Scientific computing and data visualization in Python

Adrian Kania[1]

[1]Department of Computational Biophysics and Bioinformatics

2023

Schedule

5 - semi-lectures, 10 - practical classes, one meeting = 2 hours (8:00-10:00)

- **Introduction, Linear Algebra and Probability (L1)**
- Introduction to numpy and matplotlib. Matrix operations and probability (PC1)
- Bayes Theorem and Bayesian Networks (PC2)
- **Correlation, linear and logistic regression (L2)**
- Correlation, linear and logistic regression Part I (PC3)
- Correlation, linear and logistic regression Part II (PC4)
- **Clustering and classification algorithms (L3)**
- Support Vector Machine (PC5)
- Decision trees (PC6)
- K-means and other clustering algorithms (PC7)
- **Dimensionality reduction (L4)**
- PCA, ICA and MDS for dimensionality reduction (PC8)
- **Single-cell RNA sequencing (L5)**
- Single-cell RNA sequencing analysis based on machine learning techniques (PC9)
- Projects presentation and summary (PC10)
- Practical test

Rules and requirements

- Students should be familiar with Python fundamentals and completed at least basic Mathematics course.
- Attendance is obligatory.
- Points are collected during lectures (activity, 5 points), practical classes (5 points per meeting), project (15 points = 5+10) and practical exam (15 points).
- Projects topics should be declared by the end of PC8 at the least and approved by me.
  - related to biology,
  - may be based on a publication (I will show some examples during the course),
  - should utilize introduced during this course methods,
  - finally, should consist of code (jupyter?), short summary (1/2 pages pdf) and presentation delivery during PC10 (about 10 minutes).
- Practical test will last about 70 minutes.
- Total number of points is $5 + 5 * 8$ (8 PC will be scored) $+15 + 15 = 75$.
  - $68 - 75 \rightarrow 5$
  - $60 - 67 \rightarrow 4.5$
  - $52 - 59 \rightarrow 4$
  - $44 - 51 \rightarrow 3.5$
  - $38 - 43 \rightarrow 3$
- Conslutation: Tuesday, Thursday (8:40–9:40)

What is your Python programming level?

- Define a function **seq** which calculates the $n-th$ element of the following sequence $a_n = n^2 - 1$. For example, $seq(3)$ should return 8 (because $3^2 - 1 = 8$).
- Define a function **product** which takes two lists (the same in length) as arguments and returns a number wich is a sum of the products of successive elements. For example $product([1, 2, -1], [4, 1, 0])$ should return 6 (because $1 \cdot 4 + 2 \cdot 1 + (-1) \cdot 0 = 6$).

Main goals

During this course we want to learn how to:

- read and refine data,
- analyze data,
- visualize data,
- interpret data,
- **recognize patterns**,
- generalize results,
- make predictions,

Pattern recognition means recognize patterns in data

Data = simple features, images, sounds, videos and so on (**numbers**)

Searching for patterns - example 1

Consider the following sequence:

$$5, 8, 11, 14....$$

What is the next number?

Searching for patterns - example 1

Consider the following sequence:

$$5, 8, 11, 14....$$

What is the next number?

17. Great! Additionally, we may indicate the general rule $a_n = 2 + 3n$.

Searching for patterns - example 2

Assume that:

- $340 \longrightarrow 2$,
- $331 \longrightarrow 0$,
- $781 \longrightarrow 2$,
- $111 \longrightarrow 0$,
- $549 \longrightarrow 2$,
- $949 \longrightarrow 3$

Question:

$$881 \longrightarrow ??$$

Cat

How do we know this a cat?

https://vetmed.tamu.edu/news/pet-talk/preparing-cat-for-vet-visit/

Cat

We are searching in our memory for similar objects we know

https://vetmed.tamu.edu/news/pet-talk/preparing-cat-for-vet-visit/

???

