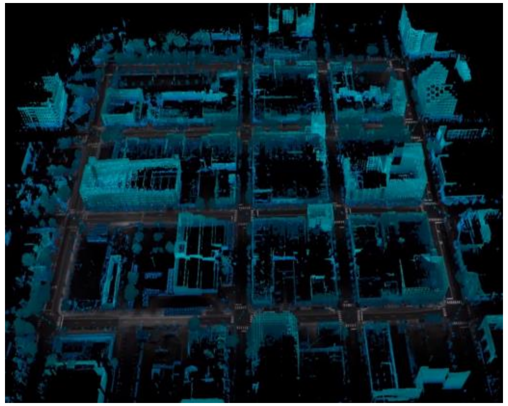
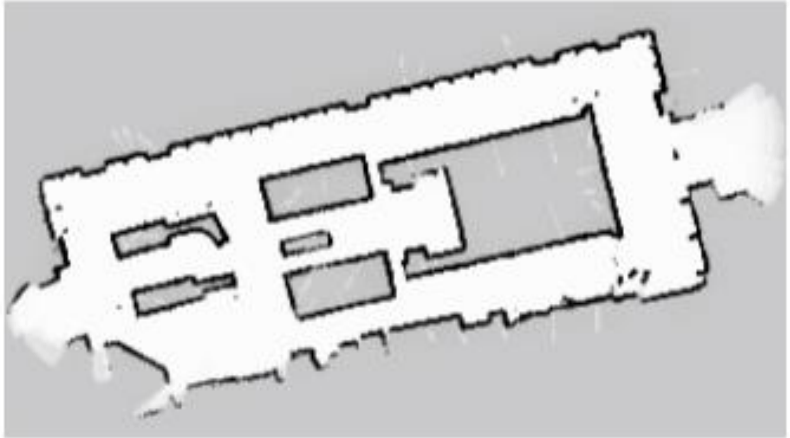
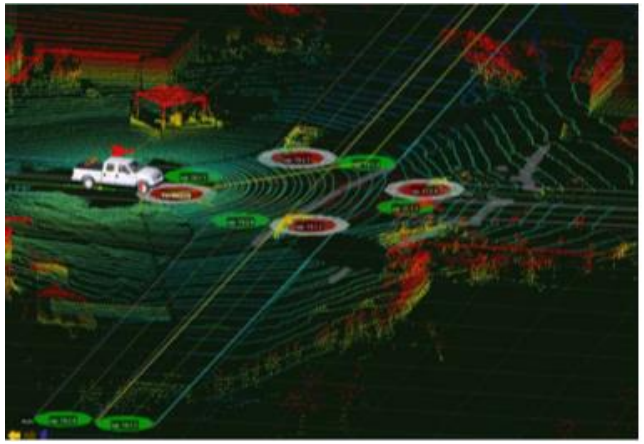
BYU Electrical & Computer Engineering





PROBABILTY REVIEW - GAUSSIANS

ECEN 633: Robotic Localization and Mapping

Some slides courtesy of Ryan Eustice.

Agenda

- State Representation and Uncertainty
- Multivariate Gaussian
- Covariance Projection/Uncertainty Propagation

Probabilistic State Estimation

- Uncertain Observations
  - Sensor noise & non-idealities
- Uncertain Beliefs
  - Derived from sensor observations
  - Approximate algorithms
- *Probabilistic State Estimation*
  - Identify the quantities (state variables) we care about.
    - E.g. “study time” versus “exam grade”
  - Determine probability for every possible simultaneous assignment

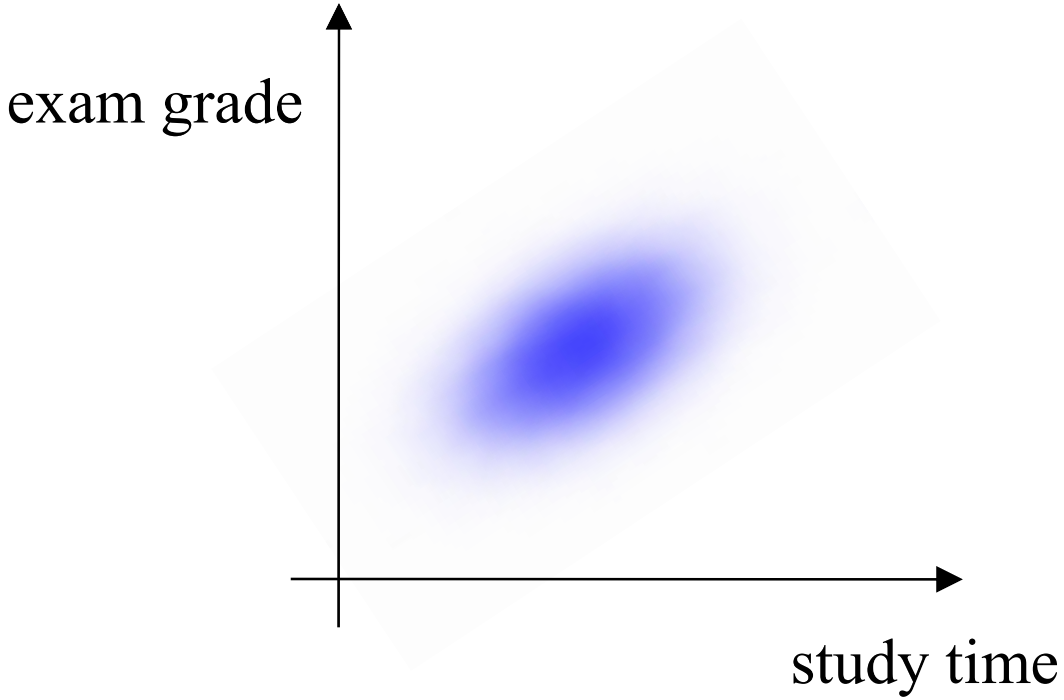
Representing State

- Represent everything we need to know in terms of a vector of quantities
  - "State vector"
  - Usually continuous-valued in this course
- The "meaning" of the variables is up to us
  - e.g., index 7 is the temperature in Seattle.
  - Bookkeeping work for us.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$$

Representing Uncertainty

- In principle, distribution of unknown quantities can be arbitrary



Common Statistics

- Expectation?

