Probability Theory
Formulary

I. Events Algebra

Random Trial
A trial whose outcome cannot be predicted in advance is called a *random trial*.

Random Event
The outcome of a random trial is called a *random event*.

Impossible and Certain Events
An event that in all repetitions of a certain random trial will never happen is called an *impossible event*. It will be denoted by $\varnothing$.
An event that in all repetitions of a certain event will always occur is called a *certain event*. It will be denoted by Ω.

Subevent
The event E_1 is called a *subevent of the event* E_2 if it always accompanies the event E_2.
We write

$$E_1 \subseteq E_2.$$

Sum of Events
E is said to be the *sum* of the events $E_i, i = 1, 2, \ldots, n$, if at least one of the events E_i occurs:

$$E := \bigcup_{i=1}^{n} E_i$$

Product of Events
E is said to be the *product* of the events $E_i, i = 1, 2, \ldots, n$, if the events E_i occurs at the same time:

$$E := \bigcap_{i=1}^{n} E_i .$$

Mutually exclusive events
The events E_1 and E_2 are said to be *mutually exclusive* if

$$E_1 \cap E_2 = \varnothing$$

Complementary Events

The events E and $\bar{E}$ are said to be *complementary* if

$$E \cup \bar{E} = \Omega \quad \wedge \quad E \cap \bar{E} = \varnothing$$

Difference

The *difference* of the events E_1 and E_2, denoted by $E_1 \setminus E_2$ is defined as the case in which E_1 occurs while E_2 does not occur.

System of Mutually Exclusive and Exhaustive Events

The events E_i, $i = 1, 2, ..., n$, form a *mutually exclusive and exhaustive* system if the following conditions are fulfilled:

1. $E_i \neq \emptyset, \qquad i = 1, 2, ..., n$
2. $\bigcup_{i=1}^{n} E_i = \Omega$
3. $E_i \cap E_j = \emptyset, \quad i \neq j, \quad i, j = 1, 2, ..., n$

Elementary or Atomic Event

An *elementary* or *atomic event* is an event which is not further genuinely decomposable.

II. Probability Algebra

***Addition rule*:**

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

Conditional probability

$$P(A / B) := \begin{cases} \dfrac{P(A \cap B)}{P(B)} & \text{if} \quad P(B) > 0 \\ 0 & \text{if} \quad P(B) = 0 \end{cases}$$

Multiplication rule

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

$$= P(A) \cdot P(B / A)$$

Dependent and independent events

The event A is said to be *independent* of B if

$$P(A) = P(A / B).$$

Otherwise, A is said to be *dependent* of B.

Multiplication rule for two independent events

Let $A, B \in S$ be two *independent* events. Then

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

Multiplication rule for n independent events

Let E_i, $i = 1, 2, ..., n$, be *independent* events. Then

$$P\left(\bigcap_{i=1}^{n} E_i\right) = \prod_{i=1}^{n} P(E_i).$$

Addition rule for n independent events

Let E_i, $i = 1, 2, ..., n$, be *independent* events. Then

$$P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \prod_{i=1}^{n} \left(1 - P(E_i)\right)$$

Total probability

Let B_i, $i = 1, 2, ..., n$, form a group of mutually exclusive and exhaustive events. For an arbitrary event A we have

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

Bayes-Theorem

Let B_i, $i = 1, 2, ..., n$, form a group of mutually exclusive and exhaustive events. For an arbitrary event $A \neq \emptyset$ we have:

$$P(B_i / A) = \frac{P(A / B_i) \cdot P(B_i)}{\sum_{i=1}^{n} P(A / B_i) \cdot P(B_i)}, \qquad i = 1, 2, ..., n.$$

Sampling schemes

Consider a finite set of N elements, $M \leq N$ elements of which have a certain property. Let us choose a sample of $n \leq N$ elements.
The probability that the sample contains $m \leq n$ elements with the above-mentioned property is in case of

