Probability: Quick and (Hopefully) Gentle Intro

Tomáš Svoboda and Petr Pošík

Vision for Robots and Autonomous Systems, Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University in Prague

March 13, 2023

Notes

Specializovaný předmět na pravděpodobnost a statistiku teprve přijde.

Outline

- Mathematics of uncertainty
- Random Experiment, Outcomes, Sample Space, Events, . . .
- Probability, Conditional Probability, Independence
- Random Variable, Expectation

Notes

Uncertainty is everywhere

- The probability of rain tomorrow is 70%.
- What are my chances to win in a lottery?
- I was tested positive for disease X, am I really sick?
- Given testimonies X, Y, and Z, is the suspect guilty?
- Unemployment changed by X, what will be the inflation?
- How will the stock prices evolve?
- We chose action X, how much will the robot move?
- What is the probability that the person on the photo is person X?
- How long will it take me to get to work if I take the tram?
- ...

We need a mathematical description ...

Notes

(Random) Experiment

Experiment :

- Vaguely: the act of observing certain feature of the world
- A procedure that
  - can be repeated many times under the same conditions and
  - has a well-defined set of possible outcomes.
- Deterministic experiment has only a single possible outcome.
- Random experiment has more than one possible outcomes.
- Before executing random experiment, we do not know the actual outcome. After execution this uncertainty vanishes.

Notes

Example 1: Three tosses of a coin (Head/Tails)

What is the probability of three heads?

Sample space (a set!) $\mathcal{S}$ of all elementary events (experiment outcomes) . How big is it?

A 3^2

B 2^3

C $2 \cdot 3$

D ∞

Events :

- A - $3\times$ head, $P(A) = ?$
- B - $3\times$ the same symbol, $P(B) = ?$
- C - at least one tail, $P(C) = ?$ ...

Notes

Pro výpočet velikosti prostoru všech elementárních jevů někdy poslouží vhodný model. Tady např. to může být $n-$bitové binární číslo.
Vyjděme z množiny elementárních jevů - $HHH, HHT, \ldots, TTT$.
Je pravděpodobnost některého elementárního jevu větší nebo menší než u ostatních? Je to tak vždy?
Jak definovat jevy A, B, C? Nešlo by jev C definovat snáze, pomocí množinových operací a jiných elementárních jevů?

Jaká je jejich pravděpodobnost? Jak by se dala spočítat $P(C)$ pomocí již známých pstí?

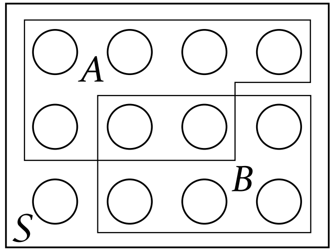
(Random) Events /(Náhodné) jevy/

Elementary events /elementární jevy/ are *all possible, mutually exclusive outcomes* of certain experiment.

The set of elementary events is called a sample space /množina elementárních jevů/, denoted as $\mathcal{S}$.

An event /jev/ is any subset of the sample space, $A \subseteq \mathcal{S}$.

- Event A occured if the experiment outcome belongs to A.
- An event is any statement about the experiment outcome for which we can decide if it occured or not.

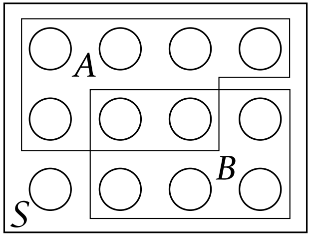


Notes

Naive probability (Bernoulli/Laplace)

$$P(A) = \frac{|A|}{|\mathcal{S}|} = \frac{\text{number of outcomes favorable to } A}{\text{total number of outcomes in } \mathcal{S}}$$

- Limited to *equally likely* outcomes/elementary events. (*Equally likely?*)
- It does not allow for infinite sample spaces, geometric probability, ...
- *Combinatorics!* Counting (variations, permutations, combinations, ...)



Notes

"Equally likely": we actually use an assumption on probability values in the definition of the probability.

