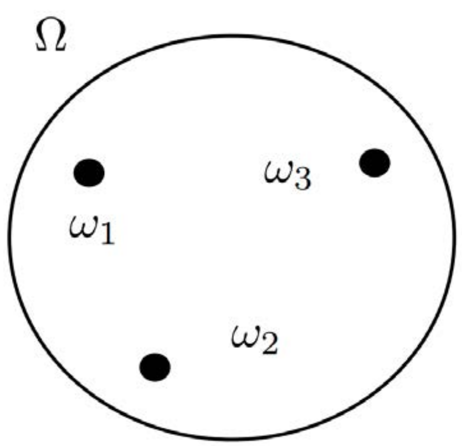
ENM 5310: Data-driven Modeling and Probabilistic Scientific Computing

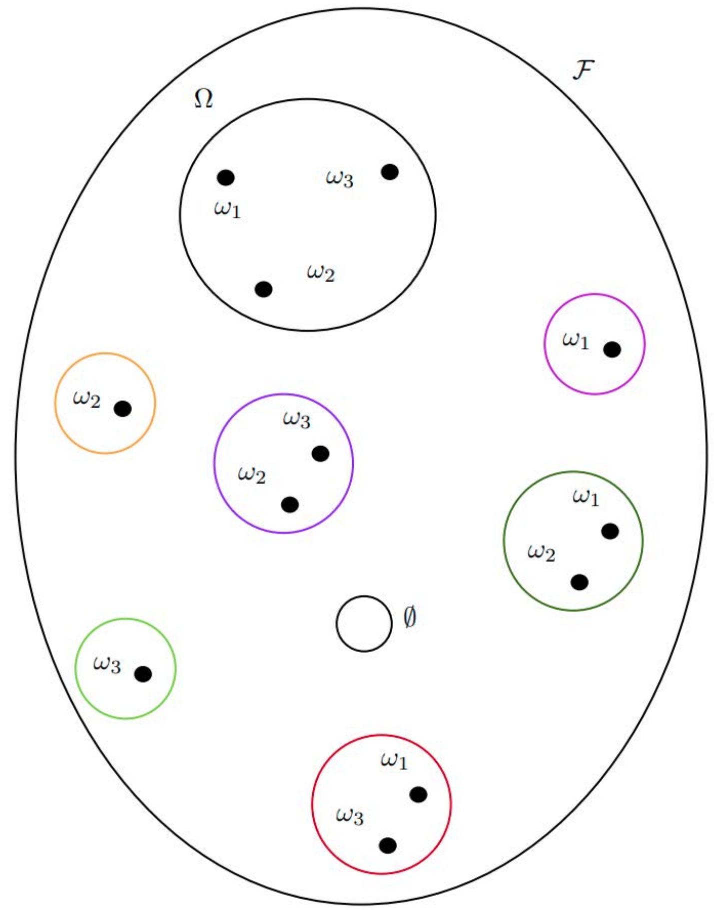
Lecture #2: Primer on Probability and Statistics