What if we don't know all categories?

https://avianreport.com/size-harpy-eagle/

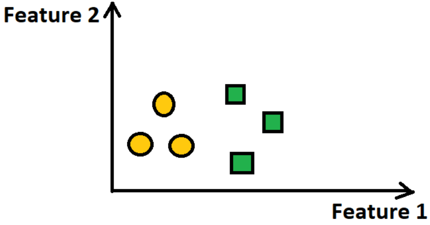
General rules:

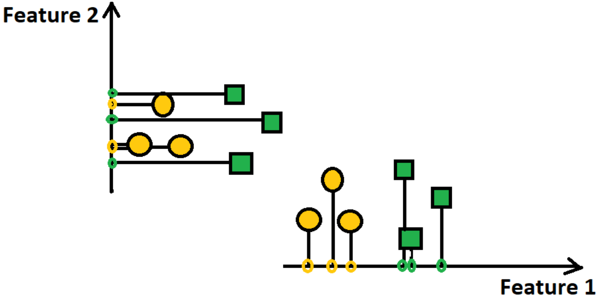
The more data, the better.

Additionally, examples should be representative to discover patterns. For example, if we consider different animals we should see many examples each of them.

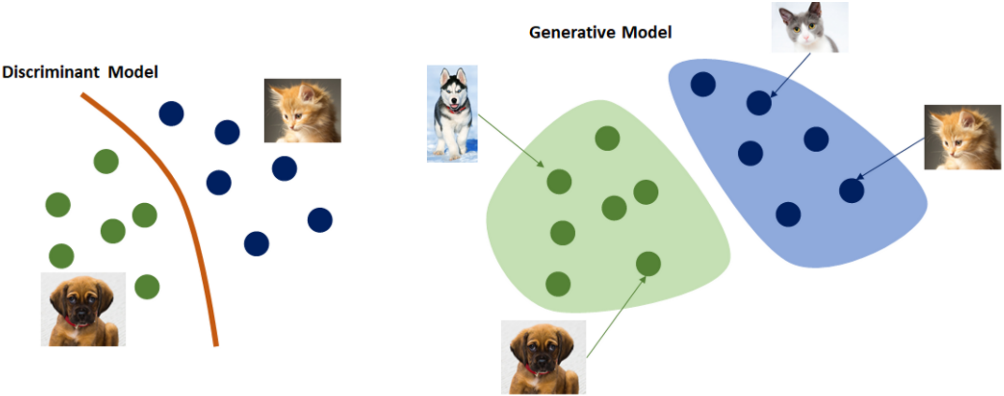
Not all features are important.

Considering a cat it is important it has big ears but the information about two eyes is less useful.





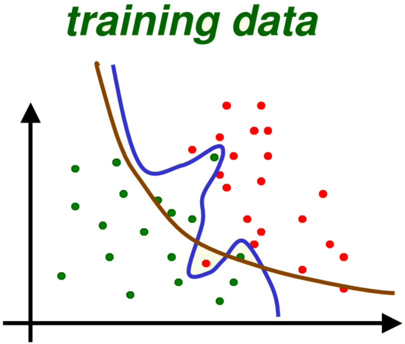
Discriminant and Generative models

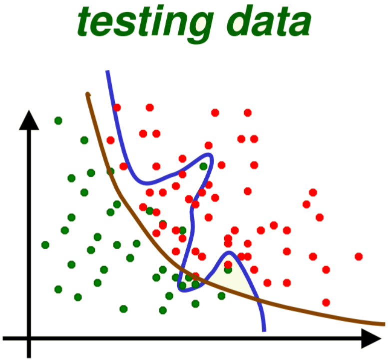


Steps

- Collect data and classify them by hand.
- Preprocess
- Extract useful features (how ???)
- Choose a model
- Train a model on collected examples
- Test the trained model on new data.
- Evaluate the model.