Events and their combinations

Important events:

- Certain event : $\mathcal{S}$, $\mathbf{1}$
- Impossible event : $\emptyset$, $\mathbf{0}$

Event combinations:

- Conjunction (A and B): $A \cap B$
- Disjuction (A or B): $A \cup B$
- Complementary event /jev opačný/ to A: $A^c = \mathcal{S} \setminus A$
- $A \Rightarrow B$: $A \subseteq B$
- Disjoint events /jevy neslučitelné/ : $A_1, \dots, A_n : \bigcap_{i \leq n} A_i = \emptyset$
- Mutually exclusive events /Jevy po dvou neslučitelné = vzájemně se vylučující/ : $A_1, \dots, A_n : \forall i, j \in \{1, \dots, n\},\ i \neq j : A_i \cap A_j = \emptyset$

Notes

Výrokovou logiku lze k popisu jevů použít místo množin, je to ekvivalentní popis. Je dobré ale obojí nemixovat.

Partition of sample space /Úplný systém jevů/

Partition of sample space $\mathcal{S}$ /Úplný systém jevů/ is composed of events $B_1, \ldots, B_n$ if they are *mutually exclusive* and $\bigcup_{i=1}^{n} B_i = \mathcal{S}$.

- The sample space $\mathcal{S}$ is its own partition by definition.
- Events $\{C, C^c\}$ form a partition: $C \cap C^c = \emptyset$ and $C \cup C^c = \mathcal{S}$.

Why is the partition of $\mathcal{S}$ an important concept?

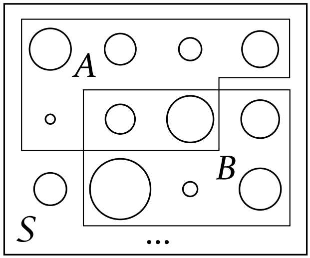
Notes

Proč je úplný systém jevů důležitý koncept?

Protože víme, že výsledkem experimentu je právě jeden z jevů v úplném systému. Proč?

Axiomatic probability (Kolmogorov)

- Sample space $\mathcal{S}$ may be infinite.
- Elementary events do not have to be equally likely.
- Axiomatic:
  1. state a set of constraints the probability function must obey
  2. find a function that fulfills them (next slides)

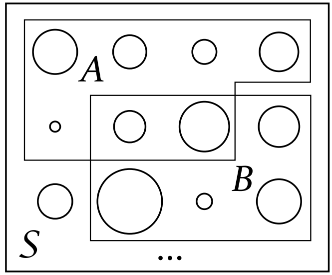


Notes

From discrete to continous, From Pebble world representation to Venn Diagrams.

Definition of probability

- Probability function /pravděpodobnostní funkce/ P is a function that assigns a real number between 0 and 1 to each event $A \subseteq \mathcal{S}$.
- P must satisfy the following axioms:



1. $P(\emptyset) = 0$, $P(\mathcal{S}) = 1$
2. For any mutually exclusive events $A_1, A_2, \ldots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

   (n may be infinite)

Notes

Interpretations of probability

Frequentist :

- Relative frequency of an event after many repetitions of random experiment.

Bayesian :

- Degree of belief that an event occurs.
- This allows us to assign probabilities to statements like "candidate A wins elections" or "suspect X is guilty", although we cannot repeat the same elections or the same crime over and over.

Notes

Example 2: Properties of P, rolling a die

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

Consider events:

- A - outcome is 6
- B - outcome is an even number

Using sets: $A \subset B$

Probability: $P(A) < P(B)$

Another event:

- C - outcome is 2 or 4

Using sets: $C = B \setminus A$

Probability: $P(C) = P(B) - P(A)$

Notes

Pro názornost nám opět dobře poslouží oblázkový svět.

Example 2: Properties of P (cont.)

Rolling a die, $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, A - outcome is 6, B - outcome is an odd number. Obvioiusly $A \cap B = \emptyset$,

$$P(A \cup B) = P(A) + P(B)$$

A pump in a power plant is backed up by another, identical pump. Event A_i means that pump i is OK. What is the probability that at least one of them is OK?