Penn
UNIVERSITY *of* PENNSYLVANIA

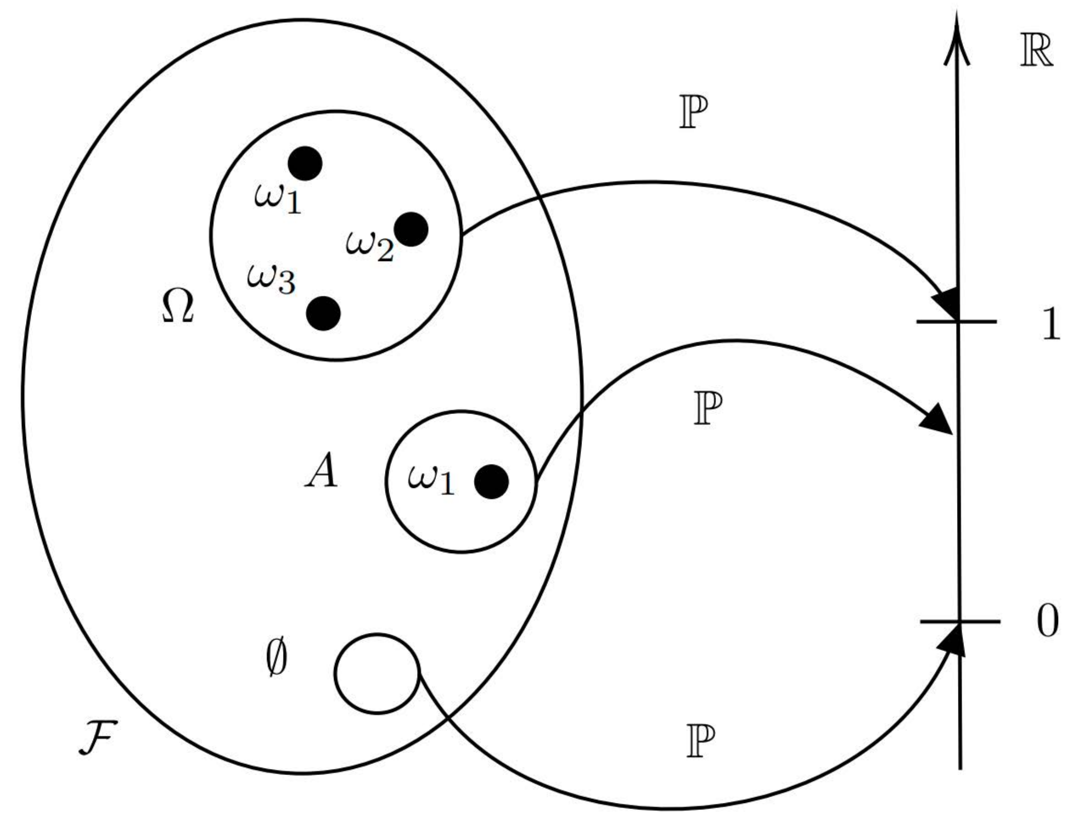
Probability spaces & random variables



Sample space



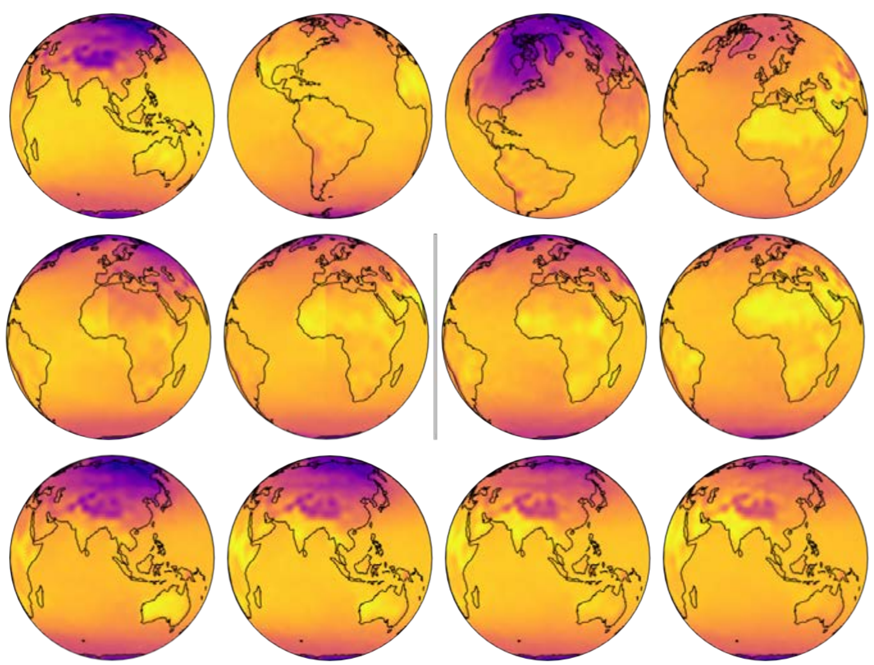
σ-algebra of events



Probability measure

vectors

matrices



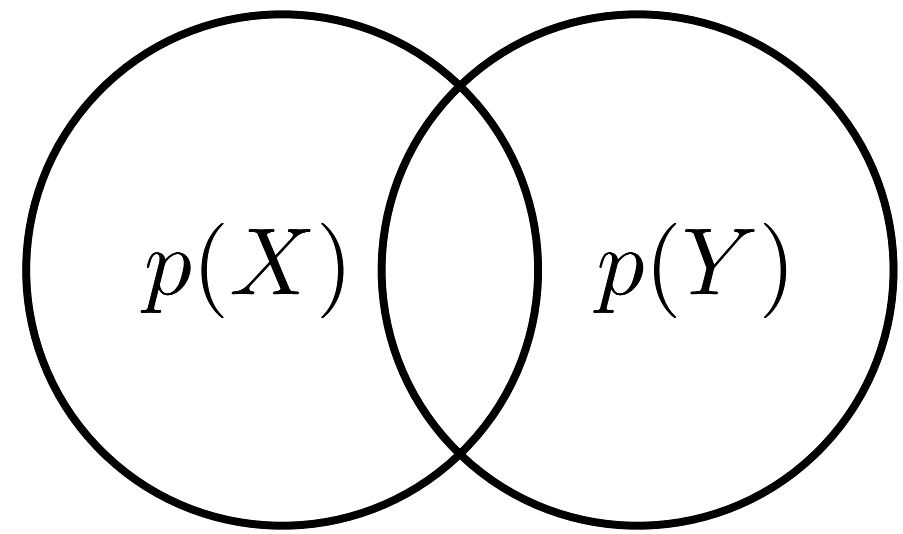
functions

Basic rules of probability

Sum rule $$p(X) = \sum_Y p(X,Y)$$

Product rule $$p(X,Y) = p(Y|X)p(X)$$

Bayes rule $$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$



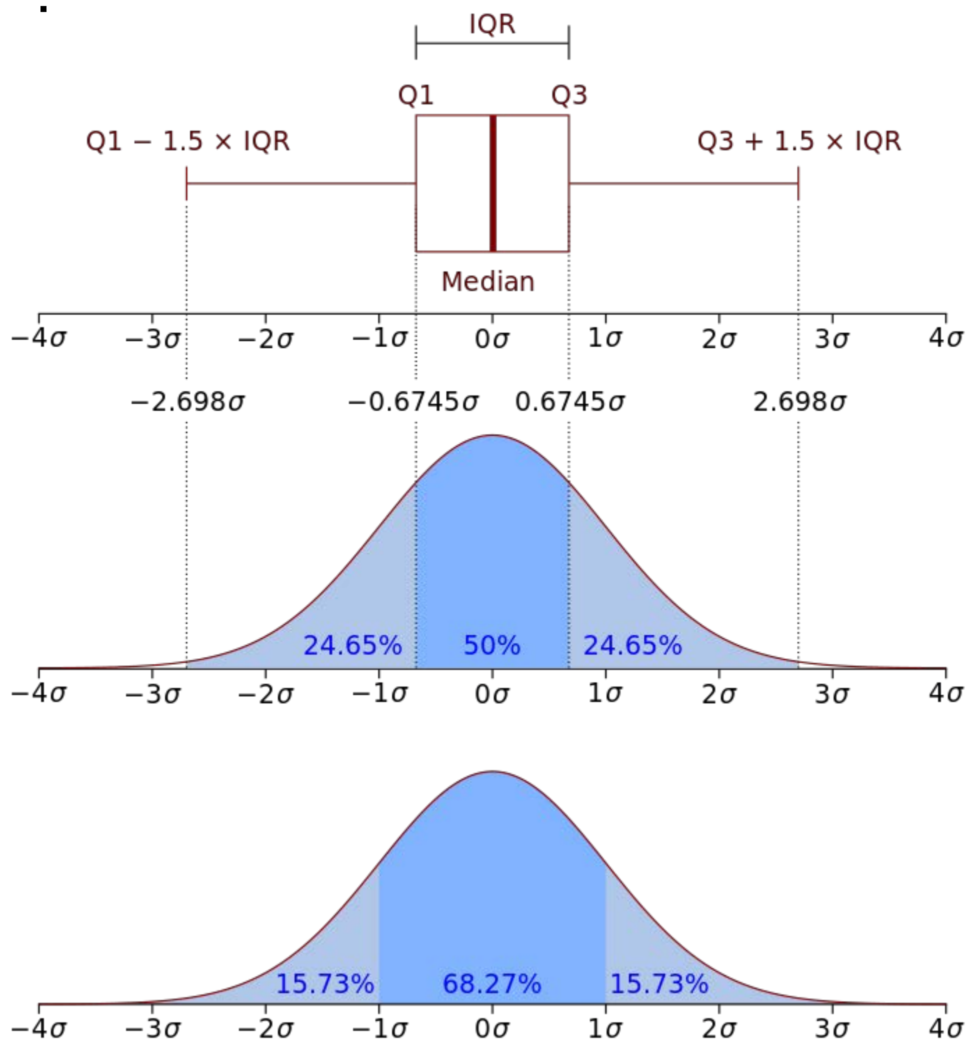
Venn diagrams

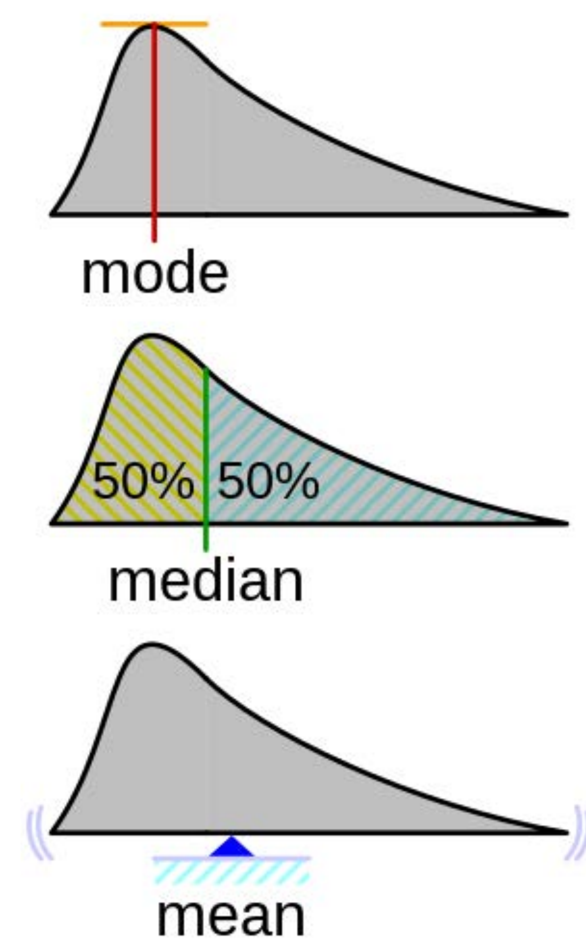