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

- Variance/Covariance?

$$\sigma_x^2 = E[(x - E[x])^2]$$

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Sigma = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^\top] = \begin{bmatrix} \sigma_{xx}^2 & \sigma_{xy}^2 \\ \sigma_{yx}^2 & \sigma_{yy}^2 \end{bmatrix}$$

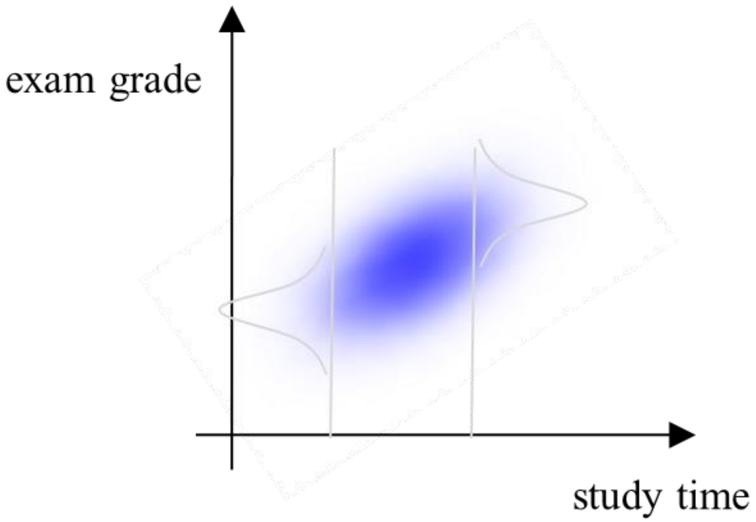
- Correlation Coefficients?

$$\rho_{xy} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y} \qquad |\rho_{xy}| \leq 1$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix}$$

Correlations

- Estimates of variables tend to become correlated over time
  - Observation: Study time is 4 hours
  - Belief about study time *and* exam grade are affected



- Distribution of exam grade depends on study time: two are correlated
  - We'll look at correlations closer later today
  - The data does not necessarily imply any *causal* relationship.

Projecting Covariances

- Suppose I know $\mathbf{x} \sim \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}$
- How do we handle $\mathbf{y} = \mathrm{A}\mathbf{x} + \mathbf{b}$ **???**

$$\Sigma_{\mathbf{yy}} = E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^\top]$$

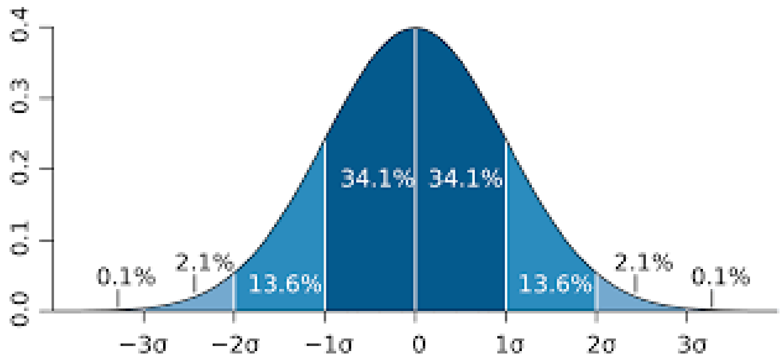
- (Algebra) ➔ $\Sigma_{\mathbf{yy}} = \mathrm{A}\Sigma_{\mathbf{xx}}\mathrm{A}^\top$

Gaussians

Gaussian (Covariance Form)

- In this class, we'll (mostly) focus on Gaussian distributions
  - For both observations and our beliefs

$$p(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$



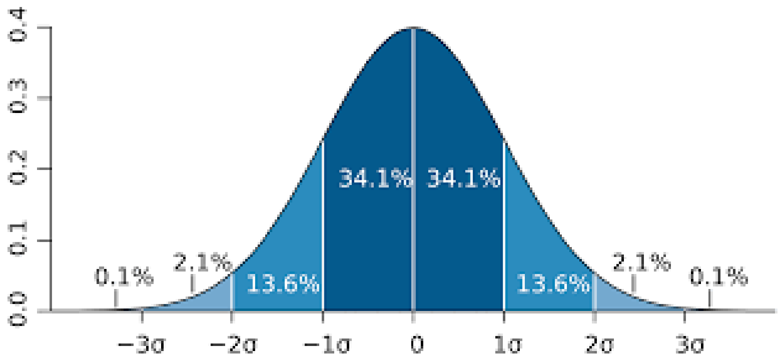
- Characterized by mean & covariance

$$\mu_{\mathbf{x}} = E[\mathbf{x}]$$

$$\Sigma_{\mathbf{xx}} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^\top]$$

Gaussian (Covariance Form)

$$p(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$



- All that gunk out front is just for normalization. Your mental model:

$$p(\mathbf{x}) = \alpha e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$

- How do the terms in the exponential, correspond to the graph?

