

ESSENTIAL BACKGROUND II

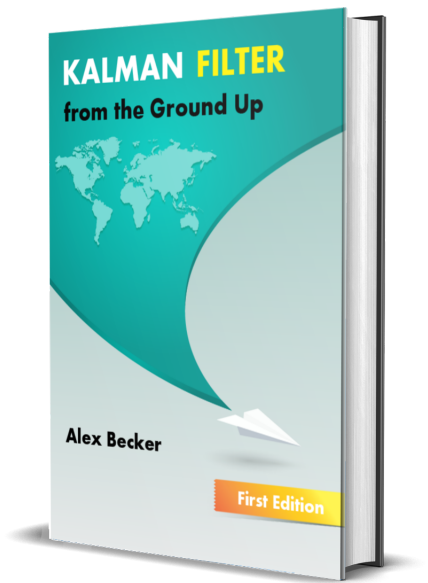
Before we tackle the multivariate Kalman Filter, we'll need to review some essential math topics:

- Matrix operations
- Expectation algebra
- Multivariate Normal Distribution

You can jump to the [next chapter](#) if you are familiar with these topics.

The notation used in this tutorial:

- Bold-face, lower-case letters refer to vectors, such as \mathbf{x}
- Bold-face, capital letters refer to matrices, such as \mathbf{A}
- Normal-face lower-case letters refer to scalars or vector elements
- Normal-face capital letters refer to matrix elements



 Get the book

MATRIX OPERATIONS

All you need to know is basic terms and operations such as:

- Vector and matrix addition and multiplication
- Matrix Transpose
- Matrix Inverse (you don't need to invert matrices by yourself, you just need to know what the inverse of the matrix is)
- Symmetric Matrix
- Eigenvalues and eigenvectors

There are numerous Linear Algebra textbooks and web tutorials that cover these topics.

EXPECTATION ALGEBRA

I extensively use the expectation algebra rules for Kalman Filter equations derivations. If you are interested in understanding the derivations, you need to master expectation algebra.

You already know what a [random variable](#) is and what an [expected value \(or expectation\)](#) is. If not, please read the [Essential background I section](#).

BASIC EXPECTATION RULES

The expectation is denoted by the capital letter E .

The expectation of the random variable $E(X)$ equals the mean of the random variable:

$$E(X) = \mu_X$$

Where μ_X is the mean of the random variable.

Here are some basic expectation rules:

	Rule	Notes
1	$E(X) = \mu_X = \sum xp(x)$	$p(x)$ is the probability of x (discrete case)
2	$E(a) = a$	a is constant
3	$E(aX) = aE(X)$	a is constant
4	$E(a \pm X) = a \pm E(X)$	a is constant
5	$E(a \pm bX) = a \pm bE(X)$	a and b are constant
6	$E(X \pm Y) = E(X) \pm E(Y)$	Y is another random variable
7	$E(XY) = E(X)E(Y)$	If X and Y are independent

VARIANCE AND COVARIANCE EXPECTATION RULES

The following table includes the variance and covariance expectation rules.

	Rule	Notes
8	$V(a) = 0$	$V(a)$ is the variance of a a is constant
9	$V(a \pm X) = V(X)$	$V(X)$ is the variance of X a is constant

10	$V(X) = E(X^2) - \mu_X^2$	$V(X)$ is the variance of X
11	$COV(X, Y) = E(XY) - \mu_X\mu_Y$	$COV(X, Y)$ is a covariance of X and Y
12	$COV(X, Y) = 0$	if X and Y are independent
13	$V(aX) = a^2V(X)$	a is constant
14	$V(X \pm Y) = V(X) + V(Y) \pm 2COV(X, Y)$	
15	$V(XY) \neq V(X)V(Y)$	

The variance and covariance expectation rules are not straightforward. I prove some of them.

RULE 8

$$V(a) = 0$$

A constant does not vary, so the variance of a constant is 0.

RULE 9

$$V(a \pm X) = V(X)$$

Adding a constant to the variable does not change its variance.

RULE 10

$$V(X) = E(X^2) - \mu_X^2$$

The proof:

Notes	
$V(X) = \sigma_X^2 = E((X - \mu_X)^2)$	
$= E(X^2 - 2X\mu_X + \mu_X^2)$	
$= E(X^2) - E(2X\mu_X) + E(\mu_X^2)$	Applied rule number 5: $E(a \pm bX) = a \pm bE(X)$
$= E(X^2) - 2\mu_X E(X) + E(\mu_X^2)$	Applied rule number 3: $E(aX) = aE(X)$
$= E(X^2) - 2\mu_X E(X) + \mu_X^2$	Applied rule number 2: $E(a) = a$
$= E(X^2) - 2\mu_X\mu_X + \mu_X^2$	Applied rule number 1: $E(X) = \mu_X$
$= E(X^2) - \mu_X^2$	

RULE 11

The proof:

Notes	
$COV(X, Y) = E((X - \mu_X)(Y - \mu_Y))$	
$= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$	
$= E(XY) - E(X\mu_Y) - E(Y\mu_X) + E(\mu_X\mu_Y)$	Applied rule number 6: $E(X \pm Y) = E(X) \pm E(Y)$
$= E(XY) - \mu_Y E(X) - \mu_X E(Y) + E(\mu_X\mu_Y)$	Applied rule number 3: $E(aX) = aE(X)$
$= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y$	Applied rule number 2: $E(a) = a$
$= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y$	Applied rule number 1: $E(X) = \mu_X$
$= E(XY) - \mu_X\mu_Y$	

RULE 13

$V(aX) = a^2V(X)$

The proof:

Notes	
$V(K) = \sigma_K^2 = E(K^2) - \mu_K^2$	
$K = aX$	
$V(K) = V(aX) = E((aX)^2) - (a\mu_X)^2$	Substitute K with aX
$= E(a^2X^2) - a^2\mu_X^2$	
$= a^2E(X^2) - a^2\mu_X^2$	Applied rule number 3: $E(aX) = aE(X)$
$= a^2(E(X^2) - \mu_X^2)$	
$= a^2V(X)$	Applied rule number 10: $V(X) = E(X^2) - \mu_X^2$

For constant velocity motion:

$$V(x) = \Delta t^2 V(v)$$

or

$$\sigma_x^2 = \Delta t^2 \sigma_v^2$$

Where:
 x is the displacement of the body
 v is the velocity of the body
 Δt is the time interval

$$V(X \pm Y) = V(X) + V(Y) \pm 2COV(X, Y)$$

The proof:

Notes	
$V(X \pm Y)$	
$= E((X \pm Y)^2) - (\mu_X \pm \mu_Y)^2$	Applied rule number 10: $V(X) = E(X^2) - \mu_X^2$
$= E(X^2 \pm 2XY + Y^2) - (\mu_X^2 \pm 2\mu_X\mu_Y + \mu_Y^2)$	
$= E(X^2) - \mu_X^2 + E(Y^2) - \mu_Y^2 \pm 2(E(XY) - \mu_X\mu_Y)$	Applied rule number 6: $E(X \pm Y) = E(X) \pm E(Y)$
$= V(X) + V(Y) \pm 2(E(XY) - \mu_X\mu_Y)$	Applied rule number 10: $V(X) = E(X^2) - \mu_X^2$
$= V(X) + V(Y) \pm 2COV(X, Y)$	Applied rule number 11: $COV(X, Y) = E(XY) - \mu_X\mu_Y$

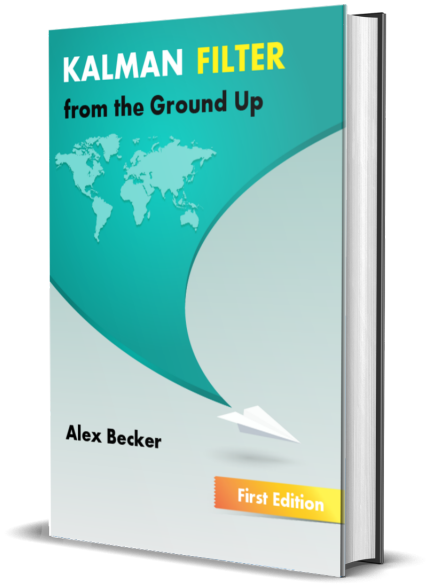
MULTIVARIATE NORMAL DISTRIBUTION

INTRODUCTION

We've seen that the Kalman Filter output is a random variable. The mean of the random variable represents the state estimate. The variance of the random variable is the estimation uncertainty. The Kalman Filter provides us with the estimate and the level of confidence of its estimate.

The one-dimensional Kalman Filter equations include four uncertainty variables:

- $p_{n,n}$ is the variance of an estimate (the current state)
- $p_{n+1,n}$ is the variance of a prediction (the next state)
- r_n is the measurement variance
- q is the process noise



 Get the book

For a multivariate Kalman Filter, the system state is described by a vector with more than one variable. For example, the object's position on the plane can be described by two variables: x – position and y – position:

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

The Kalman Filter output is a **multivariate random variable**. A **covariance matrix** describes the squared uncertainty of the multivariate random variable.

KALMAN FILTER TUTORIAL

ORDER

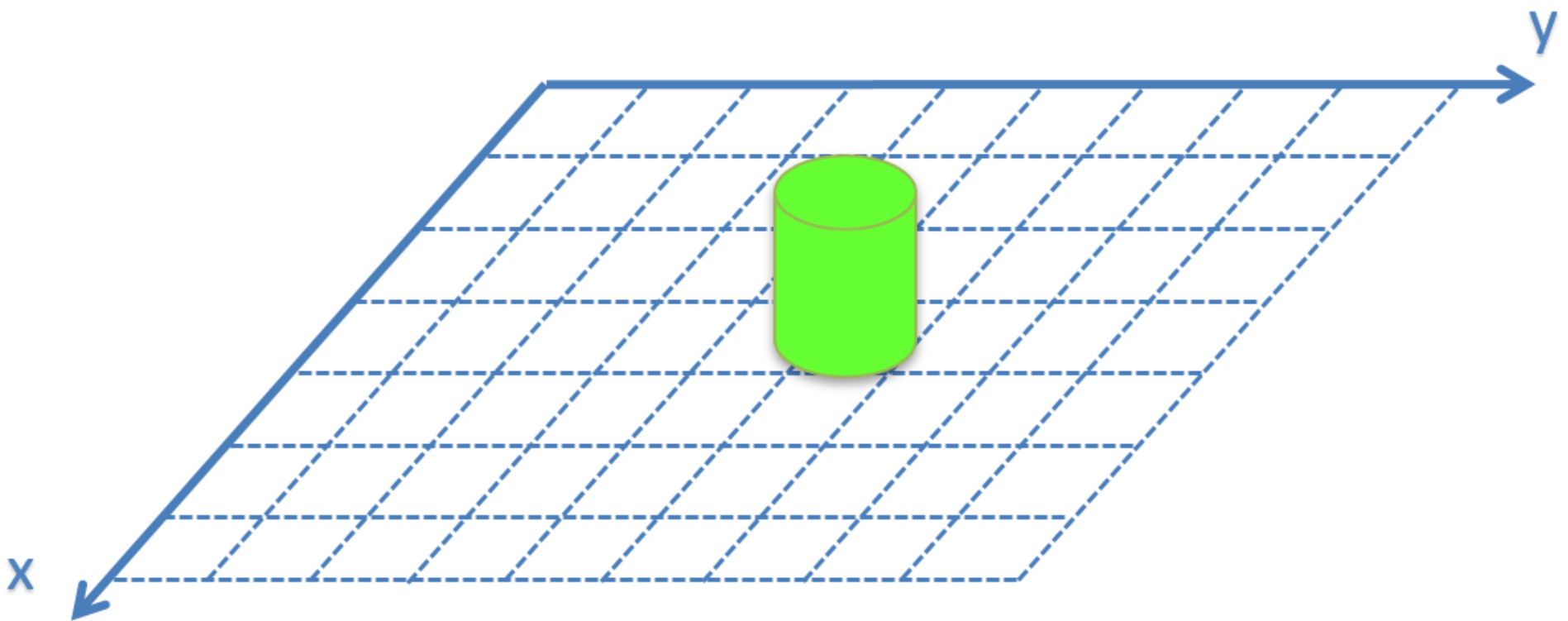
- $P_{n+1,n}$ is a covariance matrix that describes the squared uncertainty of a prediction
- R_n is a covariance matrix that describes the squared measurement uncertainty
- Q is a covariance matrix that describes the process noise

This chapter is about the **multivariate normal distribution** and the covariance matrix.