Density function

Closely related to the distribution function is the density function. Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a nonnegative function, satisfying $\int_{\mathbb{R}} f d\lambda = 1$. The function f is called a density function (with respect to the Lebesgue measure) and the associated probability measure for a random variable X, defined on $(\Omega, \mathcal{F}, P)$, is

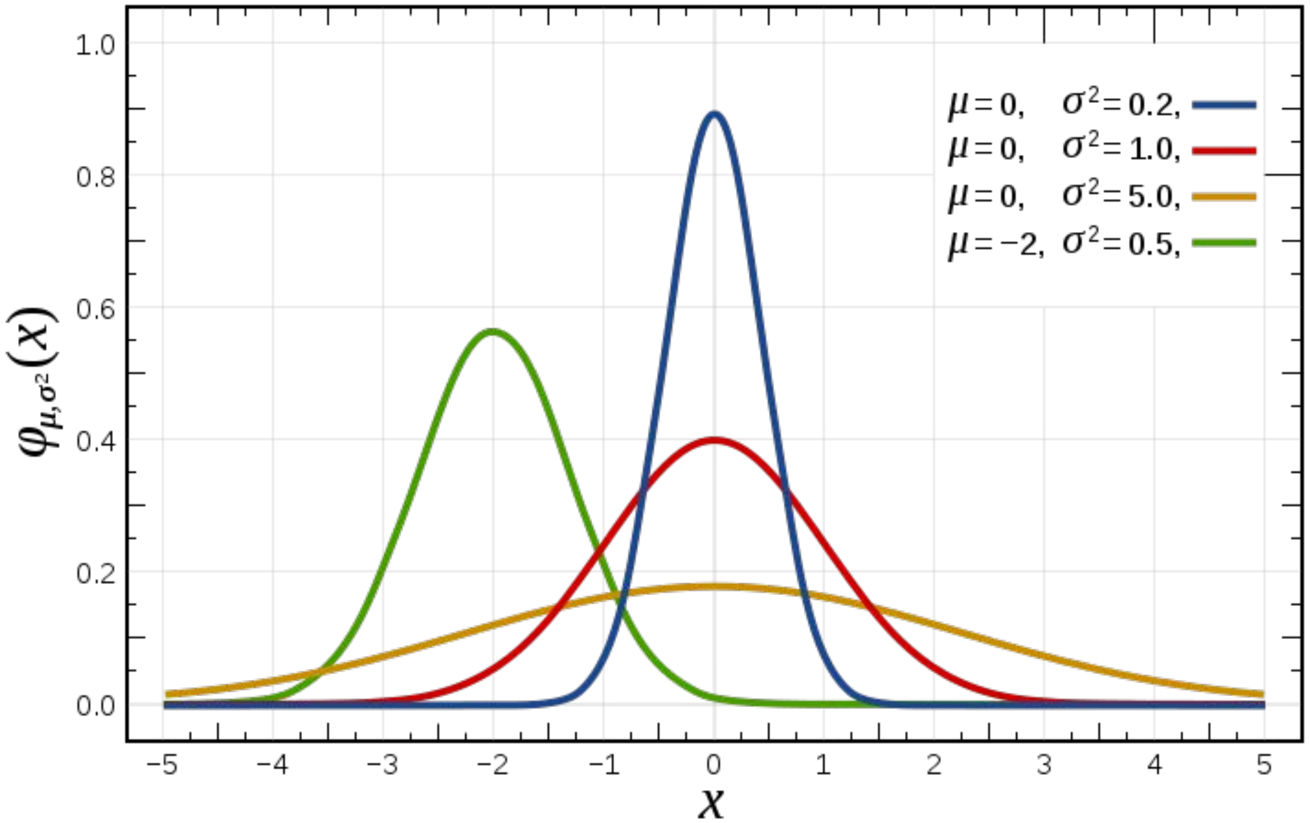
$$P(\{\omega : \omega \in A\}) = \int_A f d\lambda.$$

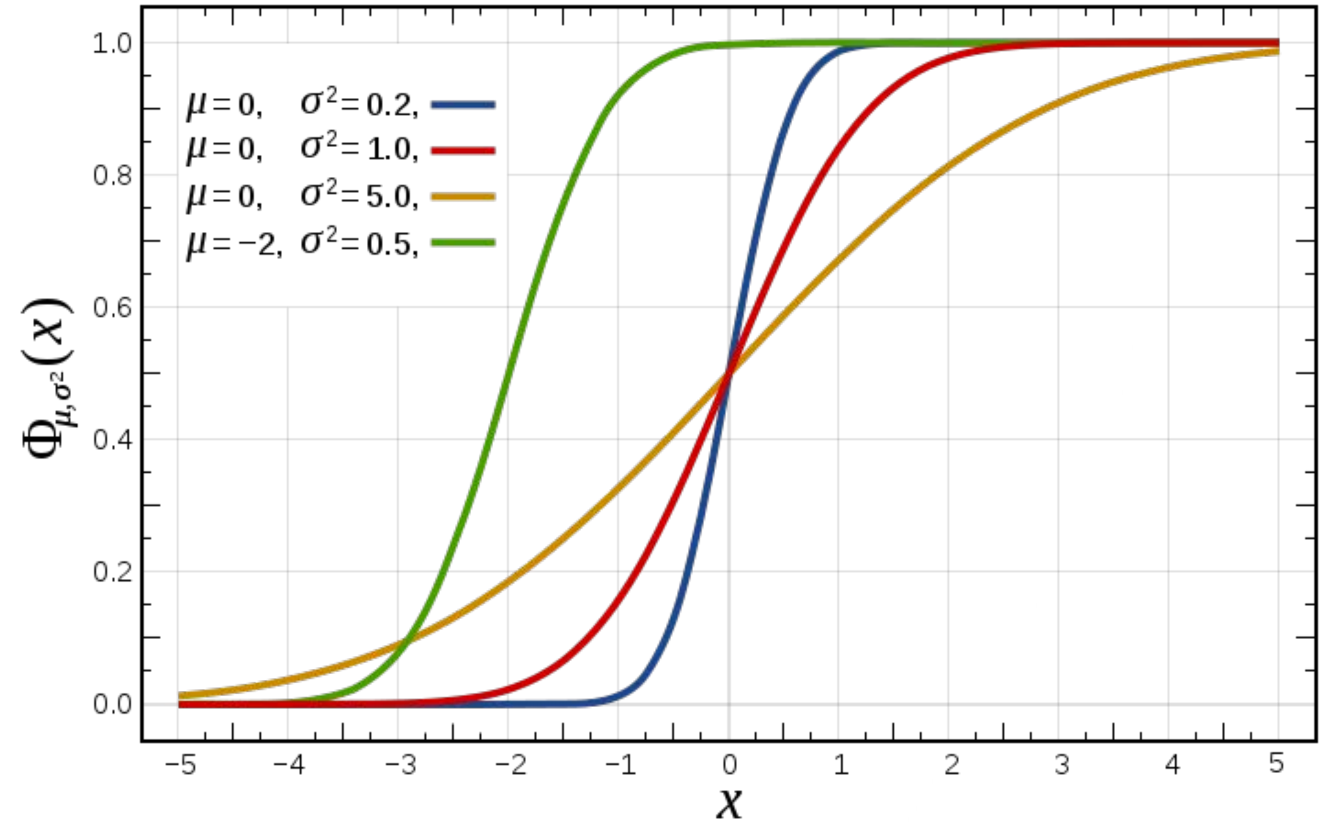
for all $A \in \mathcal{F}$.