Gaussian (Information Form)

- An alternative parameterization of the Gaussian distribution

$$
\begin{aligned}
p(\boldsymbol{\xi}_t) &= \mathcal{N}(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t) \\
&= \frac{1}{\sqrt{|2\pi\Sigma_t|}} \exp\left\{-\tfrac{1}{2}(\boldsymbol{\xi}_t - \boldsymbol{\mu}_t)^\top \Sigma_t^{-1} (\boldsymbol{\xi}_t - \boldsymbol{\mu}_t)\right\} \\
&= \frac{1}{\sqrt{|2\pi\Sigma_t|}} \exp\left\{-\tfrac{1}{2}\left(\boldsymbol{\xi}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t - 2\boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t + \boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\mu}_t\right)\right\} \\
&= \frac{e^{-\frac{1}{2}\boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\mu}_t}}{\sqrt{|2\pi\Sigma_t|}} \exp\left\{-\tfrac{1}{2}\boldsymbol{\xi}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t + \boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t\right\} \\
&= \frac{e^{-\frac{1}{2}\boldsymbol{\eta}_t^\top \Lambda_t^{-1} \boldsymbol{\eta}_t}}{\sqrt{|2\pi\Lambda_t^{-1}|}} \exp\left\{-\tfrac{1}{2}\boldsymbol{\xi}_t^\top \Lambda_t \boldsymbol{\xi}_t + \boldsymbol{\eta}_t^\top \boldsymbol{\xi}_t\right\} \\
&= \mathcal{N}^{-1}(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t)
\end{aligned}
$$

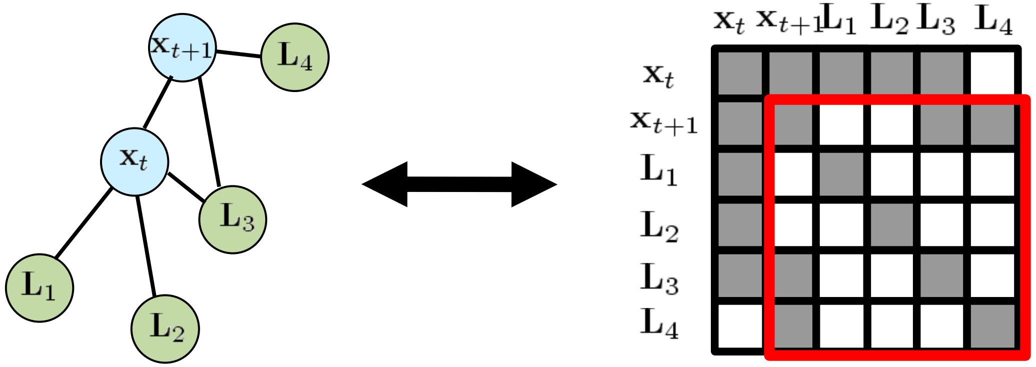
Gaussian (Information Form)

- Information matrix and vector

$$p(\boldsymbol{\xi}_t) = \mathcal{N}(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t) \\ = \mathcal{N}^{-1}(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t)$$

$$\Lambda_t = \Sigma_t^{-1} \\ \boldsymbol{\eta}_t = \Lambda_t \boldsymbol{\mu}_t$$

- Encodes a graphical model
  - Markov Random Field or Markov Net (Will talk about in a month or so)



Sparsity => Missing Edges => Available Conditional Independence

Gaussian Covariance & Information Parameterizations:

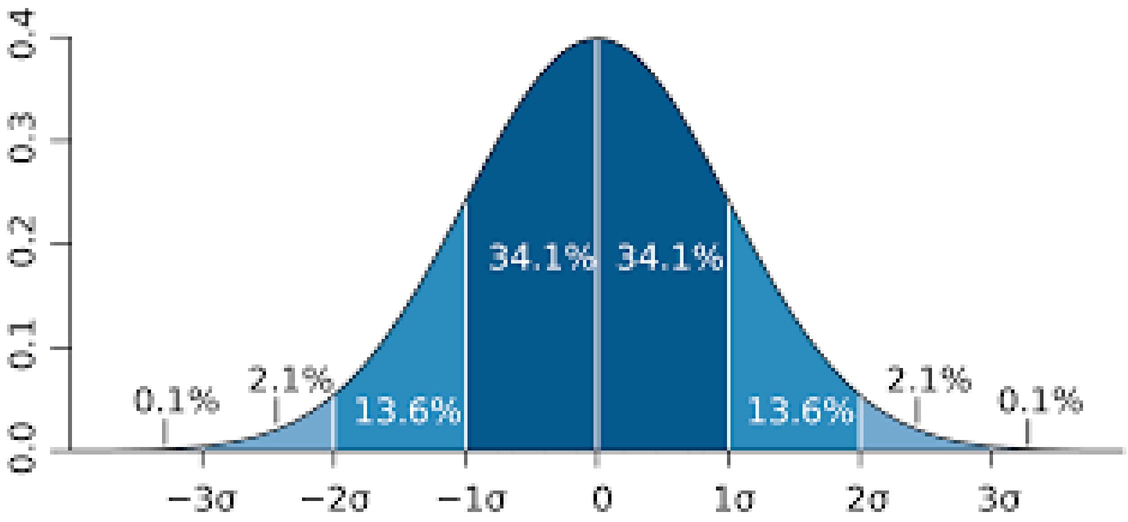
$\alpha\alpha$	$\alpha\beta$
$\beta\alpha$	$\beta\beta$

