Probabilistic Machine Learning:
1. Probabilistic refresher

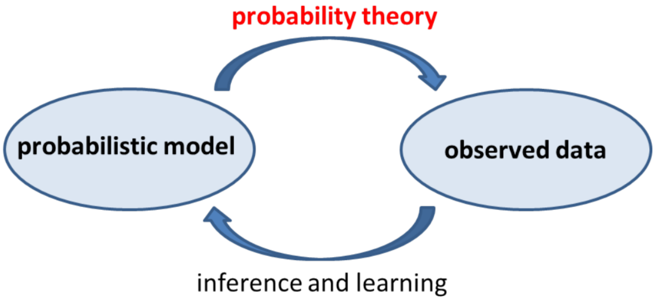
Tomasz Kajdanowicz, Przemysław Kazienko (substitution)

Department of Computational Intelligence
Wroclaw University of Science and Technology

The presentation has been inspired and in some parts totally based on Prof. Mario A. T. Figueiredo presentation at LxMLS'2017, Instituto Superior Tecnico & Instituto de Telecomunicacoes, Lisboa, Portugal.

Apropriate agreements to propagate his ideas has been acquired.

Probability theory

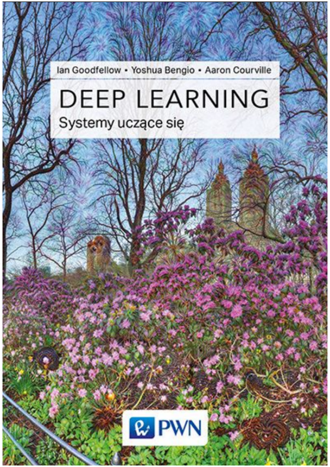


- has its origins in gambling
- great names: Fermat, Pascal, Bernoulli, Huygens, Laplace, Kolmogorov, Poisson, Cauchy, Boltzman, Bayes, Cardano, ...
- tool to handle uncertainty, information, knowledge, observations, ...
- ...thus also learning, decision making, inference, science, data science ...

Do we still need to know probability theory?

What book is this from?

Do we still need to know probability theory?



What is probability?

Example

$\mathbb{P}$(randomly drawn card is $\heartsuit$) = $13/52$
$\mathbb{P}$(getting 1 in throwing a fair die) = $1/6$

- Classical definition: $\mathbb{P}(A) = \frac{N_A}{N}$
  ...with N mutually exclusive equally likely outcomes, N_A of which result in the occurrence of A.
- Frequentist definition: $\mathbb{P}(A) = \lim_{N\to\infty} \frac{N_A}{N}$
  ...relative frequency of occurrence of A in infinite number of trials
- Subjective probability:
  ...gives meaning to $\mathbb{P}$("it will rain today"), or $\mathbb{P}$("I'll have passed the PUMa's exam next winter")

Key concepts: Sample space and events

- Sample space $\mathcal{X}$ = set of possible outcomes of a random experiment.

Example

- Tossing two coins: $\mathcal{X} = \{HH, TH, HT, TT\}$
- Roulette: $\mathcal{X} = \{1, 2, ..., 36\}$
- Draw a card from a shuffled deck $\mathcal{X} = \{A\heartsuit, 2\heartsuit, \ldots, Q\diamondsuit, K\diamondsuit\}$

- An event A is a subset of $\mathcal{X}$: $A \subseteq \mathcal{X}$ (also written $A \in 2^{\mathcal{X}}$)

Example

- exactly one H in 2-coin toss: $A = \{TH, HT\}$
- odd number in the roulette: $A = \{1, 3, ..., 35\}$
- drawn a $\heartsuit$ card: $A = \{A\heartsuit, 2\heartsuit, \ldots, K\heartsuit\}$

Key concepts: Sample space and events

- Sample space $\mathcal{X}$ = set of possible outcomes of a random experiment.