1. *Nonreplacement:*

$$P_{noplacemen\ t} = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}}$$

2. *Replacement*

$$P_{nreplacet} = \binom{n}{m} \cdot p^m \cdot q^{n-m}, \qquad \frac{M}{N} =: p, \qquad q := 1 - p$$

III. Random Variables

Distribution function
A *distribution function* is defined as:

$$F(X) = P(X < x), \quad x \in R^1$$

Important properties of the distribution function
1.
$$0 \le F(x) \le 1, \quad \forall x \in R^1 .$$
2.
$$\forall x_1, x_2 : x_1 < x_2 \quad \Rightarrow \quad F(x_1) \le F(x_2) .$$
3.
$$\forall x_1, x_2 : x_1 < x_2 \quad \Rightarrow \quad P(x_1 \le X < x_2) = F(x_2) - F(x_1) .$$
4.
$$x \to -\infty \quad \Rightarrow \quad F(x) \to 0$$

$$x \to +\infty \quad \Rightarrow \quad F(x) \to 1 .$$
5.
$F(x)$ is at least left-sided continuous and has at most a finite number of jump discontinuities.

Probability function
If X is a discrete random variable, then the function

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

defined on the outcomes of X is called the *probability function* of the discrete random variable X .
If X has the outcomes x_i, $i = 1, 2, ..., n$, then we can write:

x_i	x_1	x_2	$\cdots$	x_n
$p_i = P(X = x_i) = f(x_i)$	p_1	p_2	$\cdots$	p_n

Distribution function

$$F(x) = P(X < x) = \begin{cases} 0 & \text{for} \quad x \le x_1 \\ \sum_{i=1}^{k} p_i & \text{for} \quad x_k < x \le x_{k+1}, \quad k = 1,\ 2,\ ..., n-1 \\ 1 & \text{for} \quad x > x_n \end{cases}$$

Density Function
Let $F(x)$ be a differentiable distribution function of a continuous random variable X . The function

$$f(x) := F^{'}(x)$$

is called the (*probability) density function of* X .

Important properties of a density function:

1.

$$F(x) = P(X < x)$$

$$= P(-\infty < X < x)$$

$$= \int_{-\infty}^{x} f(t)dt .$$

2.

$$P(a \le X < b) = F(b) - F(a)$$

$$= \int_{a}^{b} f(x)dx .$$

3.

$$P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1 .$$

Expected value

The *expected value* of the random variable X , denoted by $E\left(X\right)$, is defined as

$$E(X) := \begin{cases} \sum_{i=1}^{\infty} x_i \cdot p_i & \text{when } X \text{ discrete} \\ \int_{-\infty}^{+\infty} x \cdot f(x)dx & \text{when } X \text{ continuous} \end{cases}$$

Variance or dispersion, standard deviation

The *dispersion* or *variance* of the random variable X , denoted by $D^2\left(X\right)$, is defined as

$$D^2\left(X\right) := E\left(X - E(X)\right)^2$$

i. e.

$$D^2(X) := \begin{cases} \sum_{i=1}^{\infty} \left(x_i - E(X)\right)^2 \cdot p_i & \text{when } X \text{ discrete} \\ \int_{.\infty}^{+\infty} \left(x - E(X)\right)^2 \cdot f(x)dx & \text{when } X \text{ continuous} \end{cases}$$

under the assumption that the expected value exists.

The standard deviation of the random variable X , denoted by $D\left(X\right)$, is defined as

$$D\left(X\right) := \sqrt{D^2(X)} \quad (>0)$$

$$D^2(X) := \begin{cases} \sum_{i=1}^{\infty} x_i^2 \cdot p_i - \left(\sum_{i=1}^{\infty} x_i \cdot p_i\right)^2 & \text{when } X \text{ discrete} \\ \int_{.\infty}^{+\infty} x^2 \cdot f(x)dx - \left(\int_{-\infty}^{+\infty} x \cdot f(x)\right)^2 & \text{when } X \text{ continuous} \end{cases}$$

Standardisation:
Let

$$Z = aX + b, \quad a, b = const.$$

Then

$$E(aX + b) = a \cdot E(X) + b\,.$$

$$D^2(aX + b) = a^2 \cdot D^2(X).$$

The random variable Z is called a *standardised random variable* if

$$E(Z) = 0, \qquad D^2(Z) = 1.$$

The process

$$Z = g(X) := \frac{X - \mu}{\sigma} = \frac{1}{\sigma} \cdot X - \frac{\mu}{\sigma}$$

is called *standardisation*.