COVARIANCE

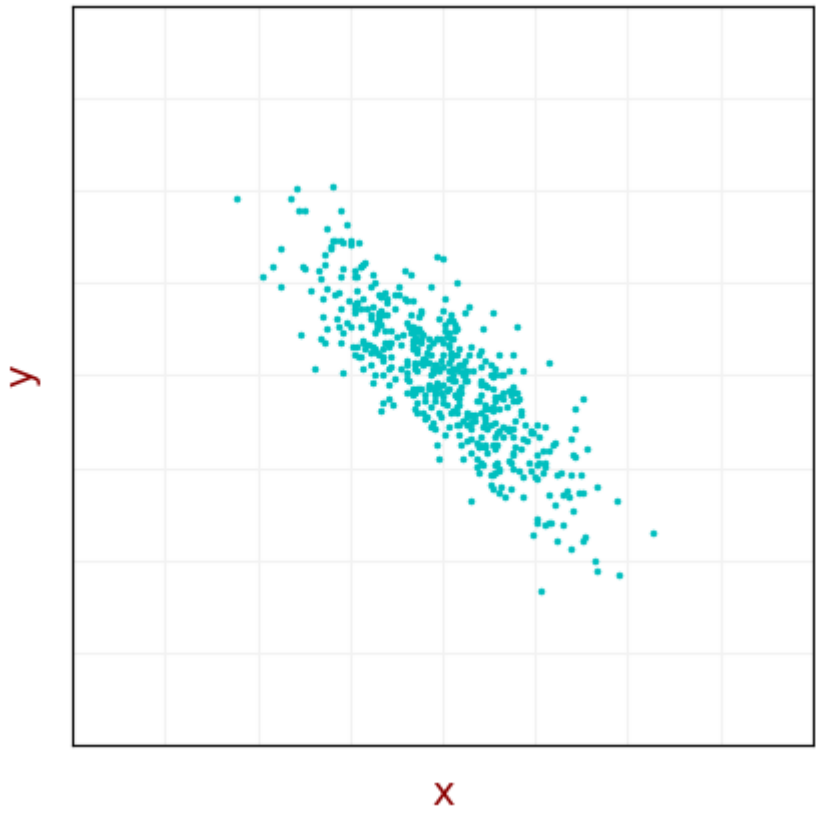
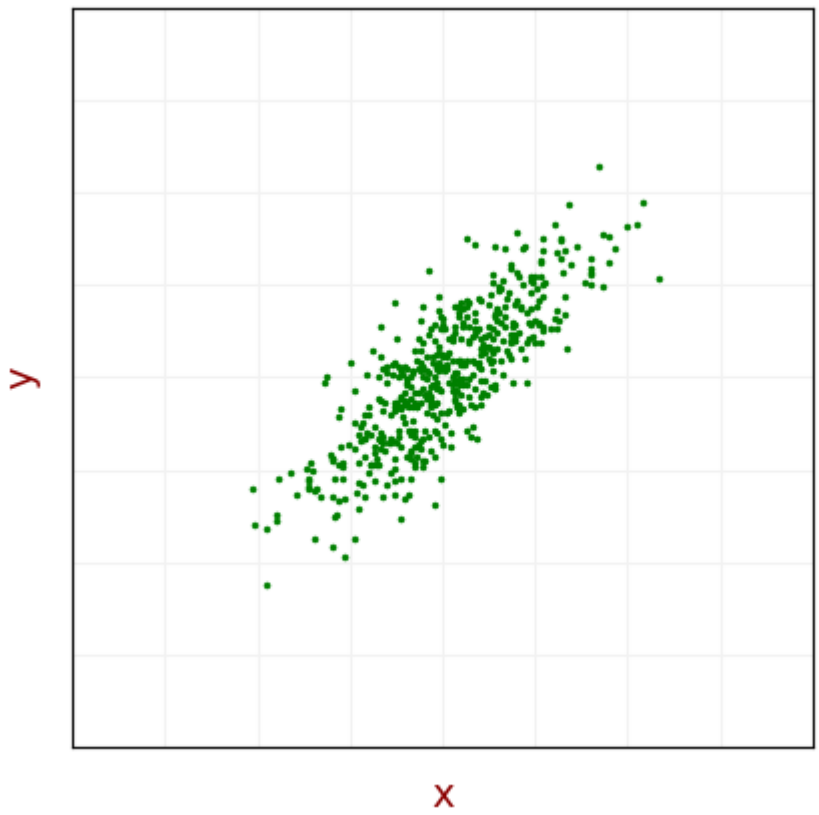
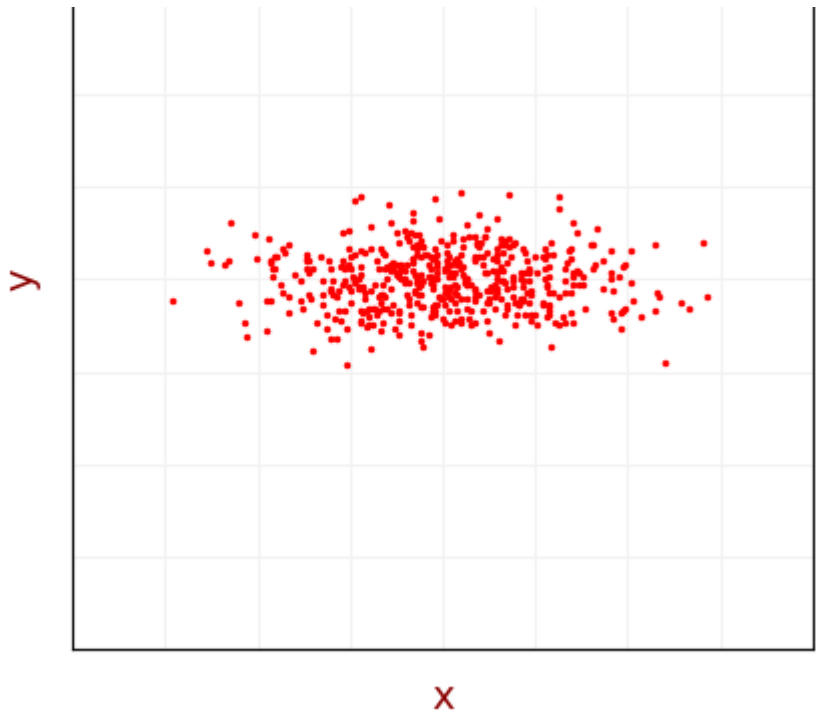
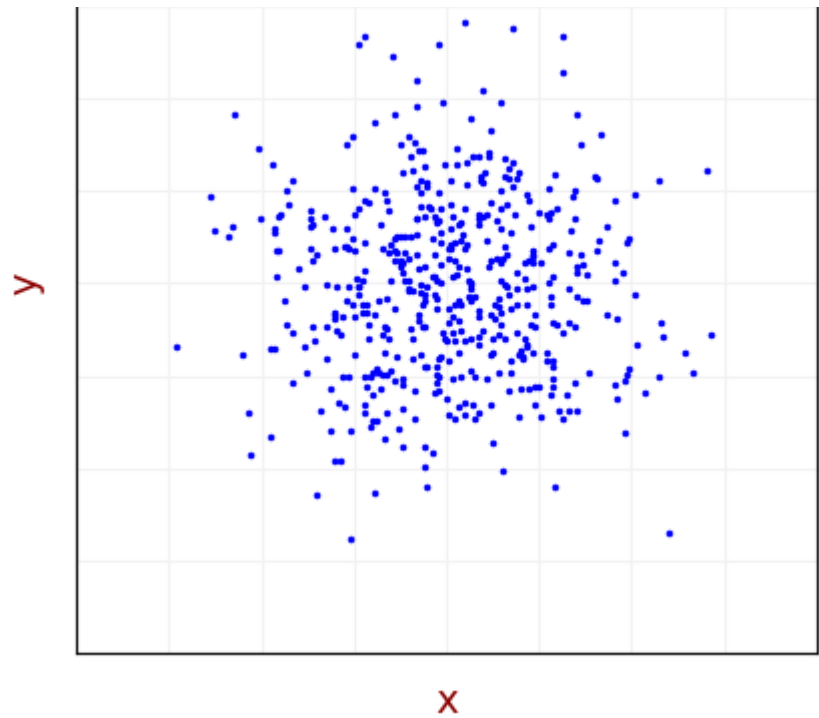
Covariance is a measure of the strength of the correlation between two or more sets of random variates.