	Covariance Form	Information Form
Marginalization $p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\beta}$	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$ $\Sigma = \Sigma_{\alpha\alpha}$ (sub-block)	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$ (Schur complement)
Conditioning $p(\boldsymbol{\alpha}\vert\boldsymbol{\beta}) = \frac{p(\boldsymbol{\alpha}, \boldsymbol{\beta})}{p(\boldsymbol{\beta})}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$ (Schur complement)	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$ (sub-block)

Standard Deviation and Mahalanobis Distance

- Standard Deviation:
  - Square root of variance

$$\sigma = \sqrt{\sigma^2}$$



- Mahalanobis Distance

$$p(\mathbf{x}) = \alpha e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$

$$q^2 = (x-\mu)^\top \Sigma^{-1}(x-\mu) \qquad q = 1,2,3, \ldots$$

Mahalanobis Distance: $q = \sqrt{(x-\mu)^\top \Sigma^{-1}(x-\mu)}$

Visualizing Gaussians

$$p(\mathbf{x}) = \alpha e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)}$$

- Find contours of constant probability

$$q^2 = (x-\mu)^\top \Sigma^{-1}(x-\mu) \qquad q = 1,2,3,\ldots$$

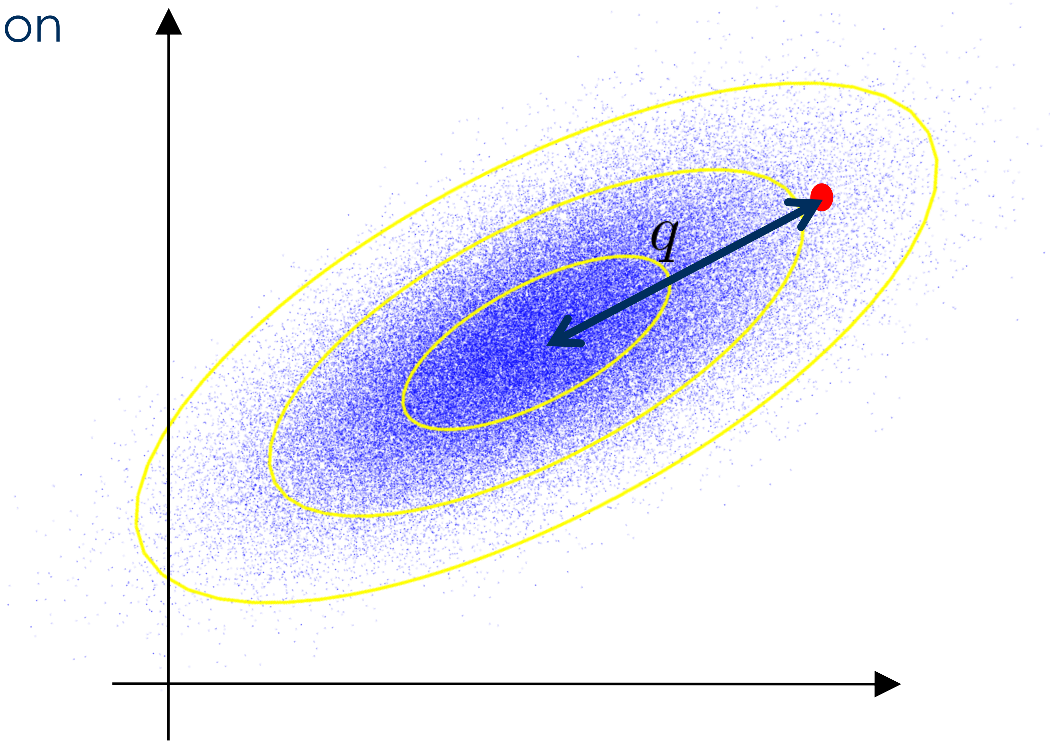
- Expand these terms, we end up with quadratic curve
  - An ellipse

Visualizing Gaussians

- Number of particles within each ellipse can be computed based on properties of Gaussian distributions
  - Distance from sample to mean (in terms of likelihood) is Mahalanobis Distance q
  - Square of q follows Chi-squared distribution

$$q^2 \sim \chi_k^2$$

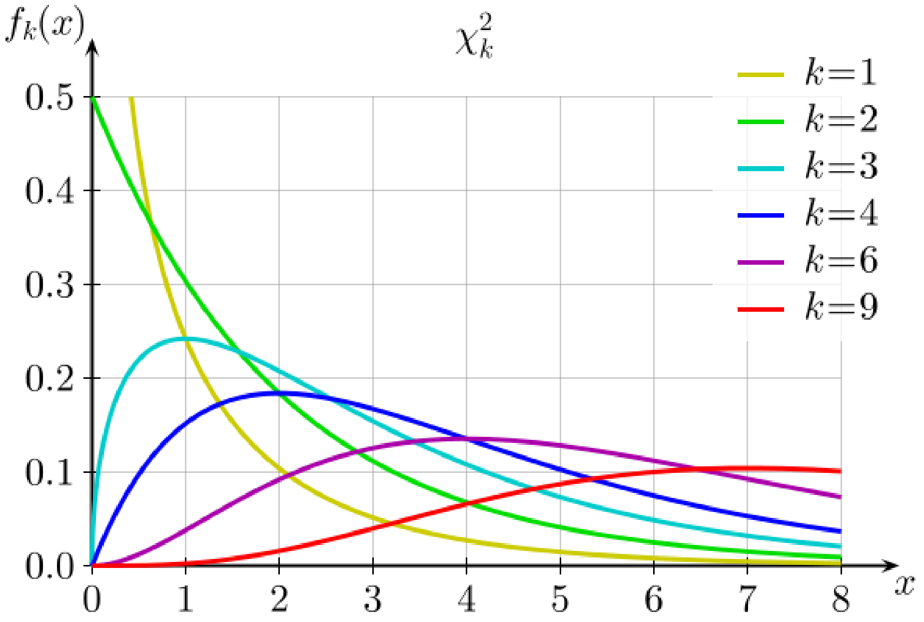
Sigma (n)	1D (k)	2D (k)
1	0.6827	0.3935
2	0.9545	0.8647
3	0.9973	0.9889

