{3x_1^2}{4}} \int_{-\infty}^{+\infty} dx_2 \, \cancel{x_2} \, e^{-\left(\frac{x_1}{2} + x_2\right)^2} = \frac{1}{N} \int_{-\infty}^{+\infty} x_1 \, e^{-\frac{3x_1^2}{4}} dx_1 \underbrace{\int_{-\infty}^{+\infty} \left(x_2 + \frac{x_1}{2}\right) e^{\left(\frac{x_1}{2} + x_2\right)^2} dx_2}$$

$$\pm \frac{x_1}{2}$$

$$\frac{1}{N} \int_{-\infty}^{+\infty} dx_1 \, x_1 \, e^{-\frac{3x_1^2}{4}} \frac{x_1}{2} \int_{-\infty}^{+\infty} e^{-\left(x_2 + \frac{x_1}{2}\right)^2} dx_2 = -\frac{\sqrt{\pi}}{N} \int_{-\infty}^{+\infty} x_1^2 \, e^{-\frac{3x_1^2}{4}} dx_1 = -1/3$$

$$\langle x_1^2 \, x_2^2 \rangle = \left(\frac{2}{3}\right)^2 + 2\left(\frac{1}{3}\right)^2 = 2/3$$

$$E[x_1 x_2^3] = E[x_1 x_2 x_2 x_2] = 3 \, E[x_1 x_2] \, E[x_2^2] = 3 \cdot \left(-\frac{1}{3}\right) 2/3 = -2/3$$

Skoltech
Skolkovo Institute of Science and Technology

Multivariate Gaussian distributions

Exercise: Joint probability distribution of the multivariate Gaussian variables

The joint pdf of two random variables x_1 and x_2 is

$$P(x_1, x_2) = \frac{1}{N} \exp(-x_1^2 - x_1 x_2 - x_2^2) \qquad -\infty < x_1, x_2 < \infty$$

Tasks:

1) What is normalisation constant N?

2) Calculate the marginal probability $P(x_1)$

3) What is the conditional probability $P(x_1|x_2)$

4) Calculate the statistical moments $E[x_1^2 x_2^2]$, $E[x_1 x_2^3]$

Hints

$$\mathbf{E}[x_i x_j x_k x_m] = \mathbf{E}[x_i x_j]\mathbf{E}[x_k x_m] + \mathbf{E}[x_i x_k]\mathbf{E}[x_j x_m] + \mathbf{E}[x_i x_m]\mathbf{E}[x_j x_k]$$

$$N = \frac{(2\pi)^{M/2}}{\sqrt{\det A}}$$

$$P(x_1, \ldots, x_M) = \frac{1}{N} \exp\left(-\frac{x_i A_{ij} x_j}{2}\right)$$