Assume a set of object location measurements on the $x - y$ plane.



Due to the random error, there is a variance in the measurements.

Let us see some examples of different measurement sets.



The two upper subplots demonstrate uncorrelated measurements. The x values don't depend on y values. The x and y values of the blue data set have the same variance, and we can see a circular distribution shape. For the red data set, the variance of the x values is greater than that of the y values, and the distribution shape is elliptic.

Since the measurements are uncorrelated, the covariance of x and y equals zero.

The two lower subplots demonstrate correlated measurements. There is a dependency between x and y values. For the green data set, an increase in x results in an increase in y and vice versa. The correlation is positive; therefore, the covariance is positive. For the cyan data set, an increase in x results in a decrease in y and vice versa. The correlation is negative; therefore, the covariance is negative.

The covariance between population X and population Y with size N is given by:

$$COV(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

Let us rewrite the covariance equation:

Notes

$$COV(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i - x_i \mu_y - y_i \mu_x + \mu_x \mu_y)$$

Open parenthesis

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i) - \frac{1}{N} \sum_{i=1}^N (x_i \mu_y) - \frac{1}{N} \sum_{i=1}^N (y_i \mu_x) + \frac{1}{N} \sum_{i=1}^N (\mu_x \mu_y)$$

Distribute

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i) - \frac{\mu_y}{N} \sum_{i=1}^N (x_i) - \frac{\mu_x}{N} \sum_{i=1}^N (y_i) + \mu_x \mu_y$$

$$\mu_x = \frac{1}{N} \sum_{i=1}^N (x_i); \mu_y = \frac{1}{N} \sum_{i=1}^N (y_i)$$

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i) - \mu_x \mu_y - \mu_x \mu_y + \mu_x \mu_y$$

$$= \frac{1}{N} \sum_{i=1}^N (x_i y_i) - \mu_x \mu_y$$

The covariance of a sample with size N is normalized by $N - 1$:

$$COV(X, Y) = \frac{1}{N - 1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

Let us rewrite the covariance equation:

Notes	
$COV(X, Y) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$	
$= \frac{1}{N-1} \sum_{i=1}^N (x_i y_i - x_i \mu_y - y_i \mu_x + \mu_x \mu_y)$	Open parenthesis
$= \frac{1}{N-1} \sum_{i=1}^N (x_i y_i) - \frac{1}{N-1} \sum_{i=1}^N (x_i \mu_y) - \frac{1}{N-1} \sum_{i=1}^N (y_i \mu_x) + \frac{1}{N-1} \sum_{i=1}^N (\mu_x \mu_y)$	Distribute
$= \frac{1}{N-1} \sum_{i=1}^N (x_i y_i) - \frac{\mu_y}{N-1} \sum_{i=1}^N (x_i) - \frac{\mu_x}{N-1} \sum_{i=1}^N (y_i) + \frac{N}{N-1} \mu_x \mu_y$	$\mu_x = \frac{1}{N} \sum_{i=1}^N (x_i); \mu_y = \frac{1}{N} \sum_{i=1}^N (y_i)$
$= \frac{1}{N-1} \sum_{i=1}^N (x_i y_i) - \frac{N}{N-1} \mu_x \mu_y - \frac{N}{N-1} \mu_x \mu_y + \frac{N}{N-1} \mu_x \mu_y$	
$= \frac{1}{N-1} \sum_{i=1}^N (x_i y_i) - \frac{N}{N-1} \mu_x \mu_y$	

Example:

Given samples:

$$\mathbf{x} = \begin{bmatrix} 2 \\ 3 \\ -1 \\ 4 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} 8 \\ 7 \\ 9 \\ 6 \end{bmatrix}$$

$$COV(X, Y) = \frac{1}{N - 1} \sum_{i=1}^N (x_i y_i) - \frac{N}{N - 1} \mu_x \mu_y =$$

$$= \frac{1}{3} (2 \times 8 + 3 \times 7 - 1 \times 9 + 4 \times 6) - \frac{4}{3} \left(\frac{(2 + 3 - 1 + 4)}{4} \frac{(8 + 7 + 9 + 6)}{4} \right) = -2.67$$

We can stack the samples in two vectors \mathbf{x} and \mathbf{y} . The covariance in vector notation is given by:

For a zero-mean random variable, the covariance is given by:

$$COV(X, Y) = \frac{1}{N - 1} \mathbf{x}^T \mathbf{y}$$

COVARIANCE MATRIX

A covariance matrix is a square matrix that represents the covariance between each pair of elements in a given multivariate random variable.

For a two-dimensional random variable, the covariance matrix is:

$$\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} VAR(\mathbf{x}) & COV(\mathbf{x}, \mathbf{y}) \\ COV(\mathbf{y}, \mathbf{x}) & VAR(\mathbf{y}) \end{bmatrix}$$

Note that the off-diagonal entries of the covariance matrix are equal since $COV(\mathbf{x}, \mathbf{y}) = COV(\mathbf{y}, \mathbf{x})$. If x and y are uncorrelated, the off-diagonal entries of the covariance matrix are zero.

For a n - dimensional random variable, the covariance matrix is:

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

Most scientific computer packages can compute the covariance matrix.

