Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

# Automotive Environment Sensing

## 02 – Introduction to probability

Olivér Törő

2019

# Event algebra

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

## Basic concepts

- **Sample space (Ω):** set of all possible events
- **Elementary events (ω):** disjoint events with a single outcome
- **Set of events $F$:** some or all subsets of Ω, that is the power set of Ω: $F \subseteq 2^{\Omega}$ and an algebra defined on it (σ-algebra)
- **Events $(\boldsymbol{A}, \boldsymbol{B}, \dots)$:** subsets of $F$, can be elementary or complex
- **Probability measure $P: F \rightarrow [0,1]$:** real valued additive function
- **An event has probability:** e.g. $P(A), P(\neg A), P(A \cap B)$ etc.
- Certain event: $P(\Omega) = 1$, impossible event: $P(\emptyset) = 0$
- The triplet $(\Omega, F, P)$ defines a probability space

# Event algebra – example



## Dice Roll

- **Sample space (Ω):** {1,2,3,4,5,6, even, odd, >3, etc}
- **Elementary events (ω):** {1,2,3,4,5,6}
- **Set of considered events ($\boldsymbol{F}$):** eg.: {∅,1,2,3,4,5,6, even}
- **Events ($\boldsymbol{A}, \boldsymbol{B}, \dots$):** {2, even, greater than 3 and odd, 4&5, etc}
- **Probability measure** $P: F \rightarrow [0,1]$**:** "favorable cases/possible cases" (Laplace)
- An event has probability: e.g. $P(A), P(\neg A), P(A \cap B)$ etc.
- Certain event: $P(\Omega) = 1$, impossible event: $P(\emptyset) = 0$
- The triplet $(\Omega, F, P)$ defines a probability space

# Event algebra – conditional probability

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- Conditional probability (definition)

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

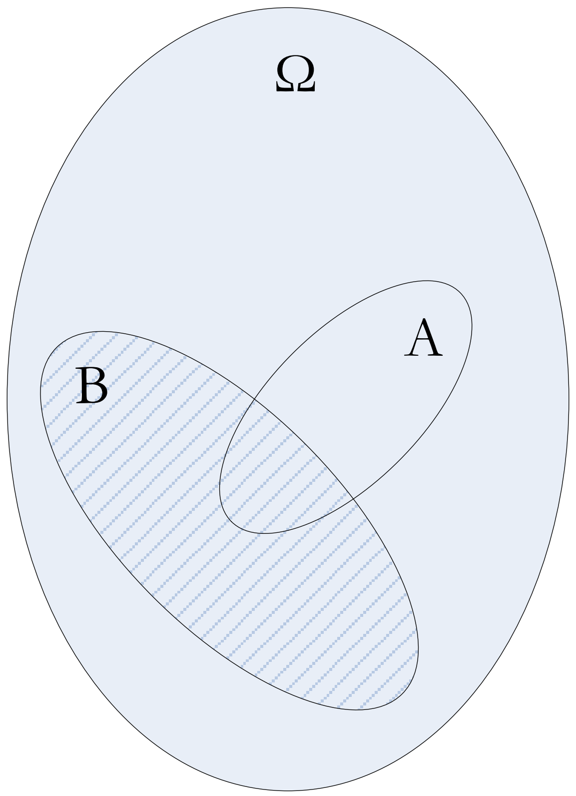
- Independent events

$$P(A|B) = P(A) \quad \text{és} \quad P(B|A) = P(B)$$

$$P(A \cap B) = P(A)P(B)$$

- Collectively exhaustive events

$$\bigcup_{i=1}^{N} B_i = \Omega \qquad B_i \cap B_j = \emptyset$$



# Correlation and causality



- Consider two events $A$ and $B$ with the following inequality

$$P(B|A) > P(B|\neg A)$$

- What does it indicate?

Dice roll example: $B =< 6 >, A =< \text{even} >$

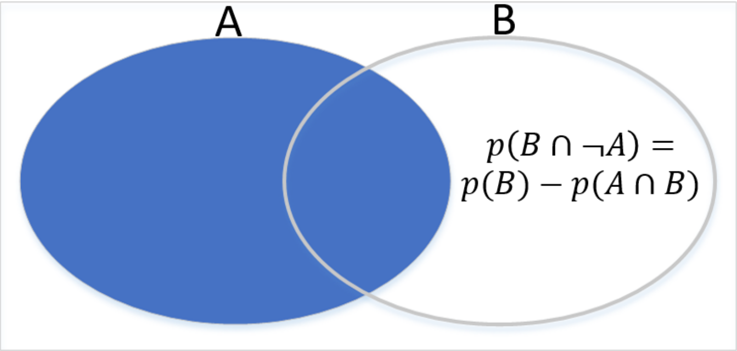
$$P(< 6 >) = 1/6 \qquad P(< \text{even} >) = \frac{1}{2}$$

LHS $\quad P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/6}{1/2} = \frac{1}{3}$ $\quad$ as expected

# Correlation and causality



RHS $$P(B|\neg A) = \frac{P(B\cap\neg A)}{P(\neg A)}$$



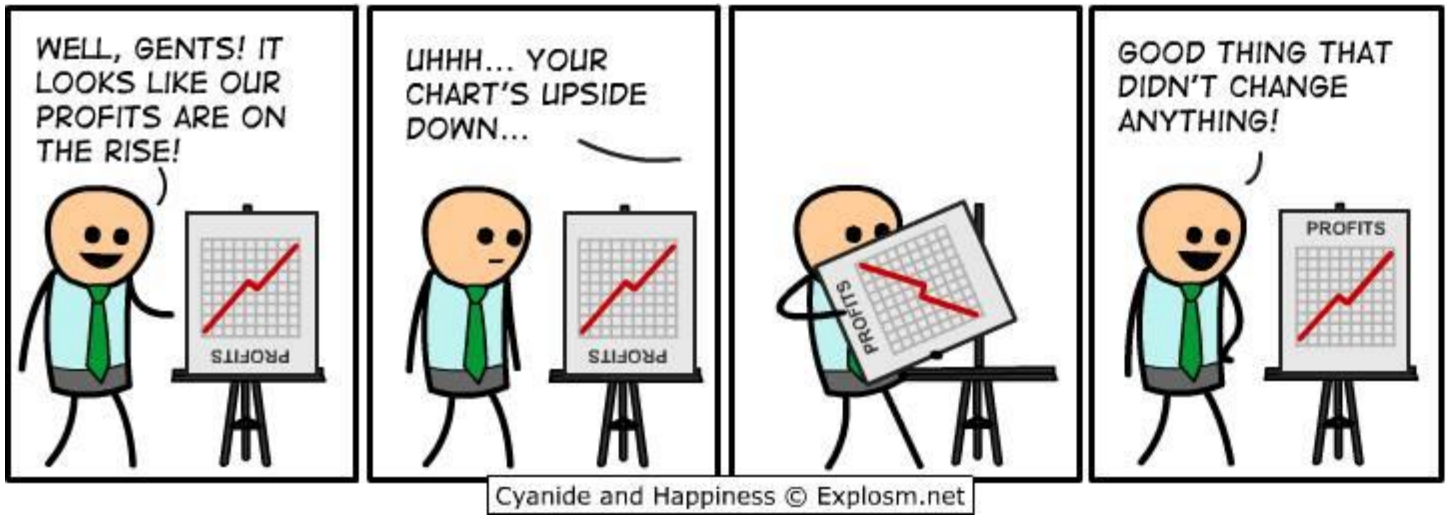

$$\frac{P(B)-P(B\cap A)}{1-P(A)} = \frac{1/6-1/6}{1-1/2} = 0$$

**cannot roll 6 and odd at the same time**

- The inequality $P(B|A) > P(B|\neg A)$ seems to indicate that event $A$ increases the probability of event $B$ and there is an asymmetric relation between them
- **The relation is symmetric actually**