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Both pumps are OK:

$$P(A_1 \cap A_2) = ?$$

Notes

Základní kontrola pravděpodobnosti libovolně složité události A: $0 \leq P(A) \leq 1$.

Jak spočítat $P(A_1 \cap A_2)$? Existuje nějaký případ, kdy to lze spočítat snadno?

Properties of probability

For any valid probability function:

- $P(A) \in \langle 0, 1 \rangle$ (definition)
- $P(\emptyset) = 0, \qquad P(\mathcal{S}) = 1$ (axioms)
- $P(A^c) = 1 - P(A)$
- If $A \subseteq B$, then $P(A) \leq P(B)$
- If $A \subseteq B$, then $P(B \setminus A) = P(B) - P(A)$
- If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$ (*aditivity*)
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Notes

Example 3: Probability of parts

If we choose a person form the population at random,

- he/she suffers from disease X and is younger than 18 years with probability 0.01,
- he/she suffers from disease X and is between 18 and 65 years with probability 0.05, and
- he/she suffers from disease X and is older than 65 years with probability 0.09.

What is the probability a randomly chosen person suffers from disease X?

Notes

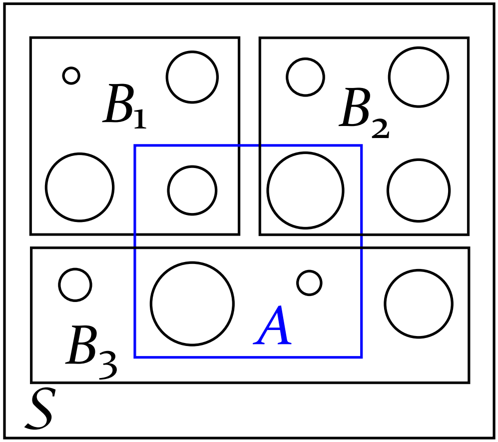
Properties of probability (cont.)

If $\{B_1, \ldots, B_n\}$ is a partition of sample space then for any event $A \subseteq \mathcal{S}$

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) \, .$$

In particular, for partition $\{C, C^c\}$

$$P(A) = P(A \cap C) + P(A \cap C^c) \, .$$



Notes

Independent events /Nezávislé jevy/

Events A and B are independent if and only iff

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

If A, B are independent, then

- $P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$,
- and pairs A, B^c and A^c, B and A^c, B^c are independent too.

Notes

Independence of events: tossing two coins

- A - head on the first coin
- B - head on the second coin
- C - different symbols on the coins

Which groups of events are independent?

A no group of events

B pairs (A, B), (B, C), (A, C)

C pairs (A, B), (B, C), (A, C) and triple (A, B, C)

D only triple (A, B, C)

Notes

Mohou být dvojice (A, A) a (A, A^c) nezávislé?

Kdy pro (A, A) platí, že $P(A \cap A) = P(A)P(A) = P(A)$?

Kdy pro (A, A^c) platí, že $P(A \cap A^c) = P(A)P(A^c) = P(A)(1 - P(A))$?

Example: Soldier or technician?

Tom likes order, is decisive, and has a good sense of justice. When he was a kid, he liked to play strategic games and shooting RPGs. He has always been interested in weapons and military equipment.

What do you think is Tom's occupation now?

A Soldier

B Technician

Notes

Nechť je identifikace povolání jev V a T. Můžeme zobecnit na hypotézu V, buď platí $H = V$ nebo $H = T$. Daný popis osobnosti nechť je E jako evidence.

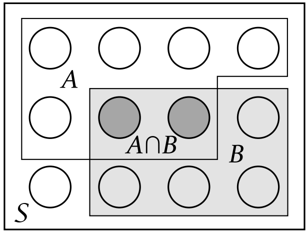
Podle Sčítání 2021 je v ČR cca 22 tis. zaměstnanců v ozbrojených silách a cca 860 tis. technických a odborných pracovníků.