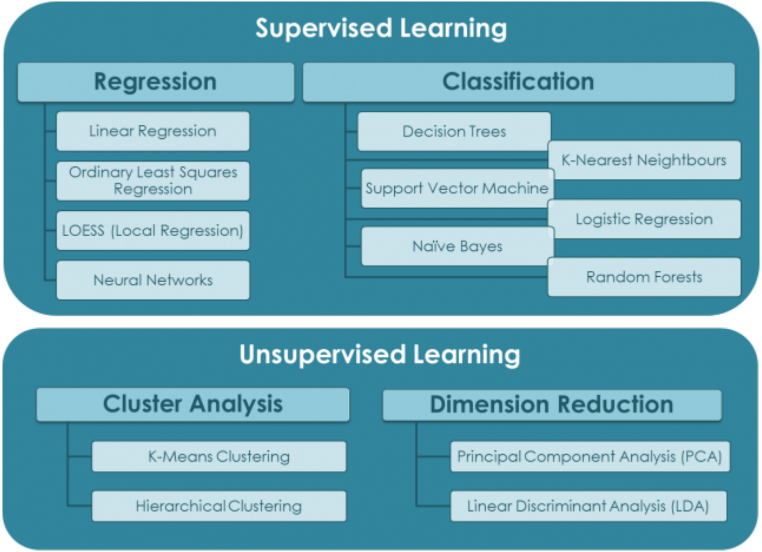
Generalization





Usually, easier models are less prone to overfitting and generalize better.

Machine Learning Algorithms



https://quantdare.com/machine-learning-a-brief-breakdown/

What is your Math level?

- Are vectors $x_1 = [2, -1, 3]$ and $x_2 = [1, 2, 0]$ orthogonal?
- Calculate the determinant of $A = \begin{bmatrix} -1 & 3 \\ 2 & 2 \end{bmatrix}$.
- Consider the following random variable and its distribution:

X	**-2**	**3**
p	**0.3**	**0.7**

Calculate EX and DX.

Summation notation

Instead of writing

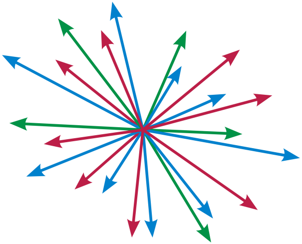
$$x_1 + x_2 + \ldots + x_n$$

we will write $\sum_{i=1}^{n} x_i$ or even shortly $\sum_i x_i$.

Examples and properities

- $x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_i x_i y_i$,
- $\sum_i x_i + \sum_i y_i = (x_1 + x_2 + \ldots + x_n) + (y_1 + y_2 + \ldots + y_n) = (x_1 + y_1) + (x_2 + y_2) + \ldots + (x_n + y_n) = \sum_i (x_i + y_i)$,
- $\sum_i x_i - \sum_i y_i = (x_1 + x_2 + \ldots + x_n) - (y_1 + y_2 + \ldots + y_n) = (x_1 - y_1) + (x_2 - y_2) + \ldots + (x_n - y_n) = \sum_i (x_i - y_i)$,
- $ax_1 + ax_2 + \ldots + ax_n = a(x_1 + x_2 + \ldots + ax_n) = a \sum x_i$,
- $\sum_i a = a + a + \ldots + a = na$,
- $x_1 - x_2 + x_3 - x_4 \ldots. = \sum_i (-1)^{i+1} x_i$.

Vectors



n-dimensional vector (may be interpreted as a list of features)

$$x = [x_1, x_2, ..., x_n]$$

and its transposition

$$x^T = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

Example: $x = [4, 5, 2, -1, 1]$

Vectors

Some definitions

- **Vector product (inner/dot product)**

$$< x, y >= x_1 \cdot y_1 + x_2 \cdot y_2 + \ldots + x_n \cdot y_n$$

- **Euclidean norm (length)**

$$|x| = \sqrt{< x, x >} = \sqrt{\sum_i x_i \cdot x_i} = \sqrt{\sum_i x_i^2}$$

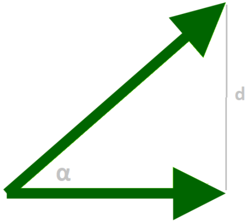
- **Angle between vectors**

$$\cos \alpha = \frac{<x,y>}{|x||y|}$$

  Vectors are **orthogonal** if $< x, y >= 0$

- **Euclidean distance**

$$d = |x - y| = \sqrt{\sum_i (x_i - y_i)^2}$$



Vectors

Example: Let's define $x = [1, -2, 3]$ and $y = [3, 6, 3]$. Then

- $< x, y >= 1 \cdot 3 + (-2) \cdot 6 + 3 \cdot 3 = 3 - 12 + 9 = 0$ (x and y are orthogonal)
- $|x| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{1 + 4 + 9} = \sqrt{14}$
- $|y| = \sqrt{3^2 + 6^2 + 3^2} = \sqrt{9 + 36 + 9} = \sqrt{54}$
- $d = \sqrt{(1-3)^2 + (-2-6)^2 + (3-3)^2} = \sqrt{4 + 64 + 0} = \sqrt{68}$