# Correlation and causality

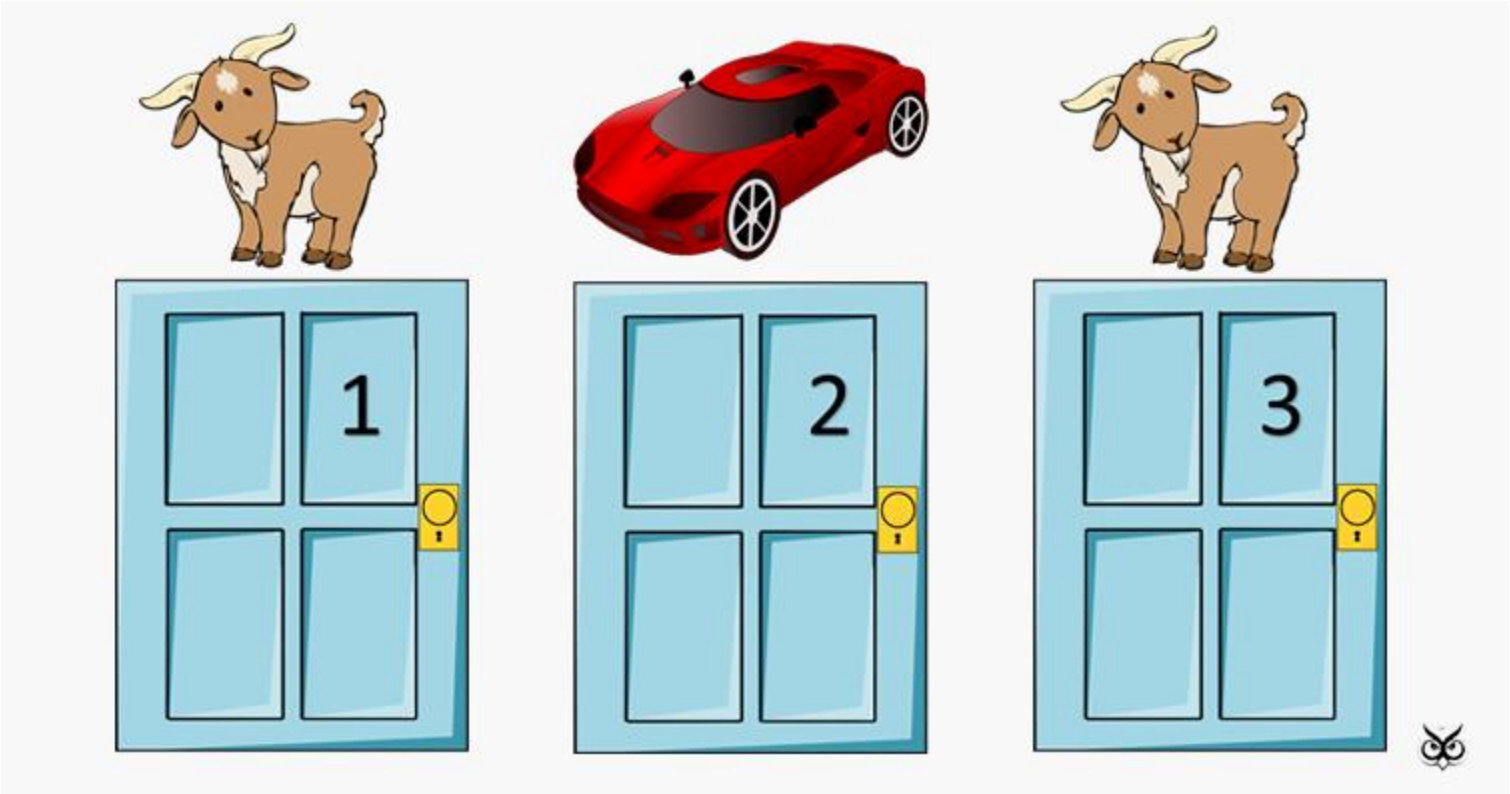





- $P(B|A) > P(B|\neg A)$ and $P(A|B) > P(A|\neg B)$ **implies the same, symmetric relation:**
  - Events $A$ and $B$ are correlated but no casual relation can be read out from these inequalities
  - Either there is a causal relation between $A$ and $B$ or there is a common cause
  - Think about: smoking – yellow finger tips – lung cancer, water level in Venice - price of bread in London

# Monty Hall problem

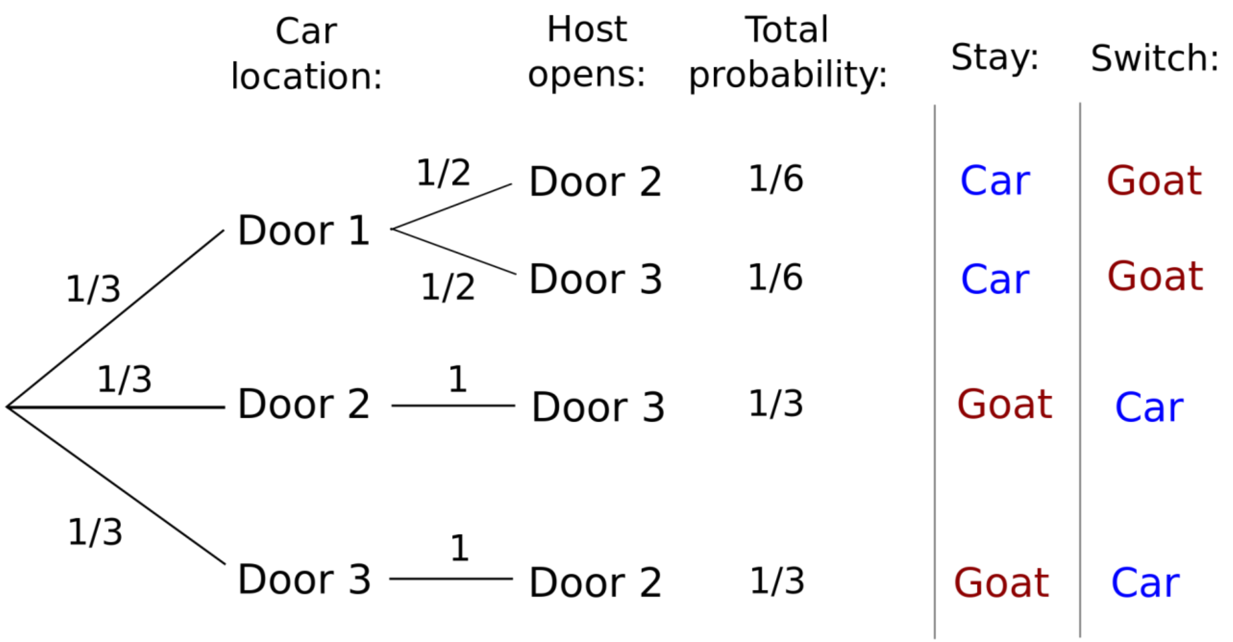




# Monty Hall problem



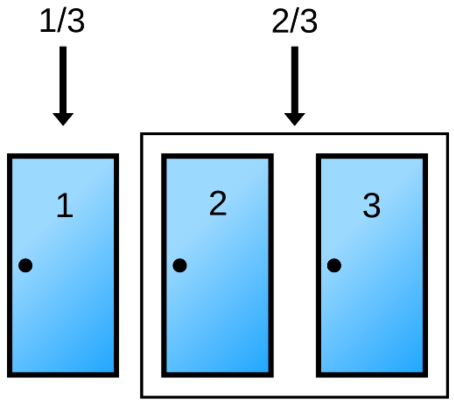
| you pick door \ the prize is behind door | 1 | 2 | 3 |
|---|---|---|---|
| 1 | Hall opens door 2 or 3 | Hall opens door 3 | Hall opens door 2 |
| 2 | Hall opens door 3 | Hall opens door 1 or 3 | Hall opens door 1 |
| 3 | Hall opens door 2 | Hall opens door 1 | Hall opens door 1 or 2 |

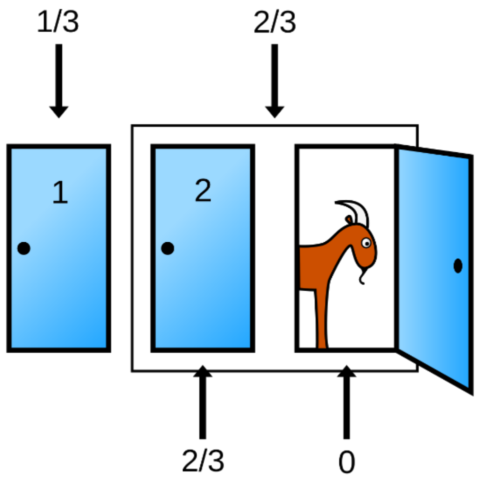


# Monty Hall problem



- So we are better off changing our mind: $\frac{1}{3} \rightarrow \frac{2}{3}$
- But why not 50-50%?
  - The situation when the host opens a door in advance and you choose from the two remaining doors is the same or not?
  - Not the same, because the action of the host depends on our choice
  - The host tells us information by opening a door



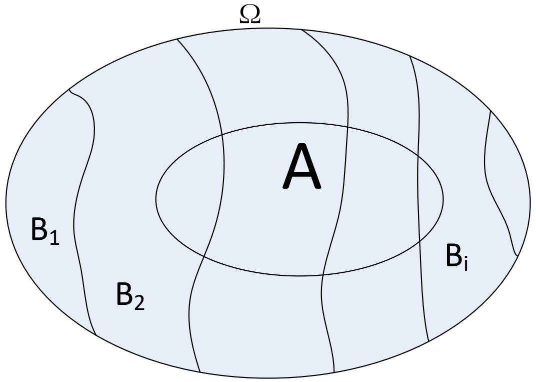



# Bayes-theorem

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- Law of total probabilities

$$P(A) = \sum_{i=1}^{N} P(A \cap B_i) = \sum_{i=1}^{N} P(A|B_i)P(B_i)$$




- Bayes-theorem

$$\boldsymbol{P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{N} P(A|B_i)P(B_i)}}$$

**Usual terminology**

Posterior: $P(B_k|A)$

Prior: $P(B_k)$

Likelihood: $P(A|B_k)$

Evidence, marginal likelihood: $P(A)$

# Bayesian inference



Application of the Bayes-theorem for hypothesis testing

- We have a prior probability, that hypothesis $H$ is true: $P(H)$
- We observe an event $E$, which is the evidence or observation and associate the probability: $P(E)$
- The likelihood that $E$ happens given $H$ is true is: $P(E|H)$
- The posterior probability that $H$ is true is given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}$$

# Hypothesis test – loaded coin

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- Someone is tossing a coin in the next room and tells us the results
- We have two hypotheses
  - The coin is loaded and produces $<\text{heads}>$ with 70% $(L)$
  - The coin is fair and does $50\% - 50\%$ $(\neg L)$
- We give probability $\mathrm{P}_0(L)$ that the coin is loaded (at the beginning)
- Based on what we hear, how shall we change our belief?
- The probabilities of the outcomes conditioned on the hypotheses are:

$$P(<\text{heads}>|L) = 0.7 \qquad P(<\text{tails}>|L) = 0.3$$
$$P(<\text{heads}>|\neg L) = 0.5 \qquad P(<\text{tails}>|\neg L) = 0.5$$