IV. Special Discrete Distribution Functions

Hypergeometric Distribution:
A discrete variable X has a hypergeometric distribution if its probability function is of the form

$$P(X = x) = p_i = \frac{\binom{M}{x} \cdot \binom{N - M}{n - x}}{\binom{N}{n}},$$

$$x = 0, 1, \ldots, n; \quad n \le M \le N\,.$$

The probability function of the hypergeometric distribution is formally equivalent to the sampling scheme *without replacement*.

Let X have a hypergeometric distribution. Then

$$E(X) = n \cdot \frac{M}{N} = n \cdot p,$$

$$D^2(X) = \frac{N - n}{N - 1} n \cdot p \cdot q$$

Binomial Distribution
A discrete variable X has a binomial distribution if its probability function is of the form

$$P(X = x) = p_i = \binom{n}{x} \cdot p^x \cdot q^{n-x}, \quad x = 0,\ 1, \ldots, n\,.$$

The probability function of the binomial distribution is formally equivalent to the sampling scheme with replacement.

Let X have a binomial distribution. Then

$$E(X) = n \cdot p,$$

$$D^2(X) = n \cdot p \cdot q.$$

Rule of thumb:
For a "sufficiently" large N, the hypergeometric distribution can be approximated by the binomial distribution. It will be recommended to use the following rule of thumb:
"If $10 \cdot n \le N$, then the hypergeometric distribution can be approximated by the binomial distribution."

Poisson Distribution:
A discrete variable X has a Poisson distribution if its probability function is of the form

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda},$$
$$x = 0, 1, \ldots, n.$$

Let X have a Poisson distribution. Then

$$E(X) = D^2(X) = n \cdot p = \lambda.$$

Rule of thumb:
The binomial distribution can be approximated by the Poisson distribution for n "sufficiently" large ($n \to \infty$) while $n \cdot p = \lambda$ remaining constant. That is why the Poisson distribution is also known as the "distribution of rare events".
It will be recommended to use the following rule of thumb:

"If

$$n \cdot p \le 10 \quad \text{and} \quad n \ge 1500p,$$

the binomial distribution can be approximated by the Poisson distribution."

V. Special Continuous Distribution Functions

A continuous variable X has *normal* (or *Gaussian*) distribution if its probability density function is of the form

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x - \mu}{\sigma}\right)$$

$$= \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in R^1$$

Let X be a normally distributed variable. Then

$$E(X) = \mu,$$

$$D^2(X) = \sigma^2.$$

The probability density function of the standardised normal distribution:

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in R^1.$$

The *standardised normal distribution function*:

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad x \in R^1.$$

Some important formulas:

$$\varphi(-x;\ 0,\ 1) = \varphi(x;\ 0,\ 1),$$

$$\Phi(-x;\ 0,\ 1) = 1 - \Phi(x;\ 0,\ 1).$$

$$P(|X - \mu| < c) = 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1, \quad c \in R^1.$$

"68-95-99.7 Rule" or "Empirical Rule":
All normal density curves satisfy the following property which follows from T. 8. 3.:

$$P(|X - \mu| < \sigma) = 0.6828$$
$$P(|X - \mu| < 2\sigma) = 0.9544$$
$$P(|X - \mu| < 3\sigma) = 0.9972.$$

Normal approximation to binomial:
The above theorem can be used to approximate binomial distribution by normal distribution. It will be recommended to use the following rule of thumb:

"If

$$n \cdot p > 5 \quad \text{and} \quad n \cdot q > 5,$$

then the binomial distribution can be approximated by normal distribution."

(Last revised: 24.10.2017)