Example

- Distance travelled by tossed die: $\mathcal{X} = \mathbb{R}_+$
- Location of the next rain drop on a given square tile: $\mathcal{X} = \mathbb{R}^2$

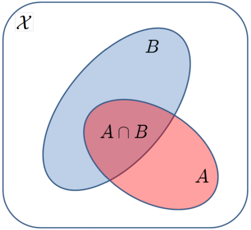
- Properly handling the continuous case requires deeper concepts:
  - Sigma algebras
  - Measurable functions
  - ... and other heavier stuff, not covered here

Kolmogorov's Axioms for Probability

- Probability is a function that maps events A into the interval $[0,1]$. Kolmogorov's axioms (1933) for probability
  - For any A, $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\mathcal{X}) = 1$
  - If $A_1, A_2, \cdots \subseteq \mathcal{X}$ are disjoint events, then $\mathbb{P}(\bigcup_i A_i) = \sum \mathbb{P}(A_i)$
- From these axioms, many results can be derived

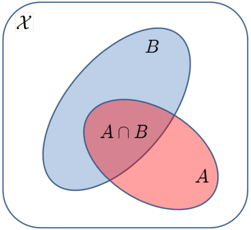
Example

- $\mathbb{P}(\varnothing) = 0$
- $C \subset D \implies \mathbb{P}(C) \leq \mathbb{P}(D)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$



Conditional Probability and Independence

- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
  (conditional probability of A, given B)
- ...satisfies all of Kolmogorov's axioms:
  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A|B) \geq 0$
  - $\mathbb{P}(\mathcal{X}|B) = 1$
  - If $A_1, A_2, \cdots \subseteq \mathcal{X}$ are disjoint,
    $\mathbb{P}(\bigcup_i A_i|B) = \sum_i \mathbb{P}(A_i|B)$
- Independence: A, B are independent ($A \perp\!\!\!\perp B$):
  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$



Conditional Probability and Independence

- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$
- Events A, B are independent $(A \perp\!\!\!\perp B) \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- Relationship with conditional probabilities:
  $(A \perp\!\!\!\perp B) \iff \mathbb{P}(A|B) = \mathbb{P}(A)$

Example

$\mathcal{X}$ ="52 cards", $A = \{4\heartsuit, 4\diamondsuit, 4\clubsuit, 4\spadesuit\}$, and $B = \{A\heartsuit, 2\heartsuit, \ldots, K\heartsuit\}$;
then
$\mathbb{P}(A) = 1/13$, $\mathbb{P}(B) = 1/4$
$\mathbb{P}(A \cap B) = \mathbb{P}(4\heartsuit) = \frac{1}{52}$
$\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{13}\frac{1}{4} = \frac{1}{52}$
$\mathbb{P}(A|B) = \mathbb{P}(''4''|''\heartsuit'') = \frac{1}{13} = \mathbb{P}(A)$

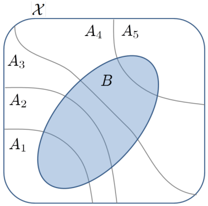
Bayes Theorem

- Law of total probability: if $A_1, \ldots, A_n$ are a partition of $\mathcal{X}$

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i) = \sum_i \mathbb{P}(B \cap A_i)$$

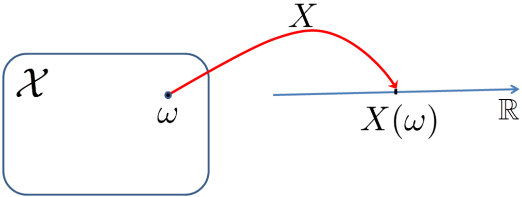
- Bayes' theorem: if $\{A_1, \ldots, A_n\}$ is a partition of $\mathcal{X}$

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



Random Variables

- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$



  - Discrete RV: range of X is countable (e.g., $\mathbb{N}$ or $\{0, 1\}$)
  - Continuous RV: range of X is uncountable (e.g., $\mathbb{R}$ or $[0, 1]$)

Example

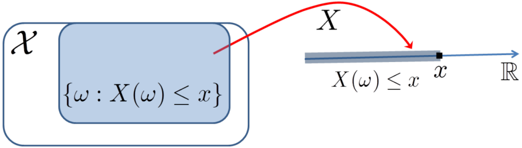
number of heads in tossing two coins, $\mathcal{X} = \{HH, HT, TH, TT\}$, $X(HH) = 2$, $X(HT) = X(TH) = 1$, $X(TT) = 0$, range of $X = \{0, 1, 2\}$