# Hypothesis test – loaded coin

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- Say the first toss gives $<\text{heads}>$ which results in:

$$P_1(L) = P_0(L| < \text{heads} >)$$

$$P_1(L) = \frac{P_0(< \text{heads} > |L)P_0(L)}{P_0(< \text{heads} > |L)P_0(L) + P_0(< \text{heads} > |\neg L)P_0(\neg L)}$$

$$P_1(L) = \frac{0.7P_0(L)}{0.7P_0(L) + 0.5(1 - P_0(L))}$$

- If we would have $<\text{tails}>$ instead:

$$P_1(L) = \frac{P_0(< \text{tails} > |L)P_0(L)}{P_0(< \text{tails} > |L)P_0(L) + P_0(< \text{tails} > |\neg L)P_0(\neg L)}$$

$$P_1(L) = \frac{0.3P_0(L)}{0.3P_0(L) + 0.5(1 - P_0(L))}$$

# Hypothesis test – loaded dice



With a concrete prior belief: $P_0(L) = 0.2$

- 1\. outcome: $< \text{heads} >$:

$$P_1(L) = \frac{0.7 \times 0.2}{0.7 \times 0.2 + 0.5 \times (1 - 0.2)} = 0.26$$

- 1\. outcome: $< \text{tails} >$:

$$P_1(L) = \frac{0.3 \times 0.2}{0.3 \times 0.2 + 0.5 \times (1 - 0.2)} = 0.13$$

# Hypothesis test – loaded dice



If we get two $<\text{heads}>$ in a row:

$$P_2(L) = P_1(L| < \text{heads} >)$$

$$P_2(L) = \frac{0.7 \times 0.26}{0.7 \times 0.26 + 0.5 \times (1 - 0.26)} = 0.33$$

- The second evidence also increases our belief but with a smaller amount
- This is a recursive process where we use the last result as prior
- We can have more than one concurrent hypotheses about a parameter (or a variable)
- In fact we can have continuously many hypotheses (from a parameter space or a state space)

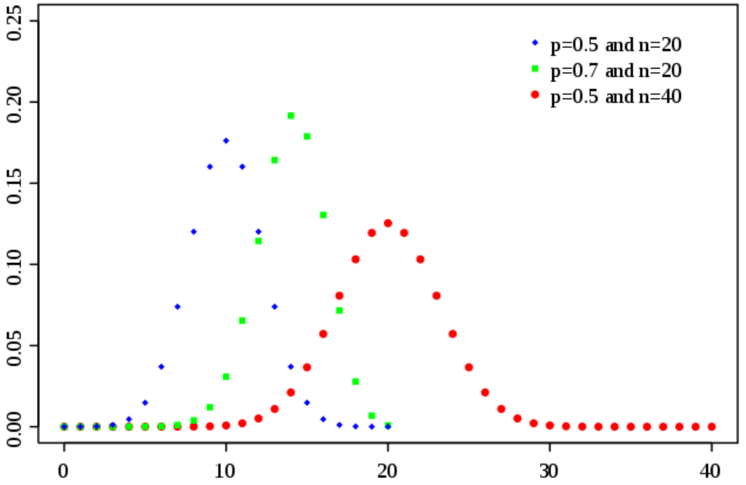
# Binomial distribution



- The probability to get $k$ success from $n$ trials is

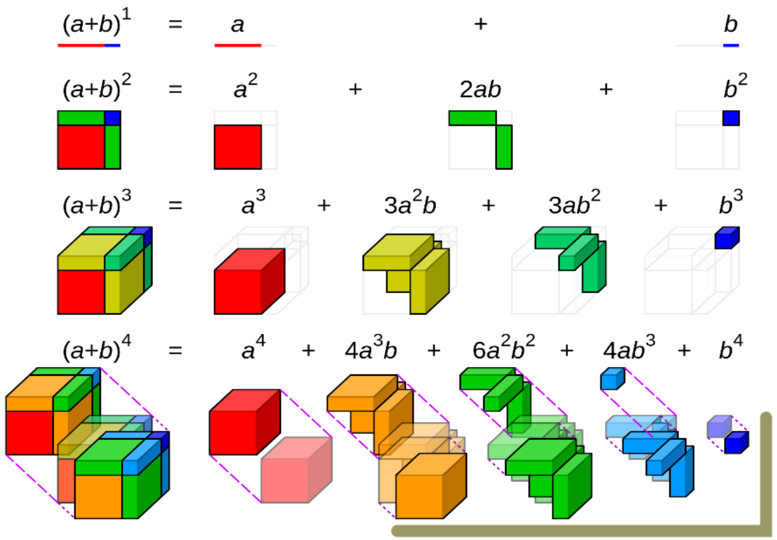
$$B(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

  - $p$ is the probability of one trial to succeed
  - $k$ is the free variable




$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$ is the binomial coefficient

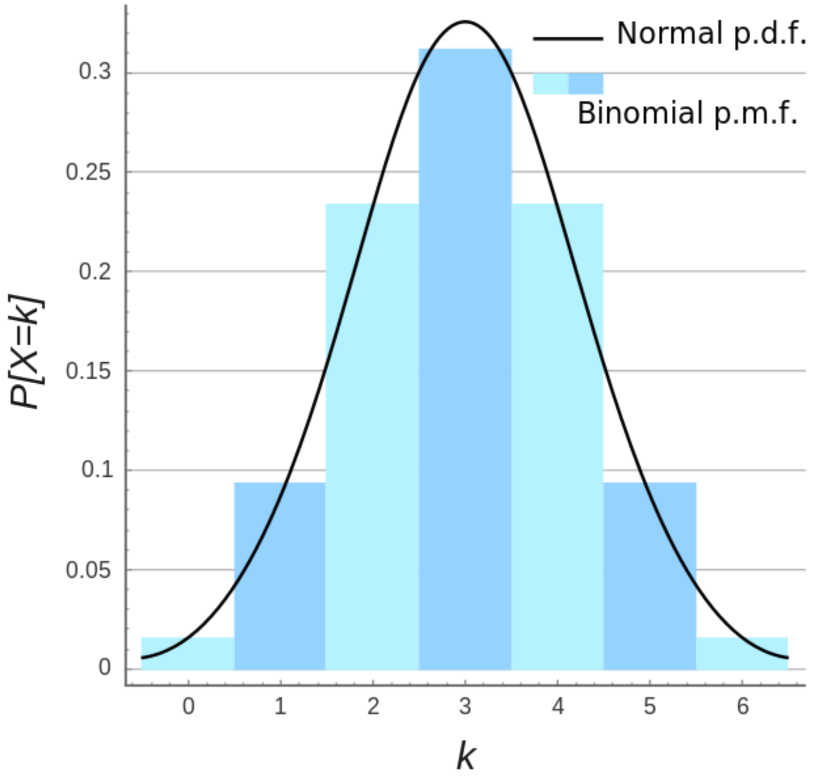
  - Pronounce: $n$ choose $k$
  - You can choose $k$ out of $n$ that many ways



# Binomial distribution



- Coin flip
  - 6 trials
  - Getting 3 heads and 3 tails is the most probable outcome
  - Increasing the number of trials will produce Gaussian-like histogram

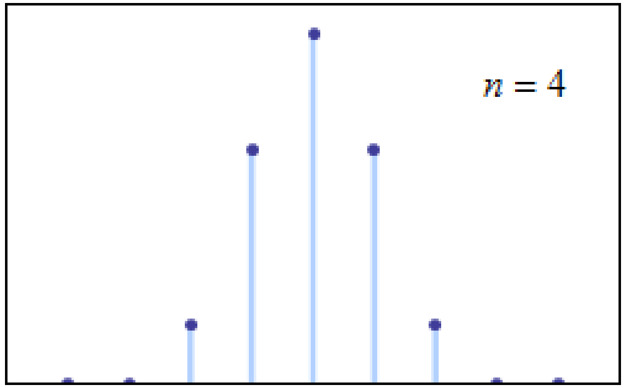


# Central limit theorem



```
%% Central limit theorem
% Dice roll
n = 1e4;

R = sum(round(6*rand(n)));
histogram(R)
```

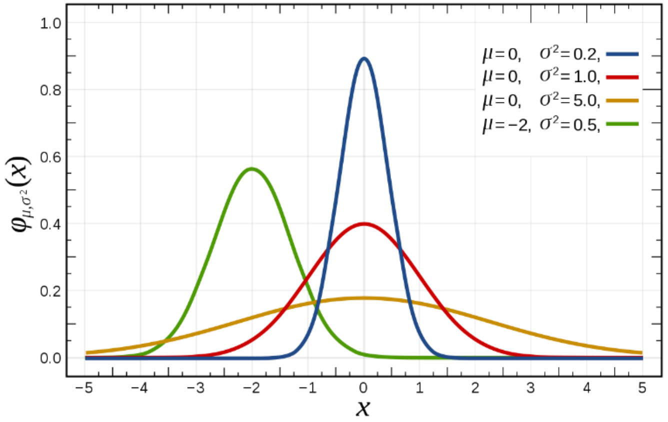



Tossing a coin n times and getting k heads

- https://phet.colorado.edu/sims/html/plinko-probability/latest/plinko-probability_hu.html