Vectors

More definitions

- Vector x is a **linear combination** of vectors $x_1, x_2, ..., x_n$ if there exist real numbers $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{R}$ such as:
$$x = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n$$

Example: $x = [1, 2, 3]$, $x_1 = [2, -3, 0]$, $x_2 = [3, -1, 3]$. Then, x is a linear combination of x_1 and x_2.

$$x = -1x_1 + 1x_2$$

$$[1, 2, 3] = -1 \cdot [2, -3, 0] + 1 \cdot [3, -1, 3]$$

$$[1, 2, 3] = [-2, 3, 0] + [3, -1, 3]$$

$$[1, 2, 3] = [1, 2, 3]$$

Vectors

More definitions

- Vector x is a **linear combination** of vectors $x_1, x_2, ..., x_n$ if there exists real numbers $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{R}$ such as:
$$x = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_n x_n$$
- Vectors $x_1, x_2, ... x_n$ are **linearly dependent** if at least one of the vectors is a linear combination of remaining vectors.
- Vectors are **linearly independent** if they are not linearly dependent.

Theorem: In $\mathbb{R}^n$ there are maximally n vectors which are linearly independent.

For example, in $\mathbb{R}^2$ vectors $[1, 2]$ and $[2, 4]$ are linearly dependent. On the other hand, $[1, 2]$ and $[2, 1]$ are linearly independent.

Matrices

Let's define two matrices $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 2 & 3 \end{bmatrix}$

Then:

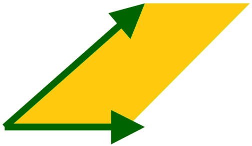
- $A + B = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 1 & 8 \end{bmatrix}$
- $5A = \begin{bmatrix} 5 & 10 & 15 \\ 0 & -5 & 25 \end{bmatrix}$
- $B^T = \begin{bmatrix} -1 & 2 \\ 0 & 2 \\ 1 & 3 \end{bmatrix}$
- $A \cdot B^T = \begin{bmatrix} 2 & 15 \\ 5 & 13 \end{bmatrix}$

Matrices

- **Inverse** of a square matrix A is matrix A^{-1} such as $A \cdot A^{-1} = I$ (identity matrix).
- **Rank** of a matrix is the number of linearly independent rows (or equivalently columns).
- **Determinant** of a square matrix A is

$$det(A) = \sum_k (-1)^{k+i} a_{ik} det(A_{ik})$$

where A_{ik} obtained from A by removing the i-th row and k-th column.



$Area = |det(A)|$

Linear systems

System of linear equations

$$\begin{cases} a_1 x + b_1 y = c_1 \\ a_2 x + b_2 y = c_2 \end{cases}$$

may be solved using matrices. If a solution exists, it may be calculated as

$$x = \frac{det(M_1)}{det(M)}$$

$$y = \frac{det(M_2)}{det(M)}$$

where $M = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$. $M_1 = \begin{bmatrix} c_1 & b_1 \\ c_2 & b_2 \end{bmatrix}$ and $M_2 = \begin{bmatrix} a_1 & c_1 \\ a_2 & c_2 \end{bmatrix}$.

Theorem (Cramer) holds for also for more than two variables. Then M is a matrix of coefficients, M_i - is a matrix where $i - th$ column was replaced by the $[c_1, c_2....]$ and the solution is $x_i = \frac{det(M_i)}{det(M)}$.

Probability

- Ω - sample space
- A - event
- $P(A) = \frac{|A|}{|\Omega|}$ (classically, there are other definitions)

Example: Toss a dice

- $\Omega = \{1, 2, 3, 4, 5, 6\}$, $|\Omega| = 6$,
- A - less than 3, $A = \{1, 2\}$, $|A| = 2$,
- $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6}$.