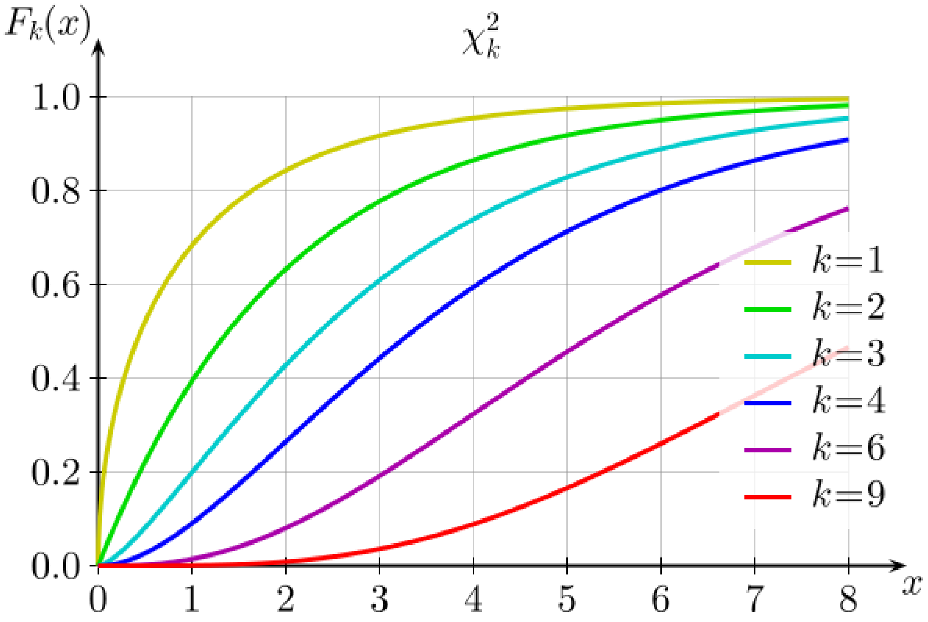


Chi-Square Distribution

- Distribution of the sum of squares of independent standard normal random variables

$$q^2 = \sum_{i=1,\cdots,k} x_i^2 \mid x_i \sim \mathcal{N},\ \forall i = 1, \cdots, k$$





Probability density function [Wikipedia] Cumulative distribution function

Visualizing Gaussians

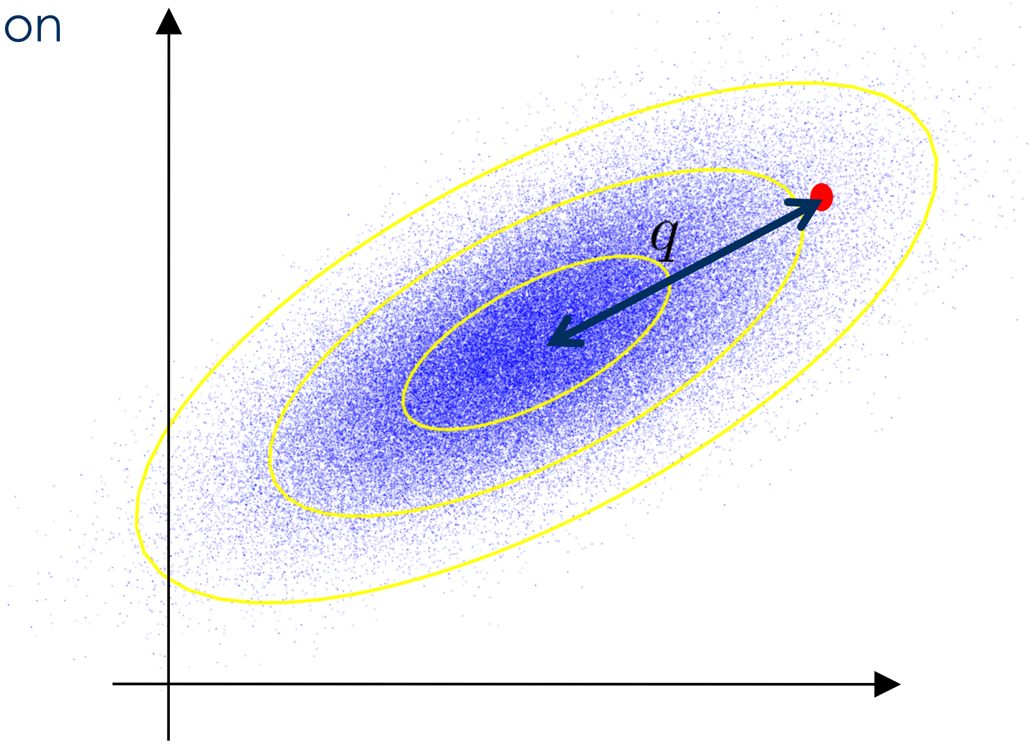
- Number of particles within each ellipse can be computed based on properties of Gaussian distributions
  - Distance from sample to mean (in terms of likelihood) is Mahalanobis Distance q
  - Square of q follows Chi-squared distribution