Python example

```
import numpy as np

x = np.array([2, 3, -1, 4])
y = np.array([8, 7, 9, 6])

C = np.cov(x,y)
print(C)

[[ 4.66666667 -2.66666667]
 [-2.66666667  1.66666667]]
```

MATLAB example

PROPERTIES OF THE COVARIANCE MATRIX

1. The diagonal entries of this covariance matrix are the variances of the components of the multivariate random variable.

$$\Sigma_{ii} = \sigma_i^2$$

KALMAN FILTER TUTORIAL

ORDER

$$tr(\mathbf{\Sigma}) = \sum_{i=1} \Sigma_{ii} \geq 0$$

3. Since $\Sigma_{ij} = \sigma_{ij} = \sigma_{ji} = \Sigma_{ji}$, the covariance matrix is symmetric:

$$\mathbf{\Sigma} = \mathbf{\Sigma}^T$$

4. The covariance matrix is positive semidefinite.

The matrix \mathbf{A} is called positive semidefinite if, for any vector $\mathbf{v} \neq 0$, $\mathbf{v}^T \mathbf{A} \mathbf{v} \geq 0$.

The eigenvalues of \mathbf{A} are non-negative.

COVARIANCE MATRIX AND EXPECTATION

Assume vector \mathbf{x} with k elements:

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

The covariance matrix of the vector \mathbf{x} is given by:

$$COV(\mathbf{x}) = E \left((\mathbf{x} - \boldsymbol{\mu}_x) (\mathbf{x} - \boldsymbol{\mu}_x)^T \right)$$

The proof:

$$\begin{aligned} COV(\mathbf{x}) &= E \left(\begin{bmatrix} (x_1 - \mu_{x_1})^2 & (x_1 - \mu_{x_1})(x_2 - \mu_{x_2}) & \cdots & (x_1 - \mu_{x_1})(x_k - \mu_{x_k}) \\ (x_2 - \mu_{x_2})(x_1 - \mu_{x_1}) & (x_2 - \mu_{x_2})^2 & \cdots & (x_2 - \mu_{x_2})(x_k - \mu_{x_k}) \\ \vdots & \vdots & \ddots & \vdots \\ (x_k - \mu_{x_k})(x_1 - \mu_{x_1}) & (x_k - \mu_{x_k})(x_2 - \mu_{x_2}) & \cdots & (x_k - \mu_{x_k})^2 \end{bmatrix} \right) = \\ &= E \left(\begin{bmatrix} (x_1 - \mu_{x_1}) \\ (x_2 - \mu_{x_2}) \\ \vdots \\ (x_k - \mu_{x_k}) \end{bmatrix} \begin{bmatrix} (x_1 - \mu_{x_1}) & (x_2 - \mu_{x_2}) & \cdots & (x_k - \mu_{x_k}) \end{bmatrix} \right) = \\ &= E \left((\mathbf{x} - \boldsymbol{\mu}_x) (\mathbf{x} - \boldsymbol{\mu}_x)^T \right) \end{aligned}$$

MULTIVARIATE NORMAL DISTRIBUTION

We are already familiar with a [univariate normal distribution](#). It is described by a bell-shaped Gaussian function:

$$p(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right)$$

The normal distribution is denoted by:

The multivariate normal distribution is a generalization of the univariate normal distribution for a multidimensional random variable.

The n - dimensional multivariate normal distribution is described by:

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

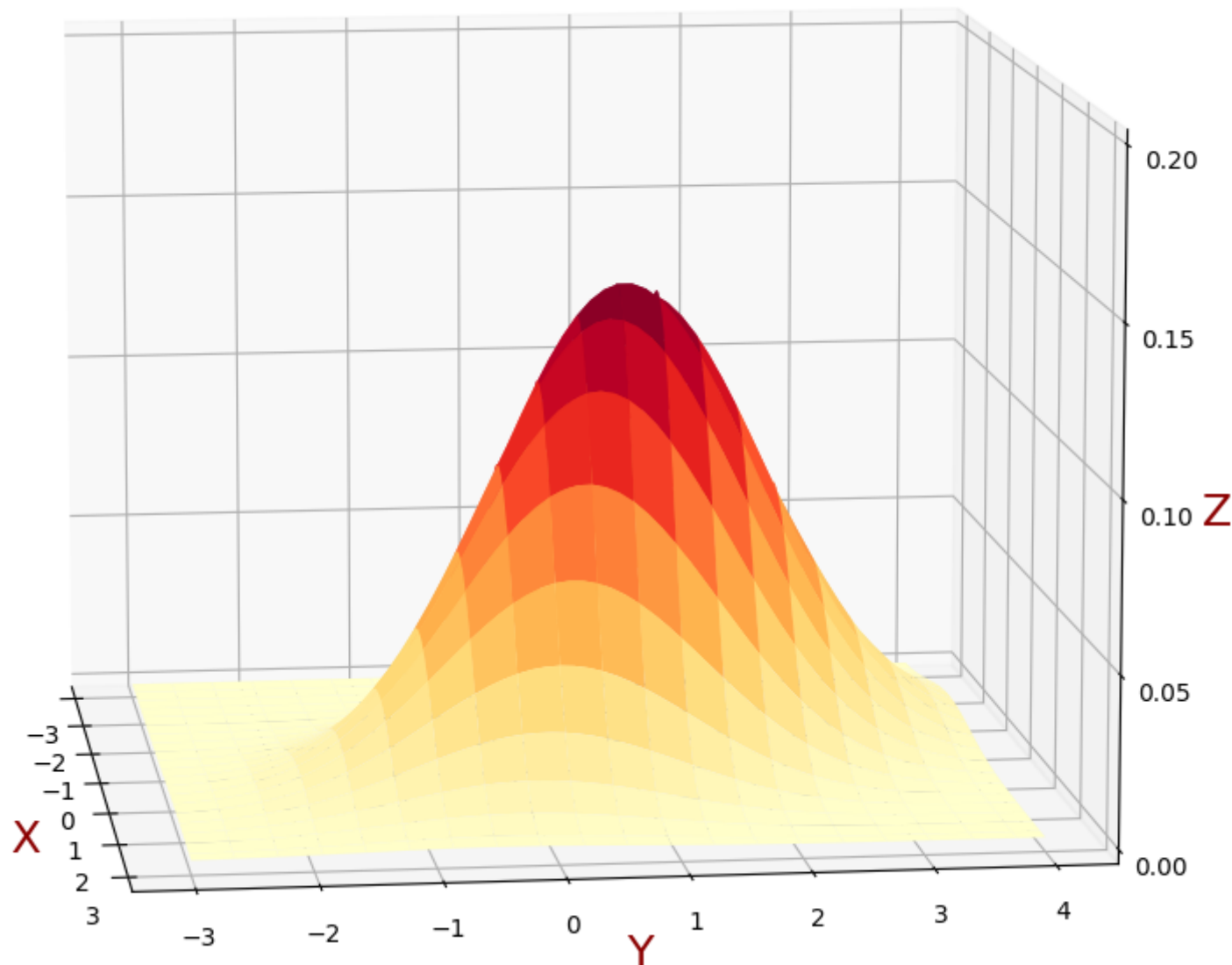
Where:

- \mathbf{x} is an n - dimensional random vector
- $\boldsymbol{\mu}$ is an n - dimensional mean vector
- $\boldsymbol{\Sigma}$ is a square $n \times n$ covariance matrix

BIVARIATE NORMAL DISTRIBUTION

A bivariate (two-dimensional) normal distribution comprises two normally distributed random variables. I want to focus on the bivariate normal distribution since it is the highest dimension of the multivariate normal distribution that we can visualize.