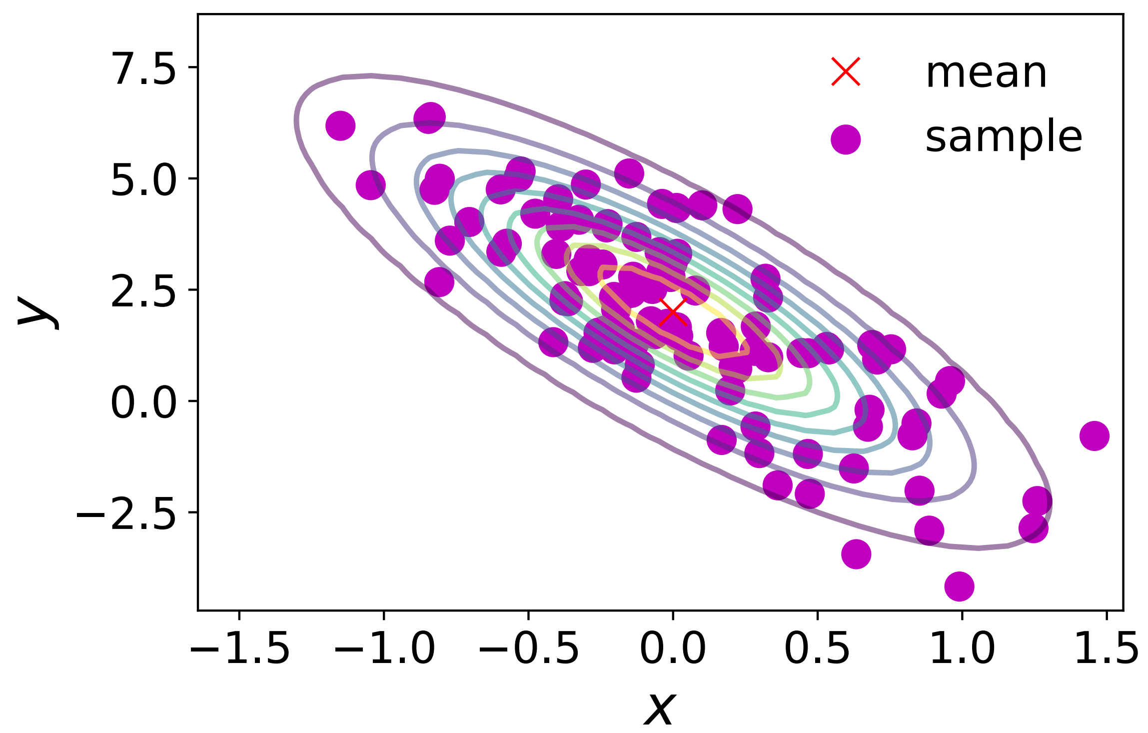
The Gaussian distribution

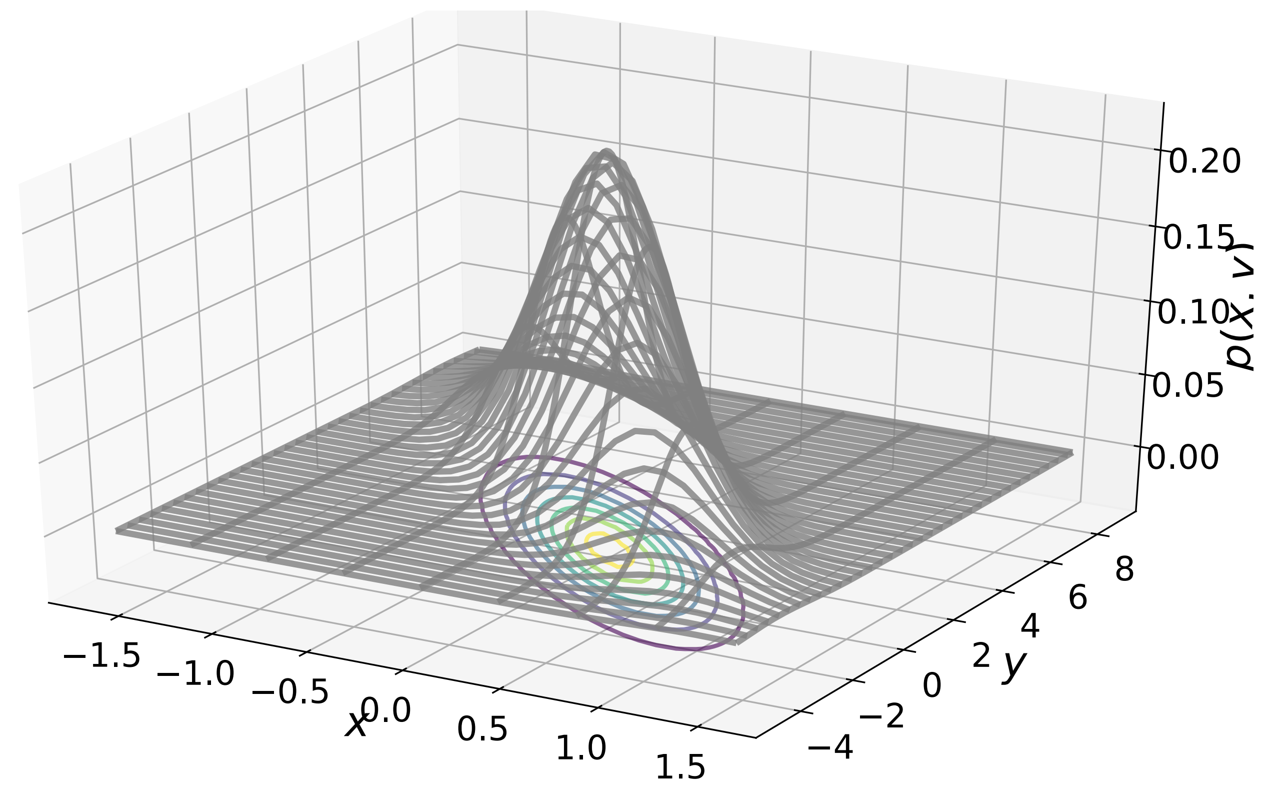




Notation	$\mathcal{N}(\mu,\,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2