Diskutujme, podle čeho jsme se rozhodovali. Nakresleme diagram s jevy V a T, znázorníme v obou jevech části, kdy platí i E.

Conditional probability

Conditional probability of event A given event B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) \neq 0.$$



- All probabilities are conditional: $P(A) = P(A|\mathcal{S})$.
- Interpretation:
  1. $P(A)$ is our current belief that event A occurs.
  2. We get a new information that a different event B occured.
  3. $P(A|B)$ is now our updated belief about A.
- Conditional probability is still a probability: it maps any event $A \subseteq \mathcal{S}$ to $\langle 0, 1 \rangle$.

Notes

Properties of Conditional Probability

- $P(\mathcal{S}|B) = 1$, $P(\emptyset|B) = 0$.
- $P(A|A) = 1$, $P(A^c|A) = 0$.
- If $B \subseteq A$, then $P(A|B) = 1$.
- If $P(A \cap B) = 0$, then $P(A|B) = 0$.
- If $A_1, \ldots, A_n$ are mutually exclusive events, then $P\left(\bigcup_{i=1}^{n} A_i \middle| B\right) = \sum_{i=1}^{n} P(A_i|B)$.
- Events A, B are independent iff $P(A|B) = P(A)$ (if $P(A|B)$ is defined).

Notes

Belief update

1. Probability $P(A)$ is our initial (prior) belief that event A occurs.
2. We learn that another event, B, occured.
3. Probability $P(A|B)$ is our updated (posterior) belief that event A occurs.

No other info about events A and B is available. Which of the following options is correct?

A $P(A|B) < P(A)$

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

C $P(A|B) > P(A)$

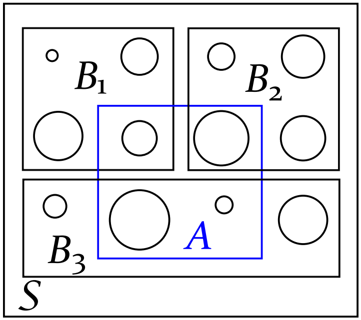
D Any of the above options can happen.

Notes

The Law of Total Probability

Let $B_1, \ldots, B_n$ be a partition of the sample space $\mathcal{S}$ (i.e., the B_i are disjoint events and their union is $\mathcal{S}$), with $P(B_i) > 0$ for all i.
Then

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



Notes

Bayes rule

Probability of the intersection of two events A and B, $P(A \cap B)$, can be expressed in 2 ways:

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

From that it follows that

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

Applying the law of total probability from previous slide:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j \in I} P(A|B_j) \cdot P(B_j)}.$$

Notes

Working with random events becomes cumbersome ...

Experiment: 3 tosses of a coin. Outcomes $s \in \mathcal{S}$. Events $A_i \subseteq \mathcal{S}$:

- three heads – $X(s) = 3$
- at least one head – $X(s) \geq 1$
- three equal symbols – $X(s) \in \{3, 0\}$
- ...

We can define each event as a set (often quite large) of outcomes s.
Or we can define a random variable :

$$X(s) = \text{number of heads in } s$$

Before the experiment, how many heads do I expect to be tossed?

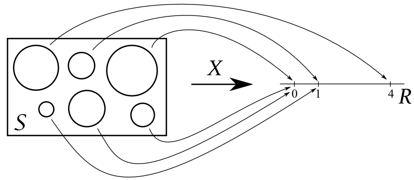
Notes

Lze uvažovat např. i o hodu 3 kostkami. Kolik existuje různých výsledků experimentu?

Jak je asi složité pracovat s jevy zahrnujícími stovky elementárních jevů?

Random Variable

Random variable (náhodná proměnná/veličina) on a probability space $(\mathcal{S}, P)$ is a function X mapping elementary events $s \in \mathcal{S}$ to real numbers $\mathbb{R}$, i.e., $X : \mathcal{S} \to \mathbb{R}$.



"Random variable is a numerical 'summary' of an aspect of the experiment."

