Applied Statistics

for Computer Science BSc, Exam

Probability Theory and Mathematical Statistics

for Computer Science Engineering BSc, Term grade

István Fazekas
University of Debrecen

2020/21 fall

This work was supported by the construction EFOP-3.4.3-16-2016-00021. The project was supported by the European Union, co-financed by the European Social Fund.

Main topics

1. Probability theory
2. Statistics

Mathematical tools: combinatorics, calculus

Computer tool: Matlab

Book:

Yates, Goodman:

Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers

Lecture 2

Probability

The scope of probability theory

Random experiment: the outcome of the experiment is not determined before performing the experiment.

Examples:
tossing a coin
rolling a die
any game of chance
observations in physics, chemistry, biology, medicine,...
any statistical observations

Probability theory studies those random experiments, which can be repeated several times.

The sample space

Consider a fixed experiment K.
Those outcomes of the experiment, which we can not divide into smaller parts are called **elementary events**.
The elementary events are generally denoted by ω (Greek omega).
The set of all elementary events is called the sample space (probability space).
Notation: Ω (Greek Omega)

Examples

1. Toss a coin. $\Omega = \{H, T\}$
2. Toss two coins. $\Omega = \{HH, HT, TH, TT\}$
3. Roll a die. $\Omega = \{1, 2, 3, 4, 5, 6\}$
4. Choose a point from the interval $(0, 1)$. Then $\Omega = (0, 1)$

Events

The subsets of Ω are called events.

Examples

1. Toss two coins. Let A denote that we obtain at least one H. Then $A = \{HH, HT, TH\}$

2. Roll a die. Let A denote that the result is even. Then $A = \{2, 4, 6\}$

Ω is called the **sure event**.
$\emptyset$ is called the **impossible event**.

Operations on events

A or $B = A \cup B = A + B$

A and $B = A \cap B = AB$

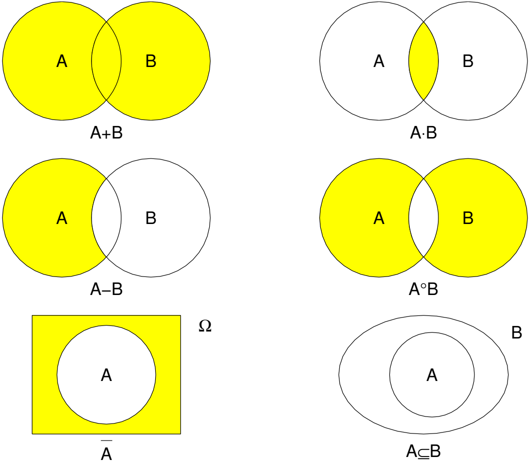
opposite of A = not $A = \bar{A}$

The difference of A and $B = A \setminus B$

Precisely one of A and B occurs
= symmetric difference of A and $B = A \circ B$.

A implies B means that $A \subseteq B$

Venn diagrams of the operations



Rules

Commutative

$$A + B = B + A, \quad A \cdot B = B \cdot A$$

Associative

$$A + (B + C) = (A + B) + C, \quad A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

Idempotent

$$A + A = A, \quad A \cdot A = A$$

Distributive

$$A(B + C) = AB + AC, \ A + (BC) = (A + B) \cdot (A + C)$$

Rules

$$\overline{\overline{A}} = A$$

$$\overline{\Omega} = \emptyset, \ \overline{\emptyset} = \Omega$$

$$A \cdot \Omega = A, \ A + \Omega = \Omega$$

$$A \cdot \emptyset = \emptyset, \ A + \emptyset = A$$

Definition. A and B are called **mutually exclusive** if $A \cap B = \emptyset$

de Morgan laws

$$\overline{A+B} = \overline{A} \cdot \overline{B}, \quad \overline{A \cdot B} = \overline{A} + \overline{B}$$

de Morgan laws for more than two events

$$\overline{\left(\bigcup_i A_i\right)} = \bigcap_i \overline{A}_i, \qquad \overline{\left(\bigcap_i A_i\right)} = \bigcup_i \overline{A}_i$$

Relative frequency

Repeat the experiment n times. The event A occurs k_A times. Then

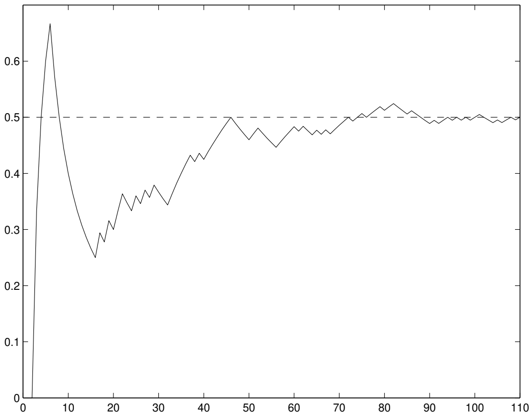
$$\frac{k_A}{n}$$

is called the relative frequency of A.

Example

Toss fair a coin 110 times. The sequence of relative frequencies of H is shown on next figure

Relative frequencies



Axioms of probability

The relative frequency is stable if n is large:

$$\frac{k_A}{n} \sim P(A),$$

where $P(A)$ is a fixed number.