Example

distance traveled by a tossed coin; range of $X = \mathbb{R}_+$

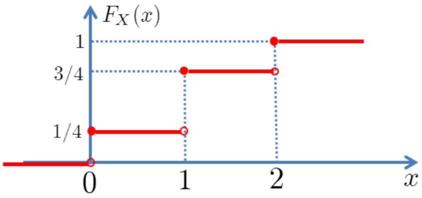
Random Variables: Distribution Function

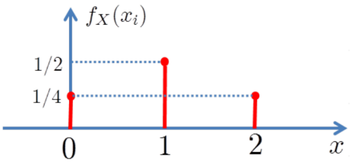
- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$



Examples

number of heads in tossing 2 coins; range(X) = $\{0, 1, 2\}$





- Probability mass function (discrete RV): $f_X(x) = \mathbb{P}(X = x)$, $F_X(x) = \sum_{x_i \leq x} f_X(x_i)$

Important Discrete Random Variables

- Uniform: $X \in \{x_1, ..., x_K\}$, pmf $f_X(x_i) = 1/K$

Examples

a fair roulette $X \in \{1, ..., 36\}$, with $f_X(x) = 1/36$
a fair die $X \in \{1, ..., 6\}$, with $f_X(x) = 1/6$

- Bernoulli RV: $X \in \{0, 1\}$, pmf $f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$

  Compact form: $f_X(x) = p^x(1 - p)^{1-x}$

Examples

an unfair coin (heads = 0, tails = 1), with $p \neq 1/2$.

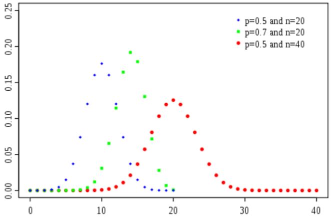
Important Discrete Random Variables

- Binomial RV: $X \in \{0, 1, ..., n\}$ (sum of n Bernoulli RVs)

  $$f_X(x) = \text{Binomial}(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

  Binomial coefficients ("n choose "x"):

  $$\binom{n}{x} = \frac{n!}{(n-x)!x!}$$



Example

number of heads in n coin tosses.

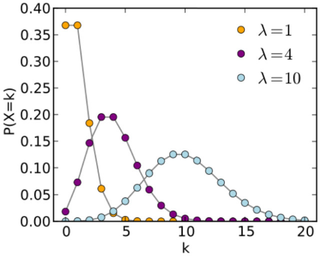
Other Important Discrete Random Variables

- Geometric(p): $X \in \mathbb{N}$, pmf
  $f_X(x) = p(1-p)^{x-1}$

Example

number of coin tosses until first heads

- Poisson(λ):
  $X \in \mathbb{N} \cup \{0\}$
  pmf $f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$



Example

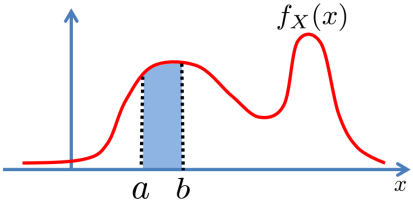
"...probability of the number of independent occurrences in a fixed (time/space) interval, if these occurrences have known average rate"

Continuous Random Variables

- Probability density function (pdf, continuous RV): $f_X(x)$

$$\int_{-\infty}^{\infty} f_X(x) = 1$$

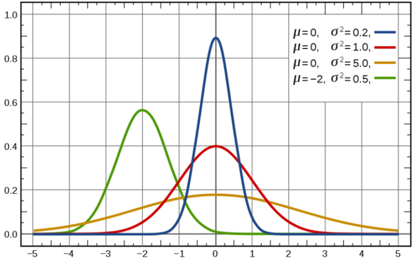
$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(x)dx$$

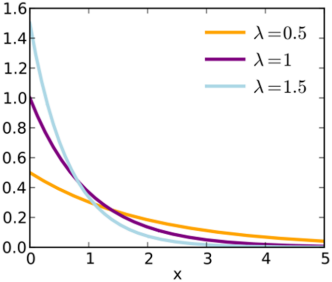


- Notice: $\mathbb{P}(X = c) = 0$

Important Continuous Random Variables

- Uniform: $f_X(x) = \text{Uniform}(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{cases}$
- Gaussian: $f_X(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$





- Exponential: $f_X(x) = \text{Exp}(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$