# Normal distribution

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems



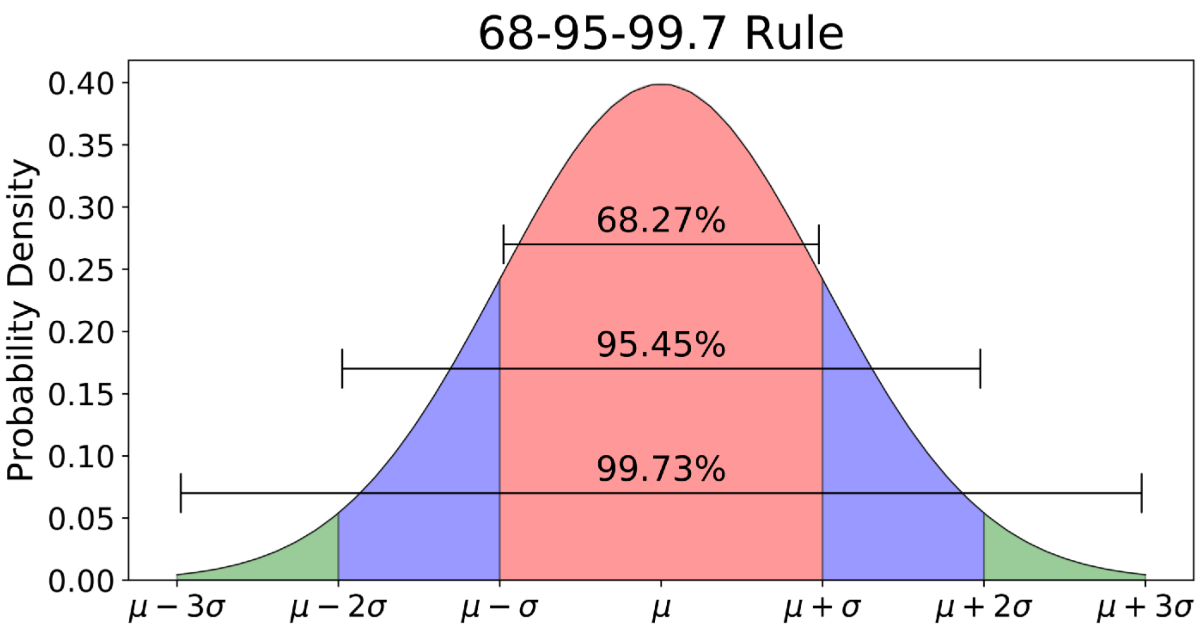

- Is the limit of the
  - Binomial distribution: $B(k;n,p) \to N(k;np,np(1-p))$
  - Poisson distribution: $P(k;\lambda) \to N(k;\lambda,\lambda)$
  - Chi-squared distribution: $\chi^2(k) \to N(k,2k)$
- Generally, the sum of independent, identically distributed random variables tends toward a normal distribution
- For a given mean and variance this is the maximum entropy distribution
  - It is the least informative distribution
  - It minimizes the information that we assume to be there
  - Physical systems generally move towards equilibrium, that is maximum entropy state
- It has nice mathematical properties

# Normal distribution

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

- Írja be az e



# Create Gaussian noise



- Usually we have a random number generator
  - We can generate a random number in the interval 0…1
  - The standard deviation is $\frac{1}{\sqrt{12}}$
  - The mean is 0.5

## Algorithm

1. Add 12 random numbers $(\mu = 6, \sigma = 1)$
2. Subtract 6 $(\mu = 0, \sigma = 1)$
3. Multiply by the desired STD
4. Add the desired mean

```
x = sum(rand(12,1e4));

x = x - 6;
x = x * 3;
x = x + 8;

histogram(x,'normalization
','pdf')
hold on
t = (-
3*sigma:0.1:3*sigma)+mu;
plot(t,normpdf(t,8,3))
```

# Gaussian vs White noise



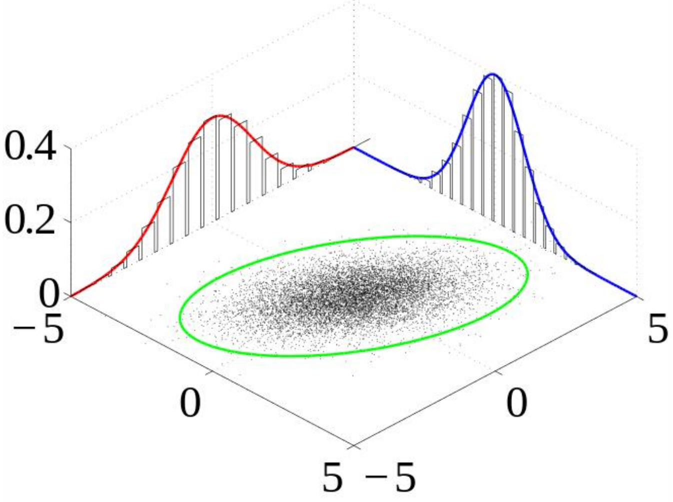
- Gaussian noise and white noise are not synonyms
  - Gaussian refers the distribution of the amplitude
  - White means that the values are not correlated in time. The intensity is the same at all frequencies and the PDF can be any
- A random signal can be white and Gaussian
  - This is a desired property
  - Tractable analytic models
  - Good approximation of real-world situations
- Additive White Gaussian Noise (AWGN)

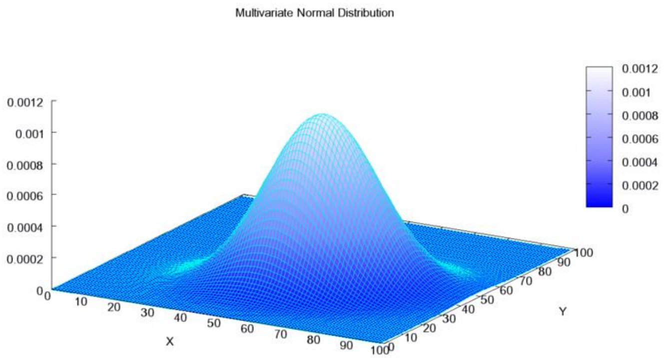
# Multivariate normal distribution



- Joint and multivariate distributions are synonyms

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right)$$



# Modelling uncertainties



- Additive noise acting on the motion and sensor model

$$\mathrm{x}_{k+1|k} = f_k(\mathrm{x}_k) + w_k$$
$$\mathrm{z}_k = h_k(\mathrm{x}_k) + v_k$$

random deterministic random

- How do we create probabilities from these random variables?
- Since x and z are usually continuous variables, the probabilities of taking specific values are zero.
- However, x and z residing in some region $S$ and $T$ have nonzero probabilities

$$P(\mathrm{x}_{k+1|k} \in S|\mathrm{x}_k) \qquad P(\mathrm{z}_k \in T|\mathrm{x}_k)$$

# Modelling uncertainties

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- The probability mass is given by integrating the probability density over a region

$$P\left(\mathrm{x}_{k+1|k} \in S \middle| \mathrm{x}_k\right) = \int_S p(\mathrm{x}|\mathrm{x}_k)\mathrm{dx} \qquad P(\mathrm{z}_k \in T|\mathrm{x}_k) = \int_T p(\mathrm{z}|\mathrm{x}_k)\mathrm{dz}$$

  - $p(\mathrm{x}|\mathrm{x}_k)$ is the probability density function associated to the uncertain motion model
  - $p(\mathrm{z}|\mathrm{x}_k)$ is the probability density function associated to the uncertain sensor model
- If the additive noise is zero mean Gaussian

$$\boxed{p(\mathrm{x}|\mathrm{x}_k) = \mathcal{N}(\mathrm{x}; f_k(\mathrm{x}_k), \sigma_w^2)}$$

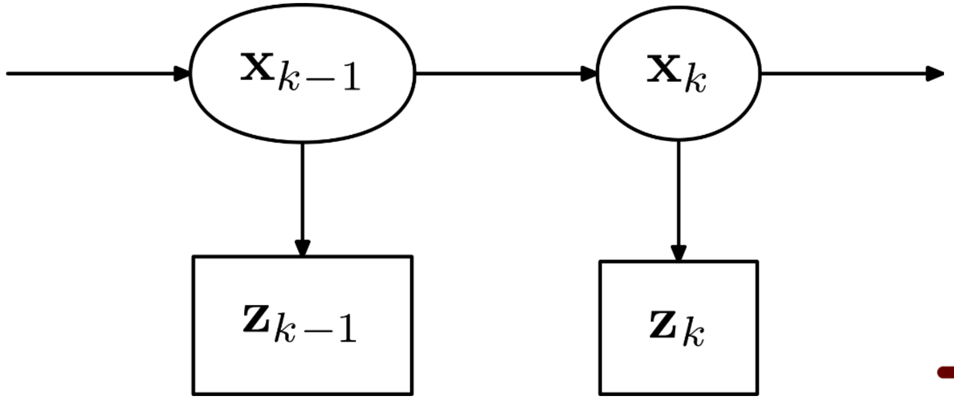
- Similarly for the sensor model

$$\boxed{p(\mathrm{z}|\mathrm{x}_k) = \mathcal{N}(\mathrm{z}; h_k(\mathrm{x}_k), \sigma_v^2)}$$

# Hidden Markov model (HMM)



- In the context of state estimation (robotics) the value to be estimated is the state (or state vector in general) of an object or an ensemble of objects
- The state in unknown to us (hidden) and possibly evolves in time: the system has dynamics
- We can observe the system and obtain a limited amount of information, for example
  - Partial observation of the state
  - Noisy measurements