$$q^2 \sim \chi_k^2$$

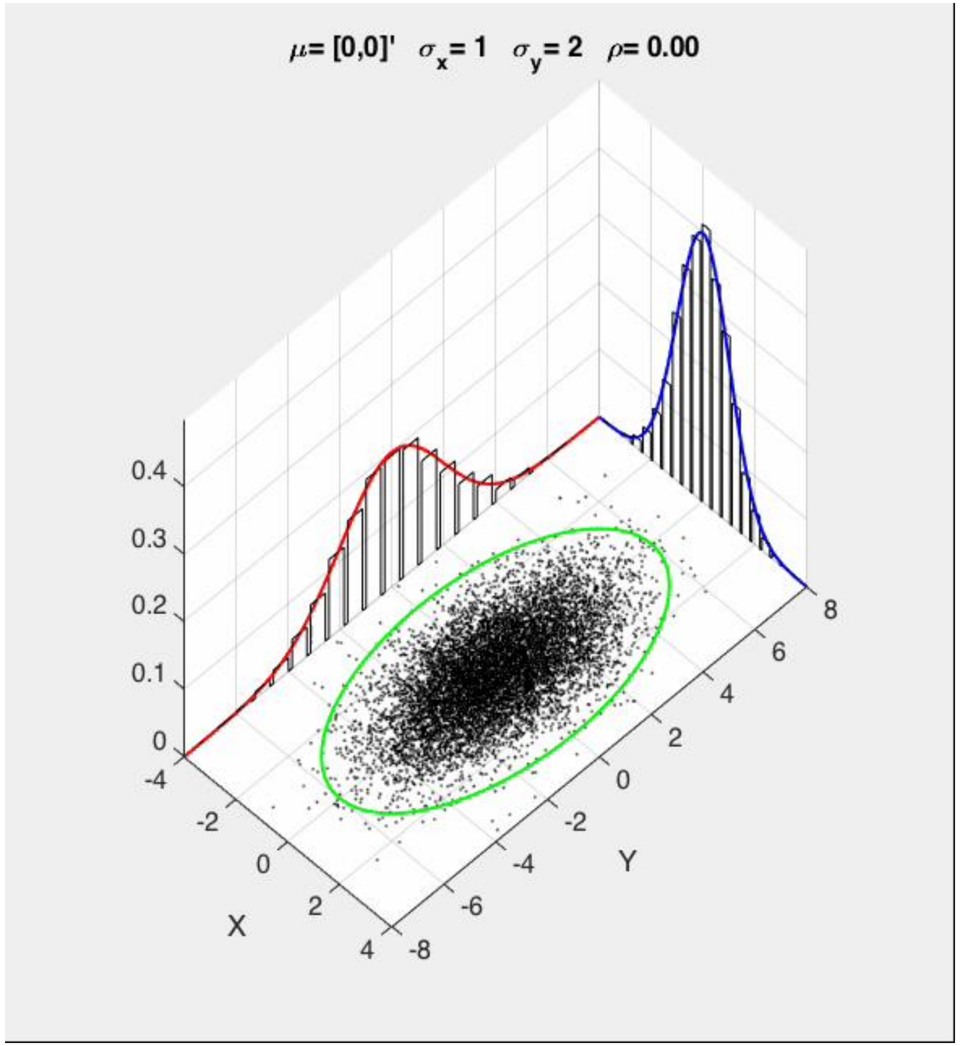
k

n

Sigma	1D	2D
1	chi2cdf(1,1)	chi2cdf(1,2)
2	chi2cdf(4,1)	chi2cdf(4,2)
3	chi2cdf(9,1)	chi2cdf(9,2)