Probability

Definition

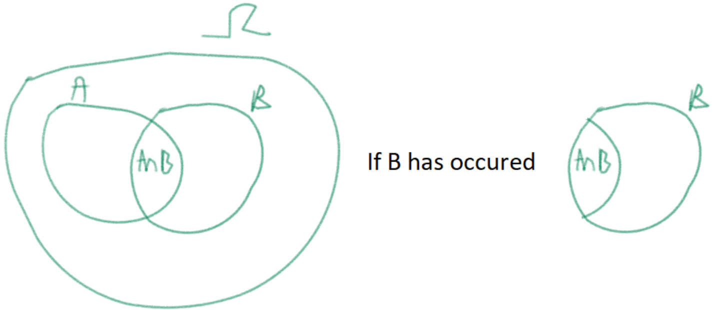
- $P(A) \geqslant 0$,
- $P(\Omega) = 1$,
- If $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Properties

- $P(\emptyset) = 0$,
- $P(A) \leqslant 1$,
- $P(A') = 1 - P(A)$,
- $A \subset B \Rightarrow P(A) \leqslant P(B)$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
- $A_i \cap A_j = \emptyset \Rightarrow P(A_1 \cup A_2 \cup ... A_n) = P(A_1) + P(A_2) + ... + P(A_n) = \sum_i P(A_i)$.

Conditional Probability

- A, B - events,
- We know that event B has occured and $P(B) > 0$.



then

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

We have new sample space and new event.

Independence

- A, B - events,
- We know that event B has occured and $P(B) > 0$.

then

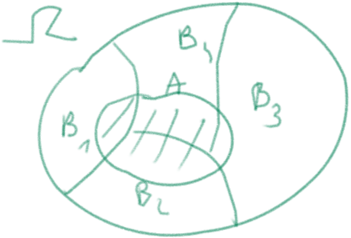
$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{P(A \cap B)}{P(B)}$$

- Events A and B are **independent** if $P(A \cap B) = P(A) \cdot P(B)$.

Theorem: If A and B independent, then $P(A|B) = P(A)$.

Law of total probability

Let's say we have a partition of Ω which is $B_1, B_2, \ldots B_n$



Then

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

as we have a sum of disjoint sets

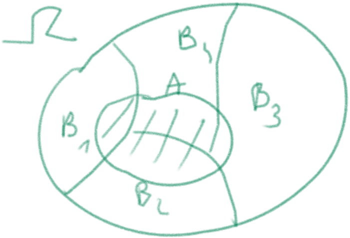
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$$

and finally

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n)$$

Bayes Theorem

Let's say we have a partition of Ω which is $B_1, B_2, ...B_n$.

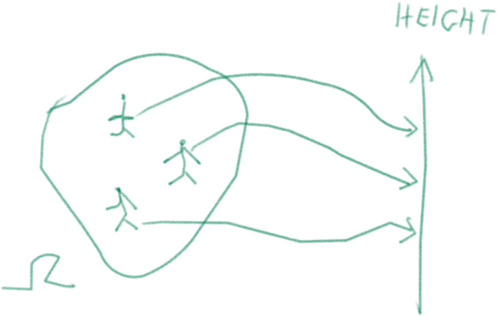


Then

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_k P(A|B_k)P(B_k)}$$

Random variables

X - random variable, $X : \Omega \longrightarrow \mathbb{R}$



X is random because of its argument.

Random variables

$P(X = a) = P(\omega \in \Omega : X(\omega) = a) = \frac{|\omega \in \Omega : X(\omega) = a|}{|\Omega|}$

Example: $X-$year of birth,

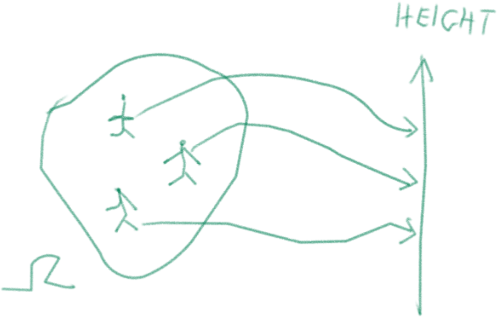
$$X : \{\omega_1, \omega_2, \omega_3\} \to \mathbb{R}.\ X(\omega_1) = 2000,\ X(\omega_2) = 2000,\ X(\omega_3) = 2003.$$

Then $P(X = 2000) = \frac{2}{3}$, $P(X < 3000) = 1$, $P(X < 0) = 0$.