# Markov assumptions

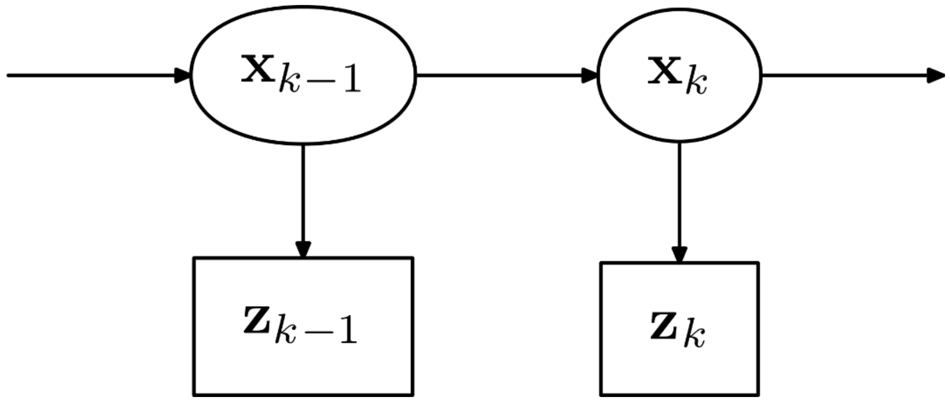


- The current state depends only on the previous state

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_0) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$$

- The measurement depends only on the current state

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_0) = p(\mathbf{z}_k|\mathbf{x}_k)$$



# Recursive Bayesian estimation (in discrete time)



- Estimate the state vector at timestep $k$ using measurements up to $k$:

$$p(\mathrm{x}_k|\mathrm{z}_{1:k}) = \frac{p(\mathrm{z}_k|\mathrm{x}_k)p(\mathrm{x}_k|\mathrm{z}_{1:k-1})}{p(\mathrm{z}_k|\mathrm{z}_{1:k-1})}$$

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{N} P(A|B_i)P(B_i)}$$

This was the Bayes-theorem

- The denominator is constant and can be expressed as

$$p(\mathrm{z}_k|\mathrm{z}_{k-1}) = \int p(\mathrm{z}_k|\mathrm{x}_k)\, p(\mathrm{x}_k|\mathrm{z}_{k-1})\mathrm{dx}_k$$

- The prior, with the help of a model of the system is obtained from the pervious posterior through the time-prediction integral (Chapman-Kolmogorov integral):

$$p(\mathrm{x}_k|\mathrm{z}_{1:k-1}) = \int p(\mathrm{x}_k|\mathrm{x}_{k-1})\, p(\mathrm{x}_{k-1}|\mathrm{z}_{1:k-1})\mathrm{dx}_{k-1}$$

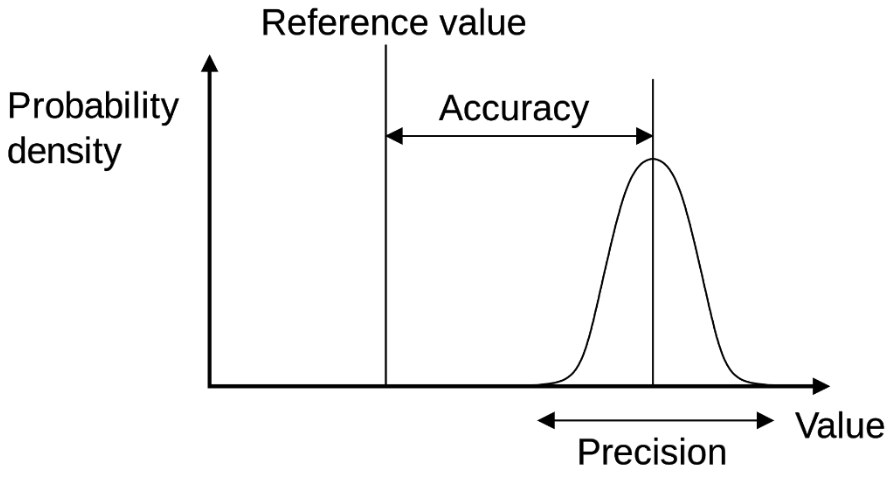
motion model previous posterior

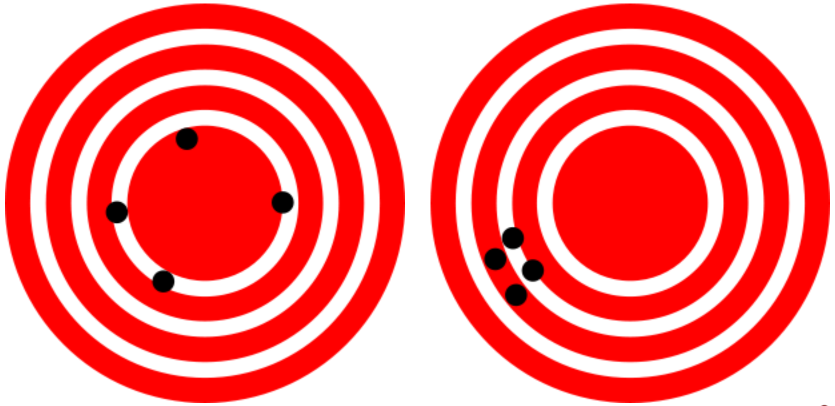
# Accuracy, precision



The quality of a sensor can be described by its precision and accuracy

- Accuracy
  - Measures the systematic error (bias)
  - Related to the mean of the measurement
- Precision
  - Measure the random error (variability)
  - Related to the variance (standard deviation) of the measurement



# Terminology in estimation



- Statistic: a function of the data
- Estimator: a function of the data that intends to describe some property of the underlying distribution
  - A statistic is not good or bad( or biased or unbiased). It is just a function
  - An estimator can be good (unbiased, minimum variance etc.). E.g.: the sample mean is an unbiased estimator of the expected value

- Filtering: estimate $x_t$ based on measurements $z_{1:t}$
- Prediction: estimate $x_{t+\tau}$ based on measurements $z_{1:t}$
- Smoothing: estimate $x_{t-\tau}$ based on measurements $z_{1:t}$

# Metric – Euclidean



Calculate "real distance" from coordinate differences

- Distance of two points in 3D: $d(P_1, P_2)$

$$d(P_1, P_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

**Euclidean metric** (in Cartesian coordinates)

Are there other ways to get a distance?

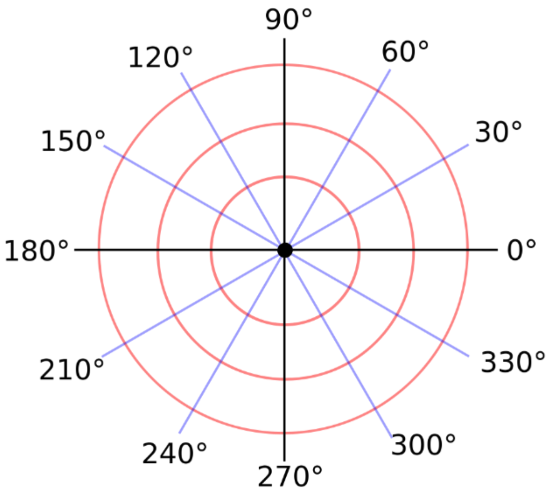
# Metric – Polar



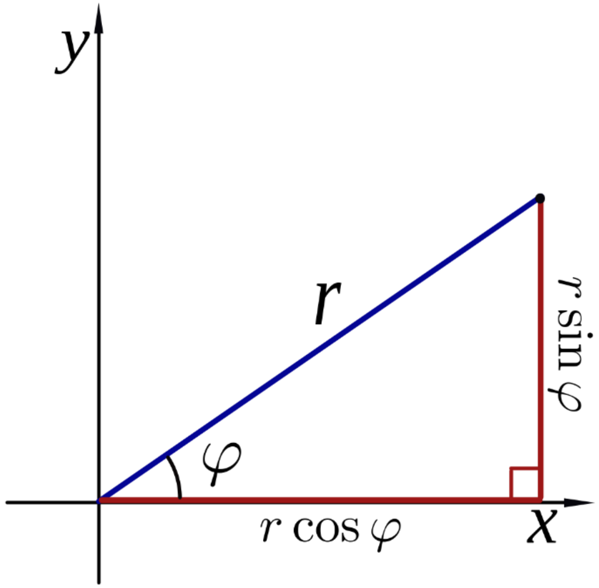
Polar coordinate system

- $x = r\cos\varphi$
- $y = r\sin\varphi$

We can also have cylindrial, toroidal, etc coordinate systems




$$d = \sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos(\varphi_1 - \varphi_2)}$$




$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

# Metric



- You can make up and use any metric if it is meaningful in a way
- Metric is not just to calculate a physical distance, it can be any "distance" that is useful
- A typical application is to measure the error between some true and measured or estimated quantities (e.g. a signal or a state vector)

- Distance between states: error metric

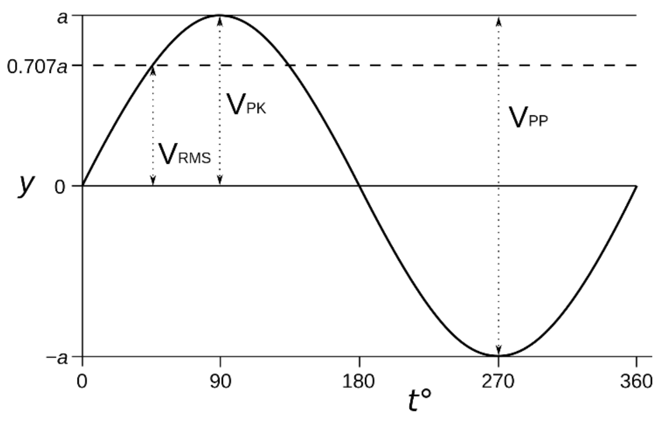
$$\mathbf{x} = [x, v_x, y, v_y] \qquad \hat{\mathbf{x}} = [\hat{x}, \hat{v}_x, \hat{y}, \hat{v}_y]$$

$$d(\mathbf{x}, \hat{\mathbf{x}}) = ?$$

# RMS – Root Mean Square



- The voltage in the wall is 230V, which is the effective value of the alternating sinusoidal signal.
- This is the RMS value of a sinusoidal signal that has 325V peak voltage.
- Sometimes we want to describe a signal with a single number to be able to easily compare them.
- Common choices: maximum (minimum) value, average value, RMS value.