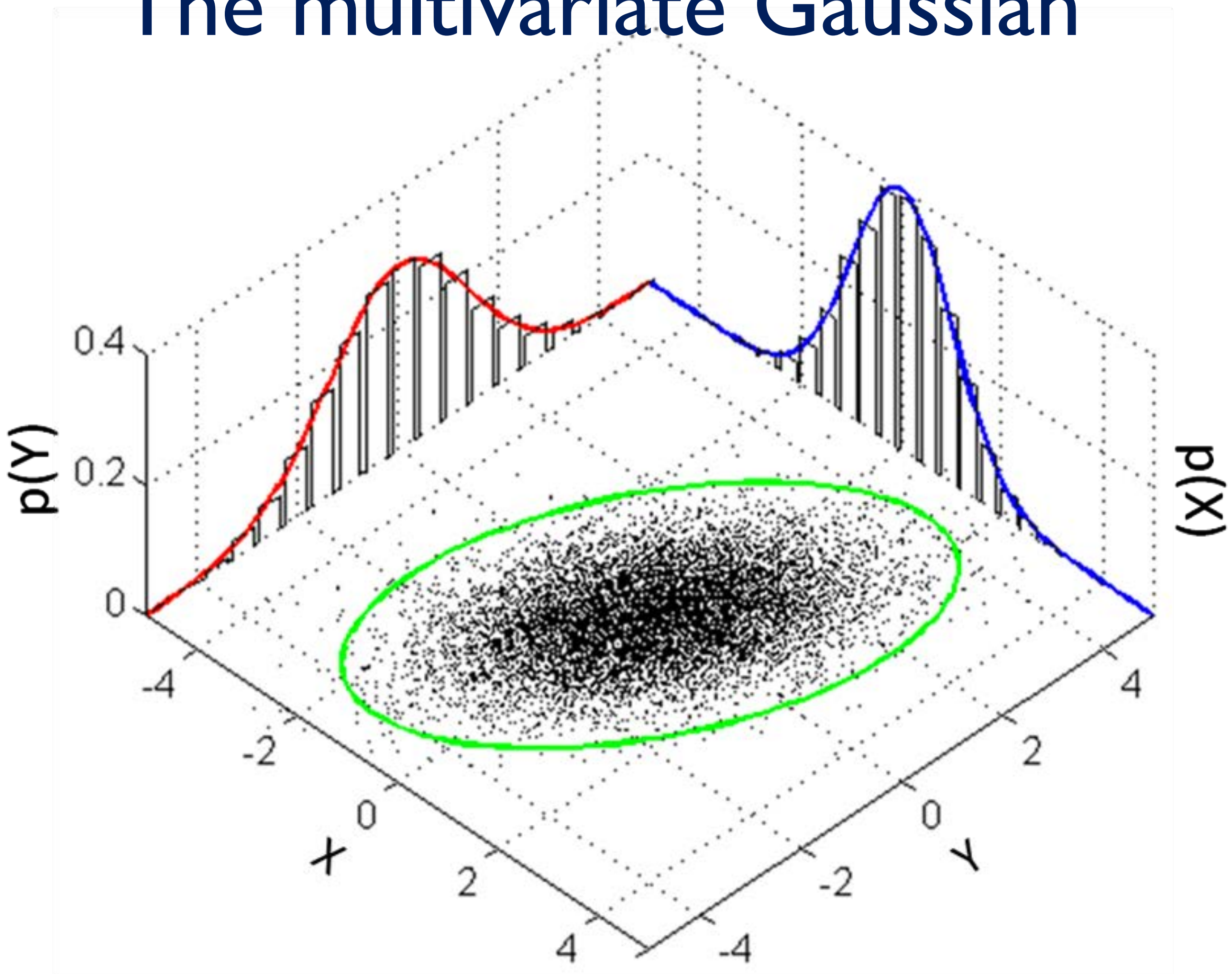
The multivariate Gaussian





$$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)$$

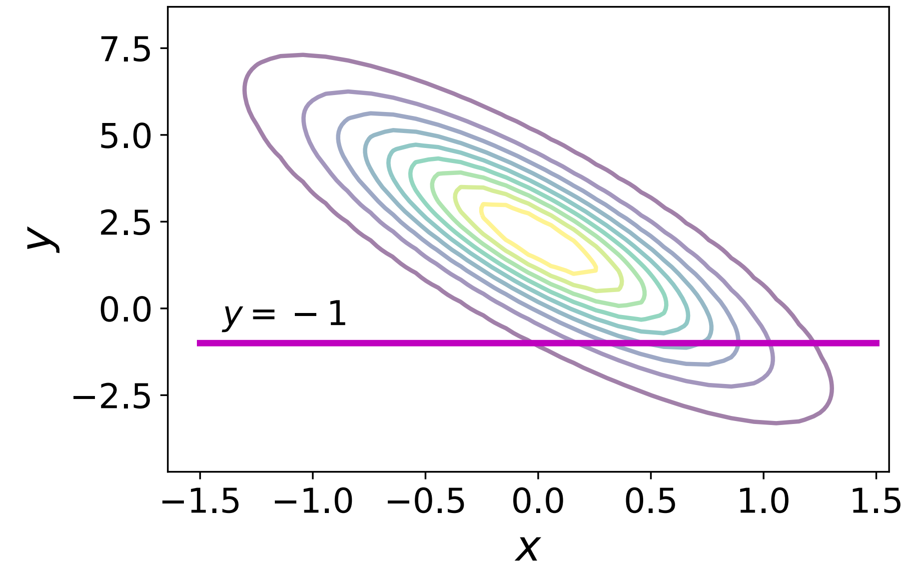
The multivariate Gaussian



Notation	$\mathcal{N}(\boldsymbol{\mu},\, \boldsymbol{\Sigma})$
Parameters	$\boldsymbol{\mu} \in \mathbf{R}^k$ — location $\boldsymbol{\Sigma} \in \mathbf{R}^{k \times k}$ — covariance (positive semi-definite matrix)
Support	$\boldsymbol{x} \in \boldsymbol{\mu} + \mathrm{span}(\boldsymbol{\Sigma}) \subseteq \mathbf{R}^k$
PDF	$\det(2\pi\boldsymbol{\Sigma})^{-\frac{1}{2}}\, e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$, exists only when $\boldsymbol{\Sigma}$ is positive-definite
Mean	$\boldsymbol{\mu}$
Mode	$\boldsymbol{\mu}$
Variance	$\boldsymbol{\Sigma}$