In practice, we are not considering single events but the behaviour of the whole population. For example, X is a random variable that describes the height of Polish people. If $P(X < 150) = 0.8$, it means 80 % of Polish people has less than 150 m in height but we don't know which ones exactly (and it is not interesting for us).

Random variables

X - random variable, $X : \Omega \rightarrow \mathbb{R}$



- **Discrete random variables** - has countable number of values (often natural numbers). Example: number of children (0,1,2...)
- **Continuous random variables** - has continuous number of values. Example: weigth (any non-negative real number).

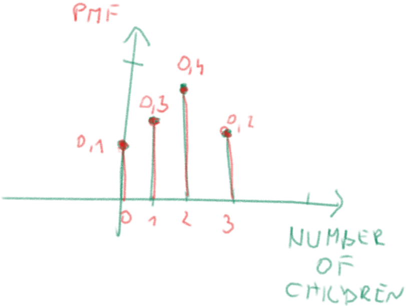
Random variables

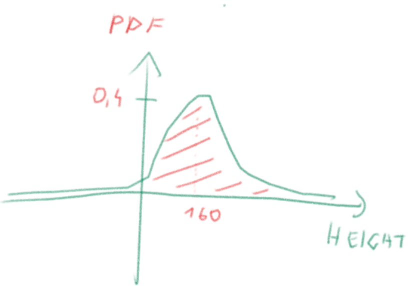
- For discrete random variables we consider **probability mass function (PMF)**

$$p(x) = P(X = x)$$

- For continuous random variables we consider **probability density function (PDF)**

$f(x)$ - describes which values are more or less probable („normalized histogram")

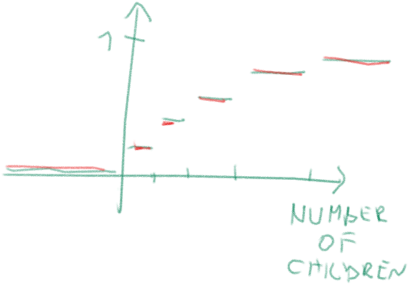


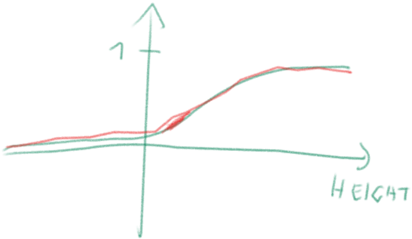


Cumultative distribution function (CDF)

Cumultative distribution function (CDF) is defined as

$$F(x) = P(X \leqslant x)$$



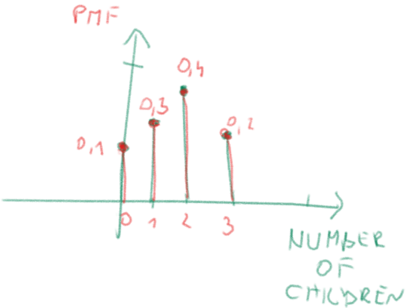


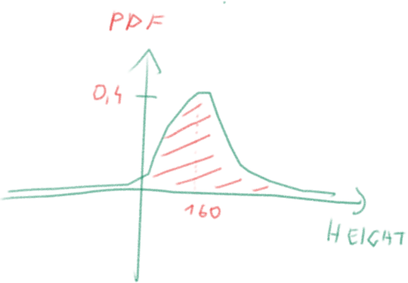
Properties:

- F is non decreasing,
- $F(x) \to 1$ if $x \to \infty$,
- $F(x) \to 0$ if $x \to -\infty$,
- $P(a < X \leqslant b) = F(b) - F(a)$.

CDF and PMF/PDF

- For discrete random variables $F(x) = P(X \leqslant x) = \sum_{t \leqslant x} P(X = t) = \sum_{t \leqslant x} p(t)$,
- For continuous random variables $F(x) = \int_{-\infty}^{x} f(t)dt$. Property: $P(a < X \leqslant b) = \int_{a}^{b} f(t)dt$.





- What is the proability of having less or equal 2 children? $F(2) = p(0) + p(1) + p(2) = 0.1 + 0.3 + 0.4 = 0.8$
- What is the probability of having a height less or equal of 160? $F(160) = \int_{-\infty}^{160} f(t)dt$.