# RMS – Root Mean Square

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- Computing the RMS of a signal in the time domain results the same as computing it in the frequency domain.
- The RMS value is invariant to the Fourier transform
  - A method to verify the result of a FFT
- It is a property of a physically existing signal, not just a property of the chosen representation
- It indicates the energy carried by the signal
  - In the context of electricity $V\_RMS^2$/RESISTANCE is the power

# RMS – RMSE



- $x_{RMS} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \cdots x_n^2)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$
- $x_{RMSE} = \sqrt{\frac{1}{n}(e_1^2 + e_2^2 + \cdots e_n^2)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{x}_1 - x_1)_i^2}$
- Sometimes RMS and STD are synonyms
- Mean squared deviation (error) is the square of RMSE

# RMS



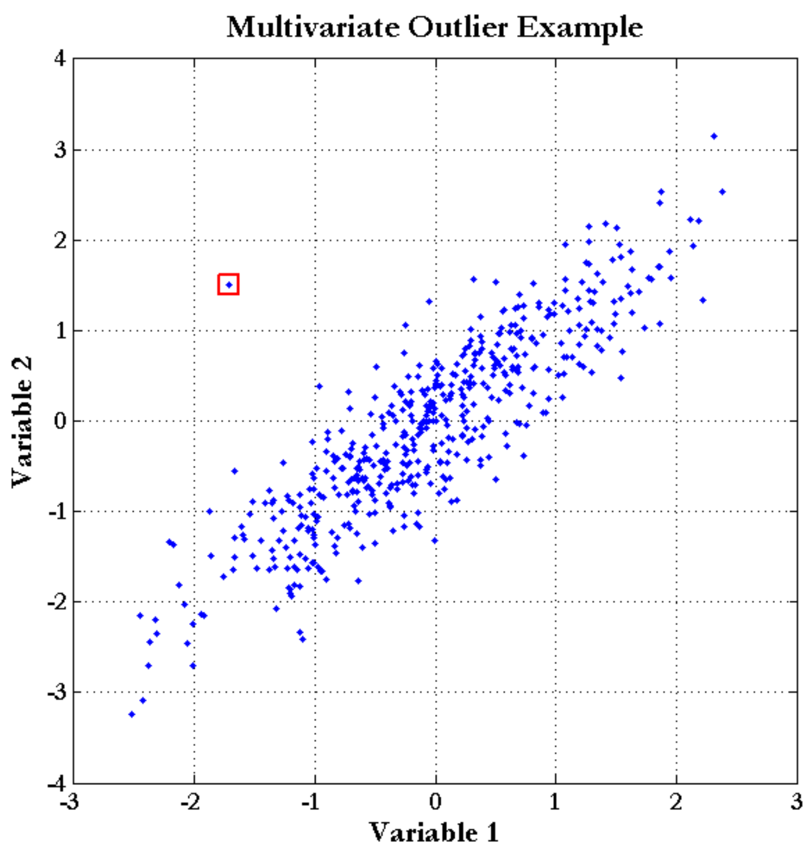
- $x$ is normally distributed random vector: $x \sim \mathcal{N}(\mu, \Sigma)$
- If $x$ describes a signal what is the expectation of the carried power?

$$\mathrm{E}[\|x\|_2^2] = \|\mu\|_2^2 + \mathrm{tr}(\Sigma)$$

# Mahalanobis distance



- What is the distance of a point to a distribution
  - Is this a meaningful question?
- Euclidean distance is always an option between points, but what point represents the distribution?
  - The **mean!**
  - Should we consider the **variance-covariance**?

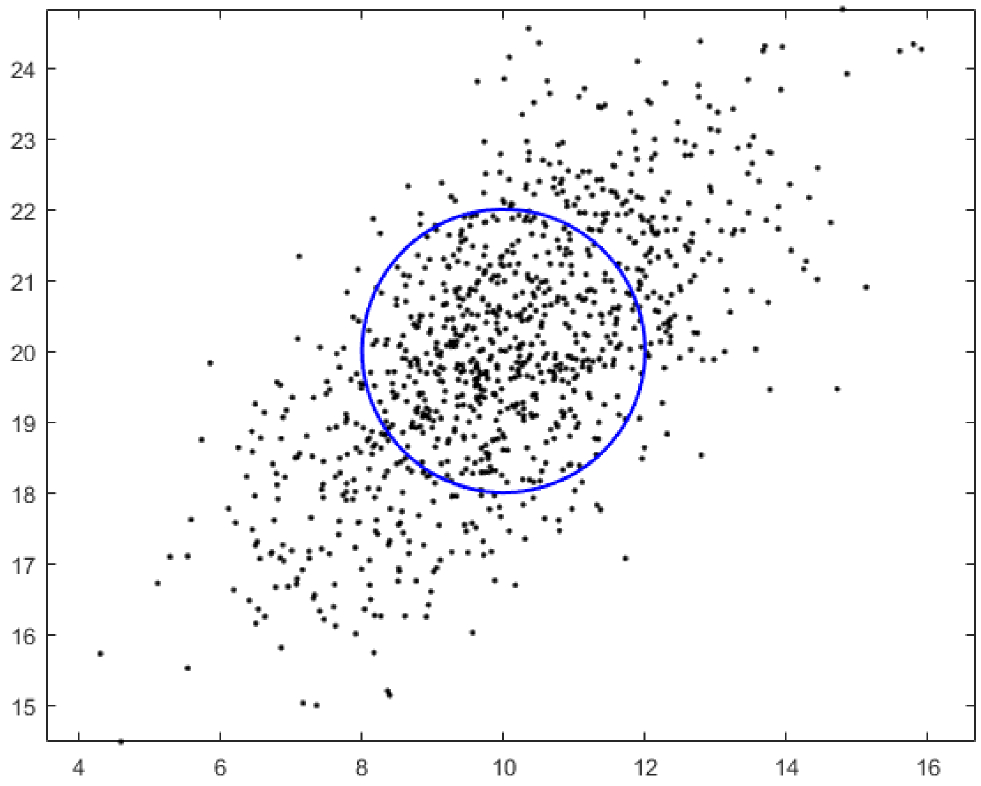


# Mahalanobis distance



```
% Generate a two dimensional Gaussian
n = 1e3;
Mu = [10;20];
Sigma = [3, 2; 2, 3];
x = mvnrnd(Mu, Sigma, n);

plot(x(:,1),x(:,2),'k.')
hold on; axis equal
% Plot a circle around the centre
(mean) with radius 2
r = 2;
cx = r * cos(0:0.01:2*pi) + Mu(1);
cy = r * sin(0:0.01:2*pi) + Mu(2);
plot(cx,cy,'b-','LineWidth',1.5)
```



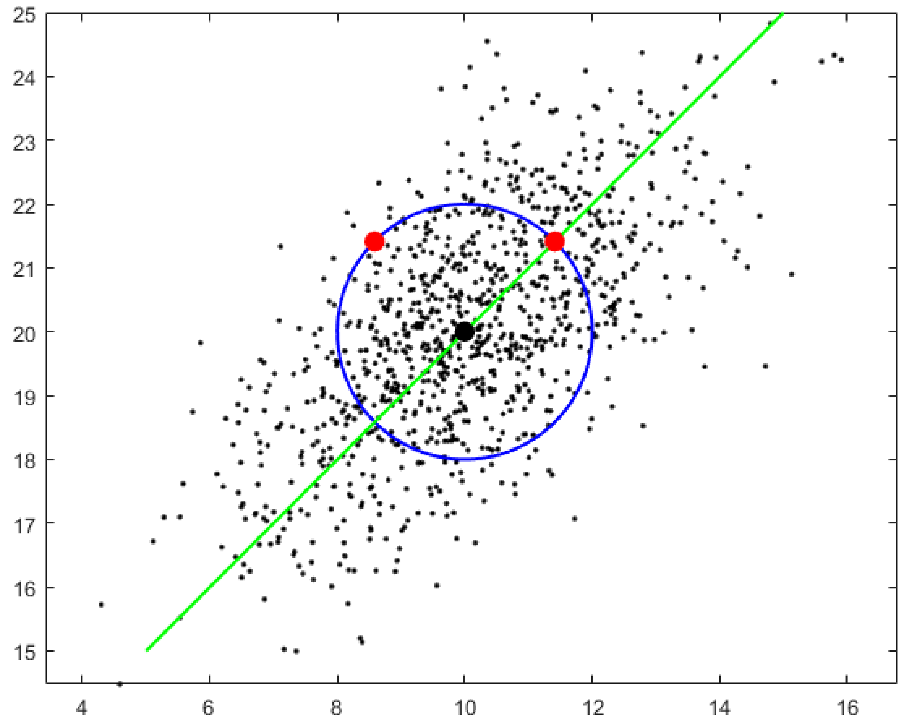
# Mahalanobis distance

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation

```
% 45 deg line
plot((-5:5)+Mu(1),(-5:5)+ Mu(2),
'g','LineWidth',1.5)

% Mean
plot(Mu(1), Mu(2),'k.','MarkerSize',32)

% Points at 45 and 135 deg
plot(r*cos(pi/4)+Mu(1),
r*sin(pi/4)+Mu(2),'r.','MarkerSize',32)
plot(r*cos(pi*3/4)+Mu(1),
r*sin(pi*3/4)+Mu(2),'r.','MarkerSize',32)
```