The following plot describes the bivariate Gaussian function.

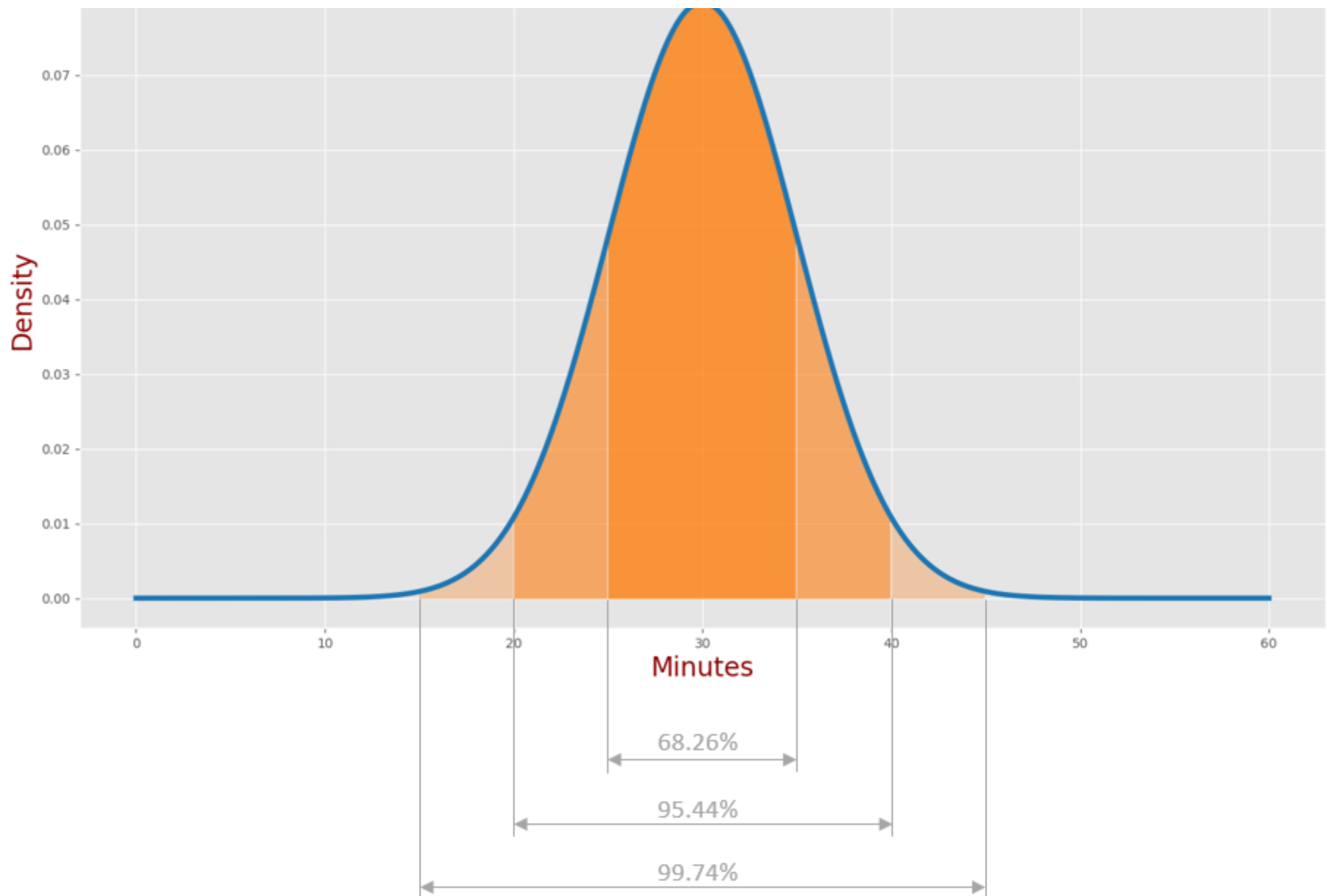


CONFIDENCE INTERVALS

A confidence interval specifies the probability that a parameter falls between a pair of values around the mean for the univariate normal distribution.

For the univariate normal distribution, the area between the 1σ boundaries and the Gaussian function is 68.26% of the total area under the Gaussian function.

KALMAN FILTER TUTORIAL

[ORDER](#)

For the univariate normal distribution, we can formulate the following statements:

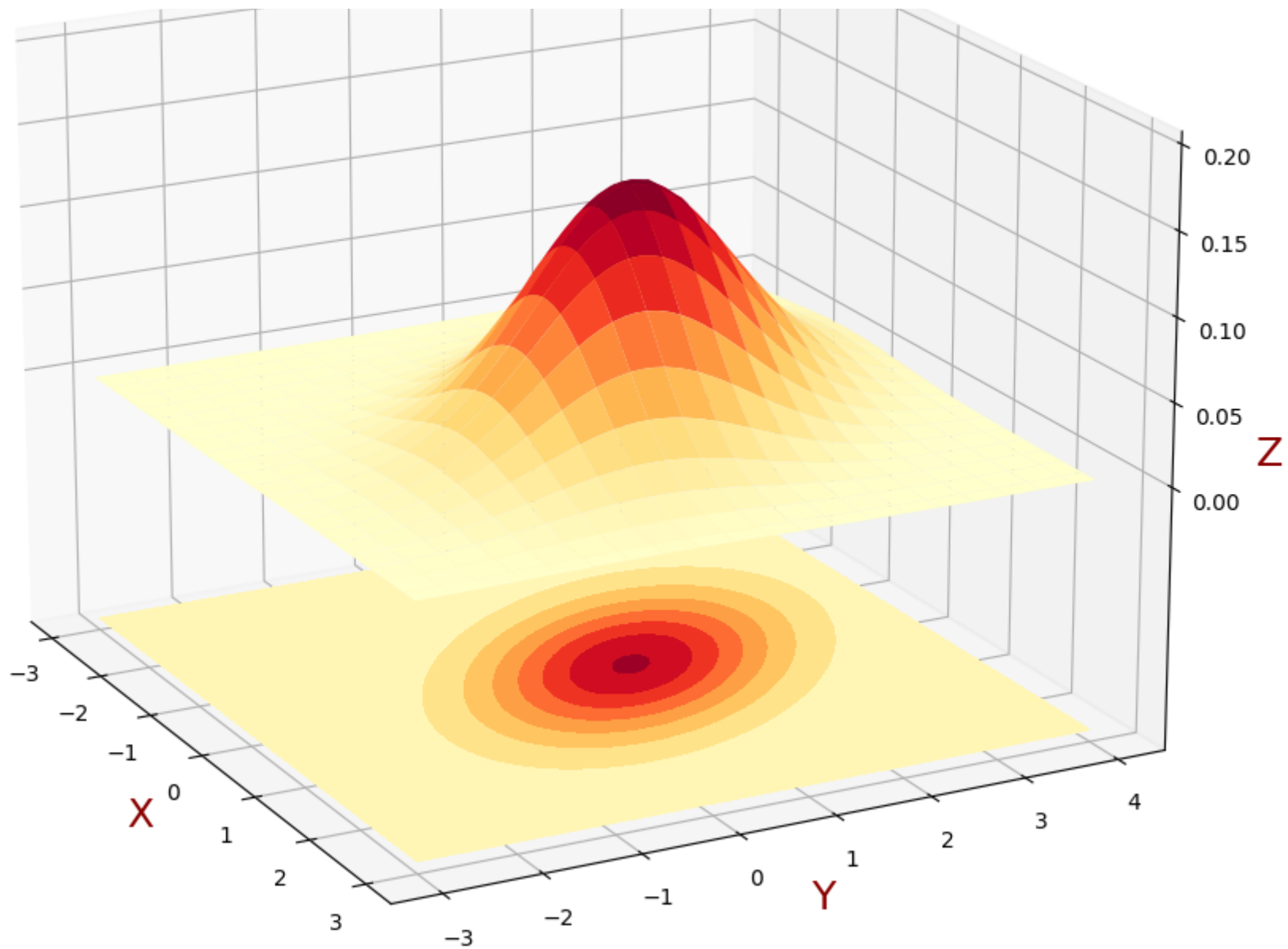
- The confidence interval for the probability of 68.26% is 1σ .
- The confidence interval for the probability of 95.44% is 2σ .
- The confidence interval for the probability of 99.74% is 3σ .

We can also find the confidence interval of any specified probability for the univariate normal distribution. You can see the explanation [here](#).

The probability of the bivariate normal distribution is a volume of the three-dimensional Gaussian function.

For example, the volume of the three-dimensional Gaussian function sliced at 1σ level is 39.35% of the total volume of the three-dimensional Gaussian function.

The projection of the three-dimensional Gaussian slice is an ellipse.

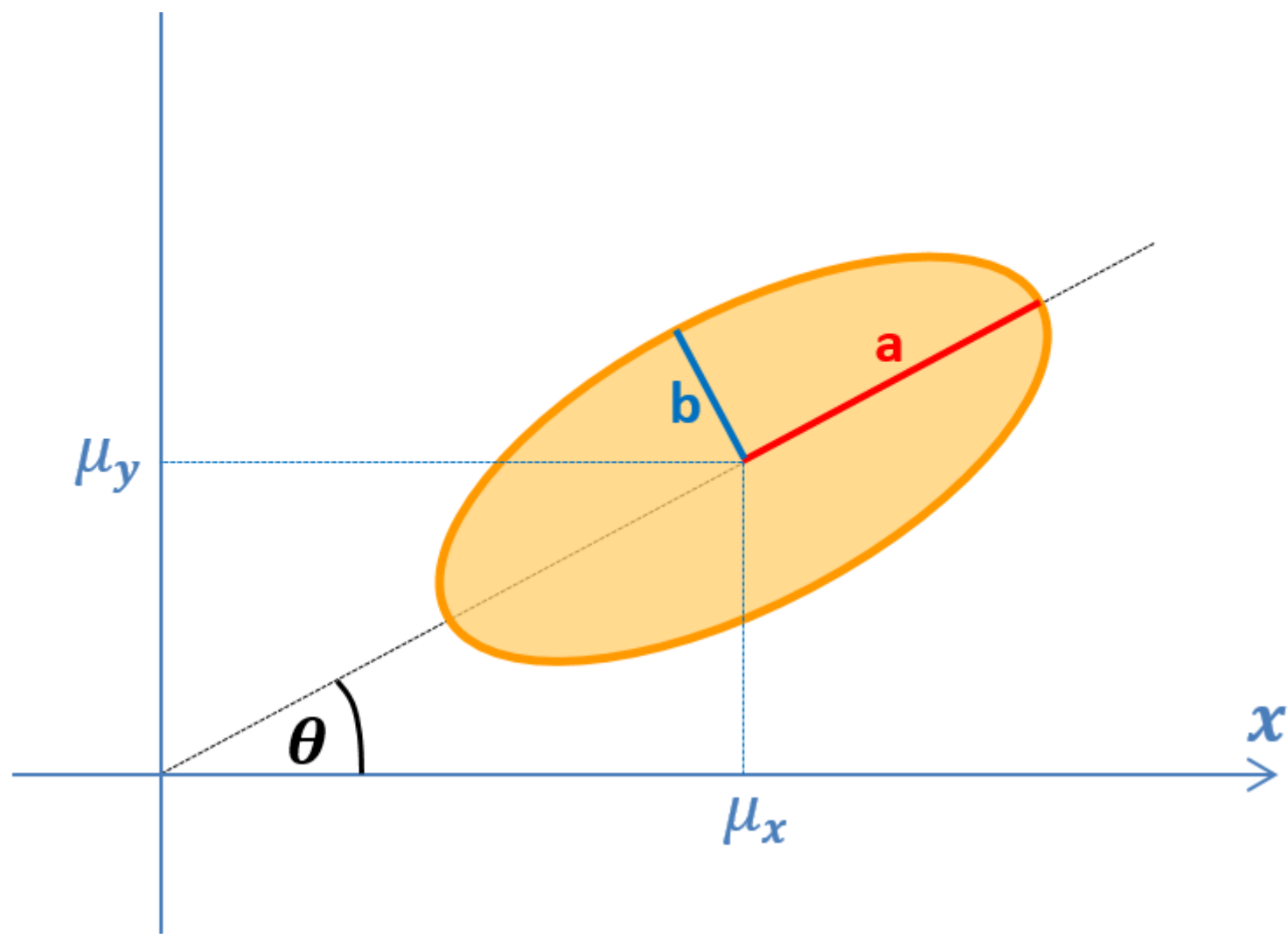


COVARIANCE ELLIPSE

First, let us find the properties of the covariance ellipse. The covariance ellipse represents an iso-contour of the Gaussian distribution and allows visualization of a 1σ confidence interval in two dimensions. The covariance ellipse provides a geometric interpretation of the covariance matrix.