Axioms of probability

$k_A/n \geq 0$ therefore let

$$P(A) \geq 0 \quad \text{for any event } A. \tag{1}$$

For the sure event: $k_\Omega/n = 1$, so let

$$P(\Omega) = 1. \tag{2}$$

If A and B are mutually exclusive, then $k_{A+B} = k_A + k_B$. So using

$$P(A+B) \sim k_{A+B}/n = k_A/n + k_B/n \sim P(A) + P(B)$$

we set

$$\boxed{P(A+B) = P(A) + P(B)}, \tag{3}$$

if A and B are mutually exclusive.

Again the axioms of probability

Non-negative

$$P(A) \geq 0 \quad \text{for any event } A. \tag{4}$$

Normed

$$P(\Omega) = 1. \tag{5}$$

Additive

$$\boxed{P(A + B) = P(A) + P(B)}, \tag{6}$$

if A and B are mutually exclusive.

Properties of the probability

Finitely additivity: if $A_1, A_2, \ldots, A_n$ are pairwise exclusive, then

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \quad (7)$$

Hint: apply mathematical induction and the axiom of additivity.

Let A and B be arbitrary events. Then the axioms imply

$$P(\emptyset) = 0. \quad (8)$$

Hint: $\Omega = \Omega + \emptyset$

$$P(A - B) = P(A) - P(A \cdot B). \quad (9)$$

Hint: $A = A \cdot B + (A - B)$

Union intersection principle

For 2 sets

$$P(A+B) = P(A) + P(B) - P(A \cdot B). \tag{10}$$

Hint: $A + B = A + (B - A)$

For 3 sets

$$P(A+B+C) =$$

$$\big(P(A)+P(B)+P(C)\big)-\big(P(A\cdot B)+P(A\cdot C)+P(B\cdot C)\big)+P(A\cdot B\cdot C).$$

Finite probability spaces

Any N element probability space can be described as follows.

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\},$$

$$P(A) = \sum_{\omega_i \in A} p_i \tag{11}$$

for any event A, where $p_1, \ldots, p_N$ is a given distribution, that is they are non-negative numbers with

$$\sum_{i=1}^{N} p_i = 1.$$

Finite distribution

$p_1, \ldots, p_N$ is called a finite (probability) distribution, (or mass function), if the numbers p_i are non-negative and

$$\sum_{i=1}^{N} p_i = 1.$$

Example.
Let $\Omega = \{a, b, c\}$.
$P(a) = \frac{1}{2},\ P(b) = \frac{1}{3},\ P(c) = \frac{1}{6}$.
Then $P(\Omega) = \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1$,
$P(\{a, b\}) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$,
$P(\{a, c\}) = \frac{1}{2} + \frac{1}{6} = \frac{4}{6}$,
$P(\{b, c\}) = \frac{1}{3} + \frac{1}{6} + \frac{3}{6}$.

Combinatorial probability

Assume that the outcomes of the experiment are equally possible. Then $p_1 = p_2 = \cdots = p_N$. As $\sum_{i=1}^N p_i = 1$, therefore

$$p_1 = p_2 = \cdots = p_N = \frac{1}{N},$$

and

$$P(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \frac{1}{N}.$$

So we obtain the classical rule to calculate probability

$$P(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega} = \frac{\text{number of favorable outcomes}}{\text{number of all outcomes}}.$$

This is the classical rule, which can be applied to solve several but not all problems.

Hypergeometric distribution

Let M red and $N - M$ white balls in a box. We chose n balls from the box without replacement. Let X denote the number of red balls chosen. Then

$$\boxed{h_k = P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}}, \tag{12}$$

$\max\{0, n - N + M\} \leq k \leq \min\{n, M\}$.

Hypergeometric distribution h_k

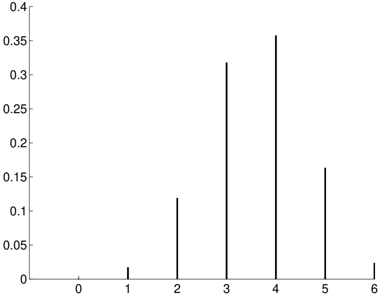


Figure: h_k for $N = 20,\ M = 12,\ n = 6$.

The sequence h_k is increasing while k is not greater than $(n+1)(M+1)/(N+2)$, then it is decreasing. If $(n+1)(M+1)/(N+2)$ is integer, then there are two maxima of the sequence h_k at $k = (n+1)(M+1)/(N+2) - 1$ and at $k = (n+1)(M+1)/(N+2)$.

Binomial distribution

Let M red and $N-M$ white balls in a box. We chose n balls from the box with replacement. Let X denote the number of red balls chosen. Then

$$b_k = P(X=k) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}, \quad k=0,1,\ldots,n.$$

Using $p = M/N$,

$$\boxed{P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\ldots,n}.$$

Binomial distribution b_k

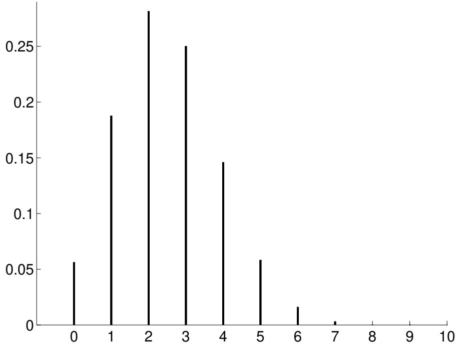


Figure: b_k for $p = 0.25,\ n = 10$

The sequence b_k is increasing while k is not greater than $(n+1)p$, then it is decreasing. If $(n+1)p$ is an integer, then there are two maxima at $k = (n+1)p - 1$ and at $k = (n+1)p$.