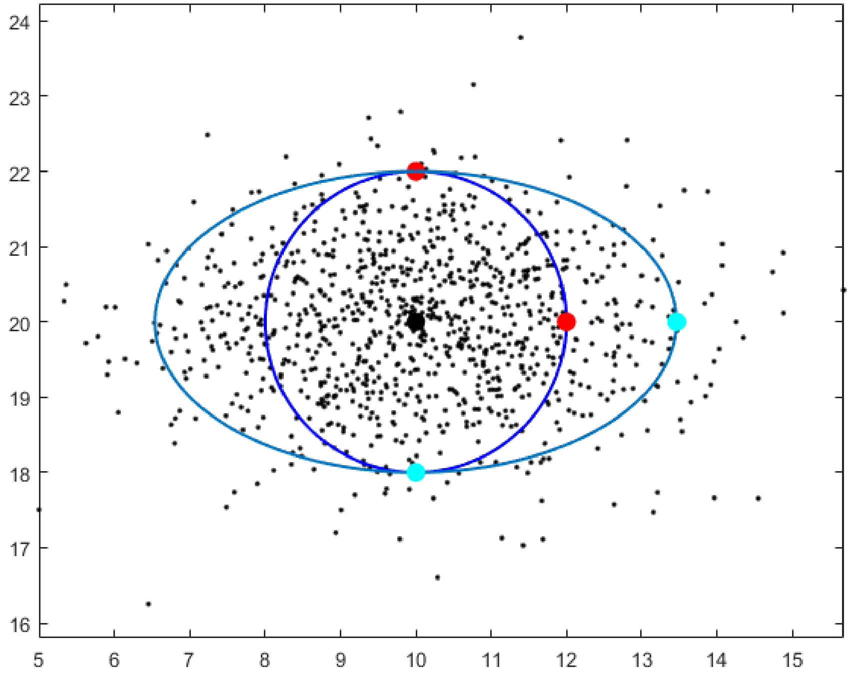



- These points are equally distant to the origin (regarding Euclidean metric)
- But one of the seems to outlie more than the other
- We should include the variances when calculating the distance!

# Mahalanobis distance



- Euclidean distance: $d = \sqrt{(x-\mu_x)^2 + \left(y-\mu_y\right)^2}$
  - Vectorized form: $d = \sqrt{(\mathbf{x}-\mu)^\intercal(\mathbf{x}-\mu)}$ with $\mathbf{x} = [x, y]^\intercal$ and $\mu = [\mu_x, \mu_y]^\intercal$
- Weighted Euclidean distance: $d = \sqrt{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2}$ Equation of an ellipse (scaled by $d$)
  - Vectorized form: $d = \sqrt{(\mathbf{x}-\mu)^\intercal \begin{bmatrix} {\sigma_x}^{-1} & 0 \\ 0 & {\sigma_y}^{-1} \end{bmatrix} (\mathbf{x}-\mu)}$
- $\Sigma^{-1} = \begin{bmatrix} {\sigma_x}^{-1} & 0 \\ 0 & {\sigma_y}^{-1} \end{bmatrix}$ Inverse of the covariance matrix



# Mahalanobis distance

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

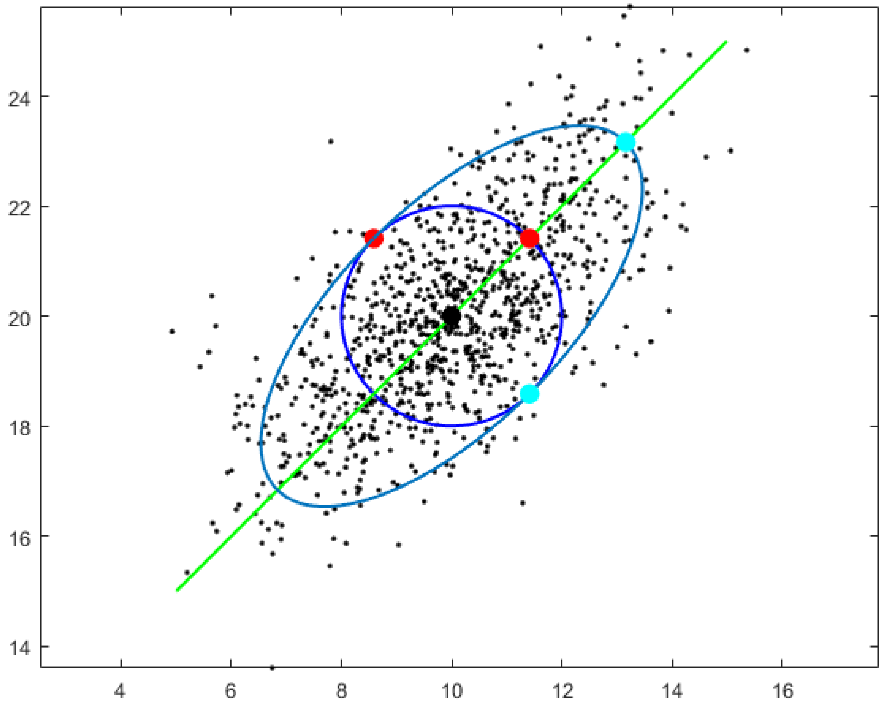
- The ellipse is the unit circle when the metric is the Mahalanobis distance
- General case (when rotated):

$$d = \sqrt{(\mathbf{x}-\mu)^{\intercal}\Sigma^{-1}(\mathbf{x}-\mu)}$$

  - Weighted scalar product:

$$(\mathbf{x}-\mu)^{\intercal}\Sigma^{-1}(\mathbf{x}-\mu)$$

  - The weight is inversely proportional to the variance: the greater the uncertainty the less we take the difference into account
  - The Euclidean metric uses no weighting (identity matrix)
  - You can make up any metric of this kind by inserting a positive definite matrix as weight. (Σ is PSD, it can be singular!)



# Classification with Mahalanobis distance



- Say we have 3 categories described by the distributions: $\mathcal{N}(\mu_i, \Sigma_i)$. The point $x$ have the following distances from the distributions:

$$D_1^2 = (x-\mu_1)^\intercal S_1^{-1} (x-\mu_1)$$

$$D_2^2 = (x-\mu_2)^\intercal S_2^{-1} (x-\mu_2)$$

$$D_3^2 = (x-\mu_3)^\intercal S_3^{-1} (x-\mu_3)$$

- To create probabilities from the distances we should normalize them. The normalization factor is

$$Z = \mathrm{e}^{-D_1^2} + \mathrm{e}^{-D_2^2} + \mathrm{e}^{-D_3^2}$$

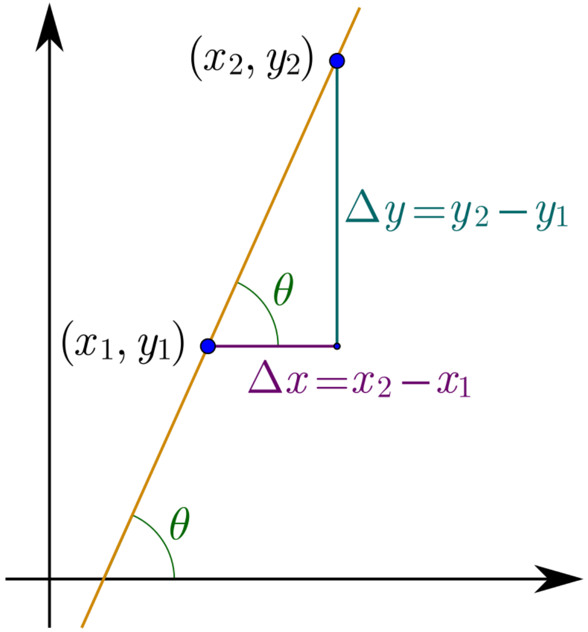
- and the probability of $x$ belonging to category $i$ is

$$\boxed{p_i = \frac{e^{-D_i^2}}{Z}}$$

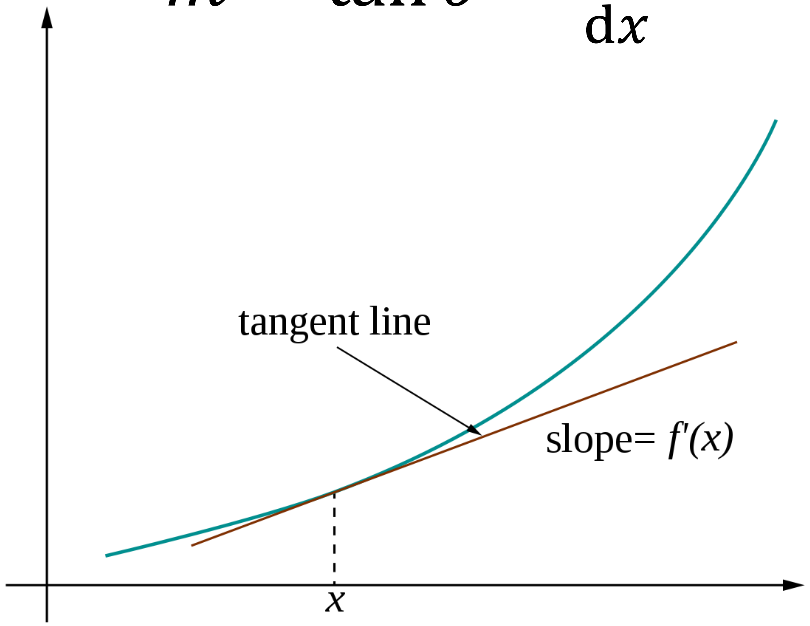
# Slope



- For a straight line
  - $m = \tan\theta = \dfrac{\Delta y}{\Delta x}$




- For a curved line
  - $m = \tan\theta = \dfrac{\mathrm{d}y}{\mathrm{d}x}$



# Linear regression



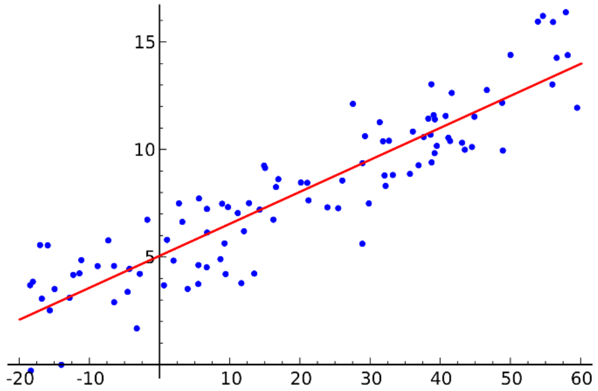
MATLAB: **`mldivide`**

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- Solve systems of linear equation: $Ax = B$
  - Can be an overdetermined system