Any ellipse can be described by four parameters:

- Ellipse center μ_x, μ_y
- Half-major axis a
- Half-minor b
- Orientation angle θ



The ellipse center is a mean of the random variable:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$$

The lengths of the ellipse axes are the square roots of the eigenvalues of the random variable covariance matrix:

- The length of the half-major axis *a* is given by the highest eigenvalue square root
- The length of the half-minor axis *b* is given by the second eigenvalue square root

The orientation of the ellipse is an orientation of the covariance matrix eigenvector that corresponds to the highest eigenvalue.

$$\theta = \arctan \left(\frac{v_y}{v_x} \right)$$

Where:

- *v_x* is the *x* - coordinate of the eigenvector that corresponds to the highest eigenvalue
- *v_y* is the *y* - coordinate of the eigenvector that corresponds to the highest eigenvalue

You can use a scientific computer package to compute the covariance ellipse parameters.

Python example

```
import numpy as np

C = np.array([[5, -2],[-2, 1]]) # define covariance matrix

eigVal, eigVec = np.linalg.eig(C) # find eigenvalues and eigenvectors

a = np.sqrt(eigVal[0]) # half-major axis length
b = np.sqrt(eigVal[1]) # half-minor axis length

# ellipse orientation angle
theta = np.arctan(eigVec[1, 0] / eigVec[0, 0])
```

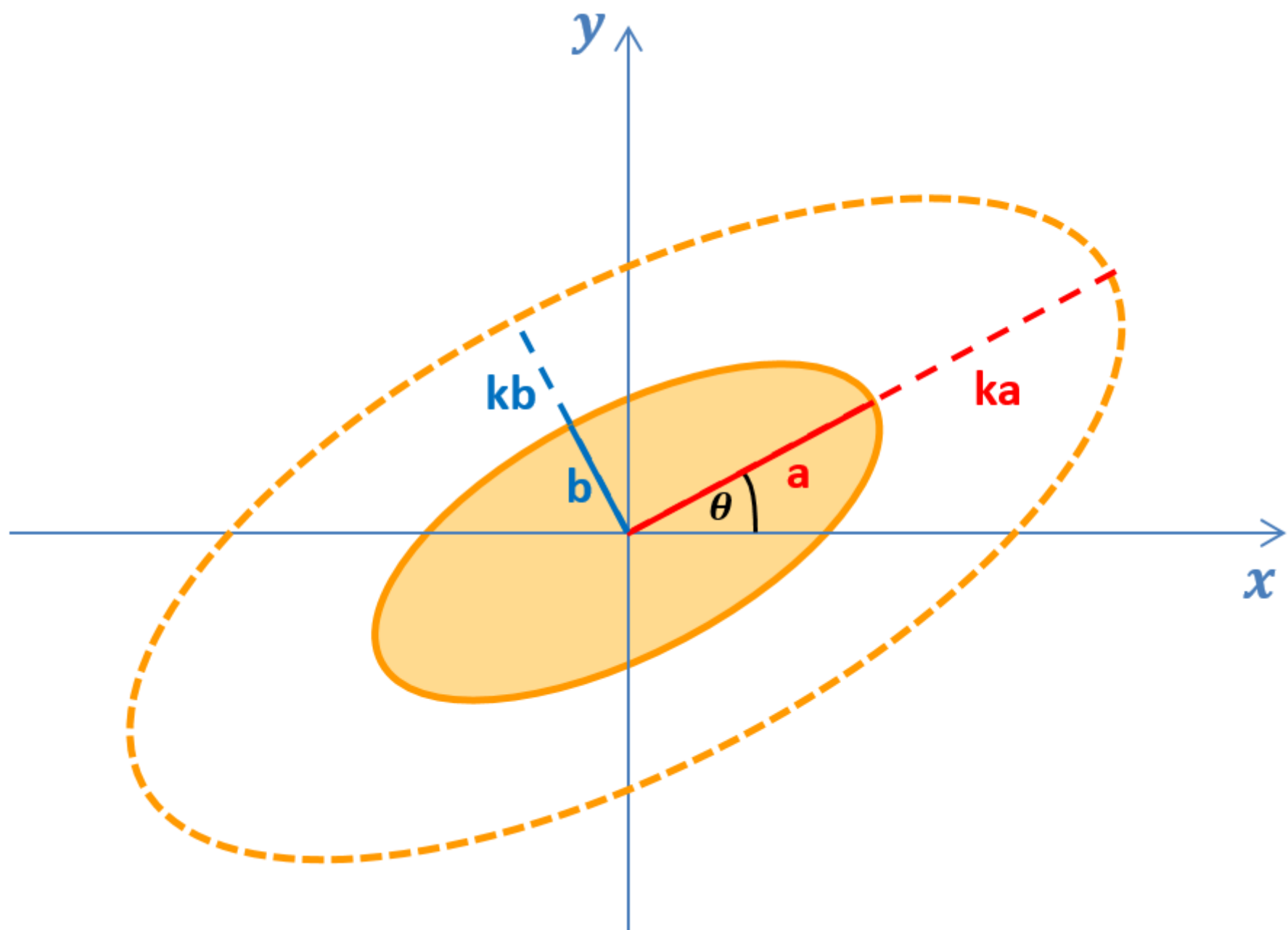
MATLAB example

CONFIDENCE ELLIPSE

In many applications, there is an interest in finding the boundaries of specific probability. For example, for 95% probability, we should find the boundary that includes 95% of the Gaussian function volume.

The projection of this boundary onto the $x - y$ plane is the confidence ellipse.

We want to find an **elliptical scale factor** k , that extends the covariance ellipse to the confidence ellipse associated with 95% probability.



Since σ_x and σ_y represent the standard deviations of stochastically independent random variables, the addition theorem for the chi-square distribution may be used to show that the probability associated with a confidence ellipse is given by:

For a covariance ellipse $k = 1$, therefore, the probability associated with a covariance ellipse is:

$$p = 1 - \exp\left(-\frac{1}{2}\right) = 39.35\%$$

For a given probability, we can find an elliptical scale factor:

$$k = \sqrt{-2\ln(1 - p)}$$

For the probability of 95%:

$$k = \sqrt{-2\ln(1 - 0.95)} = 2.45$$

The properties of the confidence ellipse associated with 95% probability are:

- Ellipse center (μ_x, μ_y) is similar to the covariance ellipse
- Orientation angle θ is similar to the covariance ellipse
- Half-major axis length is $2.45a$ – a scaled half-major axis of the covariance ellipse
- Half-minor axis length is $2.45b$ – a scaled half-minor axis of the covariance ellipse