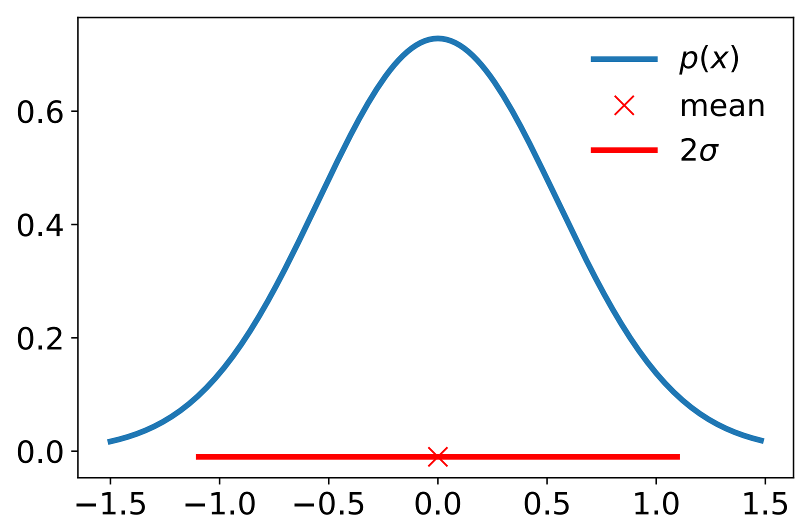
Marginals and conditionals of a Gaussian

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}\right)$$



Marginal distribution

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} = \mathcal{N}\left(\boldsymbol{x} \,|\, \boldsymbol{\mu}_x, \, \boldsymbol{\Sigma}_{xx}\right)$$

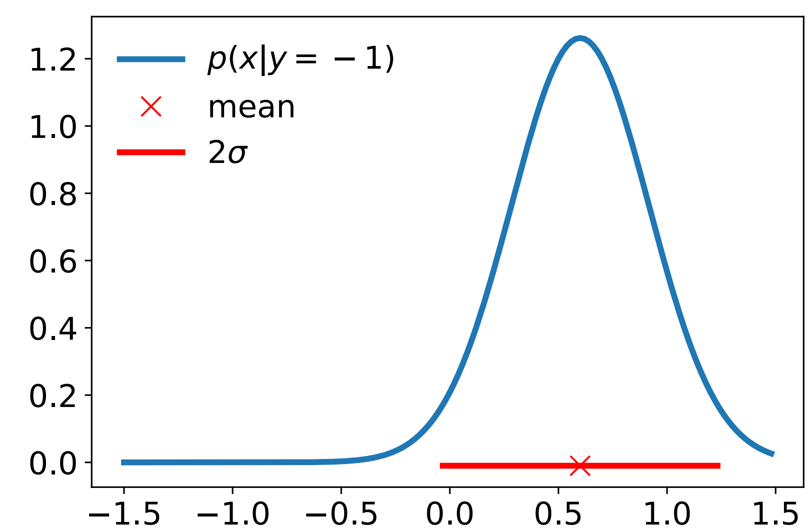


Conditional distribution

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}) = \mathcal{N}\left(\boldsymbol{\mu}_{x\,|\,y}, \, \boldsymbol{\Sigma}_{x\,|\,y}\right)$$

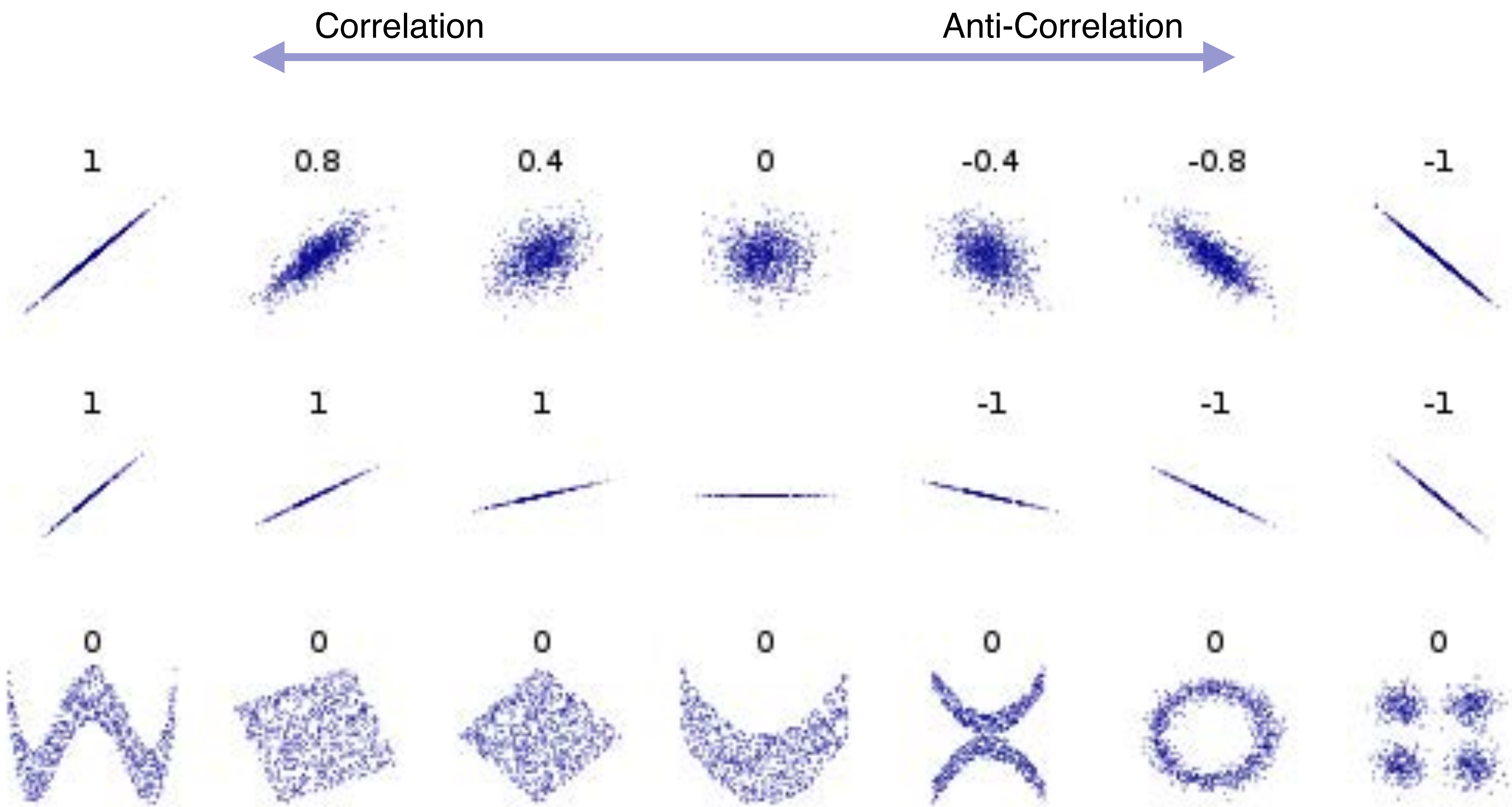
$$\boldsymbol{\mu}_{x\,|\,y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$$

$$\boldsymbol{\Sigma}_{x\,|\,y} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} \,.$$