    More data points than variables. In this case the solution is given by least-squares method
  - Usage: `x=mldivide(A,B)` or `x=A\B`

- Use to fit a line to data points
  - We have $x = [x_1, x_2, \ldots x_n]$ and $y = [y_1, y_2, \ldots y_n]$
  - We want to fit a line: $y = mx + b$
  - Now we have $x$ and $y$ and the unknown is $m$

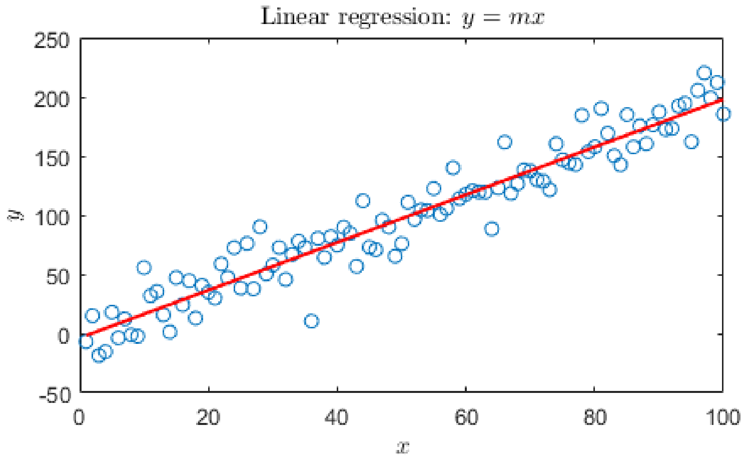


# Linear regression



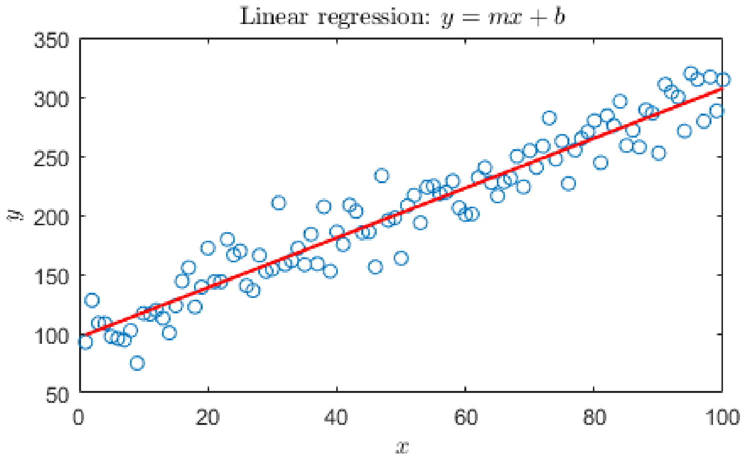
## Homogeneous

- $y = mx \;\rightarrow\; xm = y$
- Usage: `m=x\y`
- $\varphi = \tan^{-1} m$




## Inhomogeneous

- $y = mx + b \;\rightarrow\; xm + b = y$
- $X = [x, \mathbf{1}]$
- Usage: `mb=X\y`
  - `m=mb(1); b=mb(2)`



# Covariance



- Fit a line
- Determine the slope
- Compute covariance
  - `cov(x,y)`

- Play with the parameters:
  - Number of data points (1e2)
  - Range (200)
  - Noise magnitude (15)
  - Coefficient of $x$ (2)
- What are their effects?

```
% Noisy data points
n = 1e2;
x = linspace(1,200,n)';
y = 2*x +100 + 15*randn(size(x));

figure
hold on; box on
plot(x,y,'o')

% Fit a line: y = m*x+b
X = [x,ones(size(x))];
mb = X\y
m=mb(1); b = mb(2);
fi = atan(m)
plot(x,m*x + b,'r')
```

# Covariance



- Fit a line
- Determine the slope
- Compute covariance
  - `cov(x,y)`

- Play with the parameters:
  - Number of data points: **no effect**
  - Range: **increases covariance**
  - Noise magnitude: **no effect**
  - Coefficient of $x$: **increases covariance**
- **What does covariance measure?**

```
% Noisy data points
n = 1e2;
x = linspace(1,100,n)';
y = 2*x + 100 + 15*randn(size(x));

figure
hold on; box on
plot(x,y,'o')

% Fit a line: y = m*x+b
X = [x,ones(size(x))];
mb = X\y
m=mb(1); b = mb(2);
fi = atan(m)
plot(x,m*x + b,'r')
```

# Covariance



- The definition is:

$$\mathrm{cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

- For concrete data points the discrete formula is:

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

  - The range of $x$ and $y$ is in $x_i - \bar{x}$ and $y_i - \bar{y}$
  - The coefficient of $x$ effects the range of $y$

# Correlation



- To measure the pure connection between $x$ and $y$ we need to normalize the covariance with the range
  - This way we create a measure that is independent of the chosen units. Scale independent
- Definition:

$$\mathrm{r} = \mathrm{corr}(x, y) = \frac{\mathrm{cov}(x, y)}{\sqrt{\mathrm{var}(x)\mathrm{var}(y)}} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}$$

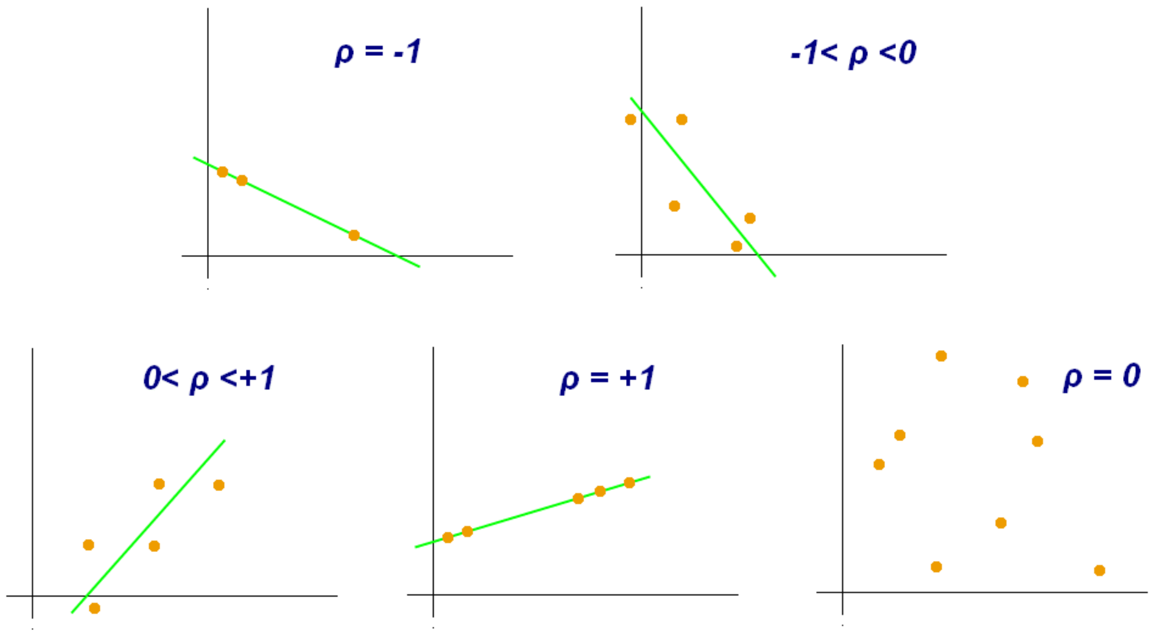
# Correlation



- The greater the correlation the more $x$ can explain $y$
  - 1: maximal correlation
  - 0: no correlation
  - -1: maximal anticorrelation

$r^2$ measures what proportion in the variance of $y$ can be explained by $x$:

- $\mathrm{var}(e) = (1 - r^2)\mathrm{var}(y)$



# Slope vs correlation

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control for Transportation and Vehicle Systems

- The slope and the correlation are the same, if $\sigma_x = \sigma_y$

$$\tan\varphi = m = \mathrm{corr}(x,y)\sqrt{\frac{\mathrm{var}(y)}{\mathrm{var}(x)}} = r\,\frac{\sigma_y}{\sigma_x}$$

- The closer the correlation to one the more perfect the linear relationship
  - The slope does not contain this information
- The slope tells how much $y$ changes with $x$
  - The correlation does not contain this information

**But the signs are the same**

- If we swap $x$ and $y$ the correlation remains the same but not the slope!