- R.v. X assigns a numerical value $X(s)$ to each possible outcome $s \in \mathcal{S}$.
- The mapping is *deterministic*; the randomness comes from outcomes of random experiment (with outcome probabilities described by probability function P).
- Before the experiment, we know neither the value of s, nor the value of $X(s)$. But we can compute the probability that X will take on a given value, or a range of values.
- After the experiment, s was realized, and the r.v. crystalizes into value $X(s)$.

Notes

Events vs Values of Random Variable

Let X be a random variable, i.e., $X : \mathcal{S} \to \mathbb{R}$.

- $X = x$ denotes the event $\{s \in \mathcal{S} : X(s) = x\}$, i.e., the event consisting of all outcomes s such that $X(s) = x$.
- $X \in \langle a, b)$ denotes the event $\{s \in \mathcal{S} : a \leq X(s) < b\}$, i.e., the event consisting of all outcomes s such that $a \leq X(s) < b$.

Notes

Discrete Random Variable

Random variable X is called discrete if the values of $X(s)$ for all $s \in \mathcal{S}$ form either

- a finite set of values $a_1, a_2, \ldots, a_n$, or
- an infinite set of countably many values $a_1, a_2, \ldots$

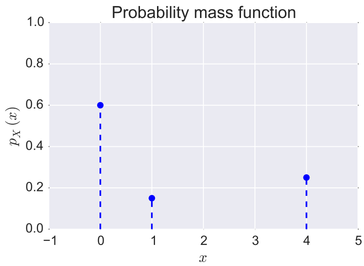
Support of X:
$\mathcal{S}_X = \{x \in \mathbb{R} : P(X = x) > 0\} = \{a_1, a_2, \ldots\}$

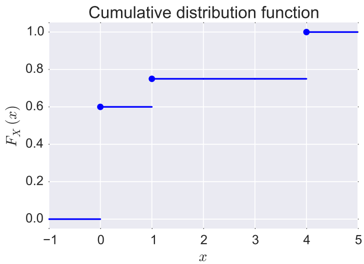
Probability Mass Function (PMF) /pstní fce/ of a discrete r.v. X is the function p_X given by

$$p_X(x) = P(X = x) = P(\{s \in \mathcal{S} : X(s) = x\}).$$

Cumulative Distribution Function (CDF) /distribuční fce/ of a discrete r.v. X is the function F_X defined as

$$F_X(x) = P(X \leq x) = \sum_{t \leq x} p_X(t).$$





Notes

Expected value

Expected value (střední hodnota) of a discrete r.v. X is denoted as $\operatorname{E} X$ and is defined as

$$\operatorname{E} X = \sum_{t \in \mathbb{R}} t \cdot p_X(t) = \sum_{t \in \mathcal{S}_X} t \cdot p_X(t).$$

For equally probable outcomes $s \in \mathcal{S}$ also $\operatorname{E} X = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} X(s)$.

Characteristics of $\operatorname{E} X$:

- $\operatorname{E} r = r$, $\operatorname{E}(\operatorname{E} X) = \operatorname{E} X$
- $\operatorname{E}(X + Y) = \operatorname{E} X + \operatorname{E} Y$, $\operatorname{E}(X + r) = \operatorname{E} X + r$, $\operatorname{E}(X - Y) = \operatorname{E} X - \operatorname{E} Y$
- $\operatorname{E}(rX + sY) = r \operatorname{E} X + s \operatorname{E} Y$
- For *independent* r.v.s: $\operatorname{E}(X \cdot Y) = \operatorname{E} X \cdot \operatorname{E} Y$.

Notes

References, further reading

[1] Joseph K. Blitzstein and Jessica Hwang.
Introduction to Probability.
CRC Press, 2nd edition, 2019.
http://probabilitybook.net/.

[2] Mirko Navara.
Pravděpodobnost a matematická statistika.
ČVUT v Praze, Praha, 2007.
https://cmp.felk.cvut.cz/~navara/stat/.

[3] N. Silver.
The Signal and The Noise.
Penguin Group, 2012.

Notes