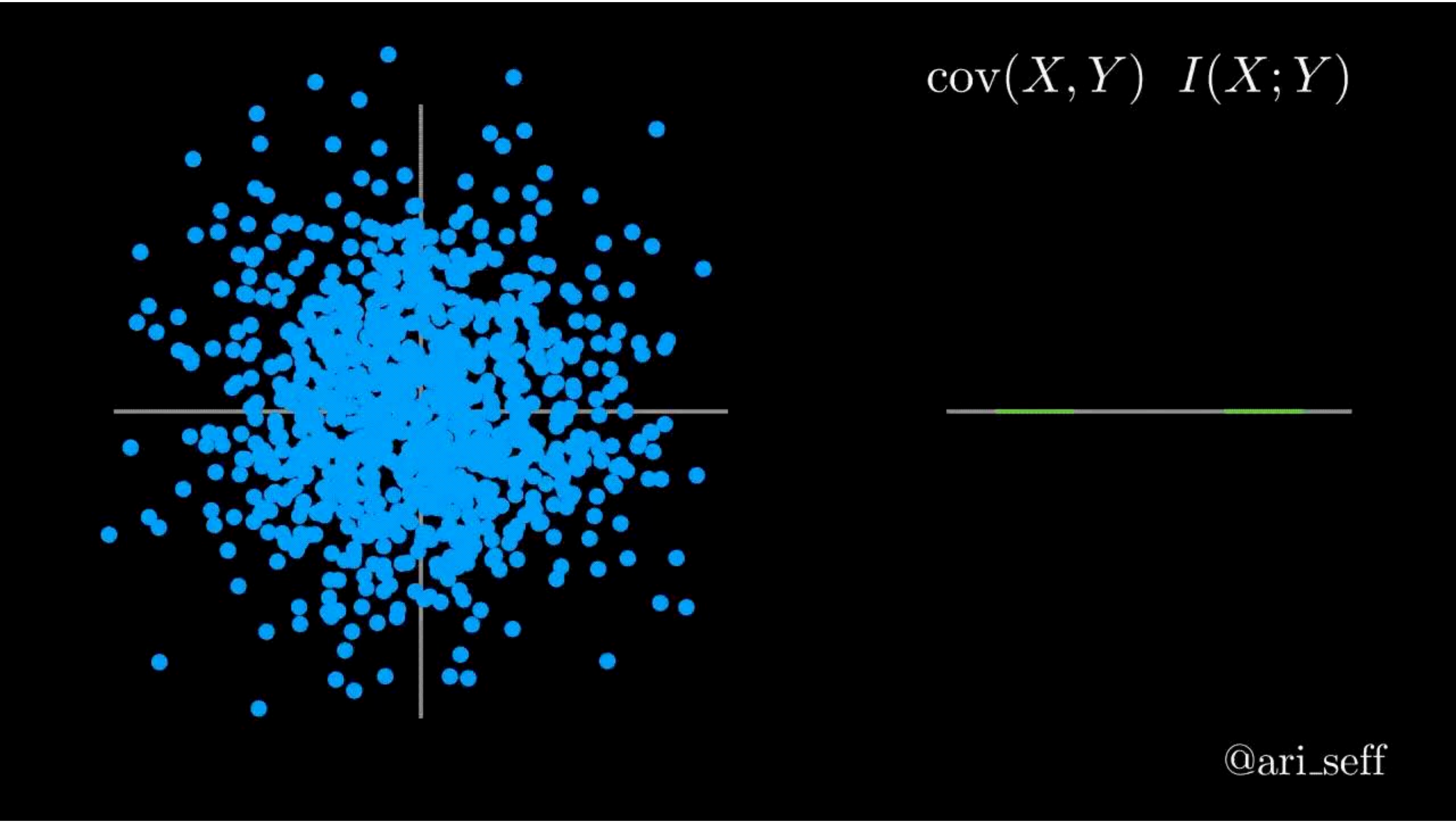


These are unique properties that make the Gaussian distribution very simple and attractive to compute with! It is essentially our main building block for computing under uncertainty.