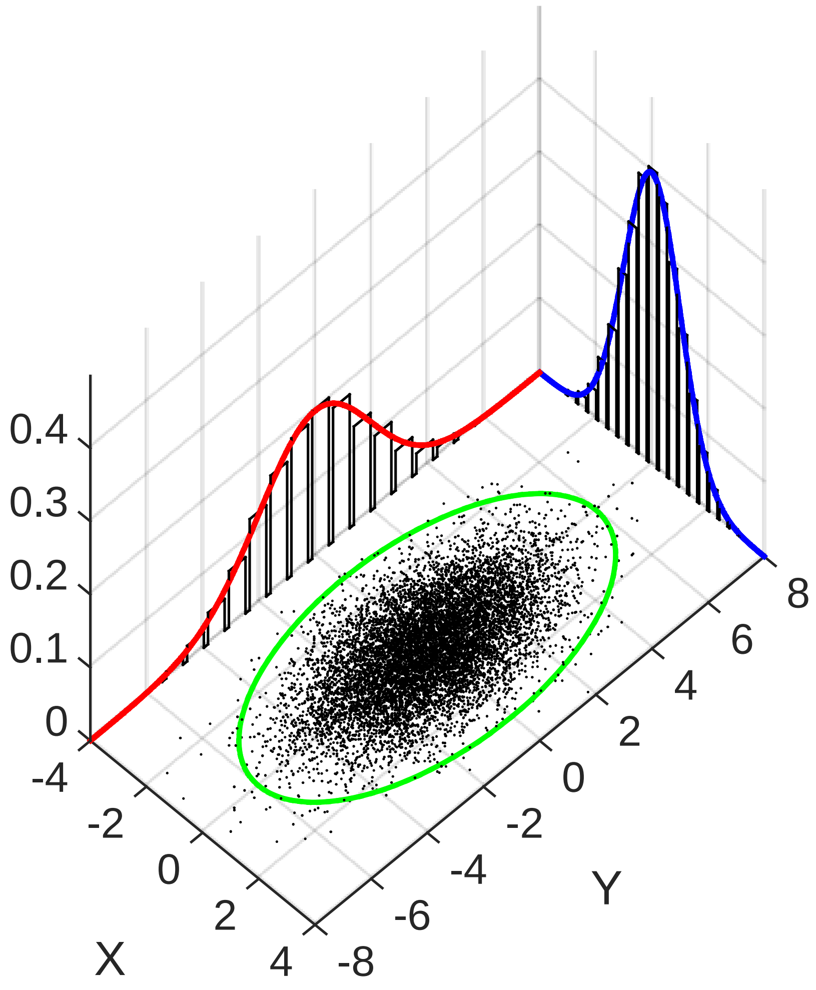


2-DOF Gaussian Correlation



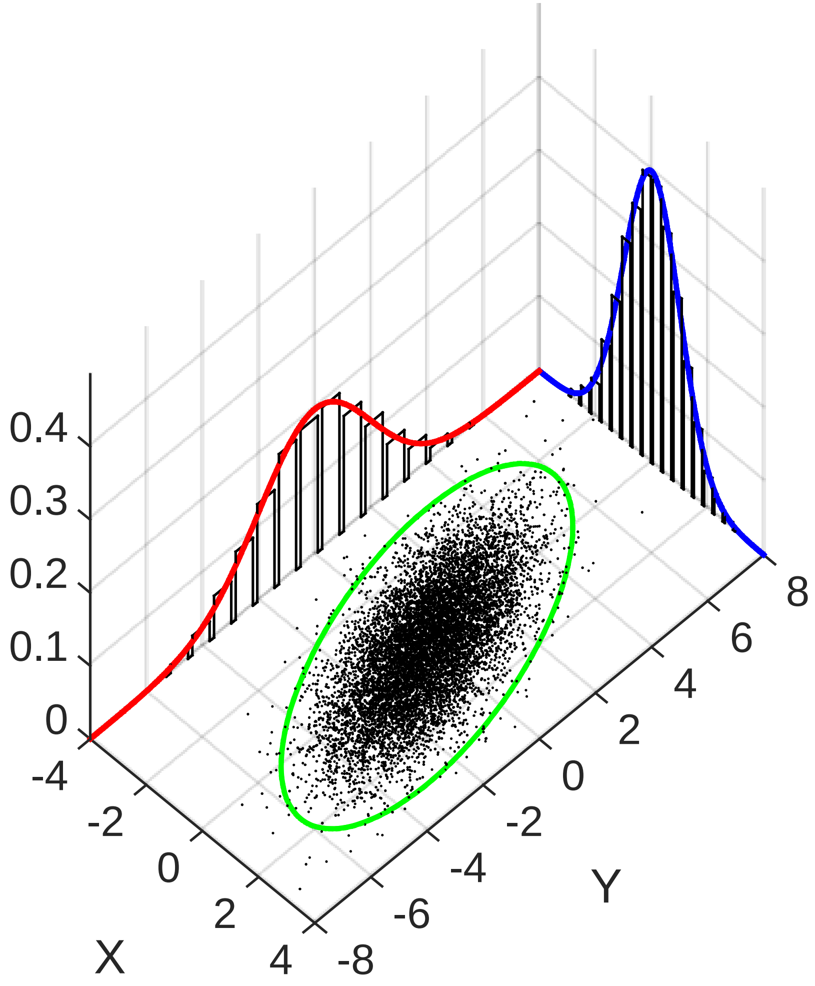
2-DOF Gaussian Correlation

- rho = 0
- Sigma =
- 1 0
- 0 4
- V =
- 1 0
- 0 1
- D =
- 1 0
- 0 4



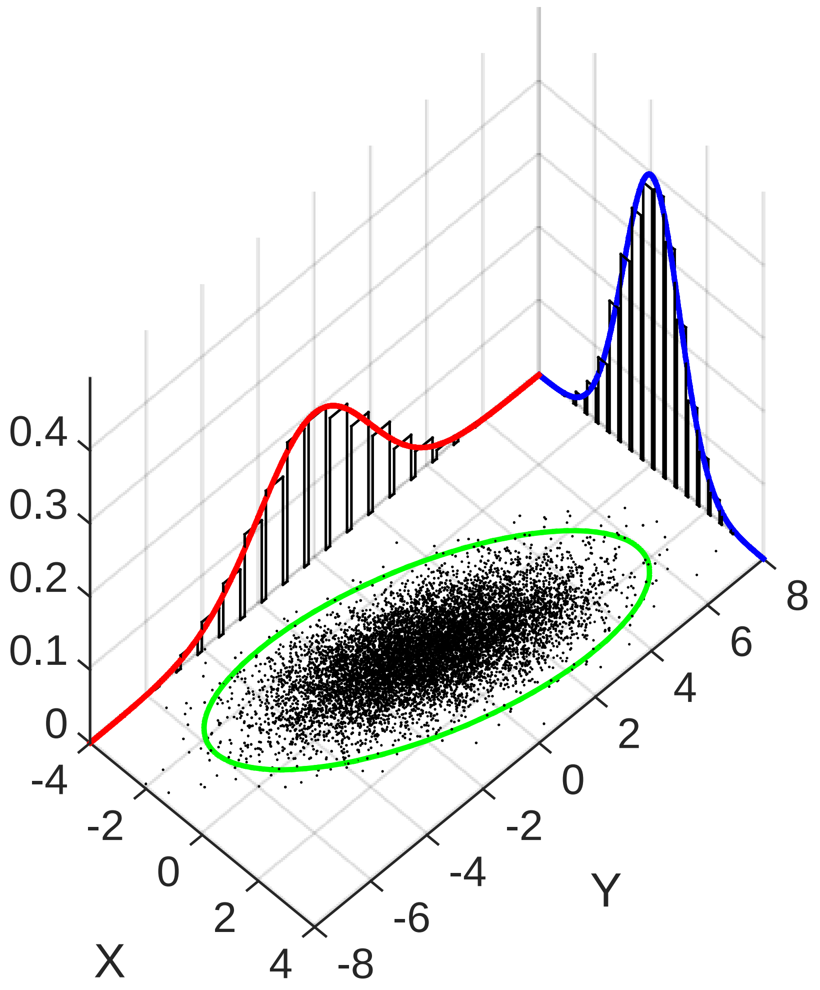
2-DOF Gaussian Correlation

- rho = -0.5000
- Sigma =
- 1 -1
- -1 4
- V =
- -0.9571 -0.2898
- -0.2898 0.9571
- D =
- 0.6972 0
- 0 4.3028



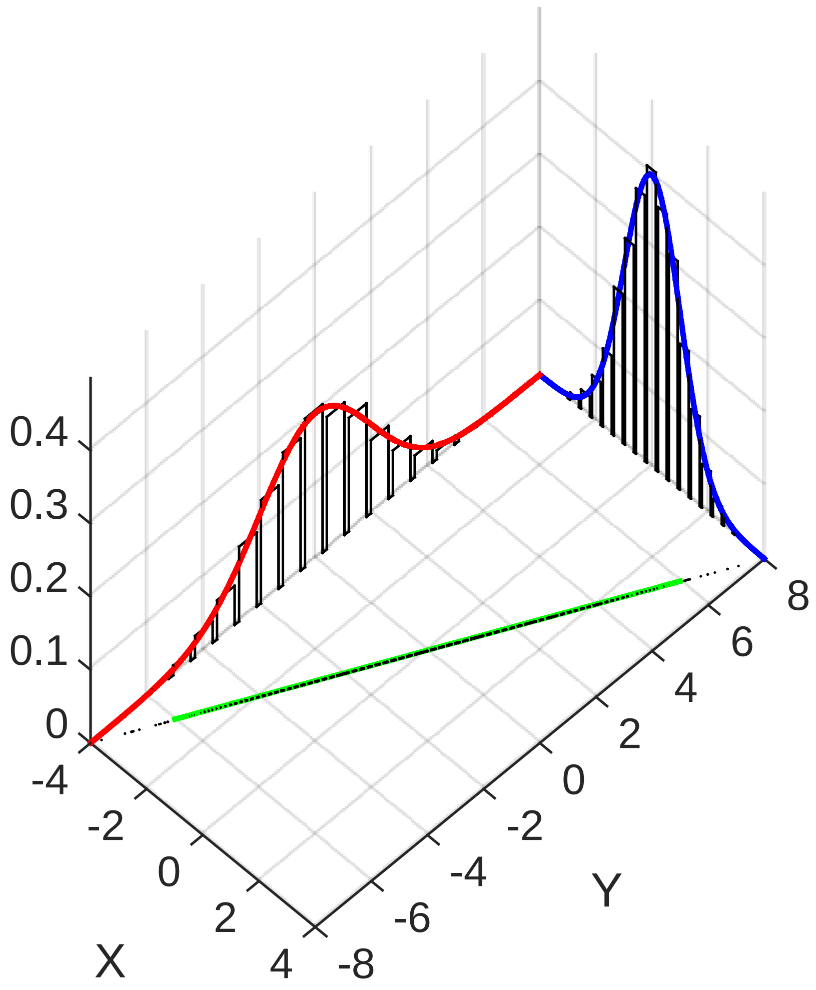
2-DOF Gaussian Correlation

- rho = 0.5000
- Sigma =
- 1 1
- 1 4
- V =
- -0.9571 0.2898
- 0.2898 0.9571
- D =
- 0.6972 0
- 0 4.3028



2-DOF Gaussian Correlation

- rho = 1
- Sigma =
-     1 2
-     2 4
- V =
-     -0.8944 0.4472
-     0.4472 0.8944
- D =
-     0 0
-     0 5



Implications of CLT

- We often estimate the state of something using many observations
  - Measuring the gravity on the moon by dropping a weight and timing the result
- Even if the distribution of each observation is non-Gaussian, their average will tend towards one.

Why use Gaussians?

- Convenience
  - Compact representation
  - Linear operations on Gaussians produce new Gaussians

- Central Limit Theorem: Distribution of the sum (or average) of N independent and identically distributed (IID) random variables approaches a normal distribution.
  - Only minor restrictions on the distribution of the individual random variables