Note that in the second case $f(x)$ is not interpreted as probability! For example $f(160)$ is not the probability of having height of 160.

Expected value and variance

Expected value EX - "mean/average/center"

- For discrete random varaible $EX = \sum_i x_i p(x_i)$,
- For continuous random variable $EX = \int_{-\infty}^{\infty} x \cdot f(x)dx$.

Variance D^2X - measures the spread around the mean.

$$D^2X = E(X - EX)^2$$

Standard deviation $DX = \sqrt{D^2X}$ - similar interpretation to variance.

Example:

X	1	2	5
p	0.1	0.4	0.5

- $EX = 1 \cdot 0.1 + 2 \cdot 0.4 + 5 \cdot 0.5 = 1 + 0.8 + 2.5 = 4.3$,
- $D^2X = (1 - 4.3)^2 \cdot 0.1 + (2 - 4.3)^2 \cdot 0.4 + (5 - 4.3)^2 \cdot 0.5 = 3.45$,
- $DX = \sqrt{3.45} = 1.86$.

Expected value and variance

Properties:

- $E(X+Y) = EX + EY$,
- $E(X+a) = EX + a$ for $a \in \mathbb{R}$,
- $E(a \cdot X) = a \cdot EX$ for $a \in \mathbb{R}$,
- $D^2(X+a) = D^2X$ for $a \in \mathbb{R}$,
- $D^2(a \cdot X) = a^2 \cdot D^2X$ for $a \in \mathbb{R}$

Independence. Random variables X and Y are independent if

$$P(X \leqslant x, Y \leqslant y) = P(X \leqslant x)P(Y \leqslant y).$$

Intuitively, we may calculate the probability for X and Y separately and they don't depend on each other. In pratice, we vey often assume that random variables are independent. Then, additional properties hold:

- $E(X \cdot Y) = EX \cdot EY$,
- $D^2(X+Y) = D^2X + D^2Y$.

Independece

Let's consider four random variables

- X_1 - describes the height in a population,
- X_2 - describes the weight in a population,
- X_3 - describes the number of children,
- X_4 - describes the salary in a population.

Which random variables seem to be independent?

Covariance

Covariance

$$cov(X, Y) = E((X - EX) \cdot (Y - EY)) = E(X \cdot Y) - EX \cdot EY$$

indicates the tendency of X and Y to vary together.

- If X and Y tend to increase together, then $cov(X, Y) > 0$,
- If X and Y tend to have inverse tendencies, then $cov(X, Y) < 0$,
- If the behavior of X does not impact Y, then $cov(X, Y) = 0$.

What is the covariance of two independent random variables?

Properties:

- $cov(X, Y) = cov(Y, X)$,
- $cov(X, X) = D^2X$.

Correlation coefficient

$$r = cor(X, Y) = \frac{cov(X, Y)}{DX \cdot DY} \in [-1, 1]$$

Random vectors

Let's consider

$$X = [X_1, X_2, ..., X_n]$$

where X_i is a random variable. Then, expected value of X is

$$EX = [EX_1, EX_2, ..., EX_n]$$

or shortly

$$\mu = [\mu_1, \mu_2, ..., \mu_n].$$

In this case, we consider a **covariance matrix**

$$\Sigma = \begin{bmatrix} E((X_1 - \mu_1)(X_1 - \mu_1)) & \dots & E((X_n - \mu_n)(X_1 - \mu_1)) \\ E((X_1 - \mu_1)(X_2 - \mu_2)) & \dots & E((X_n - \mu_n)(X_2 - \mu_2)) \\ \dots & \dots & \dots \\ E((X_1 - \mu_1)(X_n - \mu_n)) & \dots & E((X_n - \mu_n)(X_n - \mu_n)) \end{bmatrix}$$

Random vectors

Example. $X = [X_1, X_2, X_3]$. Then

$$\Sigma = \begin{bmatrix} \sigma_1^2 & c_{12} & c_{13} \\ c_{21} & \sigma_2^2 & c_{23} \\ c_{31} & c_{32} & \sigma_3^2 \end{bmatrix}$$

where

$$\sigma_i^2 = D^2 X_i$$

$$c_{ij} = cov(X_i, X_j)$$

It is a symmetric matrix.