Correlation and linear dependence

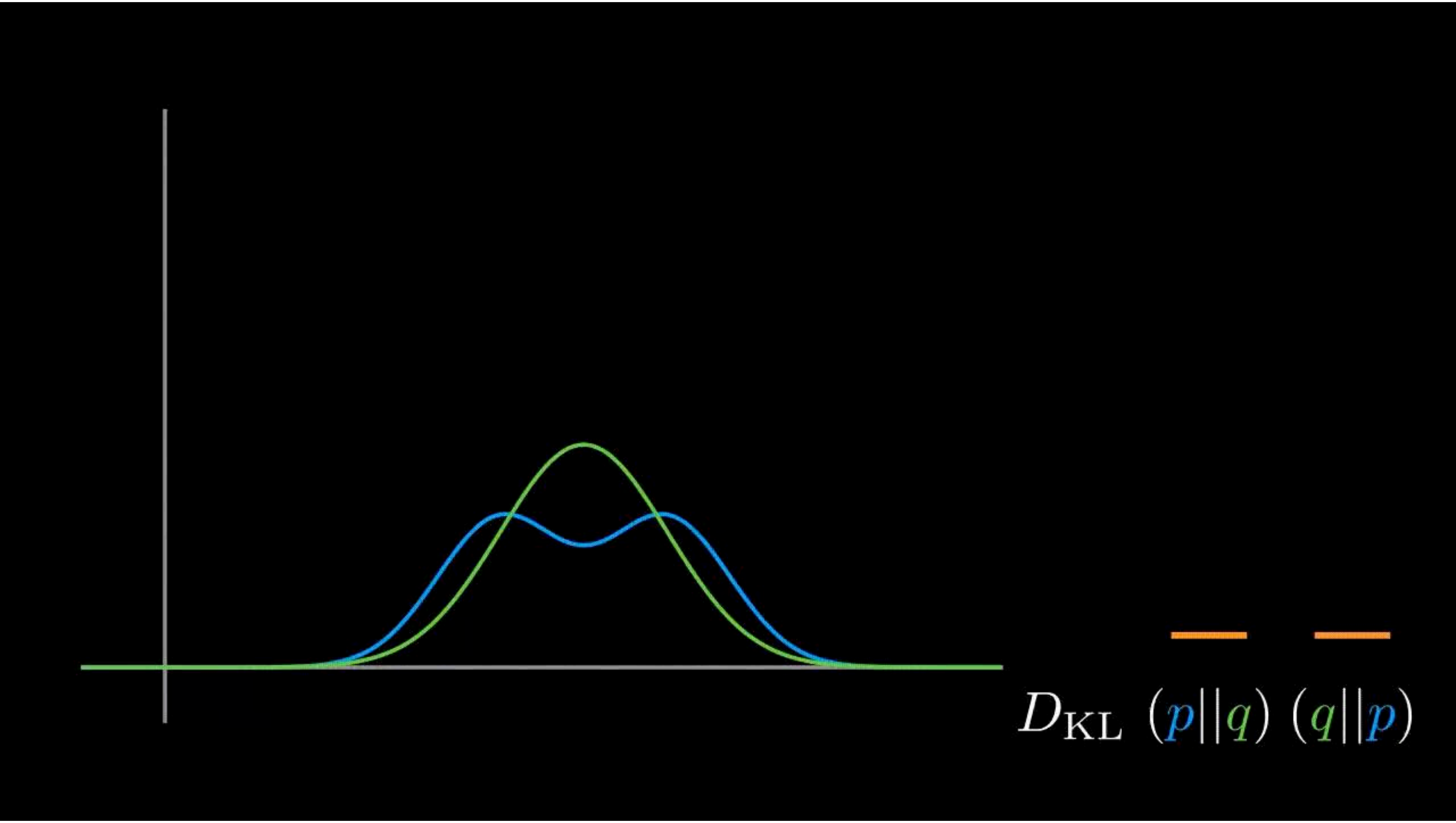


Covariance vs Mutual Information



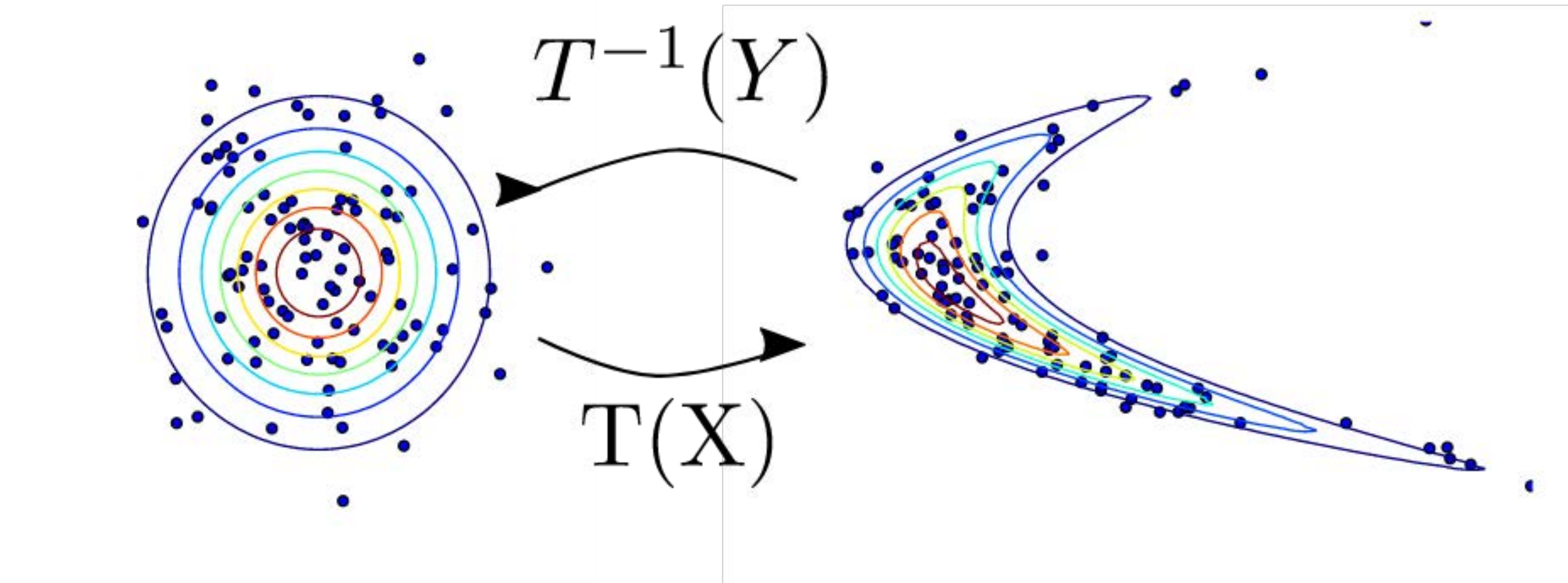
*credit: Ari Seff (Princeton)

Kullbak-Leibler divergence

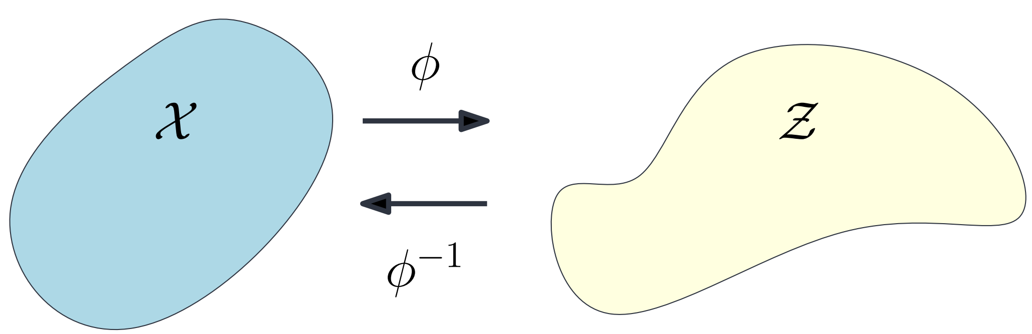


*credit: Ari Seff (Princeton)

Transformations



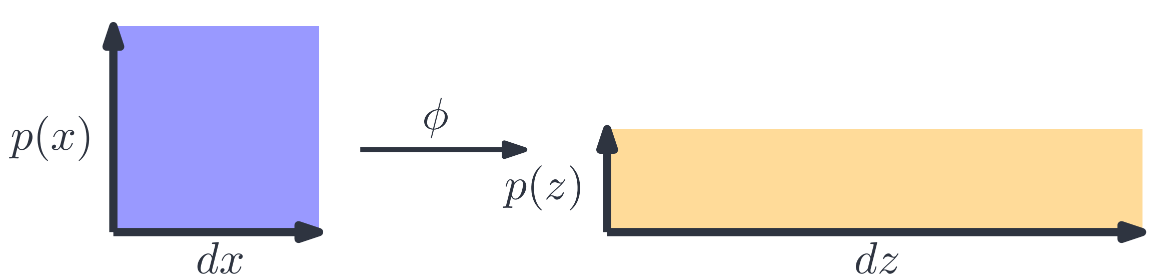
Change of variables



Key idea:

Transform random variable X into random variable Z using an invertible transformation ϕ, while keeping track of the change in distribution.

$$p_Z(\boldsymbol{z}) = p_X(\boldsymbol{x}) \left| \det \left(\frac{d\phi(\boldsymbol{x})}{d\boldsymbol{x}} \right) \right|^{-1}$$



Determinant of Jacobian

$$\left| \det \left(\frac{dz}{dx} \right) \right| = \left| \det \left(\frac{d\phi(x)}{dx} \right) \right|$$

tells us how much the domain dx is stretched to dz

Maximum likelihood estimation

$$\theta_{\mathrm{MLE}} = \arg\max_{\theta \in \Theta} p(\mathcal{D}|\theta)$$

Maximum a-posteriori estimation

$$\theta_{\text{MAP}} = \arg\max_{\theta \in \Theta} p(\theta | \mathcal{D})$$